



Единое цифровое пространство научных знаний: проблемы и решения

Сборник научных трудов



**Москва
Берлин
2021**

УДК 001:004(063)

ББК 72с51я431

Е33

Е33 **Единое** цифровое пространство научных знаний: проблемы и решения : сборник научных трудов / под ред. Н. Е. Каленова, А. Н. Сотникова — Москва ; Берлин : Директмедиа Паблишинг, 2021. — 464 с.
DOI: 10.23681/610687

ISBN 978-5-4499-1905-2

Настоящий сборник содержит материалы всероссийской научной конференции «Единое цифровое пространство научных знаний: проблемы и решения», проходившей при поддержке РФФИ в Москве в Межведомственном суперкомпьютерном центре (МСЦ) РАН – филиале ФГУ ФНЦ НИИСИ РАН с 10 по 12 ноября 2020 года.

В конференции приняли участие более 70-ти специалистов, представляющих научные организации, вузы, академические библиотеки, архивы и научные музеи Москвы, Санкт-Петербурга и др. регионов страны.

Конференция является результатом трансформации научно-практического семинара «Информационное обеспечение науки: новые технологии», который проводился в течение более 30 лет (с 1985 по 2017 гг.) под эгидой Библиотеки по естественным наукам РАН и Центральной научной библиотеки Уральского отделения РАН.

С презентациями докладов и выступлениями на конференции можно ознакомиться на сайте <http://dirmsmc.ru/konf>

УДК 001:004(063)

ББК 72с51я431

ISBN 978-5-4499-1905-2

© Коллектив авторов, текст, 2021

© Издательство «Директмедиа Паблишинг», оформление, 2021

Содержание

Архитектура единого цифрового пространства научных знаний. <i>Н. Е. Каленов, Г. И. Савин, А. Н. Сотников</i>	7
Международная платформа iScience4All: интеграция в единое цифровое пространство научных знаний стран-членов МЦНТИ. <i>Е. В. Угринович</i>	17
Роль ИКТ в задачах цифровой трансформации э-правительства. <i>И. Г. Федоров, М.-О. Никсон</i>	26
О некоторых тенденциях развития глобальной цифровой среды. <i>О. В. Сюттюренко, Е. Ю. Дмитриева</i>	51
Институты социальной памяти и проблемы перехода к цифровому обществу. <i>Д. В. Ефременко</i>	57
Теоретико-методологические основания проектирования единого цифрового пространства научных знаний. <i>Н. В. Лопатина</i>	68
О создании междисциплинарного аналитического центра для исследования дальнего чтения и digital humanities. <i>В. П. Леонов</i>	78
О функциональности единого цифрового пространства научных знаний. <i>О. М. Атаева, Н. Е. Каленов, В. А. Серебряков, А. Н. Сотников,</i>	88
Интеграция лингвистических информационных ресурсов в рамках концепции Единого цифрового пространства научных знаний. <i>А. Б. Антопольский</i>	108
Использование предметной онтологии единого цифрового пространства научных знаний в наукометрических задачах. <i>Цветкова В. А., Каленов Н. Е., Мохначева Ю. В., Митрошин И. А.</i>	129

Сеть классификационных систем как элемент обобщенной предметной онтологии единого цифрового пространства научных знаний. <i>А. В. Шапкин, В. Н. Белоозеров, Е. Ю. Дмитриева</i>	138
Создание информационного ресурса «Электронные архивы ИЭА РАН»; технологии и результаты. <i>Д. А. Функ, М. Б. Лейбов, Р. М. Галеев, Н. В. Хохлов</i>	150
Академические собрания: новый сетевой инструмент филологических исследований. <i>С. Н. Гуськов</i>	157
Вопросы цифровой интеграции научно-технических знаний архивов. <i>Г. З. Залаев</i>	165
Единое информационное пространство архивов науки в контексте модели ИСАРАН. <i>И. Н. Ильина</i>	169
Информационная система «История геологии и горного дела» как элемент единого цифрового пространства научных знаний в области истории науки. <i>И. Г. Малахова, И. П. Второв</i>	186
Развитие программного обеспечения электронной библиотеки «Научное наследие России». <i>К. П. Погорелко</i>	199
Автоматизированная система формирования и поддержки базы данных результатов научной деятельности сотрудников академических организаций. <i>С. А. Власова</i>	208
Некоторые аспекты представления 3D-объектов в едином цифровом пространстве научных знаний. <i>И. Н. Соболевская</i>	221
Формирование цифровых ресурсов для электронной библиотеки «Научное наследие России». <i>С. А. Кириллов</i>	233
Разработка тезаурусов для поиска и навигации в цифровом пространстве знаний. <i>Н. Н. Шабурова, В. Н. Белоозеров</i>	240

Российские публикации по информатике в базе данных «Web of Science». <i>А. В. Глушановский</i>	253
Классификационные индексы документов электронных каталогов естественнонаучных и технических библиотек: аспектный анализ. <i>А. Н. Сысоев, В. Н. Белоозеров</i>	270
Принципы обеспечения повторного использования данных и их применение в инфраструктурах исследовательских данных. <i>Н. А. Скворцов</i>	288
Выбор оптимального количества процессорных ядер для алгоритма многократной маркировки перколяционных кластеров на высокопроизводительных вычислительных системах. <i>С. Ю. Лапина</i>	311
Национальная исследовательская компьютерная сеть: инфраструктурно-сервисная основа единого цифрового пространства научных знаний. <i>А. Г. Абрамов, А. В. Евсеев, А. А. Гончар</i>	320
Презентация цифрового пространства знаний через сайт научной библиотеки. <i>Т. Н. Харьбина, И. А. Митрошин, Е. В. Бескаравайная</i>	334
Портал БЕН РАН как источник научных знаний. <i>Т. Н. Соловьева</i>	345
Процессы импортозамещения и конверсии: проблемы информационного обеспечения. <i>О. В. Сюттюренко, Е. Ю. Дмитриева</i>	358
Разработка единой системы наукометрического анализа на примере УрО РАН. <i>Ю. Д. Прокофьева</i>	372
Профессионал научной инфосферы как субъект единого цифрового пространства научных знаний. <i>М. А. Пекшиева</i>	392

Методика оценки потенциала мобильных приложений для информационного обеспечения научных исследований. <i>А. Ю. Герасименко</i>	415
Сохранение знаний и развитие научных школ в процессе цифровизации. <i>О. М. Атаева, В. А. Серебряков, Н. П. Тучкова</i>	427
Особенности динамики формирования цифрового кластера знаний о результативности и востребованности итогов исследований российских учёных в области медиации. <i>В. В. Арутюнов</i>	451
Экспериментальное обоснование критериев количественной оценки инновационности объекта. <i>В. К. Иванов</i>	459
Мягкие измерения и вычисления при эволюционном управлении непрерывным многостадийным производством. <i>Б. В. Палух, А. Н. Ветров</i>	476
Информационное пространство современности: реалии, проблемы и перспективы. <i>Н. А. Коровникова</i>	485
Международный научно-практический журнал «Программные продукты и системы» — более 30-ти лет в цифровом пространстве научных знаний. <i>Н. А. Семенов</i>	489
Авторы сборника	498

Архитектура единого цифрового пространства научных знаний

Н. Е. Каленов, Г. И. Савин, А. Н. Сотников

МСЦ РАН

Аннотация. Архитектура Единого цифрового пространства научных знаний (ЕЦПНЗ) определяется его функциями и задачами. ЕЦПНЗ включает совокупность подпространств, относящихся к различным научным направлениям. Единство подпространств обеспечивается унифицированными принципами построения подпространств и онтологическими связями между их объектами. Каждое подпространство включает цифровые объекты, метаданные, содержащие факты, связанные с объектами, и предметные онтологии, обеспечивающие развитый поиск и навигацию по пространству. Вся информация, отражается в ЕЦПНЗ по правилам «semantic WEB». Контент каждого подпространства включает ядро (проверенные временем достоверные научные результаты) и надстройку – новые научные результаты, прошедшие предварительную экспертизу.

Ключевые слова: научная информация, структура пространства знаний, онтологии, контент.

The Architecture of the Common Digital Space of Scientific Knowledge

N. E. Kalenov, G. I. Savin, A. N. Sotnikov

JSCC RAS

Abstract: The architecture of the Common Digital Space of Scientific Knowledge (CDSSK) is determined by its functions and objectives. CDSSK includes a set of subspaces related to various scientific fields. The unity of subspaces is provided by unified principles for constructing subspaces and ontological connections between their objects. Each subspace includes digital objects, metadata containing facts related to objects, and subject ontologies that provide advanced searches and navigation through space. All information is reflected in the CDSSK according to the rules of the «semantic WEB». The content of each subspace includes a core (time-tested reliable

scientific results) and a superstructure – new scientific results that have passed preliminary examination.

Keywords: scientific information, structure of knowledge space, ontologies, content

Современное информационное пространство содержит огромные объемы научной информации, отображаемой в публикациях и базах данных, представленных в цифровом виде. Эти объемы нарастают с увеличивающейся скоростью по всем областям науки. Если говорить об абсолютных цифрах, то в качестве примера можно привести данные о ежегодном количестве опубликованных статей в журналах и материалах конференций, отраженных в Российском индексе научного цитирования (РИНЦ) [1] по трем принципиально различным областям науки – информатике, микробиологии и языкознанию. Эти данные за последнее десятилетие по трем годам (2010, 2014 и 2018) приведены в табл. 1.

Таблица 1

Динамика роста количества публикаций

Область науки	Вид публикации	2010	2014	2018
Информатика	Статьи в журналах	24968	40293	36760
	Материалы конференций	7501	17591	36364
	Итого	32469	57884	73124
Микробиология	Статьи в журналах	10471	11607	11987
	Материалы конференций	999	2178	4213
	Итого	11470	13785	16200
Языкознание	Статьи в журналах	25900	47182	50064
	Материалы конференций	8047	20670	47745
	Итого	33947	67852	97809

Хотя динамика роста количества публикаций различна для различных научных направлений, и ее анализ не входит в задачи данной статьи, из приведенного примера видно, что во всех трех областях науки количество публикуемых статей

год от года возрастает и достаточно существенно. Чтобы быть в фарватере развития науки, исследователю необходимо обладать новейшей и точной информацией в его области исследований. Предоставление ученым той и только той информации, которая им необходима в каждый момент времени, традиционно являлось в России основной задачей научных библиотек и информационных центров. Развитие информатики, ее технической и алгоритмической базы, замена печатных изданий цифровыми породили тезис «в интернете всё есть», следствием чего явилось переключивание задачи информационного сопровождения научных исследований на плечи самих ученых. Этот подход может привести к тому, что исследователь будет тратить львиную долю своего времени и интеллектуальных усилий на ознакомление с информацией вместо того, чтобы заниматься исследованиями. С другой стороны, если он этого не сделает, отсутствие информации может привести к дублированию исследований и бессмысленной затрате времени на получение результатов, ранее уже полученных другими исследователями.

Создание Единого цифрового пространства научных знаний (ЕЦПНЗ) [2, 3], содержащего достоверную и не дублированную научную фактографическую и документальную структурированную информацию по различным разделам науки, является одним из направлений решения проблемы информационного сопровождения научных исследований. Но это — лишь одна из целей создания ЕЦПНЗ.

Общая цель создания ЕЦПНЗ — формирование и поддержка цифровой информационной среды, необходимой для решения комплекса задач развития общества, обеспечивающих:

- информационное сопровождение научных исследований;
- поддержку образовательных процессов, начиная от средней школы (тесно связанных с популяризацией нау-

ки) и кончая аспирантурой (напрямую связанных с информационным сопровождением науки);

- популяризацию науки (развитие мотивации для занятий наукой и получения соответствующего образования, одобрение финансирования науки со стороны общества).

- сохранность научных знаний;

- процессы мониторинга и управления наукой.

Таким образом, ЕЦПНЗ должно выполнять научно-информационные, образовательные, общекультурные и управленческие функции.

Соответственно, ЕЦПНЗ должно содержать элементы, ориентированные на следующие категории пользователей:

- исследователи, которым должна быть предоставлена многоаспектная ретроспективная и текущая (отфильтрованная по различным критериям, обеспечивающим ее достоверность и новизну ¹) информация по соответствующему их интересам научному направлению;

- школьники и студенты, которые должны получать проверенную временем достоверную базовую информацию различного уровня; эта информация должна включать факты со ссылками на учебные материалы, собственно тексты классических учебников (отбор их также должен осуществляться на основе независимых критериев), цифровые модели тех или иных явлений и событий;

- специалисты – аналитики и представители управленческих структур, анализирующие состояние и тенденции развития различных областей науки.

- широкая общественность, которую необходимо знакомить с наиболее интересными результатами, полученными в той или иной области науки, а также с историей научных открытий и их авторами;

- специалисты и «любители», интересующиеся историей науки и ее творцами.

¹ Фильтрация может осуществляться на основе критериев отбора источников информации, по экспертным оценкам признанных авторитетных ученых и т. п.

Задачи, решаемые в рамках создания ЕЦПНЗ, определяют его архитектуру. Пространство должно включать совокупность подпространств (ПП) по различным областям науки, построенных по единым принципам, основанным на использовании онтологических стандартов, используемых в среде semantic WEB. Контент каждого входящего в ЕЦПНЗ ПП включает три составляющих – цифровые объекты, являющиеся отражением объектов реального мира; связанные между собой и с цифровыми объектами метаданные, отражающие различные свойства каждого экземпляра цифрового объекта; предметные онтологии (тезаурусы, дополненные индексами различных классификационных систем), описывающие данную научную область. Отдельные ПП объединяются в единое пространство благодаря (а) наличию междисциплинарных связей внутри предметных онтологий и (б) наличию универсальных классов объектов, связанных с различными областями науки.

К универсальным классам относятся объекты, онтологически не связанные с какой-то конкретной областью науки. Профили их метаданных, представленных в ЕЦПНЗ, не зависят от того, к какому тематическому подпространству они относятся. К универсальным принадлежат такие классы объектов как «событие», «персона», «публикация», «организация», «архивный документ» и др.

В каждом подпространстве, наряду с универсальными классами объектов, присутствуют свои (локальные) классы, специфичные для данного научного направления. Например, такие классы, как «теорема», «уравнение», «множество», характерны для математики; «химический элемент», «реакция» – для химии; «корпус текста», «язык» – для языкознания.

Выявление локальных классов объектов, относящихся к тому или иному подпространству ЕЦПНЗ, определение профилей метаданных объектов каждого класса и видов связей между объектами одного и различных классов являются основными задачами при проектировании любого тематического подпространства ЕЦПНЗ.

Исходя из задач, решаемых в рамках ЕЦПНЗ, каждое подпространство должно включать «базис» – фундамен-

тальную, проверенную временем, информацию, относящуюся к данному научному направлению, и «надстройку» — новые научные результаты, прошедшие предварительную экспертизу. Базис представляет собой совокупность законов, постулатов, основных результатов, полученных в данной научной области, с отсылками на источники, в которых они опубликованы, и полные тексты этих источников. Контент базиса включает три взаимосвязанных уровня — образовательный, научно-популярный и фундаментальный.

- Образовательный уровень, рассчитанный, в частности, на учащихся школ, формируется на основе информации, полученной из фундаментальных учебников, и включает, наряду с фактографической информацией и полными текстами учебников, верифицированные мультимедийные ресурсы и, в частности, оцифрованные музейные объекты. Контент этого уровня является, в основном, статическим — информация обновляется достаточно редко.

- Научно-популярный, рассчитанный на более «продвинутых» пользователей, но не специалистов в данной области. Этот уровень квазистатический. Он может изменяться при появлении новых существенных для данной области знания открытий или результатов. Формирование этого уровня строится на основе энциклопедической и справочной информации, а также монографий и учебников для высшей школы. На этом уровне наряду с извлечёнными метаданными, пользователь может получить ссылки на источники и их цифровые копии в соответствии с законом об охране авторских прав.

- Контент фундаментального уровня предназначен для специалистов в данной области науки. Этот уровень включает более глубокую, по сравнению с предыдущими двумя уровнями, информацию по данному научному направлению, ядро этого контента является статическим и содержит многоаспектную фактографическую информацию со ссылками на ее источники. Если источниками являются опубликованные материалы, их полные тексты (по возможности, первые публикации, представляющие историческую

ценность) также должны входить в контент базиса данного ПП.

Контент надстройки подпространства рассчитан на исследователей, работающих в данной области науки, и содержит новые научные результаты, опубликованные в авторитетных журналах или отраженные в авторских свидетельствах. По прошествии определенного времени, по результатам оценки экспертным сообществом, отдельные составляющие надстройки могут быть переведены в базис или удалены из ЕЦПНЗ.

Технологически ЕЦПНЗ должно включать следующие взаимосвязанные составляющие: центральное ядро, совокупность локальных ядер тематических подпространств и распределенный контейнер. Центральное ядро содержит метаданные объектов универсальных классов и обобщенную предметную онтологию; локальные ядра содержат метаданные объектов локальных классов, связанные между собой и с объектами универсальных классов. Распределенный контейнер содержит цифровые копии предметов реального мира (публикаций, архивных документов, музейных артефактов, научно-популярных фильмов), они могут храниться на серверах владельцев – библиотек, музеев, архивов и др.

Создание ЕЦПНЗ является проблемой национального масштаба. Для ее решения необходимо разработать общую онтологию ЕЦПНЗ; функциональную и организационную структуру пространства; проанализировать существующие возможные источники информации и сформулировать критерии отбора объектов для включения в контент пространства; разработать технологию наполнения и актуализации контента; проработать вопросы сервисной составляющей ЕЦПНЗ и правовые вопросы использования и предоставления пользователям элементов пространства; сформировать и реализовать требования к лингвистическим, техническим и программным средствам отдельных составляющих ЕЦПНЗ и пространства в целом.

Для реализации перечисленных задач у отечественных исследователей имеется серьезный задел. В качестве основы для дальнейшего развития работ в направлении создания

ЕЦПНЗ можно рассматривать исследования, связанные с созданием и поддержкой электронной библиотеки «Научное наследие России» [4, 5], проводимые в МСЦ РАН; исследования в области семантических цифровых библиотек, проводимые в ФИЦИУ [6, 7]; исследования в области развития лингвистических средств, выполняемые специалистами ВИНТИ РАН и ИНИОН РАН [8–10]. В качестве прототипов практической реализации отдельных составляющих ЕЦПНЗ могут выступать Общероссийский математический портал MathNet [11, 12], информационная система «Соционет» [13, 14], информационная система «История геологии и горного дела» [15, 16].

В качестве источников контента ЕЦПНЗ необходимо использовать энциклопедические данные, общенаучные документальные информационные системы, такие как WEB of Science, Scopus, РИНЦ, каталоги крупнейших научных библиотек, проблемно-ориентированные фактографические базы данных, информационные системы научных музеев и архивов.

Исследования в области формирования ЕЦПНЗ проводятся в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН в рамках госзадания 0580–2021–0016 и при поддержке РФФИ (проект 20–07–00773).

Литература

1. Российский индекс научного цитирования. <https://www.elibrary.ru/> (Дата обращения 9.11.2020)
2. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии, 2019. Т. 89. № 7. С. 728–735.
3. Антопольский А. Б. и др. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.
4. Каленов Н. Е., Савин Г. И., Серебряков В. А., Сотников А. Н. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты, системы и алгоритмы, 2012. Т. 4. № 100. С. 30–40.
5. Погорелко К. П. Динамика использования электронной библиотеки «Научное наследие России» // Информационное обеспечение

науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2017. С. 192–200.

6. Атаева О. М., Серебряков В. А. Онтология цифровой семантической библиотеки LibMeta // Информатика и ее применения. 2018. Т. 12. № 1. С. 2–10.

7. Атаева О. М., Серебряков В. А., Тучкова Н. П. Расширение предметной области информационного запроса на основе онтологии знаний цифровой библиотеки LibMeta // Научный сервис в сети Интернет. 2019. № 21. С. 63–75.

8. Антопольский А. Б., Белоозеров В. Н., Каленов Н. Е., Маркарова Т. С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информационные ресурсы России. 2018. № 5 (165). С. 22–30.

9. Белоозеров В. Н., Шапкин А. В., Шуко Ю. Н. Сеть классификационных систем ВИНТИ РАН // Программные продукты, системы и алгоритмы. 2018. № 4. С. 20–23.

10. Антопольский А. Б. О создании центра лингвистических ресурсов РАН // Известия Российской академии наук. Серия литературы и языка. 2019. Т. 78. № 4. С. 5–12.

11. Жижченко А. Б., Изаак А. Д. Информационная система Math-Net.Ru. Современное состояние и перспективы развития. Импакт-факторы российских математических журналов // Успехи математических наук. 2009. Т. 64. № 4 (388). С. 195–204.

12. Общероссийский математический портал <http://www.mathnet.ru/> (дата обращения 9.11.2020).

13. Parinov S., Lyapunov V., Puzyrev R., Kogalovsky M. Semantically Enrichable Research Information System Socionet // Communications in Computer and Information Science. 2015. Т. 518. С. 147–157.

14. Соционет. <https://socionet.ru/> (дата обращения 9.11.2020).

15. Каленов Н. Е., Малахова И. Г. Интегрированный общедоступный информационный ресурс «История геологии и горного дела» // Информационные ресурсы России, 2017. № 1. С. 19–23.

16. Информационная система «История геологии и горного дела». <http://higeo.ginras.ru/> (дата обращения 9.11.2020).

References

1. Rossijskij indeks nauchnogo citirovaniya. <https://www.elibrary.ru/> (Data obrashheniya 9.11.2020)

2. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. O edinom cifrovom prostranstve nauchny`x znaniy // Vestnik Rossijskoj akademii, 2019. Т. 89. № 7. S. 728–735.

3. Antopol'skij A. B. i dr. Principy` postroeniya i struktura edinogo cifrovogo prostranstva nauchny`x znaniy (ECzPNZ) // Nauchno-
texnicheskaya informaciya. Ser. 1. 2020. № 4. S. 9–17.
4. Kalenov N. E., Savin G. I., Serebryakov V. A., Sotnikov A. N. Principy` postroeniya i formirovaniya e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Programmy`e produkty`, sistemy` i algoritmy`, 2012. T. 4. № 100. S. 30–40.
5. Pogorelko K. P. Dinamika ispol'zovaniya e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Informacionnoe obespechenie nauki: novy`e texnologii: Sbornik nauchny`x trudov. M.: BEN RAN, 2017. S. 192–200.
6. Ataeva O. M., Serebryakov V. A. Ontologiya cifrovoy semanticheskoy biblioteki LibMeta // Informatika i ee primeneniya. 2018. T. 12. № 1. S. 2–10.
7. Ataeva O. M., Serebryakov V. A., Tuchkova N. P. Rasshirenie predmetnoj oblasti informacionnogo zaprosa na osnove ontologii znaniy cifrovoy biblioteki LibMeta // Nauchny`j servis v seti Internet. 2019. № 21. S. 63–75.
8. Antopol'skij A. B., Beloozerov V. N., Kalenov N. E., Markarova T. S. O razvitii terminologicheskoy bazy` danny`x v vide kompleksa otraslevy`x informacionno-poiskovy`x tezaurosov // Informacionny`e resursy` Rossii. 2018. № 5 (165). S. 22–30.
9. Beloozerov V. N., Shapkin A. V., Shhuko Yu. N. Set` klassifikacionny`x sistem VINTI RAN // Programmy`e produkty`, sistemy` i algoritmy`. 2018. № 4. S. 20–23.
10. Antopol'skij A. B. O sozdanii centra lingvisticheskix resursov RAN // Izvestiya Rossijskoj akademii nauk. Seriya literatury` i yazy`ka. 2019. T. 78. № 4. S. 5–12.
11. Zhizhchenko A. B., Izaak A. D. Informacionnaya sistema Math-Net.Ru. Sovremennoe sostoyanie i perspektivy` razvitiya. Impakt-factory` rossijskix matematicheskix zhurnalov // Uspexi matematicheskix nauk. 2009. T. 64. № 4 (388). S. 195–204.
12. Obshherossijskij matematicheskij portal <http://www.mathnet.ru/> (data obrashheniya 9.11.2020).
13. Parinov S., Lyapunov V., Puzyrev R., Kogalovsky M. Semantically Enrichable Research Information System Socionet // Communications in Computer and Information Science. 2015. T. 518. S. 147–157.
14. Socionet. <https://socionet.ru/> (data obrashheniya 9.11.2020).
15. Kalenov N. E., Malaxova I. G. Integrirovanny`j obshhedostupny`j informacionny`j resurs «Istoriya geologii i gornogo dela» // Informacionny`e resursy` Rossii, 2017. № 1. S. 19–23.
16. Informacionnaya sistema «Istoriya geologii i gornogo dela». <http://higeo.ginras.ru/> (data obrashheniya 9.11.2020).

Международная платформа iScience4All: интеграция в единое цифровое пространство научных знаний стран-членов МЦНТИ

Е. В. Угринович

МЦНТИ

Аннотация. Описываются возможности и перспективы интеграции международной платформы iScience4All в единое цифровое пространство научных знаний стран-членов межправительственной организации «Международный центр научной и технической информации» (МЦНТИ), уникальность и преимущества реализации данного проекта, его вклад в трансформацию цифровой экономики, а также в формирование цифровой экосистемы научной и технической информации сообщества МЦНТИ.

Ключевые слова: единое цифровое пространство научных знаний, цифровая экономика, экосистемы, научно-техническая информация, открытый доступ, трансфер технологий, наука, бизнес.

International Platform iScience4All: Integration into Single Digital Space of Scientific Knowledge of ICSTI Member Countries

E. V. Ugrinovich

ICSTI

Abstract. The article describes the possibilities and prospects of integrating the international platform iScience4All into a single digital space of scientific knowledge of the member countries of the International Center for Scientific and Technical Information (ICSTI), the uniqueness and advantages of this project, its contribution to the transformation of the digital economy, as well as to the formation of a digital ecosystems of scientific and technical information of the ICSTI community.

Keywords: unified digital space of scientific knowledge, digital economy, ecosystems, scientific and technical information, open access, technology transfer, science, business.

Современную постиндустриальную экономику, которая называлась ранее инновационной, экономикой знаний, с недавнего времени именуют еще и цифровой. Появляются новые формы взаимодействия, новые технологии, происходит диффузия знаний в цифровом пространстве.

Формируется пространство цифровых экосистем на основе информационной революции и процессов глобализации экономики. Информация становится одним из основных сырьевых ресурсов, наряду с полезными ископаемыми, и преобразуется в знания. Качеством информационного «сырья» определяется качество конечного продукта.

Исследователи дают различные трактовки понятия «цифровая экосистема». Ключевыми в этом понятии необходимо считать обмен знаниями, технологии и людей, способных управлять знаниями.

Сейчас мы можем уже говорить и о феномене единого цифрового пространства научных знаний, и не только «как более узкой формулировки понятия единого российского электронного пространства знаний, которое предложено в ряде правительственных документов» [1], но и как глобальное понятие в мировом цифровом научном пространстве.

Одним из основных факторов, обеспечивающих устойчивое функционирование цифровой экономики, является управление в ней знаниями и инновациями.

Цифровая трансформация процесса производства и распространения знаний происходит с участием многочисленных акторов – представителей государства, научного сообщества, промышленности, бизнеса и гражданского общества.

Формирование экосистем становится базовым решением цифровой экономики. Наука формирует экосистему генерации знаний; бизнес формирует экосистему инновационной деятельности. Государство определяет стратегические направления развития науки и экономики и институциональные нормы деятельности для науки и бизнеса [2].

Основная проблема: разрыв коммуникаций между научным сообществом и индустрией (бизнесом). Цифро-

вая трансформация требует необходимость открытости [3]. Под влиянием технологического развития меняются традиционные каналы обмена информацией и коммуникации: виртуальные сети; журналы открытого доступа; открытые репозитории; открытые базы данных научных исследований; открытые инновации. Функционирование науки по законам бизнеса приводит к публикации недостоверной информации, а также использованию методов искусственной возгонки индексов цитирования [4].

Проблему совместимости основных понятий науки и бизнеса призвана обеспечить инновационная экосистема, обеспечивающая перевод научных знаний в бизнес-знания. Бизнес-знания обладают атрибутами, присущими всем другим видам товаров.

На многонациональной платформе межправительственной организации «Международный центр научной и технической информации» базируется и развивается цифровая экосистема научной и технической информации.

МЦНТИ – специализированная международная организация, создана на базе межправительственного соглашения в 1969 г. и зарегистрирована в ООН. В состав межправительственной организации «Международный центр научной и технической информации» (МЦНТИ) входят 22 страны из различных регионов мира. Каждая из стран обладает специфической экосистемой науки и инноваций, уникальным экономическим контекстом, индустриальной инфраструктурой. Одна из основных задач МЦНТИ – это развитие системы международной научной и технической информации.

Основными участниками экосистемы МЦНТИ являются: поставщики научно-технической информации (библиотеки, репозитории, агрегаторы, центры научно-технической информации, университеты, научно-исследовательские центры, издательства и др.), потребители НТИ (образовательные и исследовательские центры, аналитические центры, инновационные компании, исследовательские коллективы и индивидуальные исследователи и др.), органы управления (государственные и отраслевые органы управления, нацио-

нальные академии, международные организации, профильные ассоциации, специализированные компании и др.).

Мультинациональная платформа интегрирует опыт и знания из различных источников стран-членов МЦНТИ в единое цифровое пространство научных знаний, акторов и технологии; формирует среду для сопроизводства новой стоимости, создавая сетевой эффект, способствует развитию научно-технической коммуникации и партнерства.

Международная платформа МЦНТИ, в том числе, включает:

- Систему открытого доступа к мировым информационным ресурсам iScience4All [5].
- Систему коммуникации с мировыми агрегаторами открытого доступа [6].
- Систему – агрегатор данных о предлагаемых технологиях.

В настоящее время усилия МЦНТИ сконцентрированы на развитии международной цифровой информационной инфраструктуры поддержки научной, научно-технической и инновационной деятельности. Наиболее значимым нашим проектом является разработка и практическая реализация международной системы открытого доступа к научным знаниям «Цифровая наука для всех» – iScience4All. Инициатива о запуске этого проекта была одобрена решением Комитета Полномочных Представителей стран-членов Международного центра научной и технической информации, состоявшегося 19 сентября 2018 года в г. Минске, Республика Беларусь.

Актуальность проекта. Существующие сегодня проблемы хранения и трансграничного распространения новых знаний, создаваемых преимущественно за счет национальных бюджетов стран, привели к появлению различных альтернативных парадигм в сфере научной и технической информации (например, подходы, которые можно условно назвать протестными, типа Sci-Hub, а также получивший широкую международную поддержку Открытый доступ). Открытый доступ – перспективное направление, которое нуждается в серьезной институциональной поддержке. В этом смысле МЦНТИ является именно такой платформой,

на которой концепция Открытого доступа может быть реализована максимально эффективно на основе принципов международного права. Согласованные на международном уровне подходы и процедуры реализации учитывают возможности и потребности заинтересованных стран, их различия в уровне развития, экономическом укладе, языке, культуре.

Разрабатываемая система iScience4All нацелена на обеспечение надежного открытого доступа исследовательского сообщества к научной и технической информации за счет объединения на единой платформе национальных хранилищ научных данных, научных публикаций, электронных библиотечных систем и других цифровых ресурсов стран-членов Сообщества МЦНТИ.

Можно выделить три основных категории пользователей, для которых предназначена разрабатываемая система: 1) поставщики научно-технической информации – это библиотеки, репозитории, агрегаторы, научно-технические и исследовательские центры, университеты, издательства; 2) потребители информации, а именно, образовательные и научные организации, аналитические центры, инновационный бизнес, исследовательские коллективы и индивидуальные исследователи, экспертные сообщества; 3) органы власти, к которым можно отнести государственные органы управления, национальные академии и др.

Ожидаемый результат. Создаваемая система обеспечит доступность научных данных и знаний в мировом едином цифровом пространстве научных знаний за счет объединения национальных агрегаторов информационных ресурсов открытого доступа. Многоязычные источники знаний позволят расширить информационные возможности поставщиков и потребителей научно-технической информации из разных стран. При этом права на результаты интеллектуальной деятельности будут гарантированно защищены в процессе их использования и коммерциализации.

Уникальность и преимущества. Уникальность инициативы в том, что коллективный информационный ресурс для поддержки научно-технологического развития создается на платформе, зарегистрированной в ООН специализиро-

ванной межправительственной организации. Организация имеет международный статус и объединяет в своей деятельности группу стран из различных регионов мира. Страны, в свою очередь представлены национальными правительственными институтами.

Реализация международного проекта в таком формате определяет очевидные *преимущества* создаваемой системы:

- сам информационный ресурс или его элементы – при фактическом размещении на территории одной или нескольких стран-членов – будет находиться вне их национальных юрисдикций и на равных условиях принадлежать всем странам-участникам проекта. Наряду с возможностью правительств стран-участниц осуществлять контроль за соблюдением своих национальных интересов, использование ресурса будет регулироваться общепринятыми нормами международного права. Такой подход исключит риски получения какой-либо из стран необоснованных преимуществ и использования любых форм дискриминации.

- регламенты международной организации будут обеспечивать контроль качества размещаемых материалов, вырабатываемых на основе согласованных с международным сообществом критериев и требований.

- функционирование системы будет поддерживаться встроенными средствами многоязычного автоматизированного перевода, который позволит пользователям системы работать с материалами на своем родном или удобном для них языке.

- использование данной системы будет нацелено не на извлечение прибыли, а на расширение информационных возможностей исследователей разных стран и обеспечение благоприятных условий для доступа, распространения и повышения качества научной информации.

Финансирование создания и эксплуатации системы iScience4All. Основными плательщиками за обеспечение доступа к научной информации, в конечном счете, являются государства. Бюджетными средствами стран-участниц можно покрыть расходы на реализацию системы, которая будет функционировать на базе международной организации. В этом случае доступ к информации будет бесплатным для ученых из стран-участников. А прозрачность процедур финансового

контроля деятельности Центра является гарантией целевого характера использования средств «доноров». «Донорами» могут быть организации любой организационно-правовой формы и национальной принадлежности, что позволит использовать инструменты государственно-частного партнерства.

Вклад реализации проекта в трансформацию цифровой экономики. МЦНТИ является единственной в мире специализированной межправительственной организацией в сфере научно-технической информации (НТИ). Реализация на платформе Центра этого проекта будет существенным вкладом в содействие достижению национальных целей развития государств-членов. В частности, в содействии их цифровой трансформации, гармонизации стандартов и технологических регламентов международной системы НТИ, глобального взаимного подключения сетей НТИ открытого доступа.

Это поможет создать условия для ускоренного развития научных, исследовательских компетенций и проведения исследований и разработок путем обмена информацией в цифровом виде. Пользователи системы будут поддерживаться за счет построения общей межгосударственной цифровой инфраструктуры обмена информацией, единого цифрового пространства научных знаний и инноваций.

Реализация проекта iScience4All способствует распространению НТИ в целях выполнения Повестки дня ООН в сфере устойчивого развития в сообществе МЦНТИ.

Сфера приложений проекта iScience4All в перспективе может быть расширена на пространстве ЕЭС, ШОС, БРИКС, АТЭС, АСЕАН, ЕС и других интеграционных структур, в которые входят государства-члены МЦНТИ.

Также МЦНТИ активно взаимодействует с ведущими международными организациями: ООН, ЮНЕСКО, ЮНИДО, ЕС, ОЭСР, МАГАТЭ, ОИЯИ и др.; национальными органами стран-членов МЦНТИ и других стран (министерства, ведомства, научные организации, государственные и частные информационные центры и ресурсы); организациями и компаниями реального сектора экономики.

Преимуществами взаимодействия с МЦНТИ являются:

- успешный опыт международной и межправительственной деятельности в течение более 50 лет;
- высокий международный статус межправительственной организации в сфере экономической деятельности;
- сетевое взаимодействие с 22 государствами-членами и партнерами в других странах;
- многоаспектное сотрудничество с научно-исследовательскими институтами, университетами, библиотеками, предприятиями и технологическими компаниями;
- осуществление экспертизы высокого уровня в рамках национальных и международных проектов;
- опыт проведения крупных международных мероприятий и экспертных сессий;
- обеспечение международной научной, технической и бизнес-кооперации.

Литература

1. О едином цифровом пространстве научных знаний. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.
2. Башкина Е. М., Едименченко Т. М., Зубарев А. П., Скура-тов А. К. (2019). Приоритеты Стратегии научно-технического развития Российской Федерации в федеральной целевой программе научных исследований и разработок. Инновации, № 3 (245), С. 3–9.
3. Open Access 2020 Initiative // [Электронный ресурс] Режим доступа: URL: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>
4. Угринович Е. В., Мун Д. В., Попета В. В. (2016). «Прогресс и регресс, или как вернуть в научные издания научные знания». Международный журнал «Информация и инновации», сборник статей 2016 г.
5. Угринович Е. В., Мун Д. В. (2018). «iScience4All – открытая международная мультилингвистическая платформа управления научно-технической информацией». Международный журнал «Информация и инновации», т. 13, № 4.
6. Ловцов А. И., Угринович Е. В. (2019). Интеграция национальных репозиторий научной информации открытого доступа стран – членов Международного центра научно-технической информации НТИ, сер. 1. Орг. и методика информ. работы, ВИНТИ РАН, № 3, С. 6–12.

References

1. Antopolskii A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. Common digital space of scientific knowledge // Vestnik Rossijskoj akademii nauk. 2019. T. 89. № 7. С. 728–735. doi: 10.31857/S0869-5873897728-735.
2. Bashkina E. M., Edimenchenko T. M., Zubarev A. P., Skuratov A. K. Priorities of the strategy of scientific and technological development of the russian federation in the federal target program of research and development // Innovations. 2019. № 3. С. 3–9.
3. Open Access 2020 Initiative // URL: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>
4. Ugrinovich E. V., Mun D. V., Popeta V. V. Progress and Regress or how to get back scientific knowledge in scientific publication? // Information and Innovations. 2016 – № 1. С. 4–11.
5. Ugrinovich E. V., Mun D. V. iScience4All – the open international multilingual platform for the scientific & technical information management // Information and Innovations. 2018 – Т. 13 – № 4. С. 7–19.
6. Lovcov A. I., Ugrinovich E. V. (2019). Integraciya nacional'nyh repozitoriev nauchnoj informacii otkrytogo dostupa stran – chlenov Mezhdunarodnogo centra nauchno-tekhniceskoy informacii // NTI, ser. 1. Org. i metodika inform. Raboty, VINITI RAN, № 3, S. 6–12.

Роль ИКТ в задачах цифровой трансформации э-правительства

И. Г. Федоров

РЭУ им Г. В. Плеханова

М.-О. Никсон

Университет Венда, Тохояндоу, Южная Африка

Аннотация. Проведенный анализ инструментов сбора и анализа данных уровня государства, позволил сделать следующие выводы. Доминирующая сегодня парадигма «правительство как сфера услуг» провела разработчиков э-правительства по ложному пути, поэтому не следует планировать ИКТ проекты для государственного сектора под давлением со стороны лоббистов ИКТ, моды и желания следовать примеру других стран, чтобы поэкспериментировать с новыми технологиями. Курс на создание э-правительства должен быть скорректирован, с учетом новых подходов реализации инструментов э-правительства. Мы полагаем, что следующий шаг в развитии электронного правительства заключается не в переходе к следующему технологическому укладу, будь то большие данные, искусственный интеллект, интернет вещей или правительство как платформа.

Ключевые слова: электронное правительство, информационные технологии, коммуникационные технологии, информация.

The role of ICTs in the digital transformation of e government challenges

I. G. Fiodorov

G. V. Plekhanov REU

M.-O. Nixon

Vends University, South Africa

Abstract. The analysis of state-level data collection and analysis tools allowed us to draw the following conclusions. The dominant paradigm of "government as a service sector" has led the developers of e-government on the wrong path, so it is not necessary to plan ICT projects for the public sector under pressure from ICT lobbyists, fashion and the desire to follow the example of other countries to experiment with new technologies. The course for the creation of e-government should be adjusted, taking into account new approaches to the implementation of e-government tools. We believe that the next step in the development of electronic government is not to move to the next technological order, whether it is big data, artificial intelligence, the Internet of Things, or government as a platform.

Keywords: e-government transformation, information and communication technologies; big data.

Введение

Электронное правительство (э-правительство) стало сегодня модной темой, о нем принято говорить только в восторженных тонах, приписывать ему множество положительных свойств. Однако, как любой другой проект внедрения информационно-коммуникационных технологий (ИКТ), он не гарантирует, что первоначально заявленные цели будут достигнуты. Агрессивный маркетинг внушает нам, что цифровая трансформация бизнеса заключается во внедрении новых прорывных ИКТ и автоматически обеспечивает взрывной рост экономики [1]. Поэтому в центре внимания большинства публикаций по вопросу трансформации бизнеса обычно находятся чисто технологические аспекты — вчера было модно писать о социальных сетях, мобильном бизнесе и облачных вычислениях, сегодня принято говорить о больших данных, блокчейне, искусственном интеллекте, пр. Однако неверно полагать, что экономический рост происходит сам собою, без участия прочих факторов. Вспомним пузырь доткомов, образовавшийся в результате взлета акций компаний, чья бизнес-модель целиком основывается на работе в сети Интернет. Если отбросить поверхностные и очевидные причины кризиса, например, высокий уровень мошенничества и недостаточное понимание преимуществ, предоставляемых Интернетом [2], коренная проблема заключается в подмене понятий — ведение бизнеса

через интернет является только инструментом для осуществления бизнес-процесса, но не самостоятельным бизнес-процессом, способным генерировать доход от вложенного капитала [3], [4]. После краха слово дотком употребляется как синоним незрелой, непродуманной, либо неэффективной концепции бизнеса.

Аналогичная ситуация складывается в области применения ИКТ для решения задач э-правительства. В течение последних десятилетий были предприняты многочисленные попытки создать э-правительство, при этом делались бесчисленные заявления о том, что ИКТ и связанные с интернетом технологии, в конечном итоге трансформируют правительство. Несмотря на серьезные усилия и внушительные затраты, отсутствуют убедительные доказательства позитивного эффекта от внедрения ИКТ, нет свидетельств изменений и трансформации деятельности правительства [5]. Настало время задать вопрос, насколько правильно был выбран курс на создание э-правительства? Не повторяется ли ситуация с бумом электронного бизнеса, когда сообщество переоценило преимущества и не учло риски, связанные с внедрением ИКТ?

Не секрет, что ИКТ-проекты для государственного сектора часто реализовались на основе не экономических причин, а под давлением со стороны ИКТ лоббистов, моды, желания поэкспериментировать с новыми технологиями, следовать примеру других стран. В большинстве случаев вопрос об оценке ценности проекта не поднимался до его реализации. Поэтому необходимо критически подходить к практическим результатам создания э-правительства, учитывать специфику каждого государства [6].

Особый интерес представляют вопросы о роли ИКТ в задачах трансформации э-правительства. Сегодня становится понятно, что цифровая трансформация бизнеса не происходит автоматически в результате внедрения ИКТ, но напротив она является необходимой предпосылкой успеха подобных проектов и должна осуществляться до начала их реализации. Можно ожидать, что трансформация э-правительства также является пререквизитом успеха вне-

дрения ИКТ, однако в литературе этот вопрос освещен слабо, что подтверждает актуальность данного исследования.

Проблемы трансформации электронного правительства

Трансформацией принято называть четвертую, последнюю фазу в модели оценки прогресса в реализации системы э-правительства (модель зрелости) [8]. Эта модель включает четыре стадии: предоставление информации, простое взаимодействие, автоматизация административных процессов и, наконец, «трансформированное правительство» [1]. Предполагается, что трансформация происходит после полной автоматизации административных регламентов, пересекающих границы организационных подразделений [2]. Трансформации э-правительства посвящено малое число исследований. В них отмечается, что влияние трансформации э-правительства на его структуру и на протекающие процессы мало изучено [3]. Исследования, в которых рассматривается преобразующий потенциал ИКТ в государственном управлении (регулирование, предоставление услуг и разработка политики), а не в операционной эффективности, относительно редки [9]. Значение термина «трансформация» не определено явно. Словарь определяет его через понятия: преобразование, превращение [4], в свою очередь преобразование означает «изменить форму или образ» [10]. Возникает вопрос, что подразумевается под формой правительства или его образом.

Часто трансформацию э-правительства трактуют как изменение сервисов (услуг), предоставляемых государством гражданам или предприятиям [11], как сглаживание межведомственных барьеров или как интеграцию структуры правительства [14]. Многие авторы определяют трансформацию технически, как интеграцию или объединение функций, услуг или организаций [15] [18]. Однако существует мало доказательств того, что переход на электронные технологии ведет к изменению организационной структуры государственного аппарата [19]. Организационная форма и функции государ-

ства закреплены юридически, поэтому их преобразование требует законодательной инициативы [20].

Ряд исследователей отрицательно относятся к возможности трансформации э-правительства. Например, исследование, проведенное в Новой Зеландии, показало: хотя электронное правительство ввело новый стиль взаимодействия с гражданами и предприятиями, что создало иллюзию изменений и позволило извлечь операционные выгоды, в государственных структурах не произошло реальных преобразований

[21]. Многие авторы обращают внимание на пропасть между риторикой и реальностью в отношении трансформации э-правительства [24] [25]. Объяснение заключается в том, что исследования в области э-правительства были преимущественно техническими и не связаны с изысканиями в области государственного управления [26]. На начальном этапе исследований, когда ИКТ были еще новинкой, их окутывал флер моды, что определило привлекательность новых технологий в области э-правительства. Замысел заключался в том, чтобы добавить инновационные технологии в существующие сервисы правительства, и тем самым выиграть время для более серьезных изменений [27]. Поэтому на ранних этапах исследования в области электронного правительства были сосредоточены на понимании технических проблем внедрения ИКТ, вопросах использования и тиражирования. В результате, серьезный акцент делался на технологии. При этом были использованы модели функционирования, ориентированные на частный сектор. Все это отвлекло исследователей от понимания политического контекста, в котором внедряются государственные проекты. В результате, вместо реальной трансформации э-правительства, реализован его безбумажный аналог. Ориентация на технологические аспекты ИТ в контексте э-правительства позволила выявить преимущественно технические проблемы: сложности при подтверждении транзакций [30]; трудности коммуникации сквозь организационные границы [35], управление доступом и идентификация [36], обмен данными [37], интеграции

данных и транзакций [38], мобильный доступ к государственным системам [39], безопасность данных [36].

Целью трансформации, в большинстве случаев, декларируется повышение эффективности внутренних процессов э-правительства [40]. Однако пока нет актуальных результатов, демонстрирующих долговременный экономический эффект внедрения ИКТ в этой сфере [41]. Очевидно, что эффективность есть важный показатель деятельности государства, но не единственный, не следует забывать про социальные вопросы и экологию [42]. Поэтому можно сделать вывод о системной ошибке, связанной с сопределиением цели и имплементацией трансформации э-правительства [44]. Обратим внимание, что для оценки деятельности электронного правительства используются показатели, присущие частному сектору, например, экономичность, затраты, выгоды и эффективность [34], а оценка удовлетворенности пользователей осуществлялась с точки зрения веб-интерфейсов [45]. Можно утверждать, что эти исследования проводились, скорее, техницистскими специалистами, а не теми, кто имеет политические знания или опыт государственного управления и мог бы лучше понять проблемы, связанные с электронным правительством [46], [47]. Такой подход стимулировался, в немалой степени, посредством обмена «наилучшими практиками» и рекомендациями международных организаций [48]. Этот некорректный, но влиятельный анализ направил разработчиков в неправильном направлении [49]. Можно говорить о существовании, по крайней мере, трех заблуждений, связанных с обсуждаемым подходом к внедрению цифровых технологий в правительстве и государственном управлении [19]:

– Частые изменения технологий, происходящие каждые 5–10 лет, повышают расходы на модернизацию, техническое обслуживание, обеспечение безопасности и пр., что не способствует сокращению расходов на выполнение административных регламентов.

– Ориентация на пользователя ограничивается повышенным вниманием к интерфейсу, при этом упускается

из виду необходимость изменения процессов взаимодействия пользователя с государством.

–Изменение государственного управления ограничено национальной конституцией и законодательством.

Следует сделать вывод об отсутствии четкого понимания направлений трансформации э-правительства. Это ставит задачу разработать новые подходы к оценке действий э-правительства и влияния ИКТ на его работу. Поэтому в рамках данной работы мы ставим вопрос о роли ИКТ в задачах цифровой трансформации э-правительства.

Правительство реализует политику, а не предоставляет услуги

Цель правительства состоит в том, чтобы принимать, осуществлять и управлять решениями от имени сообщества, за которые оно несет ответственность, по вопросам политики, экономики, социальной сферы и экологии, которые влияют на жизнь этого сообщества в целом [50]. В отношении э-правительства сложилось спорное мнение, что оно оказывает услуги, из чего делается вывод, что применительно к государству можно использовать модель функционирования частного сектора. Конечно, любое государственное учреждение в отдельности можно рассматривать как предприятие, но на самом деле это не главное, чем занимается государство, а граждане – не клиенты. Способ достижения целей государства мы договоримся называть политикой [53]. Разработка политики предполагает выбор соответствующих средств (инструментов политики). Диапазон инструментов, доступных для достижения целей политики, огромен, на выбор могут влиять многие факторы [54], [55]. Разные правительства могут использовать разные наборы инструментов даже для достижения аналогичных целей. И наоборот, применение одного набора инструментов может иметь различные последствия в разных странах. При этом, существуют инструменты, которые подходят для реализации разных политик [56].

С каждой функцией государства связан набор инструментов реализации политики. Спектр инструментов, доступных для достижения целей политики, огромен и охватывает методы сбора и перераспределения финансовых средств, выдачу разрешений, регистрацию, принуждение, разработку законов и контроль их исполнения, регулирование, пр. На выбор инструмента могут влиять политические, экономические, социальные, экологические факторы или просто привычка или догма. Как только выбранный инструмент (или, более вероятно, набор инструментов) для реализации политики юридически закреплён соответствующим национальным законодательством, государственная администрация приступает к созданию и выполнению необходимых функций. При внимательном рассмотрении можно заметить, что большинство институтов государственного аппарата могут быть классифицированы либо как инструменты сами по себе (например, здравоохранение, транспорт или пенитенциарная служба), либо как организации, управляющие такими инструментами [19]. Таким образом, реформа государственного сектора заключается в изменении набора инструментов политики. Цифровые технологии могут изменить существующие инструменты и открыть возможность создания новых. Государственные службы решают сложную задачу по реализации политики и реформ. Трансформация системы налоговой, медицинской или энергетической политики из сложного сочетания множества взаимозависимых инструментов в реформированный набор взаимосвязанных служб, является чрезвычайно сложной задачей, требующей навыков управления программами и проектами высокого порядка и новых возможностей, чтобы заставить новую схему работы. Поэтому эксперты по цифровым технологиям должны лучше понимать специализированный и зачастую сложный политический, законодательный и административный мир, в котором они пытаются осуществить трансформацию. Речь идет не только о веб-сайтах и связанных с ними технических концепциях, но о хорошем понимании связанных с их применением политических, социальных и поведенческих последствий.

Сделаем вывод, что доминирующая сегодня парадигма «правительство предоставляет услуги» ведет ИКТ-разработчиков э-правительства по ложному пути. Становится очевидным, что выбранный путь исчерпал себя, однако за последнее время не появилось новых подходов, которые могли бы обеспечить развитие в этой сфере. Следующий шаг в развитии электронного правительства заключается не в переходе к следующему технологическому укладу, будь то большие данные, искусственный интеллект или интернет вещей. Эти распространенные техницистские подходы упускают из виду тот факт, что трансформация государственного управления означает изменение набора инструментов, обеспечивающих политические цели [57]. Пересмотр инструмента политики с учетом потенциала цифровых технологий может привести к гораздо более фундаментальным и эффективным результатам. Для этого необходимо при разработке политики использовать все текущие знания о том, что технологически возможно и имеет отношение к достижению цели политики. Кодирование существующих административных процессов в аппаратное и программное обеспечение, каким бы технологически элегантным оно не было, означает трату времени и денег [19].

Цели государства возникают на основе целей, лежащих в основе жизнедеятельности данного общества. Сегодня принята точка зрения, что цель государства заключается в обеспечении баланса трех компонентов устойчивого развития:

экономического, социального и экологического. Подразумевается принятие мер, направленных на оптимальное использование ограниченных ресурсов и применение экологичных природо-, энерго-, и материало-сберегающих технологий, для сохранения стабильности социальных и культурных систем, обеспечения целостности биологических и физических природных систем [57]. Таким образом, чтобы оценить роль ИКТ в достижении целей государства, необходимо исследовать механизмы деятельности государства, понять, что и как оно делает.

Инструменты реализации политики

К. Худ и Х. Маргетс обсуждают три возможных подхода, чтобы ответить на вопрос — что делает государство? Во-первых, можно попытаться описать, что происходит «внутри» правительства, как принимаются решения, как передаются приказы, как движется информация [58]. Этот подход приведет нас к исследованию процессов принятия решений правительством. Во-вторых, можно выявить функции правительства, перечислить все то, чем оно занимается. Однако в вопросе выделения функций э-правительства нет единства мнений. В-третьих, можно описать инструменты, которые использует правительство. Авторы полагают, что инструменты политики являются практическим средством достижения политических целей [59], [51]. Через их изучение легче понять влияние технологии на то, как правительство регулирует экономику, социальную сферу и окружающую среду [59]. Авторы предлагают ограничиться изучением инструментов политики, отложить исследование вопросов принятия решений, выявления функции правительства до следующего этапа. Инструменты политики универсальны, но могут использоваться правительствами по-разному даже для достижения одних и тех же целей в соответствии с местным политическим, правовым и социальным контекстом. По их мнению, этот подход является универсальным, подходит для анализа любого правительства, независимо от политического режима.

Предпринимались многочисленные попытки составить таксономию инструментов политики, но наиболее интересные результаты получили К. Худ и Х. Маргетс [58]. Они предположили, что правительства имеют в своем распоряжении четыре ресурса, которые они могут использовать для обнаружения изменений в окружающей среде и воздействия на нее.

Информационный ресурс, существующий просто в силу того факта, что правительство находится в центре информационной сети, по которой передаются разнообразные информационные потоки, которые могут рассматриваться как «информация» и «знание».

Ресурс власти, определяющий способность государства диктовать правила действия граждан путем разработки законов, ограничений, разрешений.

Ресурс общественных благ, определяющий способность собирать и перераспределять блага, финансовые средства льготы, гранты, кредиты, субсидии, управлять активами.

Организационный ресурс, определяющий способность государства воздействовать непосредственно на людей, направлять их действия.

Нас будет интересовать в первую очередь информационный ресурс, поскольку именно он является объектом автоматизации при внедрении ИКТ. Однако следует учитывать, что все ресурсы взаимосвязаны. Например, граждане могут предоставлять правительству информацию просто потому, что оно находится в центре всех коммуникаций, при этом информация оказывается для правительства «бесплатной». Во-вторых, правительство может использовать свои законные полномочия и требовать информацию от граждан по официальному запросу, который, сопровождается угрозой санкций за несоблюдение требований. В-третьих, информация может быть получена в обмен на блага, предоставляемые государством. В-четвертых, правительство может использовать свой организационный потенциал и не использовать человеческую мотивацию.

Инструментами К. Худ и Х. Маргеттс называют детекторы, которые позволяют собирать определенный вид ресурса в пользу государства, и эффектеры, которые распределяют данный ресурс между гражданами [58]. Выделим главную идею предлагаемого подхода — набор инструментов является инвариантом, относительно возможных преобразований э-правительства. Однако конкретный выбор инструмента реализации политики является прерогативой данного правительства и зависит от разных обстоятельств. Например, государственная служба может оказывать услуги населению напрямую, или предать право оказывать их частным поставщикам. Оно может финансировать оказание услуги напрямую или через систему страхования, причем последняя может быть государственной или частной. Сегодня основным

критерием выбора инструмента является его экономичность. Иными словами, государство принимает решение о том, кто и как будет предоставить услугу, исходя из соображений экономии своих затрат. При этом часто предполагается, что частный сектор ведет свой бизнес более эффективно, чем государственный. Последнее верно лишь отчасти. Как было показано выше, в современных условиях выбор должен делаться на основе комплексного решения, учитывающего множество факторов.

Устойчивое развитие

Понятие «устойчивое развитие» исторически возникло применительно к отношениям общества и природной среды. В настоящее время оно применяется для характеристики различных сторон жизни общества: экономики, экологии и социальных институтов. В самом общем виде, под устойчивостью экономической системы понимается ее способность к продолжительному осуществлению своей деятельности [60]. В узком смысле его можно определить, как развитие, позволяющее на долговременной основе обеспечить стабильный экономический рост, не приводящий к деградации природной среды, что гарантирует удовлетворение потребностей не только настоящего, но и будущих поколений [61], [62]. Целью является улучшение экономических, социальных и культурных условий жизни общества, интеграция отдельных сообществ в общий поток жизни нации, помочь им вносить максимальный вклад в национальное развитие [63]. С позиции математической теории управления, «устойчивое развитие» следует понимать, как способность системы функционировать в состояниях, по меньшей мере, близких к равновесию, в условиях постоянных внешних и внутренних возмущающих воздействий [64].

В нашем понимании, последнее определение не противоречит первым, поскольку государство можно рассматривать как социотехническую систему, стабильность которой (гарантия от смены социального строя) зависит от сбалансированности всех трех вышеуказанных факторов

[65]. Таким образом, гармоничное или сбалансированное развитие, при котором институциональные изменения и ориентация научно-технического прогресса, обеспечивающие рост экономики, который приводит к повышенной эксплуатации природных ресурсов, а также развитие социальных институтов, согласованы друг с другом [66]. Происходящие в государстве трансформации связаны с радикальными изменениями в технологии, которые влекут за собой изменения в экономике, государственной политике, экологии [67]. Возвращаясь к вопросу выбора инструментов осуществления политики устойчивого развития, отметим, что государство должно принимать решение с учетом необходимости обеспечить согласование трех аспектов: экономического, экологического и социального.

Очевидно, что для осуществления многофакторного анализа, обеспечивающего устойчивое развитие государства, необходимо уметь накапливать, обрабатывать и распространять большие объемы данных. Теперь нам предстоит выяснить готовность современной ИКТ осуществлять подобную деятельность.

Большие и сложные данные

Термин «большие данные» приобрел широкую популярность, однако он не имеет строгого определения, поскольку слово «большой» субъективно – нельзя провести четкую границу, отделяющую «большое» от «малого». В качестве отличительных признаков «больших данных», наравне с размером, указывают слабую структуризацию и высокую скорость их поступления. Но эти признаки также не являются достаточными, чтобы эксплицировать данное понятие.

Под слабоструктурированными данными понимают такую форму их организации, при которой их структура не может быть задана заранее, или она изменяется во времени. Чтобы определить семантику слабо структурированных данных, используется разметка в виде тегов и других маркеров [68]. Вследствие отсутствия жесткой

структуры и невозможности табличного представления, обработка слабоструктурированных данных связана с трудностью, поскольку при осуществлении поиска выполняется значительно число дисковых операций, причем, на сегодняшний день, дисковая подсистема является самым медленным элементом компьютера. Однако вряд ли стоит вводить новое понятие «большие данные» в связи с технологическими ограничениями, возникающими при обработке информации. В противном случае, надо признать, что термин «большие данные» является проявлением агрессивного маркетинга по навязыванию возможным покупателям решений о покупке ИКТ.

Мы попытаемся дать экспликацию указанного понятия, для этого обратим внимание, что данные нам интересны не сами по себе, но только поскольку они отображают свойства или динамику какого-либо объекта наблюдения. В нашем случае, мы договоримся рассматривать государство как социотехническую систему. Рассмотрим понятия «большой» и «сложный», в контексте системного анализа.

Как считает У. Р. Эшби, физические размеры не являются критерием отнесения системы к классу больших [69]. Он полагает, что система является большой с точки зрения наблюдателя, возможности которого она превосходит в каком-то аспекте, поэтому ее невозможно исследовать иначе, как по частям [70]. По мнению Ю. А. Черняка, большая система может быть описана с одной точки наблюдения, с помощью единой модели, используя один языковый формализм, с использованием единого метода моделирования. Сложность состоит в том, что модель, описывающая структуру большой системы (по сути, схема данных), имеет размер, превышающий возможности наблюдателя анализировать ее. Мы будем полагать, что объем данных, генерируемых большой системой, потребует для обработки высокой мощности вычислительного устройства, что, по-видимому, является технической трудностью, но не является принципиальным ограничением. Сложной называют систему, которая строится для решения многоцелевой, многоаспектной задачи. Ю. И. Черняк называет сложной систему, для изучения кото-

рой необходимы несколько наблюдателей, каждый из которых описывает систему со своей точки зрения, в определенном аспекте. Для ее описания необходимы несколько частных моделей (проекций), каждая из которых имеет свой язык и описывает модель наблюдения с позиции определенного представления. Важно отметить, что проекции должны быть между собой согласованы. Можно предположить, что главной проблемой изучения сложной системы является описание связей между проекциями. В их отсутствие частные модели могут оказаться не когерентными. Таким образом, главным ограничением, в этом случае, является отсутствие связей, обеспечивающих согласование проекций.

Подводя итог, мы будем называть данные большими, если они получены в результате наблюдения большой системы, которая имеет структуру, описываемую единой моделью большого размера. Иными словами, структура данных известна, хотя и имеет внушительный размер. Обработка таких данных потребует большой «грубой» вычислительной силы, но не представляет принципиальных сложностей. Сложными мы будем называть данные, характеризующие свойства или поведения сложной системы, которая не может быть описана в рамках одной модели, но требует нескольких, взаимно-согласованных моделей. Ограничением, в данном случае, мы будем полагать отсутствие описаний, характеризующих связи между отдельными моделями.

Возвращаясь к главной задаче выявления возможностей ИКТ в задачах преобразования э-правительства, мы можем заметить, что основным ограничением является не столько размер данных, сколько их многофакторность. Каждый из главных аспектов: экономика, социальные институты, экология, характеризуется множеством показателей, причем связи между ними, по большей части, не определены явно. Можно говорить о мультимодельной спецификации социотехнической системы. Как было показано выше, мерой сложности, в интересующем нас случае, выступает количество моделей и качество описаний связей между отдельными мо-

делями. Поэтому, необходимо оценить модели, лежащие в основе анализа сложных данных.

Наборы и комплекты моделей в задачах анализа данных

Критерием отнесения данных к категории сложных, мы будем считать количество моделей, необходимых для описания социотехнической системы, и качество связей между отдельными моделями. От того, насколько качественно проработаны модели и их связи, зависит успех анализа многофакторных данных. Очевидно, что частные модели, образующие мультимодельную спецификацию, должны быть согласованны и тщательно интегрированы между собой, в противном случае, они будут содержать противоречивую информацию.

Б. Тальхайм ввел понятие комплекта моделей (*model suite*) с согласованным дизайном и общими данными. Его можно представить как совокупность частных моделей, описывающих один объект моделирования с разных точек зрения. При этом, каждая из частных моделей описывает отдельные аспекты изучаемого объекта, а все вместе они дают полное для целей проектирования представление об изучаемом объекте [71]. По его мнению, комплект моделей должен содержать явные ассоциации между перспективами, включать контроллеры для обеспечения согласованности частных моделей, вводить схемы приложений для их явного обслуживания и развития, а также трассировщики для установления их согласованности. Взаимозависимости между частными моделями должны быть указаны в явном виде. Можно утверждать, что комплект моделей необходим и достаточен для разработки ИКТ. Напротив, набор моделей содержит частные модели, описывающие отдельные аспекты, поэтому требование достаточности не соблюдается.

Поэтому предлагается различать наборы моделей и комплекты [72]. Если совокупность частных моделей устроена так, что каждая отражает определенный набор отно-

шений между предметами одного, общего для всех моделей домена, ее можно называть комплектом. В случае, когда частные модели отражают отношения между предметами разных доменов, эти модели образуют коллекцию. Если, например, две модели отображают разные аспекты изучаемой системы, они могут считаться автономными и разрабатываться независимо друг от друга. В противном случае, если отображаемые аспекты этих моделей частично пересекаются, модели полагаются зависимыми, должны разрабатываться согласованно, дополнительно потребуются адаптеры, обеспечивающие когерентность данных в этих моделях.

Следует отметить, что сегодня э-правительство накапливает и обрабатываются многочисленные, но слабо связанные информационные потоки, отражающие отдельные аспекты деятельности государства, что определяет сложность их обработки и интерпретации. Во многих случаях, разные отчеты

содержат одинаковые показатели, которые могут именоваться различным способом, что затрудняет их интерпретацию. Можно заключить, что анализ данных, характеризующих устойчивость развития государства, связан с обработкой сложных данных, образованных набором слабо связанных моделей. Это определяет актуальную задачу преобразования частных моделей в комплект моделей.

Заключение

Проведенный анализ инструментов сбора и обработки данных уровня государства, позволил сделать следующие выводы. Доминирующая сегодня парадигма «правительство как сфера услуг» провела разработчиков э-правительства по ложному пути, поэтому не следует планировать ИКТ проекты для государственного сектора под давлением со стороны лоббистов ИКТ, моды и желания следовать примеру других стран, чтобы поэкспериментировать с новыми технологиями. Курс на создание э-правительства должен быть скорректирован, с учетом новых подходов реализации инструментов э-правительства. Мы полагаем, что следующий

шаг в развитии электронного правительства заключается не в переходе к следующему технологическому укладу, будь то большие данные, искусственный интеллект, интернет вещей или правительство как платформа, а в изменении набора политических инструментов, обеспечивающих всеобъемлющие политические цели. Следовательно, пересмотр инструментов политики с учетом потенциала цифровых технологий может привести к гораздо более фундаментальным и эффективным результатам. Важно отметить, что набор инструментов политики является инвариантом, относительно возможных преобразований э-правительства. При этом выбор инструмента реализации политики является прерогативой данного правительства и зависит от способности государства собирать, обрабатывать и распространять большие объемы сложных данных.

Государство можно рассматривать как социотехническую систему, стабильность которой (гарантия от смены социального строя) зависит от сбалансированности трех факторов: экономики, экологии и социальной стабильности. Происходящие в государстве институциональные изменения и ориентация научно-технического прогресса, обеспечивающие рост экономики, который приводит к повышенной эксплуатации природных ресурсов, а также развитие социальных институтов, согласованы друг с другом. Следовательно выбор инструментов осуществления политики устойчивого государства, осуществляется с учетом необходимости обеспечить согласование трех вышеуказанных аспектов. Эта задача связана с необходимостью они обрабатывать многочисленные, но слабо связанные информационные потоки, отражающие отдельные аспекты деятельности, что определяет сложность их обработки и интерпретации. Принципиальной проблемой является сложность данных, их многофакторный анализ. В качестве меры сложности данных предлагается считать количество и связность моделей, отображающих различные аспекты устройства и деятельности социотехнической системы.

Таким образом, можно сделать вывод о существовании проблемной ситуации — для выбора инструментов политики, направленной на устойчивое развитие, необходимо со-

бирать и правильно интерпретировать полные и точные данные из распределенных, разнородных источников, однако существующие методы интерпретации собираемых данных не позволяют автоматически сопоставлять информацию из разных источников. Выявленные нами ограничения, не позволяют надеяться на решение этой проблемы только путем повышения мощности используемых компьютеров. Существует объективная потребность в разработке новых подходов, методов и средств обработки сложных данных в распределенной высокопроизводительной вычислительной среде для цифровой трансформации задач планирования устойчивого регионального развития.

Задачей дальнейшего исследования станет согласование отдельных моделей, образующих мультимодельную спецификацию социотехнической системы. Предполагается разработать новые подходы к обеспечению интероперабельности информационных систем на семантическом уровне, позволяющем установить соответствие между смысловым содержанием данных в разных информационных системах. Без решения этой задачи анализ больших данных будет осуществляться в ручном режиме, что требует значительных ресурсов, делает невозможным принятие решения в реальном масштабе времени.

Работа выполнена при поддержке РФФИ (проект № 19-57-60004).

Литература:

1. Холмс Д. e.Gov. Стратегии электронного бизнеса для государства. М.: АСТ, 2004. 368 с.
2. De Bri F., Bannister F. E-government stage models: A contextual critique // Proceedings of the Annual Hawaii International Conference on System Sciences. Hawaii. 2015, с. 2222-2231.
3. Nogrask J., Vintar M. E-government and organisational transformation of government: Black box revisited? // Government Information Quarterly, 31, 108-118, Vol. 31, № 1, 2014, с. 108-118.
4. Словарь русского языка. В 4-х т. / РАН, Ин-т лингвистич. исследований; под ред. А. П. Евгеньевой. М.: Рус. яз.; Полиграфресурсы, 1999; (электронная версия: Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; под ред. А. П. Евгеньевой. 4-е изд., стер. М.:

Рус. яз.); Полиграфресурсы, 1999; (электронная версия, 1999. <http://feb-web.ru/feb/mas/mas-abc/19/ma439917.htm?cmd=0&istext=1>).

5. Трахтенберг А. Д. Переход к электронному правительству как инновация сверху // Digital Russia. 2012. URL: <http://d-russia.ru/perexod-k-elektronnomu-pravitelstvu-kak-innovaciya-sverxu.html> (дата обращения: 09.10.2020).

6. Трахтенберг А. Д. Электронное правительство: Состоится ли «изобретение государства заново»? // Научный ежегодник Института философии и права Уральского отделения Российской академии наук, № 12, 2012, с. 285–297.

7. Lee J. Search for stage theory in e-government development // Innovation and the Public Sector, Vol. 13, 2007, с. 33–48.

8. Siau K., Long Y. Synthesizing e. government stage models – a meta.synthesis based on meta. ethnography approach // Industrial Management & Data Systems, Vol. 105, № 4, 2005, с. 443–458.

9. Zwahr T., Finger M., Muller P. More than Digitisation – The Transformative Potential of E-Governance: An Exploratory Case Study // Proceedings of the 38th Hawaii International Conference on System Sciences. 2005.

10. Ушаков Н. Д., editor. Преобразование. 19393rd ed. // In: Толковый словарь русского языка. М.: Государственное издательство иностранных и национальных словарей, 1939.

11. Transformational government: enabled by technology, Cabinet Office, London, 2005. 28 с.

12. Bannister F., Connolly R. Forward to the Past: lessons for the future of e-government from the story so far // Innovation and the Public Sector, Vol. 17, № 3, 4, 2013, с. 211–226.

13. Yildiz M., Saylam A. E-government discourses: An inductive analysis // Government Information Quarterly, Vol. 30, № 2, 2013, с. 141–153.

14. Parisopoulos K., Tambouris E., Tarabanis K. An investigation of national policies on transformational government (t-Gov) in Europe // International Journal of Information Technology and Management, Vol. 13, № 4, 2014, с. 305–323.

15. Weerakkody V., Janssen M., and Dwivedi Y. K. Transformational Change and Business Process Reengineering (BPR): Lessons from the British and Dutch Public Sector // Government Information Quarterly, Vol. 28, № 3, 2011, с. 320–328.

16. Klievink B., Janssen M. Realizing joined-up government – Dynamic capabilities and stage models for transformation // Government Information Quarterly, Vol. 26, № 2, 2009, с. 275–284.

17. Irani Z., Elliman T., Jackson P. Electronic transformation of government in the U. K. : A research agenda // European Journal of Information Systems, Vol. 16, № 4, 2007, с. 327–335.

18. Layne K., Lee J. Developing fully functional e-government: a four stage model // *Government Information Quarterly*, Vol. 18, № 2, 2001, с. 122–136.
19. Waller P., Weerakkody V. *Digital Government: overcoming the systemic failure of transformation*, Brunel University, London, Working Paper 2018. 36 с.
20. Трахтенберг А. Д. Идеологический концепт электронного правительства: как работает риторика разрыва? // *Научный ежегодник Ин-та философии и права Урал. отделения РАН*, Vol. 17, № 2, 2017. С. 41–58.
21. O'Neill R. The Transformative Impact Of E-Government On Public Governance In New Zealand // *Public Management Review*, Vol. 11, № 6, с. 751–770.
22. Beynon-Davies P., Martin S. Electronic local government and the modernisation agenda: Progress and prospects for public service improvement // Beynon-Davies, P., and Martin, S. (2004), *Electronic local government and the modernisation agenda: Progress and prospects for public service improvement*. *Local government studies*, 30 (2), 214–229, Vol. 30, № 2, 2004, с. 214–229.
23. Tan C., Pan S. Managing e-transformation in the public sector: an e-government study of the Inland Revenue Authority of Singapore (IRAS) // *European Journal of Information Systems*, Vol. 12, № 4, 2003, с. 269–281.
24. Bannister F. Plus Ca Change? ICT and structural change in government // *Innovation and the Public Sector*, Vol. 19, 2012, с. 133–142.
25. Bekkers V., Homburg V. The Myths of E-Government: Looking Beyond the Assumptions of a New and Better Government // *The Information Society*, Vol. 23, № 5, 2007, с. 373–382.
26. Yildiz M. Big Questions of e-government research // *Information Polity*, Vol. 17, 2012, с. 343–355.
27. Fishenden J. Of lipstick and pigs in government // CIO. 2015. URL: <http://mobile.cio.co.uk/blogs/political-debate/of-lipstick-pigs-in-government-362448>
28. Lee J., Kim H., Ahn M. The willingness of e-Government service adoption by business users: The role of offline service quality and trust in technology // *Government Information Quarterly*, Vol. 28, № 2, 2011, с. 222–230.
29. Hu P., Brown S., Thong J., Chan F., Tam K. Determinants of Service Quality and Continuance Intention of Online Services: The Case of eTax // *Journal of the American Society for Information Science and Technology*, 60(2), 292–3, Vol. 60, № 2, 2009, с. 292–306.
30. Carter L., Weerakkody V. E-Government Adoption: A Cultural Comparison // *Information Systems Frontiers*, Vol. 10, № 4, 2008, с. 473–482.

31. Carter L., Belanger F. The utilization of e-government services: Citizen trust, innovation and acceptance factors // *Information Systems Journal*, Vol. 15, № 1, 2005se, c. 15–25.
32. Heeks R., Bailur S. Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice // *Government Information Quarterly*, Vol. 24, № 2, 2007, c. 243–265.
33. Bolivar M., Munoz L., Hernandez A. Studying E-Government: Research Methodologies, Data Compilation Techniques and Future Outlook // *Academia, Revista Latinoamericana de Administracion*, Vol. 51, № 25, 2012, c. 79–95.
34. El-Haddadeh R., Weerakkody V., Al-Shafi S. Understanding the Challenges and Complexities Influencing The Implementation and Institutionalisation of Electronic Services in the Public Sector // *Journal of Information and Management*, Vol. 50, № 4, c. 135–143.
35. Kamal M., Weerakkody V., Irani I. Analysing the role of stakeholders in the adoption of technology integration solutions in UK local government: an exploratory study // *Government Information Quarterly*, Vol. 28, № 2, 2011, c. 200–210.
36. Al-Shafi S., Weerakkody V. Adoption and Diffusion of free Wireless Internet Parks in Qatar // *International Journal of Value Chain Management*, Vol. 4, № 1, 2, 2010, c. 68–85.
37. Gil-Garcia R., Chun S., Janssen M. Government Information Sharing and Integration: Combining the social and technical // *Information Policy*, 14(1-2), 1–10, Vol. 14, № 1, 2, 2009, c. 1–10.
38. Janssen M., Tan Y. Dynamic Capabilities for Information Sharing: XBRL enabling business-to-government information exchange // *47th Hawaii International Conference on System Science (HICSS2014)*. 2014. Vol. IEEE Computer Society, c. 2104–2113.
39. Sharif A., Irani Z., Weerakkody V. Sharif, A., Irani, Z. and Weerakkody, V. (2010), Evaluating and Modelling Constructs for Decision making in E-Government // *Journal of Operational Research Society*, Vol. 61, № 1, 2010, c. 929–952.
40. Seifert J. W., Chung J. Using E-Government to Reinforce Government – Citizen Relationships: Comparing Government Reform in the United States and China // *Social Science Computer Review*, Vol. 27, № 1, 2009, c. 3–23.
41. Weerakkody V., Irani Z., Lee H., Osman I., Hindi N. E-government implementation: A bird’s eye view of issues relating to costs, opportunities, benefits and risks // *Information Systems frontiers*, Vol. 17, № 4, 2015, c. 889–915.
42. Hood C. Intellectual Obsolescence and Intellectual Makeovers: Reflections on the Tools of Government after Two Decades. *Governance // An International Journal of Policy, Administration, and Institutions*, Vol. 20, № 1, 2007, c. 127–144.

43. Bannister F., Connolly R. The great theory hunt: Does e-government really have a problem? // *Government Information Quarterly*, Vol. 32, № 1, 2015, c. 1-11.

44. Bannister F., Connolly R. ICT, public values and transformative government: A framework and programme for research // *Government Information Quarterly*, Vol. 31, № 1, 2014, c. 119-128.

45. Osman I., Anouze A., Irani Z., Al-Ayoubi B., Lee H., Balci A., Medeni T., Weerakkody V. COBRA framework to evaluate e-government services: A citizen-centric perspective // *Government Information Quarterly*, Vol. 31(1), № 1, 2014, c. 243-256.

46. Irani Z., Weerakkody V., Kamal M., Hindi N., Osman I., Anouze A., El-Haddadeh R., Lee H., Osmani M., Al-Ayoubi B. Irani, Z., Weerakkody, V., Kamal, M., Hindi, N. M., Osman, I. H., Anouze, A., El-Haddadeh, R., Lee, H., Osmani, M. and Al-Ayoubi, B. (2012), An Analysis of Methodologies Utilised in e-Government Research: A User Satisfaction Perspective // *Journal of Enterprise Information Management*, Vol. 25, № 3, 2012, c. 298-313.

47. Weerakkody V., Irani Z., Lee H., Hindi N., Osman I. A Review of the Factors Affecting User Satisfaction in Electronic Government Services // *International Journal of Electronic Government Research*, Vol. 10, № 4, 2014, c. 21-56.

48. United Nations. From E-government to E-inclusion, Department of Economic and Social Affairs Division for Public Administration and Development Management, UN Global E-government Readiness Report 2005.

49. Bannister F. The curse of the benchmark: an assessment of the validity and value of e-government comparisons // *International Review of Administrative Sciences*, 73, 171-188, Vol. 73, 2007, c. 171-188.

50. Петрунин Ю. Ю., Кириленко Г. Г., Кудина М. В., Логунова Л. Б. «КДУ», «Добросвет», 2018., 2018.

51. Howlett M. Beyond good and evil in policy implementation: Instrument mixes, implementation styles and second generation theories of policy instrument choice // *Policy & Society*, 23(2), 1-17, Vol. 23, № 2, 2004, c. 1-17.

52. Howlett M. Governance modes, policy regimes and operational plans: A multi-level nested model of policy instrument choice and policy design // *Policy Sciences*, Vol. 42, № 1, 2009, c. 73-89.

53. Anderson C. *Statecraft: An Introduction to Political Choice and Judgement*. New York: John Wiley and Sons, 1977. 318 c.

54. Schneider A., Ingram H. Behavioral Assumptions of Policy Tools // *Journal of Politics*, Vol. 52, № 2, 1990, c. 510-529.

55. Howlett M. From the 'old' to the 'new' policy design: design thinking beyond markets and collaborative governance // *Policy Sciences*, Vol. 47, № 3, 2014, c. 187-207.

56. Lascoumes P., Le Gales P. Introduction: Understanding Public Policy through Its Instruments – From the Nature of Instruments to the Sociology of Public Policy Instrumentation // *Governance*, Vol. 20, 2007, с. 1–22.
57. Open Working Group proposal for Sustainable Development Goals // United Nations. 2015. URL: <https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=1579&menu=1300> (дата обращения: 10.5.2019).
58. Hood C., Margetts H. *The Tools of Government in the Digital Age*. Basingstoke: Palgrave Macmillan, 2007. 231 с.
59. Salamon L. *The tools of government: A guide to the new governance*. New York: Oxford University Press, 2002. 688 с.
60. Бараненко С. П., Шеметов В. В. *Стратегическая устойчивость предприятия*. М.: Центрполиграф, 2004. 496 с.
61. Основные решения, принятые на конференции ООН по окружающей среде // ООН. 1992. URL: https://www.un.org/ru/documents/decl_conv/declarations/riodecl.shtml (дата обращения: 02.05.2019).
62. Kemp R., Loorbach D. *Governance for Sustainability Through Transition Management* // EAERE (European Association for Evolutionary Political Economy) Conference. Maastricht. 2003, с. 1–27.
63. О Концепции перехода Российской Федерации к устойчивому развитию : указ Президента РФ от 1 апр. 1996 г. № 440 // *Собрание законодательства РФ*, Vol. 15, 1996, с. 1572.
64. Алиева В. Г., editor. *Теория организации*. учеб. для вузов / Минобр. РФ ed. М.: Экономика, 2003. 431 с.
65. Hesketh B., Graco W. *Sociotechnical System, Aclied Psychology of* // In: *International Encyclopedia of the Social & Behavioral Sciences*. 2015. P. 23185.
66. Угольницкий Г. А. *Устойчивое развитие организаций. Системный анализ, математические модели и информационные технологии*. М.: Издательство физико-математической литературы, 2011, 320 с.
67. Geels F. From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory // *Research Policy*, Vol. 33, № 6–7, 2004, с. 897–920.
68. Слабоструктурированные данные // Национальная библиотека им. Н. Э. Баумана. 2017. URL: https://ru.bmstu.wiki/Слабоструктурированные_данные (дата обращения: 15.10.2019).
69. Эшби У. Р. *Введение в кибернетику*. М. : КомКнига, 2006. 432 с.
70. Черняк Ю. И. *Системный анализ в управлении экономикой*. М.: Экономика, 1975, 191 с.

71. Thalheim B. The conceptual framework to multi-layered database modelling // *Frontiers in Artificial Intelligence and Applications*, EJC, Proc. Maribor, Slovenia. 2010. Vol. 206, c. 118–138.

72. Fiodorov I., Sotnikov A. Model Suite and Model Set in Software Development 2018. Vol. *New Trends in Databases and Information Systems*, c. 243–252.

О некоторых тенденциях развития глобальной цифровой среды

О. В. Сунтюренко, Е. Ю. Дмитриева

ВИНИТИ РАН

Аннотация. С системных позиций рассмотрены основные тенденции развития цифровых сетевых технологий. Рассмотрены специфика и перспективы реализации нового подхода в области сетевых технологий – развитие программно-конфигурируемых сетей. Представлены основы подхода к разработке технологии смысловой навигации и поиска знаний в информационных сетях. Показано, что на основе новых сетевых технологий перспективным системным инструментом стало формирование нового направления наукометрии и библиометрии – вебометрии.

Ключевые слова: цифровая среда, сетевые технологии, информационные ресурсы, программно-конфигурируемые сети, конвергенция технологий, семантический интернет, классификационные системы, вебометрия.

Global digital environment: some trends in developing of digital network technologies

Suintuirenko O. V., Dmitrieva E. Yu.

VINITI RAS

Abstract. The main trends of digital network technologies development are considered from the system positions. The specifics and prospects of implementation of the new approach in the field of network technologies – development of software-configurable networks are considered. The basics of approach to development of technology of semantic navigation and search of knowledge in information networks are presented. It was shown that on the basis of new network technologies, the formation of a new direction of scientometrics and webometrics became a promising system tool.

Keywords: digital environment, network technologies, information resources, software-configurable networks, technology convergence, semantic Internet, classification systems, webometrics.

Современная цифровая среда постепенно становится доминирующей частью мирового информационного пространства. Базовым компонентом макроструктуры глобальной цифровой среды являются системы и сети телекоммуникаций, прежде всего Интернет: давно сложившийся и самый большой сегмент сети web 1; сегмент социальных сетей и платформ web 2; растущий в последние годы наиболее быстрыми темпами web 3 — сегмент мобильных приложений [1].

Структурная основа развивающихся сетевых технологий — это согласованный набор программных и аппаратных средств, а также механизмов передачи данных по линиям связи, достаточный для построения вычислительной сети.

Устойчивыми, и во многом определяющими, тенденциями развития сетевых технологий и современного цифрового информационного пространства являются:

А. Активная конвергенция информационных, традиционных библиотечных, компьютерных и телекоммуникационных технологий. Самоорганизация (с позиций адаптивности структуры и функциональных ролей участников) глобальной сетевой институциональной среды.

В. Интеграция разнородных информационных ресурсов (ИР) в гетерогенной цифровой среде. Применение единых стандартов метаданных и интероперабельных программных средств. Преимущественное использование унифицированных и отчуждаемых программных и технологических решений, современных сервисно-ориентированных архитектур, специализированных систем навигации к проблемно-ориентированым ИР.

С. Реализация нового подхода в области сетевых технологий — развитие программно-конфигурируемых сетей (ПКС). Система управления компьютерными сетями на основе стека протоколов TCP/IP, является устаревшей и обладающей низкой эффективностью при передаче гетерогенного трафика. В основе новой технологии — положение о разделении функции управления и передачи данных. В сетях новой архитектуры ПКС все маршрутизаторы

и коммутаторы объединяются под управлением Сетевой Операционной Системы (СОС), которая обеспечивает приложениям доступ к управлению сетью и постоянно отслеживает конфигурацию средств сети. В отличие от традиционного толкования термина СОС как операционной системы, интегрированной со стеком сетевых протоколов, в данном случае под СОС понимается программная система, обеспечивающая мониторинг, доступ, управление ресурсами всей сети, а не конкретного узла. СОС формирует данные о состоянии всех ресурсов сети и обеспечивает доступ к ним для приложений управления сетью. Эти приложения управляют разными аспектами функционирования сети, типа построения топологии, принятия маршрутизирующих решений, балансировки нагрузки и т. п.

Д. Разработка механизма (технологии) смысловой навигации и поиска знаний в информационных сетях. В настоящее время теория научно-технической информации не располагает методами индустриальной интеграции знаний, представленных в разнородных источниках. При поддержке Российского фонда фундаментальных исследований (проект № 20-07-00103) в ВИНТИ РАН ведутся работы в этом направлении на основе как интеллектуальных методов, так и автоматических методов анализа содержания классификационных систем и их соотношений. Созданы алгоритмы и программный комплекс навигации, поиска и сбора информации на основе связей, зафиксированных в онтологии научного и технического знания [2].

Е. Устойчивый тренд увеличения скорости передачи данных и пропускной способности. На сегодняшний день скорость передачи данных в суперкомпьютерных сетях – 50 Гбит/сек.

Ф. Развитие концепции «семантического интернета», в котором информации придается точно определенный смысл, что позволяет компьютерам «понимать» и обрабатывать ее на семантическом уровне (сегодня компьютеры работают на синтаксическом уровне, на уровне знаков).

Г. Разработка и внедрение новых типов электронных информационных ресурсов, сервисов и медиа. Использование специализированных ИР, например, Российского индекса научного цитирования, в системе управления наукой и образованием (для оценки результатов научной деятельности, поддержки принятия решений).

Н. Переход в цифровую форму информационно-коммуникационных каналов в сфере науки, образования и культуры, их интеграция в сети Интернет.

И. Диверсификация форм и способов организации электронной коммуникации в Интернет:

- сайты библиотек, музеев, научных и образовательных учреждений;
- электронные библиотеки, архивы, репозитории и депозитарии;
- библиографические, реферативные и фактографические БД, электронные каталоги различных видов ресурсов;
- электронные книжные магазины и издательства, электронные энциклопедии;
- социальные сети, научно-образовательные сайты научных сообществ [3].

Стремительное расширение цифровой информационной среды, развитие сетевых технологий актуализируют проблему поиска новых концептуальных подходов к структуризации информационного пространства в целях оптимизации использования информационных ресурсов, и в первую очередь в информационно-библиотечной сфере.

На основе новых сетевых технологий перспективным системным инструментом стало формирование нового направления наукометрии и библиометрии – *вебометрии*. Для сохранения библиотечной системы как институционального компонента информационной инфраструктуры вебометрическая система должна стать современным эффективным инструментом развития библиотек в цифровой среде.

Основными целями и задачами создания вебометрической системы библиотек являются:

- повышение роли и значимости публичных и научных библиотек в обществе;
- сохранение и развитие функциональной деятельности библиотек (в зависимости от их типа и вида), поддержание позитивного имиджа в мировом web-пространстве;
- совершенствование (опосредованно) состава и структуры фондов, оптимизация комплектования библиотек;
- интенсификация процессов цифровизации фондов библиотек;
- стимулирование процессов диверсификации библиотечных услуг и продуктов в цифровой среде;
- социологический мониторинг культурного и образовательного предпочтения россиян;
- формирование интегральной оценки уровня и рейтингового распределения библиотек;
- мониторинг и поддержка принятия управленческих решений.

Аналитическая постобработка вебметрической информации обеспечит выявление слабых и сильных сторон сайтов научных библиотек и анализ пользовательской аудитории.

Следует отметить, что широкое применение в структуре Государственной системы НТИ цифровых информационных ресурсов, новых информационных технологий содействует более эффективному решению задач информационного обеспечения фундаментальных и прикладных исследований, инновационной деятельности, что обеспечивается рациональным, сбалансированным развитием информационной инфраструктуры, информационных ресурсов, информационных технологий. Комплекс задач по более глубокой переработке информации, извлечения новых знаний является частью общих проблем информационного обеспечения научных исследований и разработок и имеет определенное экономическое измерение.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект № 20-07-00103.

Литература

1. Сюнтюренко О. В. Факторы-детерминанты неэффективного использования информационных ресурсов в научно-технической деятельности // Научно-техническая информация. Сер. 1. 2017. № 7. С. 1-12.
2. Syuntyurenko O. V. Determinants of the Ineffective Use of Information Resources in Scientific and Technological Activities // Scientific and Technical Information Processing. 2017. Vol. 44, № 3. P. 159-169.
3. Дмитриева Е. Ю., Сюнтюренко О. В. Актуальные задачи диверсификации технологий, информационных продуктов и услуг // Научно-техническая информация. Сер. 2. 2020. № 11. С. 1-9.
4. Сюнтюренко О. В. Перспективы использования интернет-СМИ, журналов открытого доступа и социальных медиа в научно-технической сфере // Научно-техническая информация. Сер. 1. 2015. № 6. С. 30-36.
5. Syuntyurenko O. V. Prospects for Using Online Media, Open-Access Journals, And Social Media Networks in the Field of Science and Technology // Scientific and Technical Information Processing. 2015. Vol. 42, № 2. P. 112-118.

References

1. Syuntyurenko O. V. Determinants of the Ineffective Use of Information Resources in Scientific and Technological Activities // Scientific and Technical Information Processing. 2017. Vol. 44, № 3. P. 159-169.
2. Dmitrieva E. Yu. Aktualnye zadachi diversifikatsii tekhnologii, informatsionnykh productov I uslug // Nauchno-technicheskaya informatsia. Ser. 2. 2020. N 11. S. 1-9.
3. Syuntyurenko O. V. Perspektivy ispolsovaniya internet SMI, jurnalov otkrytogo dostupa I sotsialnykh media v nauchno-technicheskoi sfere // Nauchno-technicheskaya informatsia. Ser. 1. 2015. N 6. S. 30-36
4. Syuntyurenko O. V. Prospects for Using Online Media, Open-Access Journals, And Social Media Networks in the Field of Science and Technology // Scientific and Technical Information Processing. 2015. Vol. 42, № 2. P. 112-118.
5. Syuntyurenko O. V. Faktory-determinanty neeffektivnogo ispolsovanya informatsionnykh resursov v nauchno-technicheskoi deyatelnosti // Nauchno-technicheskaya informatsia. Ser. 1. 2017. № 7. S. 1-12.

Институты социальной памяти и проблемы перехода к цифровому обществу

Д. В. Ефременко

ИНИОН РАН

Аннотация. В статье рассматриваются проблемы трансформации институтов социальной памяти в условиях перехода к цифровому обществу. В ходе трансформации выявляется разнонаправленная динамика, связанная с изменением технологической основы, модели функционирования и социальными функциями библиотек, архивов и музеев. Наряду с усилением роли части институтов социальной памяти как площадок коммуникации и публичного диалога, их интеграция в Единое цифровое пространство знаний ведет к расширению функциональных возможностей и повышению статуса институтов социальной памяти в цифровом обществе. При этом может возрасти влияние научных библиотек и центров научно-технической информации на процессы производства знания, его экспертной оценки, а также рефлексию и контроль динамики формирования цифрового общества.

Ключевые слова: институты социальной памяти, цифровое общество, публичные библиотеки, научные библиотеки, центры научно-технической информации, пространство знаний.

Institutions of Social Memory and the Problems of Transition to Digital Society

D. V. Efremenko

ISIS RAS

Abstract. The article discusses the problems of transformation of social memory institutions in the context of the transition to a digital society. During the transformation, multidirectional dynamics are revealed, associated with a change in the technological basis, functioning models and social functions of libraries, archives and museums. Along with the

strengthening of the role of some of the institutions of social memory as platforms for communication and public dialogue, their integration into the Unified Digital Knowledge Space leads to the expansion of functionality and to the upgrading of the status of social memory institutions in digital society. At the same time, the influence of scientific libraries and centers of scientific and technical information on the processes of knowledge production, its expert evaluation, as well as reflection and control of the dynamics of the formation of digital society can increase.

Keywords: institutions of social memory, digital society, public libraries, scientific libraries, centers of scientific and technical information, knowledge space.

Важнейшие институты социальной памяти — библиотеки, архивы, музеи — в эпоху тотальной цифровизации сталкиваются с фундаментальными вызовами, ответы на которые в настоящий момент далеко не очевидны, хотя общая социокультурная динамика делает неизбежным поиск этих ответов, а сами ответы, т. е. выбор определенного направления трансформации, чреватые глубокими и — возможно — необратимыми последствиями. Так, библиотеки оказываются сегодня на своеобразном перепутье — смириться с постепенной утратой прежнего статуса хранилища знаний и культурного наследия либо найти новую нишу в сетевом социуме. Опыт развитых стран показывает, что библиотеки способны успешно следовать по второму пути. По результатам исследования, проведенного компанией Gallup в конце декабря 2019 г., посещение библиотек оказалось самым популярным видом культурного досуга в США (в среднем каждый американец посетил библиотеку 10,5 раз за год, а в кино сходил 5,3 раза). Но сама востребованность библиотек объясняется исследователями как результат уже частично реализованных трансформационных стратегий, предполагающих расширение набора бесплатных услуг и обеспечение доступа к различным сервисам, включая предоставление Wi-Fi [1]. Фактически речь идет о существенном расширении функциональных ролей публичных библиотек:

— библиотека как гражданское пространство;

- библиотека как инструмент обеспечения социальной грамотности;
- библиотека как публичный форум и катализатор общественных дискуссий;
- библиотека как гражданский информационный центр;
- библиотека как читальный зал для всего сообщества (от локального до национального);
- библиотека как партнер по общественным услугам;
- библиотека как сервисный учебный центр [2].

Заметим, однако, что идя по пути расширения функциональных ролей, публичные библиотеки вступают в исторически им несвойственную конкуренцию с коммерческими институтами, предоставляющими различные услуги и несущими при этом важную социальную нагрузку – кафе, книжные магазины и т. д.

Традиционно основной задачей публичных и научных библиотек было хранение информации на различных носителях и предоставление к ней общего доступа. Несомненно, однако, что библиотеки порождают множество различных эффектов, будучи, с одной стороны, вписанными в определенный социальный контекст, а с другой – оказывая влияние на него своей деятельностью. Эти эффекты и сама библиотека как социокультурный феномен могут быть рассмотрены в различных философских и социально-теоретических ракурсах. В частности, весьма перспективным является взгляд на библиотеку и другие институты социальной памяти как на социобиотехнические системы [3; 4] или – с позиций акторно-сетевого подхода – как на гибридные объекты, являющиеся культурными артефактами, в которых биофизические и культурные качества неразрывно переплетены [5]. Если же принять во внимание, что библиотеки, архивы, музеи – это в первую очередь феномены городской культуры, то еще больше возможностей открывает междисциплинарное изучение метаболических взаимодействий между социальными, природными, техническими и информационными компонентами современного

города. Привлекательность данного подхода заключается в рассмотрении сложных обменных процессов между городскими системами, переходом одних процессов и веществ в другие, во взаимовлиянии их друг на друга, а также в анализе и учете невидимых последствий таких переходов и трансформаций для разных групп городского населения [6].

Многие публичные библиотеки стремятся ответить на вызовы современности, изменяя принципы своей работы в сторону большей открытости и инклюзивности. Это можно было наблюдать и в Москве в начале 2010-х годов, когда при поддержке столичного департамента культуры во главе с С. Капковым был запущен крупномасштабный проект реструктуризации городских публичных библиотек. В настоящее время эта трансформация значительно замедлилась в связи с сокращением поддержки со стороны городских властей, но в целом понятно, что получен достаточно ценный опыт трансформации модели обслуживания, стратегий формирования фондов, предоставления дополнительных услуг и стилистики библиотечного пространства, позволяющей использовать его для различных видов социальной активности и взаимодействия [7].

Одна из развилок, возникающая на пути такого рода трансформаций, связана с тем, останется ли приоритетной традиционная роль библиотек как институтов хранения и передачи знания, или же библиотеки всё сильнее будут ориентироваться на создание модных, удобных и востребованных публичных пространств. С этой дилеммой, пусть и с некоторыми вариациями, сталкиваются и научные библиотеки. В научных библиотеках публичное пространство отличается определенной спецификой, в первую очередь — ориентацией на решение задач популяризации научных знаний и организацию диалога между учеными и представителями широкой общественности. В то же время научные библиотеки имеют больше возможностей стать значимыми акторами пространства цифровой коммуникации. Классическая модель комплектования и хранения фондов и обслуживания пользователей все в большей степени будет

комбинироваться с развитием электронных пользовательских сервисов, а также с качественным расширением функционала, когда научная библиотека сможет взять на себя и функции агрегатора цифрового контента [8]. Разумеется, решение задач такого уровня под силу далеко не каждой научной библиотеке; преимуществами будут обладать те из них, кто еще в «аналоговую» эпоху встал на путь трансформации в многофункциональные научно-информационные центры. В случае успешной реализации такой модели саму библиотеку можно будет рассматривать не только как социобиотехническую систему, но и как систему социокиберфизическую [9], обеспечивающую новый уровень интеграции цифровых ресурсов (в не слишком отдаленном будущем – искусственного интеллекта), физических и социальных сущностей.

На общегосударственном уровне задача системной организации и трансформации институтов социальной памяти поставлена в ряде официальных документов органов власти, в которых вводится понятие Единого российского электронного пространства знаний [10]. Вместе с тем в этих документах, принятых в период 2014–2019 гг., имеются разночтения, показывающие, что у центральных органов власти пока еще нет единого понимания целей, задач, масштабов, способов консолидации и развития российского электронного (цифрового) пространства знаний. Сохраняется ведомственная разобщенность, а также специфический подход к освоению средств, зачастую не позволяющий добиться системно значимых результатов.

Построение и развитие Единого электронного (цифрового) пространства знаний (ЕЦПЗ), вне всякого сомнения, является комплексной задачей, решение которой должно стать неотъемлемой частью национальных проектов и программ, реализация которых будет охватывать сферы науки, культуры, образования, здравоохранения, демографии, развития цифровой экономики, охраны окружающей среды и др., в которых фактор знания играет ключевую роль. Следует отдавать отчет в том, что сформулированные президентом РФ В. В. Путиным в послании Федеральному собра-

нию 15 января 2020 г. установки на создание собственных технологий и стандартов «по тем направлениям, которые определяют будущее» — искусственный интеллект, генетика, новые материалы, источники энергии, цифровые технологии [11], имеют стратегическое значение. Они, в частности, отражают качественно новую глобальную ситуацию, когда Китай принял решение о создании собственной, автономной от Запада техноэкономической платформы [12], а России предстоит выбор между присоединением к одной из платформ или формированием собственных экосистем развития ключевых технологий. Принимаемые в данном контексте политические решения, очевидно, потребуют коррекции одобренного ранее набора национальных проектов и программ, в т. ч. программ развития цифровой экономики.

В технологическом аспекте формирование Единого цифрового пространства знаний означает обеспечение через «единый вход» возможности поиска данных в различных информационных (прежде всего, документальных) системах. Тем не менее, интеграция институтов социальной памяти в цифровое пространство знаний — процесс более комплексный и многозначный, чем простое объединение в той или иной форме многообразных информационных ресурсов, обеспечивающих хранение существующего и производство нового знания. Прежде всего, принципиально важна укорененность институтов социальной памяти в культуре аналоговой, доцифровой эпохи. Понятно, однако, что создание 3D-моделей музейных экспонатов, мультимедийных инсталляций, электронных копий архивных документов, периодики и книг — это лишь первый шаг трансформационного перехода, предваряющий дальнейшее полуавтономное (сохраняющее родовую связь с аналоговой основой) существование и развитие институтов социальной памяти в цифровой среде. Как отмечает Т. В. Майстрович, некоторые виды и жанры «документов существуют в электронном пространстве знаний по признаку подобия, но все больше возрастает число тех, которые либо не свойственны аналоговой среде, либо имели в ней иной статус» [13, с. 217].

Таким образом, речь идет не только о системной интеграции массивов электронных документов и ресурсов, отражающих культурное и научное наследие, но, в конечном счете, о формировании качественно новой цифровой среды [14], которая оказывает активное влияние на социокультурную динамику и одновременно становится площадкой экспертного диалога и производства нового научного знания. Не менее важно обеспечить возможность обмена данными с российскими и зарубежными информационными системами благодаря реализации принципа семантической интероперабельности (использованию технологии Semantic Web и стандартизации наборов элементов соответствующих метаданных).

Разные компоненты единого цифрового пространства знаний будут сохранять многие специфические особенности, как традиционные, так и новоприобретенные. Такие институты социальной памяти как музеи и публичные библиотеки, очевидно, будут стремиться сохранить и даже расширить функции центров культурного и гражданского просвещения, площадок для общественных дебатов и коммуникации. Такого рода активность переносится и в онлайн-сферу, чему во многом способствовали социальное дистанцирование и длительное прекращение доступа посетителей в музеи, архивы и библиотеки в период пандемии 2020 г. Научные библиотеки, как уже было отмечено выше, также становятся акторами публичной сферы (онлайн и офлайн), хотя и со смещением акцентов на задачи научного просвещения и обсуждения актуальных научных проблем. В целом можно говорить о том, что построение ЕЦПЗ будет сопровождаться абсорбцией им функций площадки сетевой коммуникации, прежде всего, научной, но отчасти и общественной. Например, если будет успешно реализован проект создания некоего российского конкурента Википедии (эта работа уже начата на базе Большой российской энциклопедии при государственном финансировании в размере 2 млрд руб.) в рамках построения ЕЦПЗ, то сложно представить, что многие статьи по проблемам истории и современной политики не станут предметом общественного интереса и дискуссий.

Рассмотрение комплекса проблем формирования ЕЦПЗ в аспекте интеграции институтов социальной памяти в цифровое общество показывает, что эти институты и ЕЦПЗ в целом станут факторами дальнейших трансформаций. Более того, крупные научные библиотеки и – в особенности – центры научно-технической информации, относящиеся к институтам социальной памяти, но имеющие и серьезный исследовательский потенциал, могут претендовать на более активное участие в развитии цифрового общества. В частности, они могут участвовать в решении ряда актуальных задач, связанных с исследованием комплексных социально-политических и социокультурных эффектов развития искусственного интеллекта и цифровых технологий [15]. В конечном счете, можно вести речь о научном анализе проблем онтологической безопасности [16] на этапе цифрового перехода, характеризующих устойчивость взаимодействий внутри социобиотехнических систем и способствующих поддержанию чувства защищенности людей и сообществ, участвующих в этих взаимодействиях. Организация такой работы предполагает наличие солидного опыта междисциплинарных исследований, поскольку речь идет об анализе широкого спектра социальных взаимодействий, в которые вовлекаются и несоциальные объекты (природные и технические), а «катализатором» изменений выступают технологии связи 5G, интернет вещей, искусственный интеллект. Также необходимы устойчивые навыки анализа гетерогенных по составу массивов научно-технической и общественно-политической информации, способность к гибкому реагированию на изменения, готовность к работе в разных форматах, включая презентацию результатов исследований в публичной сфере.

В целом процессы цифрового перехода характеризуются разнонаправленной сопутствующей динамикой, связанной с изменением технологической основы, модели функционирования и социальными функциями библиотек, архивов и музеев. Наряду с усилением роли части институтов социальной памяти как площадок коммуникации

и публичного диалога, их интеграция в Единое цифровое пространство знаний ведет к расширению функциональных возможностей и повышению статуса институтов социальной памяти в цифровом обществе. При этом может возрасти влияние научных библиотек и центров научно-технической информации на процессы производства знания, его экспертной оценки, а также на рефлексию динамики формирования цифрового общества. В общем массиве институтов социальной памяти эти компоненты и их характеристики доминирующими не будут, но если ЕЦПЗ превратится в функционирующую социоклиберфизическую систему, то благодаря наличию такого рода компонентов единое пространство знаний имеет шансы стать системой со встроенными в нее механизмами самоконтроля.

Работа выполнена в рамках проекта РФФИ / КОМФИ № 18-00-00298.

Список литературы

1. Mccarthy J. In U.S., Library Visits Outpaced Trips to Movies in 2019 // Gallup. 2020. January 24. URL: <https://news.gallup.com/poll/284009/library-visits-outpaced-trips-movies-2019.aspx> (обращение: 14.02.2020).
2. Kranich N. Libraries and civic engagement // Library and book trade almanac, ed. by D. Bogart, B. J. Turock. Medford, NJ: Information today, 2012. Pp. 75-96.
3. Яницкий О. Н. Социобиотехнические системы: новый взгляд на взаимодействие человека и природы // Социальная наука и социальная практика. 2016. Т. 4. № 3, с. 5-22.
4. Ефременко Д. В., Яницкий О. Н., Ермолаева П. А. О социобиотехнических системах // Вопросы философии. 2019. № 5, с. 138-147.
5. Latour B. Nous n'avons jamais été modernes. Essai d'anthropologie symétrique. Paris: La Découverte, «L'armillaire», 1991.
6. Gandy M. Rethinking Urban Metabolism: Water, Space and the Modern City // City: Analysis of Urban Trends, Culture, Theory, Policy, Action. 2004. Vol. 8. № 3, pp. 363-379.
7. Pape U., Smirnova A. Transforming the public sphere: The case of Moscow's city libraries // Cultural Studies. 2018. Vol. 32. № 5, pp. 772-799.
8. Антопольский А. Б. Научная информация и электронное пространство знаний. М.: ИНИОН РАН, 2020.

9. Kirikova M. Challenges in Enterprise and Information Systems Modeling in the Contexts of Socio Cyber Physical Systems // Enterprise and Organizational Modeling and Simulation: 15th International Workshop, EOMAS 2019, Held at CAiSE 2019, Rome, Italy, June 3–4, 2019, Selected Papers, ed. by R. Pergl, E. Babkin, R. Lock, P. Malyzhenkov, V. Merunka. Cham: Springer, 2019. Pp. 60–69.

10. Антопольский А. Б. Научная информация и электронное пространство знаний. М.: ИНИОН РАН, 2020.

11. Путин В. В. Послание Президента Федеральному Собранию. 15.01.2020. // Официальный сайт Президента России. URL: <http://kremlin.ru/events/president/news/62582> (обращение: 28.01.2020).

12. Bremmer I. The End of the American Order: Ian Bremmer speech at 2019 GZERO Summit. 18.11.2019. // Eurasia Group. URL: <https://www.eurasiagroup.net/live-post/end-of-american-order-ian-bremmer-2019-gzero-summit-speech> (обращение: 23.01.2010).

13. Майстрович Т. В. Роль научных электронных библиотек в расширении пространства знаний: монография. М.: ИНИОН РАН, 2020.

14. Антопольский А. Б., Босов А. В., Савин Г. И., Сотников А. Н., Цветкова В. А., Каленов Н. Е., Серебряков В. А., Ефременко Д. В. Принципы построения и структура Единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. 2020. № 4, с. 9–17.

15. Ефременко Д. В. Формирование цифрового общества и геополитическая конкуренция // Контуры глобальных трансформаций: политика, экономика, право. Т. 13. 2020. № 2, с. 25–43

16. Giddens A. Modernity and self-identity: Self and society in the late modern age. Stanford: Stanford university press, 1991.

References

1. Mccarthy J. In U.S., Library Visits Outpaced Trips to Movies in 2019 // Gallup. 2020. January 24. URL: <https://news.gallup.com/poll/284009/library-visits-outpaced-trips-movies-2019.aspx> (accessed: 14.02.2020).

2. Kranich N. Libraries and civic engagement // Library and book trade almanac, ed. by D. Bogart, B. J. Turock. Medford, NJ: Information today, 2012. Pp. 75–96.

3. Yanitsky O. N. Sociobiotechnicheskie systemy: novyj vzglyad na vzaimodeistvie cheloveka i prirody // Socialnaya nauka I socialnaya praktika. 2016. T. 4. № 3, с. 5–22.

4. Efremenko D. V., Yanitsky O. N., Ermolaeva P. O. O sociobiotechnicheskikh systemah // Voprosy filosofii. 2019. № 5, с. 138–147.

5. Latour B. Nous n'avons jamais été modernes. Essai d'anthropologie symétrique. Paris: La Découverte, «L'armillaire», 1991.
6. Gandy M. Rethinking Urban Metabolism: Water, Space and the Modern City // City: Analysis of Urban Trends, Culture, Theory, Policy, Action. 2004. Vol. 8. № 3, pp. 363–379.
7. Pape U., Smirnova A. Transforming the public sphere: The case of Moscow's city libraries // Cultural Studies. 2018. Vol. 32. № 5, pp. 772–799.
8. Antopolskij A. B. Nauchnaya informatsia i elektronnoe prostranstvo znanij – Moscow: INION RAS, 2020.
9. Kirikova M. Challenges in Enterprise and Information Systems Modeling in the Contexts of Socio Cyber Physical Systems // Enterprise and Organizational Modeling and Simulation: 15th International Workshop, EOMAS 2019, Held at CAISE 2019, Rome, Italy, June 3–4, 2019, Selected Papers, ed. by R. Pergl, E. Babkin, R. Lock, P. Malyzhenkov, V. Merunka. Cham: Springer, 2019. Pp. 60–69.
10. Antopolskij A. B. Nauchnaya informatsia i elektronnoe prostranstvo znanij – Moscow: INION RAS, 2020.
11. Putin V. V. Poslanie Prezidenta Federalnomu Sobranju. 15.01.2020. // Oficialnyj sait Prezidenta Rossji. URL: <http://kremlin.ru/events/president/news/62582> (accessed: 28.01.2020).
12. Bremmer I. The End of the American Order: Ian Bremmer speech at 2019 GZERO Summit. 18.11.2019. // Eurasia Group. URL: <https://www.eurasiagroup.net/live-post/end-of-american-order-ian-bremmer-2019-gzero-summit-speech> (обращение: 23.01.2010).
13. Majstrovich T. V. Rol nauchnykh elektronnykh bibliotek v rashirenii prostranstva znanij: monografia. Moscow: INION RAN, 2020.
14. Antopolskij A. B., Bosov A. V., Savin G. I., Sotnikov A. N., Tsvetkova V. A., Kalenov N. E., Serebryakov V. A., Efremenko D. V. Prinzipy postroenja I struktura Edinogo tsifrovogo prostranstva znanij (ETPNZ) // Nauchno-technicheskaya informatsia. Ser. 1: Organizatsia I metodika informatsionnoj raboty. 2020. № 4, s. 9–17.
15. Efremenko D. V. Formirovanie tsifrovogo obshestva I geopoliticheskaya konkurencia // Kontury globalnykh transformatsij: politika, ekonomika, pravo. T. 13. 2020. № 2, s. 25–43
16. Giddens A. Modernity and self-identity: Self and society in the late modern age. Stanford: Stanford university press, 1991.

Теоретико-методологические основания проектирования единого цифрового пространства научных знаний

Н. В. Лопатина

МГИК

Аннотация. Рассматриваются нерешенные методологические проблемы создания единого цифрового пространства знаний. Ставятся задачи и предлагаются решения в области выбора объектов информационного моделирования и отражения, диалектики единичных и системных объектов в структуре ЕЦПНЗ, отраслевого и дифференцированного подходов в проектировании ЕЦПНЗ.

Ключевые слова: цифровизация, единое пространство знаний, наука, научная информация, научно-информационная деятельность, информационное обеспечение учёных.

Theoretical and methodological bases for designing a unified digital space of scientific knowledge

N. V. Lopatina

Moscow State Institute of Culture

Abstract. Unsolved methodological problems of creating a single digital knowledge space are considered. The challenges and offers solutions in the selection area of information modelling and reflection, dialectic of individual and system objects in the structure of Common Digital Space of Scientific Knowledge (CDSSK), industry-specific and differentiated approaches in designing CDSSK.

Key words: digitalization, common knowledge space, science, scientific information, scientific and informational activity, information support of scientists.

Высокий уровень внимания к задачам проектирования единого цифрового пространства научных знаний (ЕЦПНЗ), обсуждение данного проблемного комплекса в профессиональной печати [1, 2] позволяют выявить круг

нерешенных теоретико-методологических проблем, имеющих принципиальное значение для реализации этой масштабной идеи.

Фундаментальные позиции в современных информационных науках должна занять методологическая интеллектуализация информационного пространства как одного из приоритетных трендов современного информационного развития. Вместе с тем интеллектуализация информационных потоков, интеллектуализированное представление феноменов в виртуальной среде, которые прогнозировались в начале 2000-х, не стали реальностью современной информационной среды. По мнению Н. А. Слядневой, «цель интеллектуализации информационной инфраструктуры заключается в создании системы производства и передачи на входы управляющих блоков социума смыслодержающей информации, максимально отвечающей задаче формирования и принятия управленческих решений». Цифровой экономике нужна новая инфраструктура, основанная на интеллектуальном типе хранения научного знания, позволяющем сделать обозримым то, что накопили предыдущие поколения, и то, что вносит современная наука. Эта обозримость — условие рационального использования интеллектуального ресурса государства.

В кругу актуальных теоретических задач, требующих оперативного решения в контексте проектирования ЕЦПНЗ, определение соотношения традиционного и цифрового информационного пространства; определение «кварки» единого цифрового пространства (на наш взгляд, это — информационная система); определение идеологии интеграции (метаинформационная или ресурсная).

В числе ключевых вопросов: будет ли лежать в основе ЕЦПНЗ принцип глобального охвата или требуется отбор данных, идей, документов. Если будет ставиться задача отбора, каковыми будут его основания: ориентация на потребителя (информационный маркетинг как реагирование на типовые и перспективные запросы, например, применяемая сегодня модель «наукометрический анализ — экстраполяция результатов», которая, на наш взгляд,

недостаточно информативна) или ценностный отбор (на основании каких критериев, что считать социально значимым).

Рассмотрим несколько вопросов, значимых для проектирования ЕЦПНЗ.

Объекты информационного моделирования и отражения. Проектирование ЕЦПНЗ требует определения единицы, «кварки» представления знания в информационных системах. Документоцентристская парадигма организации информационной среды, которая оставалась непреодолимой даже на протяжении четверти века информатизации, акцентировала аналитико-синтетическую переработку на моделировании и отражении лишь одного информационного объекта — документа, несмотря на обеспеченность ИТ-решениями, позволяющими осуществлять экстрагирование «смыслоёмких» фрагментов текста. Современные системы метаданных моделируют и отражают уровень документа (библиографический), но не уровень смыслов (идеографический [3]). Но современный исследователь живет в условиях высокой динамики и социальной активности, темпы которой лишают его времени не только изучать монографии, статьи и даже рефераты (как во времена научно-технической революции), но и писать и оформлять публикации. Закладывается новый тип профессионального общения, сворачивающий уровень научных текстов. Науке в цифровой экономике необходимы особые способы представления научной информации, научные сети, где идея перерастает в знание и появляется продукт — результат интеллектуальной деятельности, где оценка научной идеи приобретает совсем другие очертания.

Проектирование ЕЦПНЗ требует архивирования и метаинформационного отражения разнообразия суверенных информационных объектов: смысла, текста, документа, автора [3], контекста [4, 5], дискретных и системных объектов разного уровня [4], а также деятельностных процессов, обеспечивающих производство нового знания и продвижение его в науку и практику.

Принципиально новые контуры ЕЦПНЗ намечают потоковые информационные объекты (стриминг событий, ме-

роприятий, взаимодействий, онлайн-продукты) и их увеличивающееся значение в научных коммуникациях (что подтвердили месяцы самоизоляции). Идеология сохранения цифрового наследия, современного научного среза не может быть реализована без архивирования информационных сообщений, введённых в общественный оборот посредством социальных сетей, блогосферы, файлообменных платформ (например, YouTube). Но эти новшества требуют осмысления науками информационного цикла.

Диалектика дискретных и системных объектов в структуре ЕЦПНЗ. ЕЦПНЗ расширяет функционал представления знания посредством построения дискретных и системных моделей результатов научной деятельности. Методологические основания ЕЦПНЗ определяют целесообразность отражения не только конкретных идей, результатов интеллектуальной деятельности, научных работ, учёных и научных коллективов, но и континуума научного знания, отдельных его подмножеств и фрагментов. Дискретные модели, отражая уникальность смыслов, идей, разработок, традиционно показывают их место в отраслевом, временном и т. д. подмножестве и научно-информационном пространстве в целом, используя систему классификаций и рубрикаторов. Системная модель складывается из специально организованной суммы дискретных моделей, являясь, однако, неаддитивной этой сумме, и представляет целостную картину научного мира. Информационное свертывание-моделирование системного подмножества создаёт базу для его комплексного анализа, раскрывающего механизмы той целостности, благодаря которой наука не рассыпается на дискретные идеи, тексты, темы, а сохраняет концептуальное, содержательное единство. Одна из ключевых функций системных моделей – демонстрация развития идей, характера их социальной диффузии, их отражения в информационных потоках разных типов и видов, динамики их профессиональных и социальных оценок.

В числе вопросов, требующих теоретической разработки, выступают критерии и принципы формирования сис-

темных моделей научных знаний и подходы к визуализации (картирование, построение ландшафтов и т. д.) их представления в конгломерате информационных систем.

Отраслевой подход в проектировании ЕЦПНЗ требует учёта особенностей формирования отраслевого знания, моделирования в институциональных системах, выступающих подсистемами ЕЦПНЗ, отраслевых информационных процессов, спецификации типологического и видового разнообразия отраслевых информационных ресурсов и, самое основное, многоаспектной дифференциации информационных потребностей, стимулируемых реальными отраслевыми практиками. В данном случае речь идёт не только об организации доступа к специфическим информационным массивам, наиболее значимым для отдельных групп потребителей и сфер деятельности (например, вариативных полных текстов, официальных документов в социально-гуманитарных науках, исходных кодов в информационных науках и т. д.), но и об уровнях и степени интеллектуализации информационного свертывания. «Точность» знания в технических и естественных науках, возможность использования искусственного языка математических и химических формул для его передачи, однозначность оценок практической значимости обуславливают клишированные модели организации знаниевых массивов, неприменимые в социальных и гуманитарных науках. Мультипарадигмальность, неизбежная зависимость форм представления гуманитарного знания в любых информационных системах и массивах от идеологических установок; сложность оценивания социальной эффективности (например, отсроченное влияние на мировоззренческие и идеологические платформы, философию, шлейф трактовок); интерпретационное разнообразие; деперсонализация, обобществление идей в ходе трансфера гуманитарного знания; высокий уровень контекстности связей между фрагментами научного знания — далеко не полный перечень факторов, требующих особого подхода к проектированию ЕЦПНЗ в социально-гуманитарной сфере.

Дифференцированный подход в проектировании ЕЦПНЗ. Принципиальный вопрос проектирования ЕЦПНЗ – вопрос потребителя конгломерата информационных систем, определяющий связи между ними, идеологию, контент, интерфейс, систему сервисов, лингвистическое обеспечение.

Универсальность как ключевая характеристика доцифровых институтов и форматов сохранения и распространения научного знания, а также аналоговых проектов их воспроизведения в цифровой среде, не позволяет выйти за безальтернативную логику линейного расположения в пространстве традиционного документного массива, которая частично компенсировалась продуктами и услугами справочно-библиографической природы. Современные ИТ-решения позволяют создавать индивидуализированные информационные продукты и сервисы в рамках ЕЦПНЗ, однако, это требует разработки дифференцированных «входов» для различных категорий пользователей, моделирования разнообразия поисковых траекторий и тезаурусов пользователей, идеологий определения и сегментирования целевой аудитории. Например, дифференцированное отражение специфики информационных потребностей ученого-эксперта в конкретной области знания и студента или молодого исследователя.

Принципиальность дифференцированного подхода в проектировании ЕЦПНЗ обусловлена неоднородностью целевой аудитории потребителей «научного знания». Эта неоднородность возникает в силу того, что закладываются новые механизмы социального взаимодействия, обеспечивающие диффузию теоретических идей и разработок в современную экономику, в производственные системы и социокультурные практики и индустрии. Эвристическими теоретическими инструментами в данной ситуации выступает, например, методика анализа отраслевых информационных систем, разработанная Н. А. Слядневой. Данная методика выделяет устойчивое для каждой отрасли разнообразие групп потребителей

и дифференцирует их в зависимости от контента и структуры информационных потребностей, в т. ч., от моделей и направленности информационного поведения, от особенностей взаимодействия с внешней информационной средой. В качестве инструментария проектирования ЕЦПНЗ могут также выступать аналитические модели взаимодействия групп поддержки (стейкхолдеров) [6] и другие.

Организационные модели ЕЦПНЗ. Уже хрестоматийной стала мысль о том, что новые информационно-коммуникативные форматы изменяют человека, а постоянно изменяющаяся информационная среда формирует новые и сохраняет традиционные каноны мышления, интеллектуального творчества. Модель «институт памяти», в которой основной акцент делается на сохранение и интерпретацию документов, уступает место моделям, основанным на идеологии доступности информационного разнообразия. Трансформация моделей научных коммуникаций определяет смену парадигмы организации цифрового пространства — от субъект-объектной к субъект-субъектной, основанной на активном участии потребителя в процессах создания и актуализации контента. Индивидуализация («персонализация» [7]) инструментов информационной деятельности, достигшая уровня массового актора, кардинально изменяет логику формирования единого цифрового пространства знаний. «От самостоятельного удовлетворения информационных потребностей с помощью доступных информационных массивов представители социально-профессиональных групп, формирующих новое знание, в цифровом мире создают не только смыслы и тексты (идею и её описание/ формулу), но и «документируют» её, самостоятельно вводят в «официальное» бытие и общественное использование, придают ей определённый правовой статус» [8].

Это требует переосмысления моделей взаимодействия отношений создателя и потребителя информационных систем нового качества, например, модель «наука-блокчейн» — новый организационный формат, в основе которого идея как кварка, соединяемая единой методологией.

В основе идеи ЕЦПНЗ – интеграция и консолидация ресурсов различных организационных структур, синергетический выход на метакорпоративный уровень. Вместе с тем, опыт межорганизационных проектов, специфика производственных и деловых отношений на информационном рынке (в первую очередь, ведомственные барьеры), особенности кадров современной инфосферы (сложная структура профессиональных ресурсов информационной деятельности, многоуровневость противостояния «сильноресурсных» и «слаборесурсных» субъектов отрасли, специфика профессиональных элит) демонстрируют сложности достижения консенсуса при принятии всего спектра решений: от выбора технологического решения, единства в вопросах идеологии и архитектуры взаимодействия до лидерства в управлении финансовыми потоками и привлечёнными инвестициями. Это требует технологизированного научно обоснованного решения в русле деклараций и приоритетов цифровой экономики.

Вопросы, поднятые в этом докладе, выступают научными задачами, имеющими важное значение для теории и практики, и обладают диссертационным потенциалом в рамках группы 05.25.00 – Документальная информация действующей номенклатуры научных специальностей.

Литература

1. Антопольский А. Б., Ефременко Д. В. К вопросу о едином электронном пространстве знаний // Вестник Российской академии наук, 2018. – Т. 88. – № 2. – С. 163–170.
2. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89, № 7. С. 728–735.
3. Сляднева Н. А. Библиография в системе Универсума человеческой деятельности: Опыт системно-деятельностного анализа. М.: Изд-во МГИК, 1993. 226 с.
4. Лопатина Н. В. Создание компьютерного аналога рекомендательного библиографического пособия по художественной литературе (вопросы теории и методологии): автореф. дис. канд. пед. наук: 05.25.03. Библиоковедение и библиографоведение. М., 1996. 16 с.

5. Сумина Т. Н. Персональная документография киноискусства: теоретико-методические основания и тенденции развития: авторефер. дис. канд. пед. наук: 05.25.03. — М.: МГУК, 1996. 16 с.

6. Неретин О. П. Формирование механизма взаимодействия групп поддержки в системе стратегического управления учреждениями культуры : дис. докт. экон. наук : 08.00.05. Санкт-Петербург, 2015. 293 с.

7. Арапов М. В. Предвидимые изменения в информационной среде общества // Научно-техническая информация. Сер. 2, 1991. № 1. С. 1-5.

8. Неретин О. П., Лопатина Н. В., Зубов Ю. С. Цифровизация сферы интеллектуальной собственности: от научного обоснования к практической реализации // Научно-техническая информация. Серия 1, 2019. — № 4. — С. 18-22.

References

1. Antopol'skij A. B., Efremenko D. V. K voprosu o edinom e`lektronnom prostranstve znanij // Vestnik Rossijskoj akademii nauk, 2018, tom 88, № 2, s. 163-170 DOI: 10.7868/S086958731802007X ; [Perev. na angl.] Antopol'skii, A. B. & Efremenko D. V. / The Uniform Electronic Knowledge Space Revisited // Her. Russ. Acad. Sci. (2018) 88: 89. <https://doi.org/10.1134/S1019331618010070>

2. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. O edinom cifrovom prostranstve nauchny`x znanij // Vestnik Rossijskoj akademii nauk. 2019. T. 89, № 7. S. 728-735.

3. Slyadneva N. A. Bibliografiya v sisteme Universuma chelovecheskoj deyatel`nosti: Opy`t sistemno-deyatel`nostnogo analiza. M.: Izd-vo MGIK, 1993. 226 s.

4. Lopatina N. V. Sozdanie komp`yuternogo analoga rekomendatel'nogo bibliograficheskogo posobiya po xudozhestvennoj literature (voprosy` teorii i metodologii): avtoref. dis. ... kand. ped. nauk: 05.25.03. Bibliotekovedenie i bibliografovedenie. M., 1996. 16 s.

5. Suminova T. N. Personal'naya dokumentografiya kinoiskusstva: teoretiko-metodicheskie osnovaniya i tendencii razvitiya: avtoreferat dissertacii na soiskanie uchenoj stepeni kandidata pedagogicheskix nauk po special`nosti 05.25.03. Bibliotekovedenie i bibliografovedenie. M.: MGUK, 1996. 16 s.

6. Neretin O. P. Formirovanie mexanizma vzaimodejstviya grupp podderzhki v sisteme strategicheskogo upravleniya uchrezhdeniyami kul`tury` : dissertaciya ... doktora e`konomicheskix nauk : 08.00.05 [Mesto zashhity` : S.-Peterb. gos. e`kon. un-t]. Sankt-Peterburg, 2015. 293 s.

7. Arapov M. V. Predvidimy`e izmeneniya v informacionnoj srede obshhestva // Nauchno-texnicheskaya informaciya. Ser. 2, 1991. № 1. S. 1-5.
8. Neretin O. P., Lopatina N. V., Zubov Yu.S. Cifrovizaciya sfery` intelektual`noj sobstvennosti: ot nauchnogo obosnovaniya k prakticheskoj realizacii // Nauchno-texnicheskaya informaciya. Seriya 1, 2019. № 4. S. 18-22.

О создании междисциплинарного аналитического центра для исследования дальнего чтения и digital humanities

В. П. Леонов

БАН

Аннотация. Принципиальное значение в работах по цифровому анализу текстов (Digital Humanities – DH) приобрела концепция «дальнего чтения» (Distant Reading – DR), предложенная профессором Стэнфордского университета США Ф. Моретти. В статье рассматривается возможность на основе методологии дальнего чтения изучать большие массивы библиографических данных с использованием формализованного анализа, количественных и библиографических методов. Для координации и планирования работы специалистов гуманитарных и технических областей необходимо создание междисциплинарного аналитического центра. Он может быть организован в учреждениях РАН с участием крупнейших библиотек страны, где сосредоточены разнообразные информационные ресурсы. Междисциплинарный аналитический центр для исследования дальнего чтения и digital humanities может быть построен по плану создания единого цифрового пространства научных знаний (ЕЦПНЗ). Его структура состоит из двух частей – постоянной и динамической (базис и надстройка). Постоянная (базис) формируется из имеющихся в России ресурсов – библиотечных, библиографических, архивных, музейных, энциклопедических, словарных, т. е. базис выращивается из существующих систем, прошедших экспертизу профессионалов. Динамическая (надстройка) часть предполагает анализ новой информации, отобранной экспертами, которую необходимо изучать. Кроме этого, проект формирования междисциплинарного аналитического центра должен координироваться специальным советом из ведущих ученых, определяющих этапы его создания и поддержки. Имеются в виду трудовые и финансовые затраты на поиск, оцифровку, сохранение и использование библиографической информации, полученной методами Digital Humanities.

Ключевые слова: дальнейшее чтение, междисциплинарный аналитический центр, Франко Моретти, цифровой анализ текстов, библиографическая информация, методология дальнего чтения

On establishing an interdisciplinary analytical center for the study of distant reading and digital humanities

V.P. Leonov

RASL

Abstract. The concept of distant reading (DR) developed by Professor Franco Moretti (Stanford University, USA) is vital in publications on digital humanities (DH). The present article considers the possibility, based on distant reading methodology, to study large amount of bibliographical data using formal method, quantitative and bibliographical methods of analysis. To coordinate and plan the work of specialists in humanitarian and technical fields it is necessary to establish an interdisciplinary analytical center. It can be established in the Russian Academy of Sciences institutions with participation of the Russian largest libraries which own various information resources. An interdisciplinary analytical center for researchers of the distant reading and digital humanities can be developed on a basis of the creating the single digital space of scientific knowledge. It consists of two parts – permanent and dynamic (base and superstructure). The constant part (basis) is formed from the resources available in Russia – library, bibliographic, archival, museum, encyclopedic, dictionary, i. e. the basis is grown from existing systems that have passed the expertise of professionals. The dynamic (add-on) part involves the analysis of new information, selected by experts, which needs to be studied. In addition, the project for the formation of an interdisciplinary analytical center should be coordinated by a special council of leading scientists who determine the stages of its creation and support. This refers to the labor and financial costs of searching, digitizing, preserving and using bibliographic information obtained by the methods of Digital Humanities.

Keywords: distant reading, interdisciplinary analytical center, Franco Moretti, digital humanities, bibliographic information, distant reading methodology.

Активное использование в гуманитарных отраслях, в том числе в социологии, литературоведении, книговедении, библиографоведении, библиографии, цифровых и автоматизированных технологий, а также необходимость решения прикладных задач для изучения обстоятельств и фактов их истории и теории привели к появлению большого числа ресурсов различного назначения. Деятельность по их формированию, обеспечению к ним доступа, сохран-

ности, возможность многократного использования требует координации и управления [1–3].

Принципиальное значение в работах по цифровому анализу текстов, известное под общим названием Digital Humanities – DH¹, приобрело понятие «дальнее чтение» (Distant Reading). Его ввел в науку о литературе профессор Стэнфордского университета США Франко Моретти [4]. В качестве дополнения к пристальному чтению (close reading), предпочитающему вдумчивое чтение конкретного текста, Моретти предложил новый подход, ориентированный на исследование доступного корпуса текстов с использованием компьютерных методов. «...Решающее значение, – пишет он, – имело возрастание важности количественных методов для моих исследований в Стэнфорде, которое в конце концов привело к созданию Литературной лаборатории в 2010 году... Количественные данные... представили такое огромное количество эмпирического материала, что я оказался к нему совершенно не готов... Сейчас ... результаты этих опытов начинают обретать форму, и период отсутствия теории заканчивается. На самом деле мир Digital Humanities, цифровых гуманитарных наук, понемногу начинает ощущать потребность в обобщающей теории для нового литературного архива» [4, с. 175–176].

Как отмечают переводчики книги «Дальнее чтение» О. Собчук и А. Шеля, «работы Франко Моретти возродили интерес к макроскопическому взгляду на литературу. Это значит: объектом исследования становится не одно или несколько произведений, как раньше, а десятки, сотни и даже тысячи. В этом океане информации (океане не только букв, но и цифр) Моретти ищет закономерности» [5, с. 88].

Междисциплинарность дальнего чтения проявляется в сотрудничестве специалистов разных гуманитарных и технических областей знания. Помимо историков, философов, социологов, филологов и специалистов по информационным технологиям, в таких исследованиях уже успешно принимают участие библиотекари, библиографы, архивисты, музейные работники. А проблемы, которые поднима-

¹ Цифровые (компьютерные) методы в гуманитарных науках.

ет Ф. Моретти, оставляют возможность для выбора. Говоря о междисциплинарных исследованиях, положенных в основу ДН, логично полагать, что они станут эффективными, если будут иметь *общий объект*, на который направлены методы различных наук [6–8]. В нашем случае в качестве такого объекта предлагаю рассматривать *библиографическую информацию*, имеющую сложную видовую структуру. Полагаю, что настало время в дополнение к литературоведческим исследованиям изучать библиографическое дальнейшее чтение, под которым понимаются способы представления автоматизированной обработки библиографической информации. Этот объект объединит формализованный анализ, количественные и библиографические методы [9].

Следует иметь в виду, что в начале междисциплинарного сотрудничества каждый из представителей наук остается как бы на своей территории, и только со временем сам объект изменяет его знание и методологию. Иногда такого времени требуется достаточно много, прежде чем специалист окажется «в проблеме» [10]. «...Мне кажется, — подчеркивает филолог И. Пильщиков, — я после многих лет работы в той сфере, которую сейчас начали называть Digital Humanities, стал своего рода специалистом “по переговорам” между филологами и программистами. Я помогаю коллегам с двух сторон этого дигитального “фронта” найти взаимопонимание» [11, с. 31–32].

И далее на вопрос о поддержке государственных и профессиональных исследований в оцифровке И. Пильщиков убедительно пояснил: «... Когда государство начинает что-то поддерживать, оно регламентирует... В России был период, когда ... создание электронных библиотек поддерживали и государство, и научные фонды, государственные и частные... рубеж столетий был кратковременным золотым веком для направления Digital Humanities... Но затем поддержка крайне сузилась... В результате профессиональные электронные библиотеки оказались практически брошены без системной поддержки ... исключения единичны» [11, с. 33–34].

Используя подход и опыт Ф. Моретти по изучению фактов и истории на основе методологии дальнего чтения, в рамках библиографии целесообразно создание службы

управления оцифровкой. В качестве первого шага это может быть междисциплинарный аналитический центр [12]. Предлагается начать работу по его созданию в учреждениях РАН с участием крупных библиотек страны, где сосредоточены самые разнообразные информационные ресурсы.

Одна из результативных попыток классификации информационных ресурсов гуманитарного профиля, выявленных в процессе мониторинга академических организаций, представлена в статье А. Б. Антопольского [13]. Статистика показала, что электронные версии изданий различаются глубиной ретроспективы и полнотой: полными текстами, аннотациями, содержанием выпусков, сводными указателями. Таким образом, одно и то же издание может быть представлено на сайте учреждения и в Интернете в разных формах и с разной полнотой.

Приведу некоторые примеры. Современный опыт создания цифровых коллекций представлен в проекте Института мировой литературы МГУ им. М. В. Ломоносова «Сравнительная поэтика и сравнительное литературоведение» для изучения взаимосвязи русской и зарубежной поэзии. Уже сейчас пользователям доступно более двух тысяч стихотворений, 60 научных изданий. Общее число авторов, чьи произведения представлены в системе, — 180 исследователей и 480 поэтов и переводчиков. Суть инновации проекта — в способах их представления [14].

Всю информацию представляют четыре взаимосвязанных подсистемы:

- 1) Корпус параллельных текстов;
- 2) Цифровая библиотека комментированных изданий поэтических переводов и их оригиналов, а также книг и статей по сравнительной поэтике;
- 3) Энциклопедия «Систематизированные библиографические сведения о поэтах, переводчиках и исследователях-компаративистах»;
- 4) Тезаурус «Структурированный глоссарий терминов, встречающихся в научной литературе, с описанием их значения и примерами их употреблений».

Интерфейс системы реализован на трех языках — русском, английском и испанском.

Цель проекта — объединить возможности цифровых библиотек и параллельных корпусов и превратить полученный инструмент в семантизированный гипертекст, в «семантическую сеть».

Успешную попытку применить статистические методы дальнего чтения (Ф. Моретти) для анализа заглавий книг осуществила аспирант Высшей школы печати и медиаиндустрии Московского политехнического университета Е. А. Тутатина. Суть ее эксперимента состояла в следующем. Из отобранных 5165 заглавий была составлена библиографическая база, включающая название книги, имя автора, год издания и жанр (роман, рассказ, повесть). Книги распределялись по направлениям: детектив, сентиментальная проза, фэнтези. Проверке подвергались издания, которые за последние 20 лет входили в списки бестселлеров газеты «The New York Times» (за исключением произведений А. Кристи и А. Конан Дойла).

Гипотеза исследования заключалась в том, что заглавия бестселлеров формируют «моду» на заглавия в издательском процессе. Как только книга становится бестселлером, ее переводят на множество языков, появляются экранизации, и тогда заглавие начинают копировать. Результаты анализа представлены в виде таблиц и содержат интересную информацию для использования в издательской деятельности, т. е. позволяют понять, как заглавие влияет на популярность книги, какие заглавия являются модными, какие могут возникать при этом ошибки [15].

...Междисциплинарный аналитический центр для исследования дальнего чтения и digital humanities может быть построен по плану создания единого цифрового пространства научных знаний (ЕЦПНЗ), предложенному А. Б. Антопольским и коллегами [1, 2]. Его структура состоит из двух частей — постоянной и динамической (базис и надстройка). Постоянная (базис) формируется из имеющихся в России ресурсов — библиотечных, библиографических, архивных, музейных, энциклопедических, словарных, т. е. базис выращивается из существующих систем, прошедших экспертизу профессионалов. Динамическая (надстройка) часть предполагает анализ новой информации, отобранной экспертами, которую необходимо изучать. Кро-

ме этого, проект формирования междисциплинарного аналитического центра должен координироваться специальным советом из ведущих ученых, определяющих этапы его создания и поддержки. Имеются в виду трудовые и финансовые затраты на поиск, оцифровку, сохранение и использование библиографической информации, полученной методами Digital Humanities.

Если коллеги поддержат идею создания междисциплинарного аналитического центра, то результаты ее деятельности могут быть впечатляющими. Задачей центра на первом этапе будет написание систематического обзора новой зарубежной литературы по дальнейшему чтению и ДН, анализ и обобщение значимых результатов, а также оценка перспективности этих исследований.

Литература

1. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестн. РАН, 2019. Т. 89. № 7. С. 728–734.
2. Антопольский А. Б. О создании центра лингвистических ресурсов РАН // Известия Российской академии наук. Сер. литературы и языка, 2019. Т. 78. № 4. С. 5–12.
3. Герасименко А. Ю. Формирование единого информационно-го пространства научной библиотеки // Библиосфера, 2019. № 4. С. 78–84.
4. Моретти Ф. Дальнее чтение / Ф. Моретти ; пер. с англ. А. Вдовина, О. Собчука, А. Шели. М.: Изд-во Института Гайдара, 2016. 344 с.
5. Собчук О. (Дальнее) чтение и (культурная) эволюция / О. Собчук, А. Шеля // НЛЮ, 2018. № 2 (150). С. 88–93.
6. Таллер М. Дискуссии вокруг Digital Humanities // Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании, 2012. № 1. С. 5–13.
7. Володин А. Ю. Digital Humanities (цифровые гуманитарные науки): в поисках самоопределения // Вестник Пермского университета. Сер.: История, 2014. № 3 (26). С. 5–12.
8. Можаяева Г. В. Digital Humanities: цифровой поворот в гуманитарных науках // Гуманитарная информатика: сб. ст. / под ред. Г. В. Можаяевой. Томск, 2015. Вып. 9. С. 8–23.
9. Леонов В. П. Вавилонская библиотека в контексте дальнего чтения // Библиография и книговедение, 2020. № 1. С. 77–83.

10. Соколова Е. В. Истолковать и/или исчислить: о возможности комбинированного метода на стыке герменевтики и digital humanities / Е. В. Соколова // Современная наука о литературе: сб. ст. / РАН, ИНИОН. М., 2018. С. 142–155.

11. Маяцкий М. Digital Humanities — это что-то новое или мы уже давно этим занимаемся? Интервью с Игорем Пильщиковым // Логос, 2015. Т. 25. № 2 (104). С. 14–36.

12. Леонов В. П. Дальнее чтение как стратегия точного библиографоведения // Научные и технические библиотеки, 2019. № 10. С. 56–67.

13. Антопольский А. Б. О классификации информационных ресурсов академических организаций социогуманитарного профиля // Взаимовлияние информационно-библиотечной среды и общественных наук: сб. ст. / РАН, ИНИОН. М., 2018. С. 48–61.

14. Полилова В. С. Цифровое будущее истории мировой литературы // Троицкий вариант, 2019. № 25. С. 14.

15. Гутагина Е. А. Заглавия книг: современные тенденции в книгоиздании // Библиография и книговедение, 2020. № 1. С. 111–119.

References

1. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. O edinom cifrovom prostranstve nauchnyh znaniy // Vestn. RAN, 2019. T. 89. № 7. P. 728–734.

2. Antopol'skij A. B. O sozdanii centra lingvisticheskikh resursov RAN // Izvestiya Rossijskoj akademii nauk. Ser. literatury i yazyka, 2019. T. 78. — № 4. P. 5–12.

3. Gerasimenko A. Yu. Formirovanie edinogo informacionnogo prostranstva nauchnoj biblioteki // Bibliosfera, 2019. № 4. P. 78–84.

4. Moretti F. Dal'nee chtenie / F. Moretti ; per. s angl. A. Vdovina, O. Sobchuka, A. Sheli. M.: Izd-vo Instituta Gajdara, 2016. 344 p.

5. Sobchuk O. (Dal'nee) chtenie i (kul'turnaya) evolyuciya / O. Sobchuk, A. Shelya // NLO, 2018. № 2 (150). P. 88–93.

6. Taller M. Diskussii vokrug Digital Humanities // Istoricheskaya informatika. Informacionnye tekhnologii i matematicheskie metody v istoricheskikh issledovaniyah i obrazovani, 2012. № 1. P. 5–13.

7. Volodin A. Yu. Digital Humanities (cifrovye gumanitarnye nauki): v poiskah samoopredeleniya // Vestnik Permskogo universiteta. Ser.: Istoriya, 2014. № 3 (26). P. 5–12.

8. Mozhaeva G. V. Digital Humanities: cifrovoy povorot v gumanitarnyh naukah // Gumanitarnaya informatika: sb. st. / pod red. G. V. Mozhaevoy. Tomsk, 2015. Vyp. 9. P. 8–23.

9. Leonov V. P. Vavilonskaya biblioteka v kontekste dal'nego chteniya // Bibliografiya i knigovedenie, 2020. № 1. P. 77–83.

10. Sokolova E. V. Istolkovat' i/ili ischislit': o vozmozhnosti kombinirovannogo metoda na styke germenevtiki i digital humanities /

E.V. Sokolova // *Sovremennaya nauka o literature: sb. st. / RAN, INION. M., 2018. P. 142–155.*

11. Mayackij M. Digital Humanities – eto chto-to novoe ili my uzhe davno etim zanimaemysya? Interv'yu s Igorem Pil'shchikovym // *Logos. 2015. T. 25, № 2 (104). P. 14–36.*

12. Leonov V. P. Dal'nee chtenie kak strategiya tochnogo bibliografovedeniya // *Nauchnye i tekhnicheskie biblioteki, 2019. № 10. P. 56–67.*

13. Antopol'skij A. B. O klassifikacii informacionnyh resursov akademicheskikh organizacij sociogumanitarnogo profilya // *Vzaimovliyanie informacionno-bibliotечноj sredy i obshchestvennyh nauk: sb. st. / RAN, INION. M., 2018. P. 48–6.*

14. Polilova V. S. Cifrovoe budushchee istorii mirovoj literatury // *Troickij variant, 2019. № 25. P. 14.*

15. Tutatina E. A. Zaglaviya knig: sovremennye tendencii v knigoizdanii // *Bibliografiya i knigovedenie, 2020. № 1. P. 111–119.*

О функциональности единого цифрового пространства научных знаний

О. М. Атаева

ВЦ ФИЦ ИУ РАН

Н. Е. Каленов

МСЦ РАН

В. А. Серебряков

ВЦ ФИЦ ИУ РАН

А. Н. Сотников

МСЦ РАН

Аннотация. Объемы научной информации, представленной в мире в настоящее время, создает серьезные проблемы с ее переработкой и использованием. В этой связи возникает необходимость фильтрации научной информации и выделения из нее знаний, отличающихся новизной и уникальностью. Совокупность такой информации в цифровом виде вместе со средствами, обеспечивающими ее актуализацию, сохранность и предоставление пользователям определяется как единое цифровое пространство научных знаний. Оно состоит из совокупности тематических подпространств, относящихся к различным направлениям науки, построенных по единым принципам. Несмотря на то, что существуют отдельные примеры формализации знаний в разных предметных областях, обобщенный подход к определению цифрового пространства научных знаний отсутствует. В данной работе рассматриваются проблемы и задачи формирования такого пространства и разработки средств, обеспечивающих возможность исследования информации в нем.

Functionality of a common digital space of scientific knowledge

O. M. Ataeva

CC FRS CSC RAS

N. E. Kalenov

JSCC RAS

V. A. Serebryakov

CC FRS CSC RAS

A. N. Sotnikov

JSCC RAS

Abstract. The volume of scientific information appearing in the world today creates serious problems with its processing and use. In this regard, it becomes necessary to filter scientific information and extract from it knowledge that is novel and unique. The aggregate of such information in digital form, together with the tools to ensure its updating, preservation and provision to users, is defined as a common digital space of scientific knowledge. It consists of a set of thematic subspaces related to various areas of science, built on the same principles. Despite the fact that there are some examples of formalization of knowledge in different subject areas, there is no generalized approach to defining the digital space of scientific knowledge. This paper discusses the problems and tasks of the formation of a such space and the development of tools that provide the ability to study information in it.

1. Введение

Лавинообразное нарастание объемов научной информации, появляющейся в мире в виде печатных и цифровых публикаций, документальных и фактографических баз данных, создает серьезные проблемы с ее переработкой и использованием. Исследователю, занимающемуся той или иной научной проблемой, для ознакомления со всей информацией, относящейся к сфере его интересов, и выделения

необходимых и полезных сведений, требуется значительное время, отрывающее его от собственно научной деятельности. С другой стороны, отсутствие у него важных данных, полученных в его научной области, может привести к бессмысленной потере времени на исследования, результаты которых уже получены другими учеными.

В этой связи возникает необходимость фильтрации научной информации и выделения из нее знаний, отличающихся новизной и уникальностью. Для выхода на новый уровень в использовании возможностей бурно развивающихся сегодня технологий обработки информации, в первую очередь, необходим переход к семантически значимому представлению научных знаний, извлекаемых из информации в цифровой среде [1, 2].

Хотя каждая область науки имеет свою специфику, в современных условиях, характеризующихся мультидисциплинарностью исследований, взаимопроникновением научных направлений, необходимого эффекта можно достичь, только разработав универсальные подходы к хранению и представлению научных знаний. Эти подходы нашли свое отражение в концепции создания Единого цифрового пространства научных знаний (ЕЦПНЗ) [3, 4].

Развитие технологий позволяет шире взглянуть на определение ЕЦПНЗ и обобщить накопившийся опыт в реализации различных решений в этой области.

Далее представлен обзор основных понятий в этой области, используемых как для представления элементов пространства, так и для обеспечения доступа к ним не только для человека, но и для программных агентов, что открывает широкие возможности их обработки и использования в различных областях потребления заинтересованными участниками научной деятельности.

2. Основные определения

Под *пространством научных знаний* понимается проверенная научным сообществом система знаний из разных областей науки.

Цифровое пространство научных знаний – цифровая среда, в которую входят проверенные научным сообществом информационные ресурсы и объекты, отражающие знания из различных областей науки.

Тематическое подпространство является частью общего пространства научных знаний, ограниченной рамками определенной предметной области. Несмотря на то, что существуют отдельные примеры формализации знаний в разных предметных областях [5, 6–15], обобщенный подход к определению цифрового пространства научных знаний отсутствует.

Под Единым цифровым пространством научных знаний (ЕЦПНЗ) будем понимать цифровую среду, интегрирующую достоверную научную информацию из различных областей знаний, вместе со средствами, обеспечивающими ее актуализацию, сохранность и предоставление пользователям. ЕЦПНЗ входит составной частью в «электронное пространство знаний», формирование которого является важной проблемой, поднимаемой в ряде программных документов, связанных с развитием информационного общества и цифровой экономикой.

ЕЦПНЗ состоит из совокупности тематических подпространств, относящихся к различным направлениям науки, построенных по единым принципам.

Анализ примеров формализации пространства знаний [16, 17] в различных областях свидетельствует о том, что основными составляющими ЕЦПНЗ в целом и каждого его тематического подпространства в частности являются – онтология и контент.

Онтология включает универсальное описание структуры данных ЕЦПНЗ. Это многоуровневая система понятий, описывающая классы объектов, отражаемых в каждом подпространстве, виды связей между этими классами и их объектами как внутри одного подпространства, так и между подпространствами, а также правила отражения объектов в ЕЦПНЗ.

Контент ЕЦПНЗ – это собственно научная информация (объекты, свойства, онтологии), а также терминологическое описание области знания [15, 18–23], (тезаурусы, слова-

ри, классификаторы). В состав контента, в частности, входит совокупность цифровых копий объектов реального мира и описание профилей их метаданных [23, 24].

Локальное подпространство ЕЦПНЗ – некоторая «изолированная» часть ЕЦПНЗ (возможно, политематическая), обладающее теми же свойствами, что и общее пространство. Локальное подпространство:

- определено на некотором подмножестве данных (контента) ЕЦПНЗ;
- построено по тем же принципам, что и ЕЦПНЗ;
- предназначено для использования группой пользователей;
- может существовать в отрыве от самого ЕЦПНЗ, не нарушая при этом структуры системы знания, содержащегося в нем;
- динамически связано с основным (базовым) ЕЦПНЗ средствами импорта / экспорта.

Интеграция – это процесс связывания данных из внешних источников с контентом ЕЦПНЗ.

Конструирование ЕЦПНЗ – процесс отображения состояния предметной области в цифровое представление, доступное для исследования и машинной обработки.

Семантическая библиотека – инструмент конструирования ЕЦПНЗ, обеспечивающий:

- формирование тематического подпространства (ТП) ЕЦПНЗ, включающее автоматизированную обработку и загрузку в ТП контента с установлением всех онтологических связей между его элементами;
- накопление и редактирование новой информации;
- многоаспектный поиск данных и навигацию по контенту ЕЦПНЗ;
- формирование локального подпространства по заданным характеристикам.

3. Онтология ЕЦПНЗ

Первым шагом на пути реализации ЕЦПНЗ является разработка его онтологии. Под онтологией в данном контексте будем понимать многоуровневую систему понятий, опи-

сывающих ресурсы и объекты предметных областей, концептов, терминов и связей между ними, характеризующуюся открытой иерархической и динамичной структуризацией, ориентированную как для хранения существующих знаний, так и для извлечения новых.

Онтология ЕЦПНЗ имеет три уровня (рис. 1).

Первый уровень. Концептуализация ЕЦПНЗ. На этом уровне описываются основные понятия, составляющие ЕЦПНЗ. Первый уровень онтологии ЕЦПНЗ включает следующие понятия.

- *Информационный объект* – цифровая копия объекта реального мира или специально созданный цифровой объект, отражающий определенные свойства реального объекта.

- *Идентификатор* информационного объекта – элемент данных, позволяющий однозначно идентифицировать объект в ЕЦПНЗ.

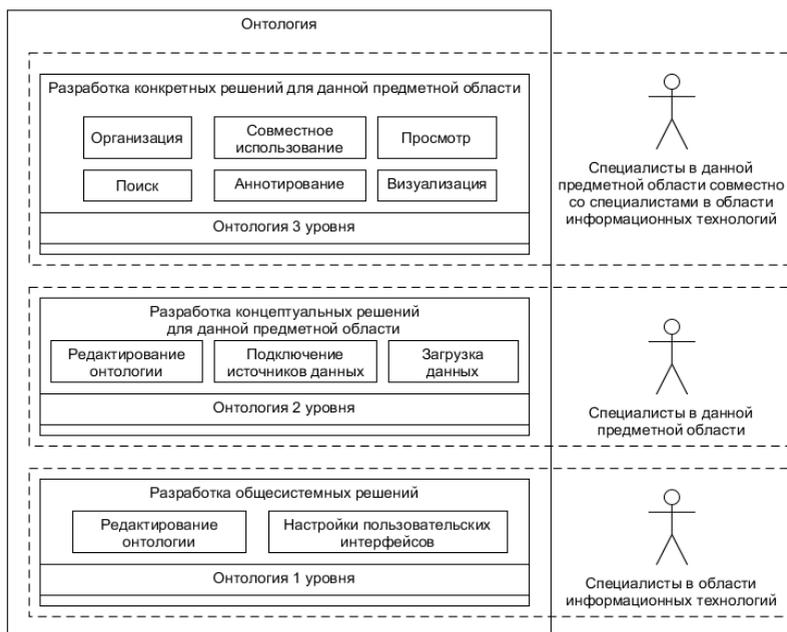


Рис. 1. Уровни онтологии ЕЦПНЗ

- *Тематическое подпространство* — совокупность информационных объектов с их свойствами и связями, относящихся к некоторому научному направлению.

- *Локальный класс объектов* — объекты, относящиеся к одному тематическому подпространству.

- *Универсальный класс объектов* — объекты, связанные с несколькими тематическими подпространствами. К универсальным классам относятся такие объекты, как персоны, публикации, документы, события, организации и т. п.

- *Атрибуты цифрового объекта* — структурированные свойства (метаданные), характеризующие объект с точки зрения задач ЕЦПНЗ.

- *Профиль метаданных* — набор атрибутов (с указанием их обязательности), описывающий объекты того или иного класса.

- *Предметная онтология (тезаурус)* тематического подпространства — совокупность индексов классификационных систем, ключевых терминов с их тезаурусными связями, относящихся к данному научному направлению.

- *Предметная онтология ЕЦПНЗ* — совокупность предметных онтологий отдельных подпространств.

- *Тезаурусные связи* — связи между двумя элементами предметной онтологии с заданным набором значений, который может устанавливаться внутри данного тематического подпространства.

- *Атрибутные связи* — связи между различными объектами как внутри одного класса, так и между объектами различных классов.

- *Тип атрибутных связей* — принимает одно из двух значений: однозначная связь (объект А связан связью b только с одним объектом В того или иного класса) и многозначная связь (объект А может быть связан связью b с несколькими объектами заданного подмножества В). Между объектом «город» и «страна» из класса «географические понятия» — однозначная связь, в то время как обратная связь «страна» — «город» — многозначная.

• Вид атрибутивных связей — конкретные отношения между различными объектами одного или различных классов, как локальных, так и универсальных. Между объектами универсальных классов «персона» и «публикация» могут быть многозначные связи вида «автор», «редактор», «составитель» и т. п. Между музейным объектом и персоной могут быть связи вида «автор модели», «автор сбора», «даритель» и т. п.

На втором уровне онтологии конкретизируются понятия первого уровня применительно к общему пространству и тому или иному тематическому подпространству, в частности, в соответствии с определением, данным на первом уровне, определяются тезаурусы конкретных предметных областей. Конкретные локальные классы объектов определяются спецификой той области науки, к которой они относятся. Для микробиологии это могут быть штаммы, коллекции и базы данных микроорганизмов; для языкознания — словари, языки, корпуса текстов и т. п., для химии — реакции, вещества, химические элементы и т. п.

На третьем онтологическом уровне устанавливаются профили метаданных объектов ЕЦПНЗ и связи между ними. В качестве примеров однозначных атрибутивных связей внутри одного класса можно привести связи публикаций различного уровня: статья (аналитический уровень) связана связью «опубликована в...» с определенным выпуском журнала (монографический уровень), который, в свою очередь, связан с описанием журнала (сводный уровень).

4. Уровни проектирования и функциональности программного обеспечения ЕЦПНЗ

Результатом проектирования ЕЦПНЗ должен являться «трехслойный» программный комплекс, слои которого соответствуют уровням онтологии пространства и обладают соответствующей функциональностью. Проектирование ЕЦПНЗ включает три уровня:

— Общесистемный (основанный на онтологии первого уровня), предназначен для работы специалистов в области информатики — системных аналитиков широкого профиля.

Он предоставляет инструментарий для формирования и редактирования основных словарей системы, типов и видов связей между объектами и т. п.

– Тематический слой программного комплекса, определяемый вторым уровнем онтологии, обладает функциональностью, обеспечивающей формирование тематических подпространств в той или иной области науки. Он ориентирован на работу авторизованных пользователей – специалистов в конкретной научной области. На этом уровне они определяют основные понятия подпространства и связи между ними, описывают и подключают необходимые источники данных, используя введенные понятия первого и второго уровней онтологии, получают доступ к пакетной загрузке данных. В рамках этого слоя формируются пользовательские интерфейсы системы, специфичные для тематического подпространства.

– Прикладной (реализация конкретных решений, принятых на первом и втором уровне), разрабатываемый специалистами в области создания программных комплексов. Этот слой программного комплекса базируется на третьем уровне онтологии и предоставляет доступ к ЕЦПНЗ внешним пользователям. Функциональность этого слоя включает обеспечение возможности многоаспектного поиска данных внутри ЕЦПНЗ, их визуализации и экспорта, навигации по пространству. На этом уровне пользователь может организовывать данные в коллекции, классифицировать их по разным признакам, формировать то или иное локальное подпространство.

5. Проблемы создания ЕЦПНЗ

Поскольку ЕЦПНЗ является принципиально новой информационной средой, на пути его реализации имеется ряд существенных проблем.

Проблемы формализации

В настоящее время отсутствуют общепризнанные подходы к определению контента ЕЦПНЗ, с учетом того, что конкретные научные знания специфичны для отдельных об-

ластей науки. Достаточно сложной является проблема выделения подпространства ЕЦПНЗ, относящегося к отдельной предметной области, поскольку отсутствует формальный общепризнанный подход к этой проблеме; казалось бы, для решения этой задачи можно было бы опираться на существующие классификационные системы (такие как ГРНТИ, УДК, МКИ и т. п.), однако, как показывает анализ [22], ни одна классификация не обеспечивает необходимой полноты и точности отражения научных знаний в той или иной области.

Лингвистические проблемы

Основой каждого тематического подпространства является предметная онтология — расширенный тезаурус со специфичными для каждой области науки видами связей между его элементами, исчерпывающим образом описывающий данную предметную область. Для формирования подобных тезаурусов и их динамической поддержки необходима разработка специальной методологии, основанной на технологиях Semantic Web.

Проблема экспертизы.

ЕЦПНЗ должно содержать достоверную научную информацию с минимальным дублированием как собственно данных, так и их источников. Соответственно, необходимо разработать механизм отбора материалов, предназначенных для включения в ЕЦПНЗ. Одной из составляющих такого механизма должен быть экспертный отбор.

Проблемы формирования контента и интеграции данных

Отсутствует унифицированный подход к интеграции данных из различных источников в единую информационную среду. В большинстве источников данных, подлежащих загрузке в то или иное тематическое подпространство ЕЦПНЗ, отсутствуют средства экспорта в каком-либо унифицированном формате. Для извлечения и загрузки в тематическое подпространство ЕЦПНЗ формализованных данных, соответствующих требованиям онтологий, в каждом конкретном случае необходимо обладать специализированными техническими знаниями и использовать специальные программные средства.

6. Вопросы интеграции данных

Рассмотрим концепцию взаимодействия программного комплекса с внешними источниками данных. Загрузка данных из внешних источников в пространство знаний (или выгрузка) осуществляются через ее онтологию. В число первоочередных задач по созданию средств интеграции данных ЕЦПНЗ входят следующие.

Разработка онтологии процесса интеграции. Она должна быть связана с онтологией ЕЦПНЗ и включать такие понятия, как: Отображение, Источник, Связь и другие.

Разработка требования к источнику. Интегрируемый источник данных должен быть снабжен семантическим слогом. Каждый объект в источнике должен обладать набором свойств для его однозначной идентификации и уникальным идентификатором URI.

Построение отображения. Отображение строится на уровне концептуальных понятий предметной области, при этом:

- вводится набор стандартных операций для очистки и приведения данных;
- вводится набор правил, по которым объект из источника можно представить в рамках понятий предметной области.

7. Инструментарий ЕЦПНЗ

Основные технологические требования к инструментарию ЕЦПНЗ;

- универсальность — возможность описания разных типов ресурсов;
- структурированность — поддержка связей между ресурсами;
- адаптируемость — возможность портирования на различные платформы;
- динамичность — возможность адаптации системы к эволюции ресурсов и настройки интерфейсов пользователя.

Функциональность инструментария ЕЦПНЗ

Основная функциональность должна обеспечивать решение следующих задач:

- формирование многоуровневой модели предметной области;

- интеграция источников данных;

- предоставление пользовательских и машиночитаемых интерфейсов.

Список функций:

- создание/просмотр/редактирование:

- информационных ресурсов и их экземпляров;

- тезауруса;

- коллекций информационных объектов;

- подключение источников данных;

- пакетная загрузка данных, составляющих контент библиотеки;

- атрибутивный/семантический/полнотекстовый поиск;

- предоставление данных в машиночитаемом формате;

- выделение дополнительных связей между экземплярами;

- информационных ресурсов;

- поддержка семантических меток для описания тематической направленности.

Основные принципы реализации инструментария поддержки локального подпространства ЕЦПНЗ

Функционирование ЕЦПНЗ основывается на следующих принципах персональной открытой семантической цифровой библиотеки [27]:

1. Библиотека должна поддерживать возможность использования медийных ресурсов или ссылки на них при описании объектов, включая текст, изображения, 3D-модели, аудио- и видеофайлы или любую их комбинацию.

2. Типы используемых ресурсов и связи между ними должны быть описаны средствами системы в рамках определенных понятий, составляющих семантическое описание ресурсов контента библиотеки. Набор этих понятий тематически ограничивается терминами предметной области из некоторого тезауруса.

3. Библиотека должна являться интеграционным узлом, предоставляя возможность связывания данных локального подпространства с данными из разных источников, в частности, представленных в виде, соответствующем требованиям Linked Open Data (LOD). Также должна обеспечиваться возможность извлечения данных этой библиотеки в структурированном формате.

4. Пользователи библиотеки должны иметь возможность организовывать свои локальные подпространства коллекции по интересующему их научному направлению, добавляя новые термины в предметный тезаурус, таким образом уточняя область своих интересов. Также пользователи должны иметь возможность осуществлять поиск не только среди объектов в рамках системы, но и по источникам данных без необходимости использования специальных знаний для поисковых запросов. Персональная открытая семантическая цифровая библиотека должна поддерживать:

- ввод и редактирование разнородных цифровых объектов (таких как текст, изображение, аудио- и видеофайлы, 3D-модели объектов и явлений);

- настраиваемые возможности импорта и экспорта данных, представленных в форматах, соответствующих требованиям Semantic Web;

- многоаспектный поиск данных и навигацию по связанным объектам;
- визуализацию данных с учетом специфики предметных областей;
- авторизацию пользователей с настраиваемыми правами доступа к тем или иным данным;
- защиту информации от несанкционированного доступа;
- возможность оперативного восстановления данных при технических сбоях.

8. Первоочередные задачи по проектированию ЕЦПНЗ

1. Разработка онтологии первого уровня ЕЦПНЗ, содержащей общие для различных научных областей понятия, на основе которых можно отразить знания из некоторой области науки в ЕЦПНЗ.

2. Выработка требований к интеграции источников данных и обеспечение решения задачи интеграции данных на основе разработанной онтологии.

3. Выработка конкретных требований к семантической библиотеке, используемой как инструмент построения подпространства ЕЦПНЗ и обеспечивающей навигацию по нему. Такая библиотека должна быть снабжена средствами интеграции источников данных; должна поддерживать механизмы связывания данных библиотеки с данными из других источников; обладать настраиваемыми пользовательскими интерфейсами; обеспечивать импорт / экспорт данных в машиночитаемых форматах.

9. Заключение

Разработка и поддержка ЕЦПНЗ является задачей национального масштаба. К ней должны быть привлечены ведущие организации в области информационных технологий (в части определения онтологии первого уровня), и в различных областях науки (для разработки онтологий

второго уровня отдельных тематических подпространств), основные организации-хранители научных ресурсов (библиотеки, архивы, музеи).

В последние годы создан определенный задел в области разработки онтологии первого уровня для ЕЦПНЗ. В рамках проекта, поддержанного РФФИ по конкурсу «КОМФИ» (18-00-00376 (К)) силами сотрудников МСЦ РАН (филиала ФГУ ФНЦ НИИСИ РАН), ФИЦ «Информатика и управление», ИНИОН РАН и БЕН РАН разработаны концепция формирования ЕЦПНЗ и основные элементы его архитектуры, создана модель предметной онтологии на примере микробиологии и языкознания [1, 2, 20]. В рамках государственного задания № 0580-2021-0016 специалистами МСЦ РАН отрабатывается технология формирования цифровых объектов различного вида, моделируются процессы интеграции данных в рамках электронной библиотеки «Научное наследие России» [18–20, 23].

Специалистами ФИЦ «Информатика и управление» РАН при поддержке РФФИ (проекты № 20-07-00324, 18-00-00297) ведутся активные исследования по созданию и развитию открытой семантической библиотеки LibMeta [25–29].

Литература

1. Губанов Н. И., Губанов Н. Н., Волков А. Э. «Критерии истинности и научности знания» *Философия и общество*, 2016 № 3 (80). С. 78–95.
2. Ильин В. В., Калинин А. Т. *Природа науки: Гносеологический анализ*. М.: Высшая школа, 1985. 230 с.
3. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // *Вестник Российской академии наук*, 2019. — Т. 89. — № 7. С. 728–735.
4. Антопольский А. Б. и др. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // *Научно-техническая информация*. Сер. 1, 2020. — № 4. — С. 9–17.
5. Муромский А. А., Тучкова Н. П. Представление математических понятий в онтологии научных знаний // *Онтология проектирования*, 2019. Т. 9. № 1 (31). С. 50–69.
6. Гуревич И. Б., Трусова Ю. О. Тезаурус и онтология предметной области «Анализ изображений» // *Всероссийская конф. с междунар. участием «Знания — Онтологии — Теории» (ЗОНТ-09)*.

Новосибирск: Институт математики им. С. Л. Соболева СО РАН, 2009. 10 с.

7. Hlava M. M. K. The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services., 2014. Т. 6. № 3. С. 1–80.

8. Hlava M. M. K. The Taxobook: Principles and practices of building taxonomies, part 2 of a 3-part series // Synthesis Lectures on Information Concepts, Retrieval, and Services, 2014. Т. 6. № 4. С. 1–164.

9. Hlava M. M. K. The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services, 2014. Т. 6. № 4. С. 1–156.

10. Бездушный А. Н., Жижченко А. Б., Кулагин М. В., Серебряков В. А. Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек. Программирование, 2000, 4, с. 3–14.

11. Ахлестин А. Ю., Лаврентьев Н. А., Фазлиев А. З. Систематизация научных графических ресурсов по молекулярной спектроскопии // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18–23 сентября 2017 г., г. Новороссийск). М.: ИПМ им. М. В. Келдыша, 2017. С. 34–42.

12. Сотников А. Н. и др. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы, 2012. № 4. С. 6

13. Елизаров А. М. и др. Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Докл. РАН. 2016. Т. 467. № 4. С. 392–395.

14. Митрофанова О. А., Константинова Н. С. Онтологии как системы хранения знаний // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». 2008. С. 54

15. Муромский А. А., Тучкова Н. П. О тезаурусе для предметной области «Обыкновенные дифференциальные уравнения». Вычисл. центр им. А. А. Дородницына РАН, 2004. 55 с.

16. ГОСТ 7.0–99 Межгосударственный стандарт ГОСТ 7.0–99 «Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения» (введен в действие постановлением Госстандарта РФ от 7 октября 1999 г. N 334-ст).

17. Костин В. В. Обзор семантических моделей, описывающих научные публикации и научно-исследовательскую деятельность // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. 2014.

18. Kalenov N., Sobolevskaya I., Sotnikov A., Kirillov S. The use of 3D visualization technology web-collections for the formation of virtual

exhibitions // CEUR Proceedings of the 21st Conference on Scientific Services & Internet, 2020. Vol. 2543. P. 382–390.

19. Kalenov N., Sobolevskaya I., Sotnikov A. Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // CEUR Proceedings of the 21st Conference on Scientific Services & Internet, 2020. Vol. 2543. P. 391–398.

20. Kalenov N., Sobolevskaya I., Sotnikov A. Hierarchical Representation of Information Objects in a Digital Library Environment // Communications in Computer and Information Science, 2019. Vol. 1093. P. 93–104.

21. Dextre Clarke S. G., Zeng M. L. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modeling // Information Standards Quarterly (ISQ). 2012. Т. 24. № 1.

22. Цветкова В. А., Харьбина Т. Н., Мохначева Ю. В., Бескаравайная Е. В., Митрошина И. Ю. Особенности совмещения классификационных систем и формирования массива ключевых слов для определения пространства знаний по микробиологии // Научные и технические библиотеки, 2019. № 11. С. 25–43. DOI 10.33186/1027-3689-2019-11-25-43.

23. Каленов Н. Е., Соболевская И. Н., Сотников А. Н. Иерархические уровни представления информационных объектов в среде электронных библиотек // Информатика и инновации, 2018. Т. 13, – № 2. С. 25–31.

24. Gruber T. Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS), 2007. Т. 3. № 1. С. 1–11.

25. Антопольский А. Б., Атаева О. М., Серебряков В. А. Среда интеграции данных научных библиотек, архивов и музеев «LibMeta» // «Информационные Ресурсы России», 2012. № 5. С. 8–12.

26. Серебряков В. А., Атаева О. М. Основные понятия формальной модели семантических библиотек и формализация процессов интеграции в ней // Программные продукты и системы, 2015. № 4. С. 180–187.

27. Атаева О. М., Серебряков В. А. Персональная открытая семантическая цифровая библиотека LibMeta. Конструирование контента. Интеграция с источниками LOD // Информ. и её примен., 2017. Т. 11. № 2. С. 85–100.

28. Атаева О. М. Информационная модель семантической библиотеки LibMeta // Программные продукты и системы, 2016. № 4. С. 36–44.

29. Атаева О. М., Серебряков В. А. Онтология цифровой семантической библиотеки LibMeta // Информатика и её применения, 2018. Т. 12. С. 2–10.

References

1. Gubanov N. I., Gubanov N. N., Volkov A. E. «Kriterii istinnosti i nauchnosti znaniya» *Filosofiya i obshchestvo*, no. 3 (80), 2016, pp. 78–95: <https://cyberleninka.ru/article/n/kriterii-istinnosti-i-nauchnosti-znaniya>
2. Il'in V. V., Kalinkin A. T. *Priroda nauki: Gnoseologicheskij analiz*. M.: Vysshaya shkola, 1985. 230 s.
3. Antopolskiy A. B., Kalenov N. E., Serebryakov V. A., Sotneykov A. N. Common digital space of scientific knowledge // *Vestn. Ros. akad. nauk*. 2019. T. 89. № 7. S. 728–735.
4. Antopolskiy A. B. and Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy *Nauchno tekhnicheskaya informaciya*. Ser. 1. 2020. № 4. S. 9–17.
5. Muromskij A. A., Tuchkova N. P. Predstavlenie matematicheskikh ponyatij v ontologii nauchnyh znaniy // *Ontologiya proektirovaniya*. 2019. T. 9. № 1 (31). S. 50–69.
6. Gurevich I. B., Trusova YU. O. Tezaurus i ontologiya predmetnoj oblasti “Analiz izobrazhenij” // *Vserossijskaya konf. s mezhdunar. uchastiem “Znaniya – Ontologii – Teorii” (ZONT-09)*. Novosibirsk: Institut matematiki im. S. L. Soboleva SO RAN, 2009. 10 s.
7. Hlava M. M. K. *The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2014. T. 6. № 3. C. 1–80.
8. Hlava M. M. K. *The taxobook: Principles and practices of building taxonomies, part 2 of a 3-part series // Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2014. T. 6. № 4. C. 1–164.
9. Hlava M. M. K. *The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2014. T. 6. № 4. C. 1–156.
10. Bezdushnyj A. N., ZHizhchenko A. B., Kulagin M. V., Serebryakov V. A. Integrirovannaya sistema informacionnyh resursov RAN i tekhnologiya razrabotki cifrovyyh bibliotek. *Programmirovanie*, 2000, 4, s. 3–14.
11. Ahlyostin A. YU., Lavrent'ev N. A., Fazliev A. Z. Sistematizaciya nauchnyh graficheskikh resursov po molekulyarnoj spektroskopii // *Nauchnyj servis v seti Internet: trudy XIX Vserossijskoj nauchnoj konferencii (18–23 sentyabrya 2017 g., g. Novorossijsk)*. M.: IPM im. M. V. Keldysha, 2017. S. 34–42. URL: <http://keldysh.ru/abrau/2017/39.pdf> doi:10.20948/abrau-2017–39
12. Sotnikov A. N. i dr. Principy postroeniya i formirovaniya elektronnoj biblioteki «Nauchnoe nasledie Rossii» // *Programmnye produkty i sistemy*. 2012. № 4.
13. Elizarov A. M. i dr. Ontologii matematicheskogo znaniya i rekomendatel'naya sistema dlya kolekcij fiziko-matematicheskikh dokumentov // *Dokl. RAN*. 2016. T. 467. № 4. S. 392–395.

14. Mitrofanova O. A., Konstantinova N. S. Ontologii kak sistemy hraneniya znaniy // Vserossijskij konkursnyj otbor obzorno-analiticheskikh statej po prioritetnomu napravleniyu «Informacionnotelekomunikacionnye sistemy». 2008. S. 54.

15. Muromskij A. A., Tuchkova N. P. O tezauruse dlya predmetnoj oblasti «Obyknoennye differencial'nye uravneniya». Vychisl. centr im. A. A. Dorodnicyna RAN, 2004. S. 55.

16. GOST 7.0-99 Mezhsudarstvennyj standart GOST 7.0-99 «Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu. Informacionno-bibliotechnaya deyatel'nost', bibliografiya. Terminy i opredeleniya» (vveden v dejstvie postanovleniem Gosstandarta RF ot 7 oktyabrya 1999 g. N 334-st).

17. Kostin V. V. Obzor semanticheskikh modelej, opisvyvayushchih nauchnye publikacii i nauchno-issledovatel'skuyu deyatel'nost' // Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kolekcii. 2014.

18. Kalenov N., Sobolevskaya I., Sotnikov A., Kirillov S. The use of 3D visualization technology web-collections for the formation of virtual exhibitions // CEUR Proceedings of the 21st Conference on Scientific Services & Internet, 2020. Vol. 2543. P. 382-390.

19. Kalenov N., Sobolevskaya I., Sotnikov A. Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // CEUR Proceedings of the 21st Conference on Scientific Services & Internet, 2020. Vol. 2543. P. 391-398.

20. Kalenov N., Sobolevskaya I., Sotnikov A. Hierarchical Representation of Information Objects in a Digital Library Environment // Communications in Computer and Information Science, 2019. Vol. 1093. P. 93-104.

21. Dextre Clarke S. G., Zeng M. L. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modeling // Information Standards Quarterly (ISQ). 2012. T. 24. № 1.

22. Tsvetkova V., Kharybina T., Mokhnacheva Yu., Beskaravaynaya E., Mitroshina I. Combining classification systems and building the array of keywords for defining the space of microbiological knowledge. Scientific and Technical Libraries. 2019;(11):25-43. (In Russ.) <https://doi.org/10.33186/1027-3689-2019-11-25-43>

23. Kalenov N. E., Sobolevskaya I. N., Sotnikov A. N. Ierarhicheskie urovni predstavleniya informacionnyh ob»ektov v srede elektronnyh bibliotek // Informaciya i innovacii, 2018. T. 13, – № 2. S. 25-31. DOI:10.31432/1994-2443-2018-13-2-25-31

24. Gruber T. Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). 2007. T. 3. № 1. C. 1-11.

25. Antopol'skij A. B., Ataeva O. M, Serebryakov V. A. Sreda integracii dannyh nauchnyh bibliotek, arhivov i muzeev «LibMeta» // «Informacionnye Resursy Rossii» № 5, 2012.

26. Serebryakov V. A., Ataeva O. M. Osnovnye ponyatiya formal'noj modeli semanticheskikh bibliotek i formalizaciya processov integracii v nej // Programmnye produkty i sistemy. 2015. № 4. S. 180–187.

27. Ataeva O. M., Serebryakov V. A. Personal'naya otkrytaya semanticheskaya cifrovaya biblioteka LibMeta. Konstruirovanie kontenta. Integraciya s istochnikami LOD // Inform. i eyo primen., 11:2 (2017), 85–100.

28. Ataeva O. M. Informacionnaya model' semanticheskoy biblioteki LibMeta // Programmnye produkty i sistemy. 2016. № 4. S. 36–44.

29. Ataeva O. M., Serebryakov V. A. Ontologiya cifrovoj semanticheskoy biblioteki LibMeta // Informatika i eyo primeneniya. 2018. T. 12. S. 2–10.

Интеграция лингвистических информационных ресурсов в рамках концепции Единого цифрового пространства научных знаний

А. Б. Антопольский

ИНИОН РАН

Аннотация. Описываются результаты исследований по проблемам интеграции российских лингвистических информационных ресурсов (ЛИР). Результаты включают концепцию создания Центра ЛИР, данные инвентаризации электронных ЛИР, методику оценки ЛИР. В ходе исследований были разработаны проекты нескольких информационных систем, основанных на интеграции ЛИР, а также онтология поисковых терминов, используемых в различных ЛИР.

Ключевые слова: лингвистические ресурсы, мониторинг, экспертиза, инфометрия, поисковые термины, онтологии.

Integration of linguistic information resources within the concept of a unified digital space of scientific knowledge

A. B. Antopol'skij

INION RAS

Abstract. The article describes the results of research on the problems of integration of Russian linguistic information resources (LIR). The results include the concept of creating a LIR center, electronic LIR inventory data, and LIR evaluation methodology. In the course of the research, projects were developed for several information systems based on the integration of LIR, as well as an ontology of search terms used in various LIR.

Keywords: linguistic resources, monitoring, expertise, infometry, search terms, ontologies.

Введение

Исследование проблем интеграции научно-информационных ресурсов учреждений РАН

по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства РАН является, с одной стороны, самостоятельным проектом, а с другой — частью комплексного проекта РФФИ «Разработка фундаментальных основ, методов и средств обеспечения конвергенции естественнонаучных и социогуманитарных ресурсов как составляющих единого цифрового пространства знаний» КОМФИ.

Проект по интеграции лингвистических информационных исследований включает как общие для комплексного проекта исследования, такие как разработка стратегии и методики оценки научных информационных ресурсов, так и специальные исследования социогуманитарного информационного пространства знаний, включая каталогизацию лингвистических информационных ресурсов учреждений РАН. Важной частью проекта является интеграция сведений по языкознанию, отраженных в информационно-поисковых языках лингвистических информационных ресурсов, в единую онтологию научного знания с целью их конвергенции с естественнонаучными знаниями.

Конкретные задачи, которые решались в проекте:

- Анализ зарубежного опыта по учету, и интеграции лингвистических информационных ресурсов;
- Разработка стратегии по интеграции лингвистических информационных ресурсов в рамках концепции создания Центра лингвистических ресурсов РАН;
- Проведение инвентаризации, классификации и каталогизации информационных ресурсов РАН по лингвистике;
- Разработка методики оценки гуманитарных научных ресурсов РАН на примере ресурсов по языкознанию;
- Разработка проекта справочной информационной системы по языкознанию для информационного обеспечения методического руководства и экспертизы научных исследований в области языкознания РАН;
- Разработка проекта автоматизированной информационной системы «Русский язык» в качестве первого этапа

формирования Центра лингвистических информационных ресурсов;

- Разработка Онтологии поисковых терминов по лингвистике (ОПТЕЛ) на основе тезаурусов, рубрикаторов, словарей метаданных и других поисковых терминов по лингвистике как основы для разработки онтологии научного знания по языкознанию.

Анализ зарубежного опыта

Цифровые (электронные) лингвистические информационные ресурсы (далее – ЛИР) стали в последние десятилетия важнейшим инструментом как научной деятельности в области теоретического и прикладного языкознания, так и практики применения компьютерных языковых технологий в индустрии обработки текста и устной речи (далее – NLP). Создание, поддержка, распространение, обмен ЛИР стали в последние годы мейнстримом международной коллаборации в компьютерной лингвистике и индустрии NLP. Далее мы приведем перечень нескольких наиболее известных международных структур, основная цель которых – организация совместной деятельности по созданию и использованию ЛИР в науке и в индустрии NLP.

ELRA – Европейская ассоциация языковых ресурсов [1].

OLAC, сообщество открытых языковых архивов [2].

CLARIN – Общие языковые ресурсы и технологическая инфраструктура [3].

LDC – Консорциум лингвистических данных [4].

ISCA – Международная ассоциация речевого общения [5].

SIL – Летний институт лингвистики [6].

В качестве примера национальной службы ЛИР можно привести французскую систему ORTOLANG – Открытые ресурсы и инструменты для обработки языка [7], объединяющую более 200 лингвистических служб университетов и компаний.

Более подробное описание функций и ресурсов перечисленных структур имеется в работе [8].

Концепция создания Центра ЛИР

Быстрое развитие информационных технологий в лингвистике как науке, а также необходимость решения различных прикладных задач, использующих методы и средства компьютерной лингвистики (таких как обработка текста, анализ и синтез речи, автоматический перевод, обучение языкам и др.) привели к созданию большого числа электронных ЛИР различного назначения. Деятельность по их формированию, обеспечению доступа, сохранности, открытости, возможности повторного использования требует оптимизации, координации и системного подхода, т. е., обобщенно говоря — управления деятельностью в сфере ЛИР.

Управление деятельностью в сфере ЛИР должно стать частью Цифровой системы управления сервисами научной инфраструктуры коллективного пользования (АС УСНИКП), создание которой предусмотрено в Национальном проекте «Наука». Создание системы управления ЛИР обеспечит значительный экономический эффект. Управление деятельностью в сфере ЛИР предлагается реализовать на основе создания Центра лингвистических ресурсов РАН при Институте русского языка РАН.

Предполагается, что Центр ЛИР должен решать следующие задачи:

1. Мониторинг, учет, каталогизация российских ЛИР.
2. Экспертиза ЛИР, в том числе обеспечение экспертной функции РАН в отношении создания ЛИР как результатов научной деятельности.
3. Координация деятельности по созданию ЛИР, стратегическое планирование деятельности по созданию ЛИР.
4. Интеграция и агрегация ЛИР (по отдельным типам ЛИР).
5. Организация открытого доступа к ЛИР и его поддержка.
6. Архивирование и обеспечение сохранности ЛИР.
7. Разработка стандартов, типовых программных средств, метаданных и методик по созданию ЛИР.

8. Создание и поддержка централизованных сервисов для создателей и пользователей ЛИР, в том числе, монетизированных.

9. Организация и выпуск журнала по проблемам прикладной и компьютерной лингвистики, а также ЛИР.

10. Организация площадки для обмена опытом в области создания и использования ЛИР.

11. Создание онтологии научного знания по лингвистике, в том числе в рамках создания Единого российского электронного пространства знаний.

12. Взаимодействие с универсальными российскими системами научной информации.

13. Взаимодействие с международными и зарубежными национальными центрами лингвистических ресурсов.

Одной из центральных проблем интеграции информационных ресурсов является определение возможности и целесообразности интеграции или агрегации ЛИР с целью создания централизованных многофункциональных ресурсов или сервисов. В концепции рассмотрены следующие типы ЛИР:

1. Библиографии, библиотечные каталоги, описи, каталоги ссылок.

2. Электронные коллекции и библиотеки полных текстов (книги, диссертации, отчеты, труды конференций и др.).

3. Периодические, продолжающиеся издания и архивы периодики.

4. Корпуса текстов.

5. Лексикографические ИР.

6. Этнолингвистические и социолингвистические БД.

7. Лингвистические географические системы, атласы.

8. Электронное представление памятников письменности.

9. Активные ресурсы (алгоритмы, процессоры, программы).

10. Грамматики.

11. Описания языков и комплексные лингвистические сайты.

12. Информационные языки.
 13. Энциклопедии, справочники, реестры языков.
 14. Сведения об отдельных персонах (сайты и страницы ученых, личные фонды, биобиблиографии).
 15. Списки, перечни, указатели персон.
 16. Медиаресурсы.
 17. Сайты учреждений – владельцев ЛИР.
 18. Сайты-сателлиты и ресурсы во внешних АИС.
- Очевидно, для одних типов ресурсов будет целесообразна интеграция, для других – агрегация, для третьих – достаточно будет их каталогизации. Обоснование этих подходов содержится в работе [8].

Навигатор информационных ресурсов по языкознанию (НИРЯЗ)

НИРЯЗ содержит перечень ресурсов по языкознанию, а также специфических лингвистических ресурсов, создаваемых учреждениями РАН и республиканских академий наук. Особенностью НИРЯЗ является то, что он включает сведения как о традиционных ресурсах (библиотечных, архивных, музейных), так и об электронных (электронные библиотеки, базы данных, информационные системы, сайты, социальные сети и др.). В НИРЯЗ включены сведения о коллекциях и массивах документов и данных, и лишь в отдельных случаях – сведения об отдельных документах. Сведения о конференциях и других научных мероприятиях, диссертациях, архивных фондах и описях ограничиваются ссылками на страницы, где эти ресурсы представлены. Сведения о персональных страницах ученых-лингвистов представлены выборочно и могут быть удалены или дополнены по желанию субъектов информации – ученых, работающих в учреждениях РАН.

На данном этапе НИРЯЗ содержит сведения о приблизительно 1,2 тыс. лингвистических информационных ресурсов. Тестовая версия НИРЯЗ размещена по адресу

<http://niryaz2.alexo.beget.tech/>

В НИРЯЗ предусмотрены следующие виды навигации и поиска ресурсов:

1. Поиск по владельцам информационных ресурсов осуществляется в разделе *Владельцы ИР*. Предъявляется алфавитный список кратких наименований учреждений. При клике на название учреждения приводится список ресурсов, которым оно владеет, а также другие данные об учреждении.

2. Навигация по типам ресурсов осуществляется в разделе *Типы ИР*. Предъявляется список из 15 типов ресурсов, часть которых разделена на виды. Типология ресурсов унаследована от исходной версии навигатора НИРОН [8]. При выборе типа или вида ресурсов предъявляется список наименований ресурсов, относящихся к этому типу и виду.

3. Навигация по тематике осуществляется в разделе *Тематика ЛИР*. Пользователю предъявляется тематический рубрикатор, разработанный на основе раздела «16 Языкознание» ГРНТИ, но значительно модернизированный. Один ресурс может иметь от 1 до 3 рубрик.

4. Языки и группы языков. В разделе *Языки и группы языков* пользователю предъявляются алфавитные перечни конкретных языков, а также генеалогических, ареальных и других групп языков, которым посвящены ресурсы, представленные в НИРЯЗ.

5. Лексический поиск по наименованию ресурсов и владельцев. В поисковом поле нужно указать слово или фрагмент слова. В результате поиска пользователю предъявляется список ресурсов, а также владельцев, наименования которых содержит это слово.

Во всех случаях найденные наименования ресурсов содержат активные ссылки, кликнув на которые можно получить сведения об этих ресурсах, или перейти на веб-сайт, где содержится ресурс или сведения о нем.

Методика оценки качества ЛИР

В данном исследовании проводился анализ ЛИР как компонента онлайн-информационных систем. Анализируются следующие параметры:

Дизайн:

- целостность стилевого оформления;
- читабельность основного материала;
- сбалансированность верстки страниц и не перегружена ли она информацией;
- сбалансированность цветов страниц;
- наличие и уместность рекламы, не мешает ли она работе.

Контент:

- релевантность заявленной тематике;
- актуальность;
- уникальность;
- представительность (объем представленной информации);
 - полнота (в какой степени покрывается заявленная тематика);
 - засоренность (оценка доли ненужной информации);
 - частота и регулярность обновления;
 - наличие и корректность метаинформации, в частности библиографических описаний к представляемым материалам;
 - логичность, полнота и удобство структуры всего контента, её соразмерность объему контента.

Функциональность:

- навигация;
- наличие и удобство использования меню;
- наличие и удобство указателей;
- наличие карты сайта;
- наличие гипертекстовых ссылок, корректность их работы;
- возможность доступа к любым представленным материалам за минимальное число кликов;

- доступ к необходимым навигационным средствам с любой страницы сайта;

- в многостраничных материалах (статьях, книгах ...) простой переход на определенную страницу, в начало/конец материала, к структуре/содержанию.

Поиск:

- наличие атрибутивного поиска;
- наличие полнотекстового (лексического) поиска во всем контенте, в заданном разделе;
- наличие комбинированного поиска (атрибутивного и полнотекстового одновременно);
- репрезентативность списка результатов поиска и возможность их сортировки;
- экспорт результатов поиска.

Интерактивность:

- наличие обратной связи;
- наличие ссылок на социальные сети;
- наличие подписки на новости;
- возможность ставить закладки, делать комментарии, выделять текст и т. п.

Юзабилити:

- работоспособность сайта на разных платформах;
- доступность сайта через различные браузеры;
- адаптивность сайта;
- разнообразие форм/форматов представления материала на сайте;
- время реакции системы на действия пользователя;
- наличие настроек (шрифты, масштаб, повороты ...);
- ограничение доступа (регистрация, плата ...);
- наличие сведений о новых поступлениях;
- наличие сопроводительной информации (цели/задачи проекта, контакты подсказки...).

Технические аспекты:

- надежность функционирования сайта (показатель отказов);
- количество повторных обращений;
- безопасность обмена данными между сайтом и пользовательским устройством;
- защита данных от несанкционированного доступа;
- поведение пользователей на сайте, включая время, проведенное на сайте и глубину просмотра.

Перечисленные параметры были экспериментально проверены несколькими экспертами на ряде лингвистических ресурсов различных типов.

Кроме перечисленных характеристик универсального характера, были разработаны некоторые специальные характеристики для оценки именно ЛИР. Описание этих характеристик представлено в работе [9].

Также в ходе исследования для всех включенных в НИРЯЗ ЛИР была получена количественная оценка при помощи Индекса качества сайтов (ИКС), разработанного компанией Яндекс [10].

Проект справочной информационной системы по языкознанию

Одним из результатов исследований по данному гранту была разработка проекта справочной информационной

системы (СИС) по языкознанию, предназначенной для информационного обеспечения процессов научно-методического руководства, а также экспертизы научных исследований по языкознанию в РАН. Проект представлен в виде Технического задания на СИС. Он направлен на согласование руководству ИНИОН РАН и ИРЯ РАН. Здесь приводятся общие принципы создания СИС.

Задачи разработки

Предлагаемая СИС должна содержать информацию, необходимую и достаточную для справочно-информационного обслуживания ученых, а также для научно-методического руководства научными исследованиями в области языкознания в учреждениях РАН, в том числе, для:

- разработки программ НИР;
- подготовки и экспертизы заявок на гранты, проекты, темы НИР;
- оценки результатов НИР, а также учреждений и научных подразделений;
- минимизации дублирования при проведении НИР и создании лингвистических информационных ресурсов;
- проведения наукометрических исследований;
- координации работ по оцифровке публикаций по языкознанию.

Общие требования к СИС

СИС должна создаваться на основе открытых источников. При наполнении СИС должен быть исключен сбор обязательной дополнительной отчетности от учреждений и научных работников. Администратор СИС, однако, имеет право просить учреждения и научных работников проверить и дополнить имеющуюся у него информацию.

Учреждения и научные работники РАН должны иметь возможность корректировать и дополнять информацию, размещенную в СИС. Корректировка и дополнение информации в СИС должна осуществляться в электронном виде через модератора — администратора СИС.

СИС должна быть размещена в Интернете и быть доступна для свободного и бесплатного доступа без регистрации.

Информация в СИС должна размещаться под лицензией Creative Commons Attribution (CC-BY) и быть доступна для скачивания и некоммерческого использования со ссылкой на источник.

Актуализация информации, размещенной в СИС, должна осуществляться не реже, чем 2 раза в год.

Источники наполнения СИС

Источниками информации для наполнения СИС являются:

- интернет-сайты институций, персон, периодических изданий, мероприятий;
- государственные учетные информационные системы Минобрнауки, ВАК, Рособнадзора, государственных научных фондов;
- информационные системы открытого доступа;
- данные, предоставляемые добровольно учреждениями и научными работниками.

Перспективы развития СИС

Предлагаемая СИС может в перспективе развиваться в нескольких направлениях.

Развитие контента как переход от наименований информационных объектов к их аннотациям и далее — к полным текстам, т. е. превращение СИС в тематический раздел Репозитория открытого доступа по общественным наукам, описанного в [11].

Увеличение глубины ретроспективы информации в СИС, вплоть до создания справочной системы по истории академического языкознания в России.

Создание сводной библиографической БД по языкознанию на основе учтенных библиографических перечней, БД АИСОН по языкознанию, профильных библиотечных каталогов, учет наличия полнотекстовых электронных публикаций и создания на этой основе электронной библиотеки и инструмента координации оцифровки библиотечных фондов РАН по языкознанию.

Расширение тематики: сначала создание СИС «Филология», с включением данных по литературоведению

и фольклористике, потом СИС «История». В результате СИС будет соответствовать по тематике задачам ОИФН РАН.

Расширение сферы охвата: учет научных результатов в области языкознания всех научных учреждений России (включая вузовскую и прикладную науку). Реализация этого варианта потребует существенного увеличения ресурсов.

Проект автоматизированной информационной системы «Русский язык»

В качестве первого этапа реализации Центра ЛИР предлагается создание автоматизированной информационной системы (АИС) «Русский язык».

Постановка задачи.

АИС «Русский язык» должна включать следующие разделы:

- сводную библиографическую БД;
- навигатор информационных ресурсов;
- электронную библиотеку.

Сводная библиографическая БД и навигатор информационных ресурсов в перспективе должны включать сведения об информационных объектах по русистике, создаваемых в российских научно-образовательных организациях, а также частными лицами, в том числе:

- публикации;
- неопубликованные научные и образовательные документы;
- цифровые материалы и информационные системы;
- архивные материалы.

Электронная библиотека должна включать полные тексты произведений по русистике, разрешенные к свободному распространению.

Основным способом создания АИС является оцифровка, организация и представление в доступ ресурсов, добровольно предоставляемых участниками системы, а также агрегация данных, свободно распространяемых в Интернете.

Задачи АИС «Русский язык»

На основе использования АИС «Русский язык» могут решаться разнообразные задачи, в том числе:

— текущее и ретроспективное информационное обслуживание ученых-русистов, а также преподавателей и студентов по специальности «Русский язык»;

— координация и научно-методическое руководство исследованиями по русистике;

— экспертиза научных и квалификационных исследований по русистике;

— наукометрические и науковедческие исследования, выполнение функций «карты российской науки» применительно к области русистики;

— оценка научных организаций и отдельных ученых в области русистики;

— возможность ведения кабинетов для учреждений, личных кабинетов, персональных списков трудов и других сервисов для ученых;

— координация оцифровки документов, относящихся к данной тематике;

— документальное сопровождение АИС «Русская корпусная грамматика» <http://rusgram.ru/>

— использование в системе Антиплагиат.

Основные принципы создания АИС «Русский язык»

ИРЯ РАН осуществляет научно-методическое и организационное руководство проектом, а также участвует в наполнении контента АИС.

ИНИОН РАН реализует программную и технологическую поддержку АИС, а также участвует в наполнении контента АИС.

ИРЯ РАН и ИНИОН РАН передают в АИС все ресурсы соответствующего профиля, созданные ранее, в том числе в рамках гранта «Интеграция научно-информационных ресурсов учреждений РАН по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства». При этом все переданные ресурсы могут использоваться в других информационных системах.

Отношения и взаимные обязательства ИРЯ РАН и ИНИОН РАН определяются Соглашением о сотрудничестве.

Проектная и технологическая документация по АИС утверждается двумя сторонами.

АИС «Русский язык» должна быть открыта для участия других научных и образовательных организаций.

АИС «Русский язык» размещается в Интернете, является открытой и бесплатной для пользователей информационной системой, функционирующей под лицензией Creative Common CC-BY.

Для контента АИС должно быть обеспечено архивное долговременное хранение.

Основные принципы проектирования АИС «Русский язык»

Сфера охвата АИС определяется специальным документом, входящим в состав проектной документации. В этом документе сферу охвата следует определить по следующим параметрам:

- по государственной принадлежности источников;
- по глубине ретроспективы;
- по составу информационных объектов;
- по языку.

Следует определить возможность и целесообразность включения в АИС иностранных и иноязычных публикаций, архивных, учебно-образовательных, популярных материалов.

Для установленной сферы охвата АИС должна обладать максимальной полнотой.

Навигатор Интернет-ресурсов, в отличие от созданного ранее Навигатора НИРЯЗ, должен охватывать все научные и образовательные ресурсы России по русистике. В то же время сервисы и форматы данных НИРЯЗ могут быть использованы в АИС «Русский язык».

Информационные объекты, размещаемые в трех разделах АИС, связываются гиперссылками.

Для индексирования и идентификации информационных объектов АИС «Русский язык» используются общие лингвистические средства для всех трех разделов АИС. В АИС «Русский язык» создается единая поисковая система.

Окончательные требования к прикладному программному обеспечению АИС «Русский язык» будут сформулиро-

ваны позже с учетом системного программного обеспечения, которое будет установлено в новом здании ИНИОН РАН. Однако, учитывая тенденции развития форм представления научной информации в Интернете, целесообразно ориентироваться на технологии и стандарты Семантического веба и связанных открытых данных (LOD).

Источники комплектования Сводной библиографической БД

Сводная библиографическая БД должна агрегировать библиографические записи опубликованных и неопубликованных документов по русистике из следующих источников:

- БД АИСОН;
- электронные каталоги ведущих российских библиотек (РГБ, РНБ, Президентская библиотек им. Б. Н. Ельцина, БАН, ГПТБ СО РАН);
- электронные каталоги библиотечно-библиографических агрегаторов (Национальная электронная библиотека, Либнет, ЭКБСОН, АРБИКОН);
- электронные каталоги ведущих информационных систем научно-технической информации (ЕГИСУ НИОКТР, Научная электронная библиотека E-library, Электронная библиотека диссертаций, КиберЛенинка, «Истина» и др.);
- персональные перечни трудов ведущих русистов, свободно доступные в Интернете;
- библиографические описания из оцифрованных каталогов ИНИОН РАН и ИРЯ РАН;
- библиографические описания из избранных оцифрованных библиографических указателей по русистике.

Источники комплектования Навигатора по русистике

Навигатор по русистике должен формироваться на основе:

- навигатора информационных ресурсов по языкознанию, созданного совместно ИНИОН РАН и ИРЯ РАН [11];
- каталогов лингвистических ресурсов доступных в Интернете;

— поиска профильных информационных ресурсов в Интернете;

— навигатор по русистике должен включать все ресурсы, относящиеся к русскому языку, включая созданные в учреждениях академической, вузовской и прикладной науки России.

Источники комплектования Электронной библиотеки по русистике

— произведения, оцифрованные из фондов ИНИОН РАН;

— произведения, права на распространение которых принадлежат ИРЯ РАН и другим возможным участникам системы;

— произведения, имеющие статус общественного достояния;

— произведения, доступные в Интернет под свободной или открытой лицензией;

— произведения, предназначенные обеспечить документальное сопровождение АИС «Русская корпусная грамматика» (для них, возможно, придется решать правовые вопросы).

Онтология поисковых терминов по лингвистике

Инструментом, который должен обеспечить навигацию и поиск в совокупности интегрированных и агрегированных ЛИР Центра ЛИР, должна стать Онтология поисковых терминов по лингвистике (ОПТЕЛ).

ОПТЕЛ содержит лексику и парадигматику 55 словарей информационно-поисковых языков из 30 различных ЛИР, созданных в различных учреждениях РАН и доступных в настоящее время в Интернет. Это тезаурусы, рубрикаторы, теги корпусов, словари лексикографических и географических информационных систем по лингвистике, реестров языков. В ОПТЕЛ реализованы функции поиска и визуализации терминов и сведений из них, включая поиск лексических пересечений.

Подробное описание ОПТЕЛ содержится в работе [12]. Там же доступны перечни словарей и ЛИР, вошедших в пилотную версию ОПТЕЛ. В настоящее время пилотная версия ОПТЕЛ размещена в Интернете [13].

Тезаурус по лингвистике передан также для использования в комплексном проекте по созданию онтологии для Единого цифрового пространства научных знаний на базе технологии LІВМЕТА.

Заключение

В докладе представлены результаты исследований 2019–2020 гг., посвященные проблемам интеграции ЛИР и созданию различных инструментов, предназначенных для управления ЛИР. Эти результаты содержатся также в многочисленных публикациях автора, но в наиболее полном и систематизированном виде отражены в монографии [8].

Однако реализация предлагаемых проектных решений должна быть увязана с общей политикой развития информационного пространства науки в России. Автор и его коллеги неоднократно предлагали свое видение этой политики [14, 15]. К сожалению, Министерство науки и высшего образования до сих пор не сформулировало принципы этой политики, хотя это было необходимо сделать в рамках Национального проекта «Наука».

Статья отражает результаты исследований по гранту РФФИ 18-00-00298 «Интеграция научно-информационных ресурсов учреждений РАН по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства РАН».

Литература

1. European Language Resources Association// [Электронный ресурс]. URL: <http://www.elra.info/en/>
2. OLAC, the Open Language Archives Community// [Электронный ресурс]. URL: <http://www.language-archives.org/>

3. CLARIN – European Research Infrastructure for Language Resources and Technology // [Электронный ресурс]. URL: <https://www.clarin.eu/>
4. Linguistic Data Consortium (LDC) // [Электронный ресурс]. URL: <https://www.ldc.upenn.edu/>
5. International Speech Communication Association // [Электронный ресурс]. URL: <https://www.isca-speech.org/iscaweb/index.php/about-isca>
6. SIL International // [Электронный ресурс]. URL: <https://www.sil.org/>
7. Outils et Ressources pour un Traitement Optimisé de la Langue // [Электронный ресурс]. URL: <https://www.ortolang.fr/>
8. Научная информация и электронное пространство знаний : монография / А. Б. Антопольский ; под науч. ред. Д. В. Ефременко ; ИНИОН РАН, Фундам. б-ка. Москва : ИНИОН, 2020. 313 с.
9. Маркарова Т. С. К вопросу о разработке информационной системы Центра лингвистических ресурсов и методических рекомендаций по оценке и отбору ресурсов. Препринт // [Электронный ресурс]. URL: http://inion.ru/site/assets/files/1206/markarova_preprint.pdf
10. Что такое индекс качества сайта (ИКС) // [Электронный ресурс]. URL: <https://yandex.ru/support/webmaster/site-quality-index.html>
11. Кузнецов А. В. О Программе по созданию на базе ИНИОН РАН автоматизированной библиотечной информационной системы по общественным наукам // Электронный ресурс. URL: http://hist-phil.ru/events/8703_kontseptsiya_Programmyi_po_sozodaniyu_na_baze_INION_RAN_avtomatizirovannoy_bibliotechnoy_informatsionnoy_sistemyi.pdf
12. Антопольский А. Б., Савчук С. О., Тамеев А. А. О разработке онтологии поисковых терминов по лингвистике // Информационные ресурсы России. 2020. № 4. С. 2–7.
13. Онтология поисковых терминов по лингвистике [Электронный ресурс]. URL: <http://db.inion.ru/optel/>
14. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) / Антопольский А. Б., Босов А. В., Савин Г. И., Сотников А. Н., Цветкова В. А., Каленов Н. Е., Серебряков В. А., Ефременко Д. В. // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2020. № 4. С. 9–17.
15. Антопольский А. Б. Научная информация и цифровое пространство знаний: постановка задачи для России. Наука и научная информация, 2020. 3 (1). С. 8–17.

References

1. European Language Resources Association// URL: <http://www.elra.info/en/>
2. OLAC, the Open Language Archives Community// URL: <http://www.language-archives.org/>
3. CLARIN – European Research Infrastructure for Language Resources and Technology// URL: <https://www.clarin.eu/>
4. Linguistic Data Consortium (LDC) // URL: <https://www ldc.upenn.edu/>
5. International Speech Communication Association// URL: <https://www.isca-speech.org/iscaweb/index.php/about-isca>
6. SIL International// URL: <https://www.sil.org/>
7. Outils et Ressources pour un Traitement Optimisé de la LAN Gue // URL: <https://www.ortolang.fr/>
8. Nauchnaya informaciya i elektronnoe prostranstvo znaniy : monografiya / A. B. Antopol'skij ; pod nauch. red. D. V. Evremenko ; INION RAN, Fundam.b-ka. Moskva : INION, 2020. 313 s.
9. Markarova T. S., K voprosu o razrabotke informacionnoj sistemy Centra lingvisticheskikh resursov i metodicheskikh rekomendacij po ocenke i otboru resursov. Preprint// [Elektronnyj resurs]. URL: http://inion.ru/site/assets/files/1206/markarova_preprint.pdf
10. Chto takoe indeks kachestva sajta (IKS) // [Elektronnyj resurs]. URL: <https://yandex.ru/support/webmaster/site-quality-index.html>
11. Kuznecov A. V. O Programme po sozdaniyu na baze INION RAN avtomatizirovannoj bibliotechnoj informacionnoj sistemy po obshchestvennym naukam // Elektronnyj resurs. URL: http://hist-phil.ru/events/8703_kontseptsiya_Programmyi_po_soz0daniyu_na_baze_INION_RAN_avtomatizirovannoy_bibliotechnoy_informatsionnoy_sistemyi.pdf
12. Antopol'skij A. B., Savchuk S. O., Tameev A. A. O razrabotke ontologii poiskovyh terminov po lingvistike // Informacionnye resursy Rossii. 2020. № 4. S. 2–7.
13. Ontologiya poiskovyh terminov po lingvistike [Elektronnyj resurs]. URL: <http://db.inion.ru/optel/>
14. Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy (ECPNZ) / Antopol'skij A. B., Bosov A. V., Savin G. I., Sotnikov A. N., Cvetkova V. A., Kalenov N. E., Serebryakov V. A., Efremenko D. V // Nauchno-tehnicheskaya informaciya. Seriya 1: Organizaciya i metodika informacionnoj raboty. 2020. № 4. S. 9–17. DOI: 10.36535/0548-0019-2020-04-2
15. Antopol'skij A. B. Nauchnaya informaciya i cifrovoe prostranstvo znaniy: postanovka zadachi dlya Rossii. Nauka i nauchnaya informaciya. 2020;3(1):8–17. URL: <https://doi.org/10.24108/2658-3143-2020-3-1->

Использование предметной онтологии единого цифрового пространства научных знаний в наукометрических задачах

В. А. Цветкова

ИНИОН РАН

Н. Е. Каленов

МСЦ РАН

Ю. В. Мохначева , И. А. Митрошин

БЕН РАН

Аннотация. Формирование Единого цифрового пространства научных знаний (ЕЦПНЗ) является одной из приоритетных задач современного периода и подразумевает использование унифицированных подходов к отбору информации, её представлению, навигации, терминологическому наполнению. При построении предметной онтологии научного направления – «Микробиология» использован метод статистического подсчёта частотного распределения ключевых слов в информационных массивах, сформированных в результате тематического поиска информации по микробиологии. Гипотеза исследования: частота встречаемости КС может являться индикатором интенсивности развития научной темы. Результаты исследований позволили предположить, что размер пула ключевых слов в рубриках может служить индикатором, показывающим широту спектра разнообразия исследований, их объектов и методов.

Ключевые слова: пространство знаний, классификаторы, цитатный анализ, частотное распределение, ключевое слово, предметная онтология.

Using the subject ontology of the unified digital space of scientific knowledge in scientometric tasks

V. A. Tsvetkova

ISSS RAS

Abstract. The Formation of a Common digital space of scientific knowledge (CDSSK) is one of the priorities of the modern period and involves the use of unified approaches to the selection of information, its presentation, navigation, and terminological content. When constructing an subject ontology of the scientific direction – «Microbiology» – we applied a method of statistical calculation of the frequency distribution of keywords (KW) in information arrays formed as a result of thematic search for information on Microbiology. We assume that the frequency of KW occurrence can be considered as indicator of the scientific topic development intensity. The results of the study show that the size of the pool of keywords in categories can serve as an indicator that shows the breadth of the spectrum of research diversity, methods and objects: the more keywords in a category, the greater the diversity of research directions.

Keywords: knowledge space, classifiers, citation analysis, frequency distribution, keywords, subject ontology.

В числе приоритетных задач современного периода рассматриваются подходы к формированию Единого цифрового пространства научных знаний (ЕЦПНЗ), что обусловлено активным переходом к формированию цифровой экономики, реализацией программ по осуществлению цифровизации страны в тех сферах, в которых фактор знания играет ключевую роль. В работе [1] детально рассмотрено Единое российское цифровое пространство научных знаний. Под онтологией понимается формальное описание некоторой области знания, которое может быть использовано для автоматической обработки компьютером [2].

Исследование направлено на создание единого пространства знаний и на оценку возможных направлений его развития на примере тематического направления «Микробиология», в основу которого положены информационно-лингвистические элементы. Исследования выполнены при поддержке грантов РФФИ № 18-00-00294комфи и № 18-00-00372комфи.

Формирование единого пространства научных знаний подразумевает использование унифицированных подходов к отбору информации, её представлению, навигации, терминологическому наполнению. В основу предложенной методологии положен анализ классификационных систем, описывающих данное научное направление, и установление связей между ними [3]; анализ публикаций, отраженных в базах данных цитирования по данному научному направлению, с выделением массивов ключевых терминов, установление связей между ними и с индексами классификационных систем, в первую очередь, таких как: Государственный рубрикатор научно-технической информации (ГРНТИ), Рубрикатор отраслей знаний ВИНТИ РАН, Универсальная десятичная классификация (УДК), Международная классификация изобретений (МКИ) и др. Проблема заключается в том, что перечисленные классификаторы во многом не согласуются между собой и значительно отличаются друг от друга по множеству критериев [4].

На основе особенностей и логики существующих классификаторов был разработан локальный базовый рубрикатор (ЛБР) по микробиологии, основные принципы и технология создания которого были подробно изложены в публикациях [3–4]. При разработке классификатора были установлены тезаурусные связи его рубрик с индексами существующих классификационных систем (КС). Установленные связи между терминами ЛБР и элементами КС принимают одно из значений: «эквивалентны» (тождественны), «содержит...» (шире), «входит в...» (уже), «пересекается с...». Последний вид связи эквивалентен принятому в классификационных системах указанию «см. также...». Локальные рубрики, сформулированные на языке современных терминов микробиологии, использовались для формирования глоссария понятий, являющихся основой предметной онтологии. Глоссарий включает ключевые термины (КТ), основой для отбора которых являлась база данных Web of Science Core Collection (WoS CC).

Микробиология – очень широкая и одна из самых быстроразвивающихся областей знания. Непрерывающееся

эволюционирование этого научного направления приводит к тому, что из классификаторов какие-то коды исчезают вообще, а какие-то появляются впервые. Кроме того, анализируемая предметная область не всегда вписывается в таксономию, предлагаемую существующими классификаторами.

При построении онтологии научного направления «Микробиология» использовался метод статистического подсчёта частотного распределения ключевых слов (КС) в информационных массивах, сформированных в результате тематического поиска информации по микробиологии. Как уже отмечено, в качестве основной информационной базы была использована Web of Science Core Collection. При построении структуры рубрикатора и составлении перечня КС учитывалась

частотность повторов того или иного термина в документальных массивах. В дальнейшем эти КС легли в основу глоссария терминов (понятий) для разработки онтологии.

На следующем этапе были построены деревья классификации понятий (иерархии классов), которых в онтологии может быть несколько. Именно эти понятия и устанавливают основные связи между классами. Всё это даёт возможность построить алгоритмы для реализации метода автоматизированного построения онтологии конкретной предметной области. Затем эксперты предметной области, для которой разрабатывается онтология (в нашем случае — «Микробиология»), определяют правила логических выводов, предоставляющие возможность использовать полученные данные, представленные в онтологии, и извлекать из неё новые знания.

Для слежения за динамикой терминологии тематического направления очень хорошо показал себя подход, разработанный на основе библиометрических методов. При этом в качестве единицы анализа выступают КС, или некоторая совокупность слов, выражающих проблему (предмет, отрасль, направление и др.). По изменениям частоты встречаемости единиц анализа в текстах можно делать выводы

об изменениях в научных направлениях, выявлять наиболее релевантные изучаемому предмету документы.

В исследовании использовался пул публикаций, сформированный из наиболее цитируемых документов по микробиологии. Наиболее цитируемые работы рассматривались как наиболее востребованные научным сообществом. Из этих документов в отдельный массив отбирались авторские КС и определялась частотность их встречаемости. Все КС были переведены на русский язык. Таким образом, был заложен фундамент для формирования онтологии.

Массив публикаций с наибольшей цитируемостью формировался по следующей схеме. Сначала был сформирован полный пул сведений о документах по микробиологии за необходимый период, затем производилось нисходящее ранжирование публикаций по их цитируемости. Суммарная цитируемость всего массива документов была разделена на 3 части, в которых суммарная цитируемость документов была приблизительно равна между собой. Для последующего анализа была отобрана первая (верхняя) треть от массива публикаций, на чью долю приходилась треть от всех ссылок. Эти публикации мы принимали как наиболее цитируемые, и дальнейшая работа производилась на этом массиве. Специалистами-микробиологами были исключены «шумовые» термины, после чего все КТ были переведены на русский язык, данные были загружены в модифицированную систему терминологических словарей [5]. Таким образом, был заложен фундамент для формирования предметной онтологии.

По нашему мнению, частота встречаемости КС может являться индикатором интенсивности развития научной темы. Однако частота встречаемости ни в коей мере не может служить критерием оценки значимости научного направления [6].

Результаты исследований на основе данных за 2018–2019 гг. следующие. Было выявлено 5865 документов за 2018–2019 гг. Публикации учитывались однократно с удалением дублированных записей. По числу документов выделяется рубрика «Генетика дрожжей и микроскопических грибов»: она вобрала в себя наибольшее количество документов. Видимо, это вы-

звано развитием программ генетического картирования, целью которых служит получение бактериальных ферментов и лекарств, изучения антибиотической резистентности, развитие биотехнологии. Кроме того, в документах этой рубрики содержится значительная доля уникальных ключевых слов (УКС) – 83 %¹.

К числу активно развивающихся рубрик можно также отнести «Почвенную микробиологию»: 250 документов, в которых присутствует 92 % УКС. На третьем месте – «Геомикробиология»: 248 документов с 79 % УКС. На общем фоне также выделяются рубрики: «Взаимоотношение возбудителя и хозяина» – 235 документов с 80 % УКС и «Генетика бактерий» – 230 документов с 80 % УКС.

Можно предположить, что доля УКС в рубриках может служить индикатором, показывающим широту спектра разнообразия исследований, методов, организмов: чем больше доля таких слов, тем более разнообразными являются исследования по направлению. Проблема выявления наиболее активно развивающихся научных тем является очень актуальной. Слежение за динамикой терминологии научных направлений (в данном случае – Микробиологии) позволяет делать выводы об происходящих внутри направления изменениях. Показательным является анализ такой динамики у наиболее цитируемых публикаций, как наиболее востребованных научным сообществом. На примере разработанного нами рубрикатора «Микробиология» мы показали один из возможных подходов к определению наиболее активно развивающихся научных тем по основе метода частотного распределения ключевых слов.

Информация, сформированная на этом этапе, открывает широкие возможности для проведения наукометрических исследований данной области науки, базирующихся

¹ Уникальное ключевое слово – ключевое слово, учтённое только один раз, несмотря на возможную множественность упоминаний в разных документах. Доля УКС = кол-во УКС / количество всех КС в документах (включая повторы одних и тех же КС в разных статьях) ×100 %.

на сравнительном анализе терминологии, используемой в отдельных тематических направлениях. В качестве единицы анализа выступают КТ, отражающие разные аспекты исследований по данной проблеме (предмет, теоретические результаты, аспекты применения и др.). По изменениям частоты встречаемости единиц анализа в текстах можно делать выводы об изменениях в научных направлениях, выявлять наиболее релевантные изучаемому предмету документы. По нашему мнению, частота встречаемости КТ в репрезентативном массиве публикаций является одним из индикаторов интенсивности развития научной темы. Однако частота встречаемости ни в коей мере не может служить критерием оценки значимости научного направления [7].

Анализ на основе количественных оценок публикаций по различным разделам научной темы данного научного направления, динамика повторяемости и пересечения ключевых терминов, их изменение во времени (путем анализа документов, опубликованных за различные интервалы годов) позволяют делать выводы о наиболее и наименее динамично развивающихся научных темах, их взаимопроникновении, практическом применении результатов теоретических исследований [8].

Сравнительный анализ различных направлений исследований, относящихся к микробиологии, позволил выделить наиболее активно развивающиеся направления, которые характеризуются большим количеством уникальных (неповторяющихся) КТ и т. п. Работы в этом направлении предполагаются продолжить в направлении расширения анализируемой информационной базы за счет привлечения данных из Scopus и РИНЦ.

Исследования выполнены при поддержке грантов РФФИ № 18-00-00294комфи и № 18-00-00372комфи.

Литература

1. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.

2. Трусова Ю. О., Белоозеров В. Н. Представление классификационных систем в виде онтологий. Обзор // Науч.-техн. информ. Сер. 1. Орг. и методика информ. работы. 2015. № 11. С. 34–38.

3. Антопольский А. Б., Белоозеров В. Н., Каленов Н. Е., Маркарова Т. С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информационные ресурсы России, 2018. № 5 (165). С. 22–30.

4. Цветкова В. А., Мохначева Ю. В., Харыбина Т. Н., Бескаравайная Е. В., Митрошин И. А. Пространство знаний: подходы к извлечению знаний из научных текстов // Информационные ресурсы России. 2019. № 2. С. 31–34.

5. Kalenov N., Senko A. Interactive System of Terminological Dictionaries as One of the Elements in the Ontology of Scientific Knowledge // Software Journal: Theory and Applications. 2019. № 4. <http://swsysweb.ru/en/interactive-system-of-terminological-dictionaries.html> (дата обращения 15.07.2020).

6. В. А. Цветкова, Т. Н. Харыбина, Ю. В. Мохначева, Е. В. Бескаравайная, И. Ю. Митрошина Особенности совмещения классификационных систем и формирования массива ключевых слов для определения пространства знаний по микробиологии // Науч. и техн. б-ки. 2019. № 11. С. 25–43.

7. Mengyang Wang, Lihe Chai Three new bibliometric indicators/approaches derived from keyword analysis // Scientometrics. 2018. V. 116. P. 721–750.

8. Цветкова В. А., Каленов Н. Е., Мохначева Ю. В., Митрошин И. А. Предметная онтология единого цифрового пространства научных знаний как источник наукометрических исследований // Информационные ресурсы России. 2020. № 5. С. 47–49.

References

1. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. O edinom cifrovom prostranstve nauchny`x znaniy // Vestnik Rossijskoj akademii nauk. 2019. T. 89. № 7. S. 728–735.

2. Trusova Yu. O., Beloozerov V. N. Predstavlenie klassifikacionny`x sistem v vide ontologij. Obzor // Nauch.-texn. inform. Ser. 1. Org. i metodika inform. raboty`. 2015. № 11. S. 34–38.

3. Antopol'skij A. B., Beloozerov V. N., Kalenov N. E., Markarova T. S. O razvitii terminologicheskoy bazy` danny`x v vide kompleksa otraslevy`x informacionno-poiskovy`x tezaurusov // Informacionny`e resursy` Rossii, 2018. № 5 (165). S. 22–30.

4. Tsvetkova V. A., Moxnacheva Yu.V., Xary`bina T. N., Beskaravajnaya E. V., Mitroshin I. A. Prostranstvo znaniy: podxody`

k izvlecheniyu znaniy iz nauchny`x tekstov // Informacionny`e resursy` Rossii. 2019. № 2. S. 31–34.

5. Kalenov N., Senko A. Interactive System of Terminological Dictionaries as One of the Elements in the Ontology of Scientific Knowledge // Software Journal: Theory and Applications. 2019. № 4. <http://swsysweb.ru/en/interactive-system-of-terminological-dictionaries.html> (data obrashheniya 15.07.2020).

6. Tsvetkova V. A., Xary`bina T. N., Moxnacheva Yu. V., Beskaravajnyaya E. V., Mitroshina I. Yu. Osobennosti sovmeshheniya klassifikacionny`x sistem i formirovaniya massiva klyuchevy`x slov dlya opredeleniya prostranstva znaniy po mikrobiologii // Nauch. i texn. b-ki. 2019. № 11. S. 25–43.

7. Mengyang Wang, Lihe Chai Three new bibliometric indicators/approaches derived from keyword analysis // Scientometrics. 2018. V. 116. P. 721–750.

8. Tsvetkova V. A., Kalenov N. E., Moxnacheva Yu. V., Mitroshin I. A. Predmetnaya ontologiya edinogo cifrovogo prostranstva nauchny`x znaniy kak istochnik nauko-metricheskix issledovanij // Informacionny`e resursy` Rossii. 2020. № 5. S. 47–49.

Сеть классификационных систем как элемент обобщенной предметной онтологии единого цифрового пространства научных знаний

А. В. Шапкин, В. Н. Белоозеров, Е. Ю. Дмитриева

ВИНИТИ РАН

Аннотация. В статье представлена работающая в ВИНТИ сеть классификационных систем с точки зрения её внутреннего устройства и обеспечения возможностей ее применения для решения задачи навигации при поиске на интегральном информационном пространстве разнородных документальных ресурсов. Основное внимание уделено особенностям модели данных системы классификационных схем, вопросам сопоставления независимых классификаторов, инструментальным средствам ведения системы взаимосвязанных классификаций и обеспечения доступа к ним.

Ключевые слова: классификационные системы, сопоставление классификаций, база данных классификаций, ключевые слова, классификационные рубрики, смысловые связи терминов.

A network of classification systems as an element of a generalized subject ontology of a common digital space of scientific knowledge

A. V. Shapkin, V. N. Beloozerov, E. Y. Dmitrieva

VINITI RAS

Abstract. The article presents the network of classification systems operating in VINITI from the point of view of its internal structure, and the possibilities of its application for solving the problem of navigation when searching for heterogeneous documentary resources in the integrated information space. The main attention is paid to the features of the data model of the classification scheme system, the issues of matching independent classifiers, tools for maintaining the system of interrelated classifications and providing access to them.

Keywords: classification systems, classifications matching, classification database, keywords, classification headings, semantic relationships of terms.

Постановка задачи

Навигация в Едином цифровом пространстве научных знаний (ЕЦПНЗ), содержащем разнородные ресурсы, предполагает, в частности, создание прозрачных для пользователей средств направления поиска при переходе от ресурса одного тематического подпространства к ресурсу другого — с адекватным воспроизведением тематического содержания запроса. Классификационные системы, применяемые при создании информационных ресурсов документного типа, могут рассматриваться как один из инструментов навигации, при условии, что они сведены в единую сеть, где между рубриками отдельных классификаций установлены смысловые связи. Такая сеть позволяет переводить индекс одной классификации в индексы других классификаций и, тем самым, формулировать команды поиска в различных базах данных при выполнении запросов пользователей в интегральном информационном пространстве, объединяющем разнородные ресурсы.

В ВИНИТИ РАН с 1999 года функционирует Система классификационных схем (СКС) [1], в которой реализована идея построения сети связанных тематических рубрик, освещающих содержание документов в аспекте различных систем классификации знания. Исходно её построение и развитие были связаны с задачей тематической разметки в институте потока документов при обработке литературы: в нее были включены классификаторы, используемые при индексировании публикаций — ГРНТИ, Рубрикатор ВИНИТИ, УДК. Они составляют основу СКС.

Потребности индексирования разнородных документов и информационной навигации заставили расширять наполнение СКС и включить в нее сторонние системы. В частности, в СКС вошли: Номенклатура специальностей научных работников (ВАК), Международная классификация

патентов (МПК), Классификатор проектов РФФИ, Библиотечно-библиографическая классификация (ББК), Mathematics Subject Classification (MSC – классификация математических дисциплин Американского математического общества), схема областей науки и техники Организации экономического содействия и развития (ОЭСР), рубрикации библиографических баз данных Web of Science, Scopus и другие. Все эти системы представляют собой иерархии тематических рубрик, основанные на родовидовых отношениях понятий. Независимо от принятого для этих систем обозначения их категориальной принадлежности будем их называть **классификаторами**.

Таким образом, в СКС находится множество классификаторов, которые помогают с разных сторон отражать содержание информационных ресурсов, предоставляемых международными, национальными, отраслевыми службами, и могут использоваться в практике информационной навигации, индексирования и поиска документов. Классификаторы различаются как по сфере приложения (наука, техника, экономика, образование, управление хозяйством и пр.), так и по своему устройству. Некоторые классификаторы имеют отклонения от иерархического построения, используют оригинальные методы кодирования рубрик, специфические связи рубрик и т. п.

При построении базы данных СКС и организации соответствующих услуг на ее основе с целью применения в качестве навигационного инструмента было необходимо решить ряд задач теоретического и прикладного характера. К таким задачам относятся:

- разработка строгой формальной модели классификационной схемы с учетом иерархических и полииерархических связей рубрик, с аппаратом управления жизненными циклами рубрик;

- загрузка в базу данных важнейших национальных и международных классификаций информационных ресурсов; оснащение рубрик предметными описаниями;

- создание формального языка сопоставления независимых классификаций с аппаратом отражения смысловых связей между понятиями;
- разработка сервисных средств для ручного сопоставления классификаций; организация работы экспертов по сопоставлению классификаций;
- разработка методов автоматизированного сопоставления классификаций (с использованием результатов статистической обработки баз данных, предметных описаний рубрик и других источников сведений);
- обеспечение средств автоматизированного ведения взаимосвязанных классификаций в общей системе; безусловное поддержание целостности данных и непротиворечивости взаимных ссылок;
- сопряжение СКС с поисковой системой информационного портала, выполняющего функции интеграции разнородных ресурсов при обработке пользовательских запросов.

Все эти задачи практически решены средствами реляционной модели, поддерживаемой СУБД Microsoft-SQL-Server. Она обеспечивает достаточные для наших целей методы поддержки безусловной целостности данных и непротиворечивости взаимных ссылок.

Представление классификатора в реляционной базе данных СКС

Общее описание базы данных классификаторов представлено в [2]. Основные объекты в базе данных СКС — это самостоятельные классификационные схемы, каждая из которых представлена отдельной КС-таблицей, которая имеет уникальное имя и содержит множество описаний рубрик классификатора.

Рубрика является минимально адресуемым объектом в пределах конкретной классификационной схемы. Идентификаторами рубрик являются первичные ключи (Primary Key) записей КС-таблицы — иерархические шифры, постро-

енные в соответствии с подчиненностью рубрик в классификаторе. В пределах каждой КС-таблицы иерархия рубрик строится через атрибут «рубрика-родитель». Обязательность, единственность и наличие рубрики-родителя строго контролируется при помощи стандартных средств SQL.

Таким образом, классификатор в СКС — это древовидная структура, включающая рубрики, находящиеся в отношениях иерархического подчинения. *Шифр рубрики* и *шифр рубрики-родителя* — системообразующие атрибуты в КС-таблице. Помимо них обязательными атрибутами рубрик являются: *оригинальный шифр* в данном классификаторе; *оригинальное название* (на языке классификатора); параметры жизненного цикла — *дата введения*, *дата отмены*, *статус* (текущее состояние: «проектируется», «действует», «предлагается для отмены», «отменена»).

Набор других элементов описания рубрики зависит от характера конкретного классификатора и может насчитывать несколько десятков. Среди них можно отметить перевод оригинального названия на другой язык, альтернативные шифры, примечания по методике индексирования, числовые показатели исползуемости.

Главенство отношения «родитель-потомок» не означает его единственность. Часто в классификации необходимо отражать более сложные взаимосвязи понятий, не укладывающиеся в строгую иерархическую схему. Например, в ГРНТИ наряду с основной иерархией между рубриками существуют полииерархические связи, т. е. подчинение одного понятия двум или более классам, размещенным в разных местах иерархии. В модели данных СКС предусмотрены специальные свойства рубрик, позволяющие представлять дополнительные отношения между понятиями: прямые связи, аспекты, дополнительные тексты, предметные описания. Эти отношения реализованы в виде отдельных таблиц, которые связаны с основными КС-таблицами классификаторов.

Прямые связи рубрики — это указания «эквивалентно ...», «см. ...», «смотри также ...», «отсылка от ...», «дейст-

вует взамен ...» и другие. Прямые связи формализованы, их целостность строго поддерживается средствами реляционной модели. Виды прямых связей стандартизованы.

Аспект – особый случай формального представления сложных взаимоотношений, когда содержание рубрики необходимо уточнить по каким-то направлениям, которые отсутствуют в иерархическом развитии данной рубрики. Часто аспекты можно выделить из примечания к рубрике; иногда через аспекты реализуется метод фасетной классификации – в дополнение к основной иерархической.

Некоторые классификации требуют дальнейших усложнений модели. Например, УДК изобилует пояснительными текстами к рубрикам – это методические указания, примеры сложных индексов, указания по развитию и пр. Они имеют свободную форму, достигают довольно больших размеров и не всегда поддаются формализации. Для этих случаев модель данных СКС предусматривает возможность оснащения рубрик *дополнительными текстами* стандартизованных типов.

Предметное описание рубрики – список дескрипторов, характеризующих данную рубрику. Каждому дескриптору может быть присвоен *вес* в качестве численной меры его значимости в предметном описании рубрики.

Взаимные связи классификаторов реализованы через аппарат отображения. В базе данных СКС сопоставление двух классификаторов (А и Б) представляется парой *отображений*: отображение А на Б, плюс отображение Б на А. Отображение одного классификатора на другой – это дополнительный объект СКС, который характеризуется идентификаторами двух систем рубрик (отображаемой и отображающей) и используемым словарем отношений. Поддерживаются следующие отношения сопоставленных рубрик: «*эквивалентно*», «*включает*», «*входит в ...*», «*пересекается с ...*». На уровне СУБД запись отображений строго формализована; ссылочная целостность поддерживается стандартными средствами и специализированными процедурами SQL.

База данных СКС и ее наполнение

Крупные структурные блоки базы данных СКС — это массив метаданных, собственно классификационные таблицы и связанный с ними, терминологический массив.

Метаданные СКС в формализованном виде содержат информацию о составе и построении системы в целом: список классификационных схем, их свойства и особенности представления, попарные отображения классификаторов друг на друга. Также система включает множество процедур проверки целостности базы данных как хранилища взаимосвязанных классификационных схем.

Кроме того, метаданные включают сведения об информационных ресурсах, использующих те или иные классификаторы, с указаниями их Интернет-адресов и формальных правил формирования поисковых запросов, таких, например, как перечни поисковых полей, операторы усечения, объединения терминов в запросах.

Метаданные необходимы для управления базой данных, для программирования пользовательских интерфейсов и для обеспечения совместимости СКС с внешними системами при предоставлении услуг по внешним запросам.

Массив классификационных таблиц включает собственно КС-таблицы классификаторов и связанные с ними дополнительные таблицы, как это изложено выше.

В настоящее время СКС содержит 18 самостоятельных классификаторов, плюс 60 дополнительных (годовых) версий. В системе установлено 156 попарных отображений между классификаторами.

Ведение классификационной схемы предполагает ввод новых рубрик, редактирование описаний рубрик, установление прямых связей между рубриками, построение отображений на другие классификаторы, экспорт данных во внешние файлы. Выполнение этих функций, а также многих сопутствующих действий, обеспечивает специализированное *клиентское приложение*, работающее с SQL-сервером и установленное на рабочих местах сотрудников, осуществ-

ляющих ведение классификаторов в корпоративной сети ВИНИТИ.

Терминологический массив служит для поддержки нормализованных предметных описаний рубрик всех классификаторов, представленных в СКС. Его основу составляет таблица терминов (слов и словосочетаний), которые могут использоваться в качестве ключевых слов, приписываемых рубрикам.

Между терминами можно устанавливать попарные связи. Поддерживаются следующие виды отношений терминов:

- английский перевод,
- синоним / эквивалент / альтернативный вариант,
- сокращение,
- аббревиатура,
- по смыслу «уже чем ...»,
- по смыслу «шире чем ...»,
- по смыслу «имеет отношение к ...»,
- принадлежит области знания ...

Исходно терминологический массив наполнен ключевыми словами из поисковых образов документов баз данных ВИНИТИ РАН. При этом сформированы предметные описания рубрик Рубрикатора ВИНИТИ. Впоследствии список терминов пополнялся и корректировался научными редакторами при создании и ведении списков ключевых слов для координатного индексирования документов в технологии формирования выпусков и базы данных Реферативного журнала ВИНИТИ. Существенное пополнение терминологический массив получил в 2014–2015 гг. в результате работы группы экспертов над предметными описаниями рубрик ГРНТИ и УДК [3].

Установление развитой системы смысловых отношений между терминами, входящими в предметные описания взаимосвязанных рубрик, приведёт к построению политематический тезаурусно-классификационной системы, которая позволит разрабатывать технологии навигации и поиска информации как по классификационным индексам доку-

ментов, так и по ключевым словам и, в перспективе, по полным текстам.

В последнее время в СКС осуществлена загрузка данных из системы «ТЕРМИН» [4]. Это список терминов с родовидовыми отношениями между ними, а также связями терминов с рубриками ГРНТИ, УДК и ББК.

В настоящее время в терминологическом массиве СКС содержится более 430 тыс. слов и словосочетаний; между ними установлено около 24 тыс. связей. С рубриками различных классификаторов в СКС имеется более 2,5 млн связей.

Сопоставление независимых классификационных схем является нетривиальной задачей, поскольку, по сути дела, — это сопоставление между собой объектов, понятий, явлений, относящихся к разным предметным областям. И предметные области, и принципы классификаций могут различаться настолько существенно, что сопоставление на первый взгляд кажется бессмысленным. Однако для классификаторов документальных систем в областях научно-технической информации, образования, культуры (по крайней мере, для них) постановка задачи сопоставления представляется оправданной. Это подтверждается многолетним функционированием и практической востребованностью системы соответствий ГРНТИ с такой специфической системой как УДК, которые встроены в эталонные таблицы рубрикатора.

В настоящее время на практике опробованы три источника сведений для сопоставления классификационных схем и, соответственно, выделены четыре типа сопоставлений.

1. *Базисное сопоставление.* Построение некоторого классификатора исходно базируется на других классификаторах более общего характера. При этом автор новой классификации изначально знает, как ее рубрики связаны с базовой классификацией, и может занести эти сведения в систему без каких-либо дополнительных усилий.

2. *Экспертное сопоставление* — получаемое в результате «ручной» обработки, когда эксперты рассматривают рубрику за рубрикой одной схемы и сопоставляют их с рубриками другой схемы. Как правило, прямого соответствия между рубриками установить не удается и приходится использовать

специальный язык квалификаторов и комбинированных кодов, позволяющий в какой-то степени формально представлять подмножества отображающих рубрик и характеризовать степень соответствия.

3. *Индуктивное сопоставление* — получаемое в результате статистической обработки множества фактов, извлекаемых из баз данных соответствующих информационных систем. При этом используется наличие многопрофильного индексирования документов (по разным классификациям). Обработывая большие массивы описаний документов, можно получать не только пары связанных рубрик, но и количественно оценивать значимость связей. Полученные результаты могут быть использованы напрямую или в качестве гипотез представлены специалистам для принятия окончательных решений.

4. *Логическое сопоставление* — сопоставление двух лексических единиц, получаемое в результате формального логического анализа их смысловых связей с другими элементами СКС. Этот метод использовался для установления связей между рубриками классификаторов на основе их связей с ГРНТИ [5]. Общий алгоритм для этой процедуры описан (см. [6]) группой исследователей Университета Тренто (Италия).

В дополнение к перечисленным методам следует отметить, что весьма продуктивным может быть автоматизированное сопоставление классификаторов на основе сравнения предметных описаний рубрик.

Заключение

Реализованная в ВИНТИ РАН сеть взаимосвязанных классификаций и ключевых терминов, описывающих содержание информационных ресурсов в широкой области научной и технической информации, может служить основой для лингвистического обеспечения информационного поиска в едином пространстве научного знания [7] и в его тематических подпространствах. Таким образом, описанная здесь технология интеграции лингвистических средств информационного поиска входит в состав работ по проекту РФФИ 20-07-00103 «Разработка методологии навигации и поиска знаний

в гетерогенной сетевой среде на основе универсального интеллектуального конвертера метаданных».

Литература

1. Шапкин А. В. Современные технологии в производстве информационных продуктов // Научно-техническая информация. Сер. 2. 2003. № 12. С. 1–18.
2. Гиляревский Р. С., Шапкин А. В., Белоозеров В. Н. Рубрикатор как инструмент информационной навигации. Санкт-Петербург : Профессия, 2008. 352 с.
3. Каленов Н. Е., Белоозеров В. Н. Формирование терминологических словарей по лексике классификационных систем // Научно-техническая информация. Сер. 1. 2015. № 3.– С. 60–70.
4. Kalenov N., Senko A. Interactive System of Terminological Dictionaries as One of the Elements in the Ontology of Scientific Knowledge // Software Journal: Theory and Applications. 2019. № 4. <http://swsysweb.ru/en/interactive-system-of-terminological-dictionaries.html> (дата обращения 15.07.2020).
5. Белоозеров В. Н., Шабурова Н. Н. Метод сопоставления библиографических классификаций на основе соответствий с ГРНТИ (на примере УДК и ББК) // Научно-техническая информация. Серия 2. 2016, № 10. С. 13–24.
6. Giunchiglia F., Soergel D., Maltese V., Bertacco A. Mapping large-scale knowledge organization systems : Technical Report # DISI-09–29 / University of Trento. May 2009. <https://clck.ru/SJCuK>
7. Антопольский А. Б. [и др.]. Принципы построения и структура Единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.

References

1. Shapkin A. V. Sovremennyye tehnologii v proizvodstve informacionny`x produktov // Nauchno-texnicheskaya informaciya. Ser. 2. 2003. № 12. S. 1–18.
2. Gilyarevskij R. S., Shapkin A. V., Beloozerov V. N. Rubrikator kak instrument informacionnoj navigacii. Sankt-Peterburg : Professiya, 2008. 352 s.
3. Kalenov N. E., Beloozerov V. N. Formirovanie terminologicheskix slovarej po leksike klassifikacionny`x sistem // Nauchno-texnicheskaya informaciya. Ser. 1. 2015. № 3. S. 60–70.
4. Kalenov N., Senko A. Interactive System of Terminological Dictionaries as One of the Elements in the Ontology of Scientific Knowledge // Software Journal: Theory and Applications. 2019. № 4. <http://swsysweb.ru/en/interactive-system-of-terminological-dictionaries.html>

web.ru/en/interactive-system-of-terminological-dictionaries.html (дата обращения 15.07.2020).

5. Beloozerov V. N., Shaburova N. N. Metod sopostavleniya bibliograficheskix klassifikacij na osnove sootvetstvij s GRNTI (na primere UDK i BBK) // Nauchno-texnicheskaya informaciya. Seriya 2. 2016, № 10. S. 13–24.

6. Giunchiglia F., Soergel D., Maltese V., Bertacco A. Mapping large-scale knowledge organization systems : Technical Report # DISI-09-29 / University of Trento. May 2009. <https://clck.ru/SJCuK>

7. Antopol'skij A. B. [i dr.]. Principy` postroeniya i struktura Edinogo cifrovogo prostranstva nauchny`x znaniy (ECzPNZ) // Nauchno-texnicheskaya informaciya. Ser. 1. 2020. № 4. S. 9–17.

**Создание информационного ресурса
«Электронные архивы ИЭА РАН»; технологии
и результаты**

***Д. А. Функ, М. Б. Лейбов,
Р. М. Галеев, Н. В. Хохлов***

ИЭА РАН им. Н. Н. Миклухо-Маклая

Аннотация. В статье приводится краткое описание сфер применения современных цифровых технологий для решения научных задач Института этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН (ИЭА РАН). В первую очередь, это создание электронных архивов фото- и видеоматериалов, создание электронной базы данных реконструкций. Второе важнейшее направление — это формирование электронной библиотеки публикаций по тематике Института, а также участие в таких важных совместных проектах как создание электронной библиотеки «Научное наследие России». Одной из перспективных областей применения новейших IT-технологий в исследованиях института является 3D-сканирование объектов исследования разного масштаба: от зубов до древних крепостей, что существенно расширяет возможности научного анализа материалов, а также является стимулом к созданию соответствующих баз данных.

Ключевые слова: электронный архив, электронная библиотека, IT-технологии, этнология, антропология.

**Creation of an information resource: Electronic
archives
of the IEA RAS. Technology and results**

***D. A. Funk, M. B. Leybov,
R. M. Galleev, N. V. Khokhlov***

N. N. Miklukho-Maklay IEA RAS

Abstract. The article provides a brief description of the use of modern digital technologies to solve the scientific problems of the Mikluho-Maklay Institute of Ethnology and Anthropology RAS (IEA RAS). The results of application of IT- technologies in various scientific areas of the IEA RAS are considered. First of all, it is the creation of electronic archives of photo and video materials, «paper» archive, creation of an electronic database of reconstructions. The second most important area is the building of an electronic library, as well as participation in such important joint projects as the creation of the electronic library «Scientific Heritage of Russia.» One of the promising applications of modern IT technologies in the institute's research is 3D scanning of research objects of different scales: from the teeth of primitive people to the ancient fortresses of Crimea, which significantly expands the possibilities of scientific analysis of materials, as well as is an incentive to create appropriate databases.

Keywords: electronic archive, digital library, IT technologies, ethnology, anthropology.

Современный научный процесс немислим без применения цифровых информационных технологий. Интернет и сопутствующие ему скоростные поисковые инструменты создали небывалые, практически безграничные возможности для расширения информационного поля человека, послужили мощнейшим стимулом и двигателем многих сфер его деятельности. Ученые, безусловно, одними из первых оценили возможности, предоставленные им IT-технологиями. Скорее всего, этнологи и антропологи не были пионерами применения этих технологий, тем не менее, в настоящее время понимание важности интеграции в «цифровой мир» очевидно всем без исключения, и этот процесс идет семимильными шагами, охватывая все сферы научных интересов сотрудников Института этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН (далее — ИЭА РАН). Об этом, в частности, свидетельствует появление персональных сайтов научных сотрудников института, а также сайтов различных его подразделений. Помимо официального сайта <http://iea-ras.ru/> в институте есть десять ассоциированных сайтов, освещающих отдельные аспекты научной и общественной деятельности сотрудников и подразделений института в целом.

В настоящее время в ИЭА РАН современные цифровые технологии применяются в трех основных стратегических

направлениях. Первое — это сохранение научных архивов, второе — обеспечение доступности научных данных и их распространения, и третье — непосредственно научно-исследовательская деятельность.

Одним из важнейших направлений «цифровой жизни» института стало создание информационного ресурса, содержащего оцифрованные архивы и соответствующую базу данных. Объемы накопленной за 85 лет существования института информации на традиционных носителях — бумаге, фото пленке, кино- и видеопленке, которые хранятся в архивах, весьма значительны. Так, в фотоархиве хранятся фотоматериалы в виде стеклянных пластинок, черно-белых и цветных пленок и отпечатков различного формата в объеме не менее 3 млн кадров. Самые старые из них сделаны в экспедициях графа А. А. Бобринского 1895–1896 годов. Сохранность этих материалов в среднем хорошая, хотя примерно 10 % из них требуют серьезной ретуши при подготовке к печати.

«Бумажный» архив содержит главным образом экспедиционные отчеты и, согласно приблизительным оценкам, в нем около 1 000 000 страниц. Среди них — уникальные материалы, такие как, например опросные листы этнографической комиссии ОЛЕЭА, заполненные от руки в конце XIX века. Особое место среди рукописных полевых материалов занимают такие уникальные коллекции, как, например записи фольклора народов Севера Б. О. Долгих, самозаписи героического эпоса последнего шорского сказителя Владимира Таннагашева (из полевых материалов Д. А. Функа), объемом более 2 тысяч страниц, и многие другие.

В архиве кино- и видеоматериалов сосредоточено около 20 000 часов полевых записей, не считая законченных этнографических фильмов. Началом съемочных работ, ставшим исходным пунктом в создании киноматериалов Архива, можно считать сентябрь 1946 года, когда в Хорезмской археолого-этнографической экспедиции Института этнографии АН СССР для решения вопроса о местонахождении древних оросительных сооружений и средневековых торго-

вых путей была впервые применена аэрофоторазведка с использованием кино- и фотосъемок.

География кино- и аудиоматериалов Архива ИЭА РАН — от Чукотского полуострова до Самарканда, от Ганга до Таймыра, от Индонезии до Кубы. Из-за несовершенства технологий все фильмы и незавершенные кинопроекты на 16-ти и 35 миллиметровой киноплёнке, аудиозаписи образцов фольклора различных этнических групп долгие годы были неизвестны и абсолютно недоступны не только широкой аудитории, но и узким специалистам — этнологам и фольклористам, хотя потребность в них существовала всегда. Для их систематизации и научного анализа был создан центр аудиовизуальных исследований (с 2019 года — Центр визуальной антропологии) ИЭА РАН.

Начиная с 2001 года, основной деятельностью Центра стало восстановление и перевод архивных кино- и аудиоматериалов на цифровые носители. При тесном сотрудничестве с Всероссийским государственным архивом кино- и фото документов ((ВГАКФД) г. Красногорск) материалы Архива были классифицированы, аннотированы и переведены на современные носители. За годы работ было оцифровано 87,5 часов киноматериалов архива ИЭА РАН. Среди них — как законченные фильмы, произведенные на базе и при участии Института в 1946–1991 гг. (Россия, Куба, Индонезия, Эфиопия и др.), так и рабочие материалы, снятые во время экспедиций. Часть киноматериалов аннотирована и передана на ответственное хранение в ВГАКФД для изучения и использования в образовательных, просветительских целях. Чтобы обеспечить доступность этих материалов, был создан интерактивный архив киноматериалов экспедиций Института на базе YouTube («АРХИВ ЦЕНТРА ВИЗУАЛЬНОЙ АНТРОПОЛОГИИ ИЭА РАН»). Для оцифровки архива киноматериалов Института создана специальная установка. Оцифровка проходит в режиме покадрового сканирования, что дает возможность использования как отдельных кадров высокого разрешения, так и создания секвенции кадров и небольших клипов для размещения на сайтах Института. Основной формат оцифровки JPEG, TIFF, TGA

с разрешением конечного кадра от 1920×1080 HDV до формата 6К — 3000×2000 пикселей на дюйм. Хранение видеоматериалов осуществляется на цифровых носителях большой емкости, так как один ролик размера 10 мин составляет в формате 6К около 72 Гб.

Отдельная задача института — оцифровка книжного фонда библиотеки ИЭА РАН. Среди прочего, в ней сосредоточены малотиражные периодические издания и книги по этнографии и физической антропологии, многие из которых представляют непреходящую культурную и научную ценность.

Цели, которые мы преследуем, создавая цифровые копии материалов архивов и книжных фондов, очевидны:

— во-первых, нам необходимо решить задачу вечного хранения материалов, которые подвержены естественным процессам деградации и разрушения;

— во-вторых, мы намерены предоставить доступ к этим материалам всем заинтересованным лицам: профессиональным ученым, общественным деятелям и любителям, интересующимся предметом. Накопленный материал — результат деятельности нескольких поколений талантливых, неординарных, инициативных ученых. Архивы содержат бесценный материал, который должен стать доступным общественности.

В настоящее время работа по оцифровке активно ведется во всех трех архивах, их материалы регулярно пополняют базу данных, содержание которой доступно на ассоциированных сайтах — фотоархива (photo.iea.ras.ru) и архива центра визуальной антропологии ИЭА РАН (<https://www.youtube.com/channel/UC55QUpVbty5R6T8CM6ff9Cg>, <https://www.facebook.com/sharer/sharer.php?u=https%3A%2F%2Fwww.facebook.com%2Fstefania.zini%2Fposts%2F10157627295218215&display=popup&ref=plugin&src=post>).

Фольклорные материалы народов Севера и Сибири, как в аудиоформате, так и рукописные, частично представлены в не имеющей анало-

гов открытой полнофункциональной базе данных «Корпусы ИЭА РАН»: <http://corpora.iea.ras.ru/corpora/>.

Оцифрованные книги поступают в базу данных и их можно увидеть на сайте института в разделе «книжная полка» и на сайте «Научное наследие России» <http://e-heritage.ru/index.html>

Важную роль в становлении и развитии программы института по цифровизации архивов играет многолетнее плодотворное сотрудничество с Межведомственным суперкомпьютерным центром (МСЦ) РАН, благодаря которому институт получил доступ к современной сканирующей технике и программному обеспечению.

В 2019 году была разработана Стратегия развития института на период до 2024 года. Важное место в данной программе отведено развитию 3D-технологий в научно-исследовательской деятельности института. Программа развития была одобрена Министерством образования и науки, и в 2019 году было получено бюджетное финансирование на расширение приборной базы института, позволявшее проводить современные научные антропологические исследования. В настоящее время приборная база состоит из двух сканеров: портативного ручного метрологического сканера Artec Spider и 3D-контактного сканера высокой точности MicroScribe. Этого достаточно как для того, чтобы осуществлять оцифровку скульптурной коллекции Института, так и для того, чтобы проводить новейшие морфометрические исследования. Планируется приобретение специализированного оборудования для оцифровки одонтологической коллекции Института.

Широкое применение 3D-технологии получили в рамках работы Лаборатории антропологической реконструкции. С помощью 3D-сканирования и рентгеновской компьютерной томографии получают цифровые высокоинформативные копии палеоантропологических находок, которые в дальнейшем печатаются (материализуются) с помощью различных аддитивных технологий. На полученных копиях ведутся работы по пластической реконструкции лица

по черепу, или же восстановление облика производится в виртуальной среде.

Отдельно стоит отметить использование технологий рентгеновской томографии, которое нашло широкое применение в палеоантропологических исследованиях. Ввиду отсутствия в Институте самого аппарата (компьютерного томографа) было налажено тесное сотрудничество Института с российскими разработчиками программного обеспечения Инобитек (Воронеж), благодаря чему используются специализированные программы и алгоритмы для извлечения научных данных.

Академические собрания: новый сетевой инструмент филологических исследований

С. Н. Гуськов

*ИРЛИ РАН (Пушкинский Дом)
НИУ ВШЭ*

Аннотация. В статье рассматривается новый электронный ресурс, созданный учеными Пушкинского Дома при участии ВШЭ. Электронная библиотека академических собраний сочинений русских писателей включает около 250 томов. Ресурс обеспечивает сохранность и доступность авторитетных изданий классики, удобство их использования в научной и педагогической работе. Сетевой адрес: www.russian-literature.org

Ключевые слова: академическое издание; полное собрание сочинений; ПСС; классика; литература; комментарий; текстология; полнотекстовая информационная система.

При финансовой поддержке РФФИ. Проект «Гончаров. Служебный круг». № 18-012-00598 А.

Academic editions of Russian writers. Digital library of Pushkinskij Dom

S. N. Guskov

*IRLI RAS
NRU HSE*

Abstract. The article discusses a new electronic resource created by scientists from the Pushkin House with the participation of the HSE. The electronic library of academic collected works of Russian writers includes about 250 volumes. The resource ensures the safety and availability of authoritative editions of the classics, ease of use in scientific and pedagogical work (www.russian-literature.org).

Key words: academic edition; complete works; PSS; classic; literature; a comment; textual criticism; full-text information system.

Выход в свет Академических собраний сочинений — один из наиболее значимых для филологической науки результатов деятельности Института русской литературы (Пушкинского Дома) РАН. Академическое полное собрание сочинений и писем является эталонным типом филологического издания. Его отличительные особенности — полнота и точность представления всего литературного, критического, публицистического, служебного и эпистолярного наследия писателя; учет всех рукописных и печатных источников текста, исчерпывающий научно-справочный аппарат. Тип современного академического издания сформировался в стенах Пушкинского Дома в процессе работы над полными собраниями А. С. Пушкина и Н. В. Гоголя в 1930-х — 1950-х гг. С тех пор сотрудниками Пушкинского Дома подготовлено более 40 академических собраний сочинений (включая переиздания), в том числе собрания А. С. Пушкина, М. Ю. Лермонтова, А. С. Грибоедова, Н. В. Гоголя, В. Г. Белинского, И. С. Тургенева, М. Е. Салтыкова-Щедрина, Ф. М. Достоевского и других.

Пушкинский Дом в настоящее время продолжает выпускать академические собрания сочинений, но в то же время работает над электронной адаптацией уже вышедших изданий. В рамках этой работы была создана электронная библиотека **«Академические собрания сочинений»**. Электронная библиотека включает многотомные собрания сочинений и писем русских писателей XVIII–XIX вв., подготовленные как в Институте русской литературы Российской академии наук, так и в Институте мировой литературы, адаптированные для цифровой среды за счет расширенных поисковых и сервисных опций. О возможностях цифровых издания (на примере 90-томного собрания сочинений Л. Н. Толстого) [1].

Создание цифровой библиотеки было начато при поддержке Программы Секции литературы и языка ОИФН РАН «Язык и информационные технологии» (2015–2017) и продолжается и сейчас. В 2020 г. ресурс был отмечен в ряду достижений академических институтов в отчете Президента РАН А. М. Сергеева.

В настоящий момент в библиотеке представлены академические собрания сочинений В. Г. Белинского, А. И. Герцена, Н. В. Гоголя, И. А. Гончарова, Н. А. Добролюбова, Ф. М. Достоевского, В. В. Капниста, М. Ю. Лермонтова, Н. А. Некрасова, Ф. Прокоповича, А. С. Пушкина, А. Н. Радищева, М. Е. Салтыкова-Щедрина, И. С. Тургенева, Г. И. Успенского, Н. Г. Чернышевского — всего около 250 томов.

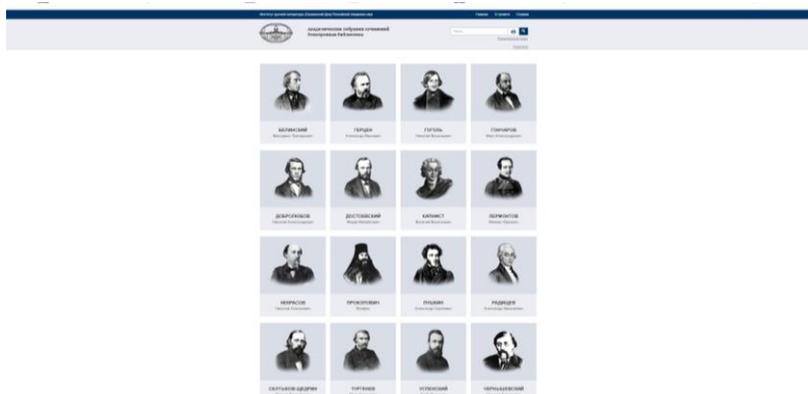


Рис. 1 Главная страница сайта www.russian-literature.org

Создавая библиотеку собраний сочинений, мы учитывали опыт предшественников, прежде всего принципы создания ФЭБ. См. о них [3].

Подача текстов

Представление текстов в библиотеке ориентировано, в первую очередь, на их использование в научной и научно-образовательной работе. Пользователь имеет возможность использовать для цитирования распознанный текстовый слой и, в то же время, сверять его по точной pdf-копии печатного издания с оригинальной пагинацией и полиграфическим оформлением. Реализована также возможность моментального оформления библиографических

ссылку как для печатных, так и для электронных публикаций.

Разметка текстов

Структура разметки учитывает специфику академических изданий, оснащенных обширным комментаторским и справочным аппаратом, включающим:

- рукописные и печатные варианты и редакции;
- историко-литературный и реальный комментарий к тексту;
- перевод иноязычных фрагментов произведения;
- указатели (именные указатели, указатели произведений, географические указатели и т. д.).

Соответствующие варианты и редакции, комментарии, переводы и т. д. привязаны к текстам произведений и писем, что значительно облегчает внутреннюю навигацию. Пользователь имеет возможность параллельного просмотра основного текста и рукописных или печатных вариантов и редакций — таким образом, открываются большие возможности как для изучения, так и для преподавания текстологических сюжетов. В структуре также реализована возможность параллельного просмотра основного текста и комментариев к нему.

Поиск

Пользователю доступен как простой поиск по слову или словосочетанию во всем массиве представленных собраний сочинений, так и поиск по сочетанию нескольких параметров.

Можно искать по одному или нескольким собраниям сочинений, по одному или нескольким авторам, а также исключить из поиска определенных авторов или собрания. Можно ограничить поле поиска отдельным произведением или определенным томом.

Имеется ценная возможность добавления поисковых параметров. Например, вы можете добавить произвольное количество раз поисковый параметр «**Автор**» и таким образом искать слово/словосочетание по интересующей Вас вы-

борке авторов. Чтобы найти, скажем, все упоминания Вольтера в собраниях сочинений В. Г. Белинского и Н. В. Гоголя, нужно задать в поле «Текст» имя «Вольтер», в поле «Автор» – «Белинский», затем добавить поисковый параметр «Автор» и задать в нем «Н. В. Гоголь». Остальные параметры для удобства можно удалить. Они вновь появятся при формировании следующего поискового запроса.

Поиск на сайте пока реализован без учета морфологии, поэтому рекомендуется вносить в поле поиска неизменяемую часть слова.

Форма поисковой выдачи

Полужирным будут показаны названия произведений, в которых найдено искомое слово/словосочетание. Клик на название откроет страницы, на которых встречается искомое слово. На каждой странице искомое слово будет выделено цветом.

Нажав на значок книжки или цепочки, пользователь получит в буфер обмена точную библиографическую или электронную ссылку на найденную страницу.

Указатели

В нескольких собраниях (Лермонтова, Радищева, Некрасова, Достоевского) уже подготовлены такие важнейшие инструменты исследовательской работы, как указатели [3]. Электронные указатели подготовлены на основе существующих в печатных изданиях указателей имен, произведений, географических названий. Была проведена их дополнительная проверка и синхронизация. В электронной форме значительно выросла эффективность работы с указателями, поскольку пользователь имеет возможность сразу видеть все контексты упоминания того или иного имени в текстах всех произведений и писем. Эта опция предоставляет новые возможности для формирования и уточнения биографических списков современников писа-

теля или просто лиц, им упоминаемых.

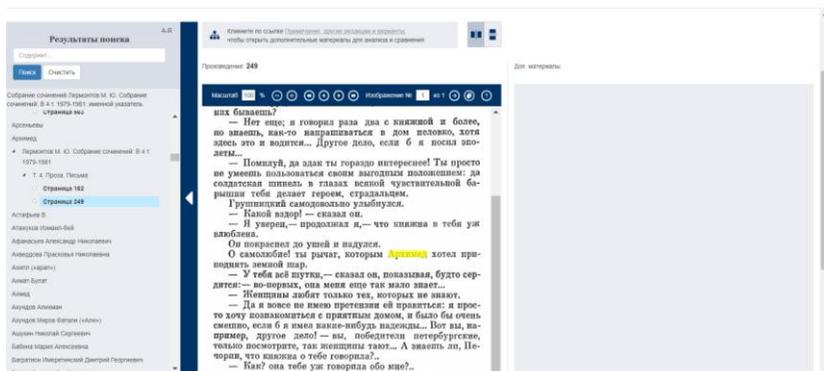


Рис. 2. Именной указатель к Собранию сочинений М. Ю. Лермонтова. В частности, она была использована для создания информационного ресурса. «Гончаров. Служебный круг». В перспективе планируется создание сводного сквозного указателя ко всем академическим изданиям. В подготовке указателей принимают участие сотрудники Пушкинского Дома и студенты ВШЭ.

В 2020 году библиотека пополнилась электронным ресурсом на базе собрания сочинений А. И. Герцена в 30-ти томах (35 книгах). Как и другие собрания сочинений в библиотеке, «электронный Герцен» дает возможность осуществлять лексический поиск, цитировать по точной pdf-копии, мгновенно оформлять библиографические ссылки и т. д.

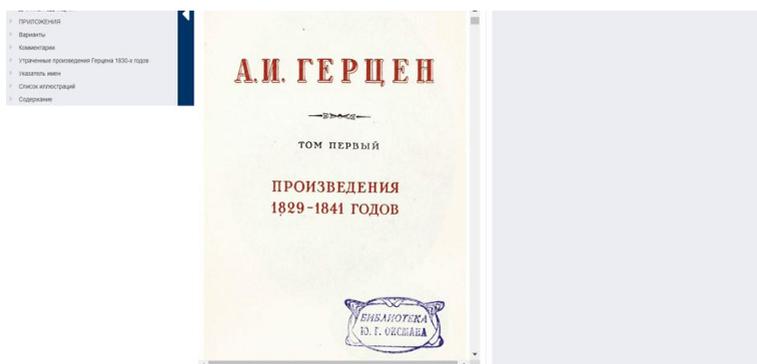


Рис. 3. Электронное собрание сочинений А. И. Герцена

Но у электронного собрания сочинений Герцена есть важное отличие. Его источником стали печатные экземпляры собрания сочинений Герцена, принадлежавшие Ю. Г. Оксману, с его экслибрисом, а также с пометами и маргиналиями. По стандартному регламенту обработки оцифрованных изданий карандашные пометы на полях считаются дефектом и устраняются при окончательной очистке сканов. В данном случае именно этот «дефект» сделал наше электронное издание уникальным, сопоставимым с электронными публикациями архивных источников. Подробнее об этом см. [4].

Работа выполнена при финансовой поддержке РФФИ (проект № 18-012-00598 А).

Литература

1. Bonch-Osmolovskaya A. A., Skorinkin D., Orekhov B., Pavlova I. S., Kolbasov M. G. Tolstoy semanticized: Constructing a digital edition for knowledge discovery // Web Semantics. 2019. № 59. (<https://doi.org/10.1016/j.websem.2018.12.001>)
2. Orekhov B., Fischer F. The 91st Volume – How the Digitised Index for the Collected Works of Leo Tolstoy Adds A New Angle for Research // Digital Humanities 2018: Book of Abstracts / Libro de resúmenes. Mexico : Red de Humanidades Digitales A. C., 2018.
3. Peschio J., Pil'shchikov I., Vigurskii K. Academic Digital Libraries Russian Style: An Introduction to The Fundamental Digital Library

of Russian Literature and Folklore // Slavic & East European Information Resources. 2005. VoReferencesl. 6. № 2-3. P. 45-63

4. Гуськов. С. Н. Маргиналии Ю. Г. Оксмана в собрании сочинений А. И. Герцена // Русская литература. 2020. № 4. С. 268 – 274.

References

1. Bonch-Osmolovskaya A. A., Skorinkin D., Orekhov B., Pavlova I. S., Kolbasov M. G. Tolstoy semanticized: Constructing a digital edition for knowledge discovery // Web Semantics. 2019. № 59. (<https://doi.org/10.1016/j.websem.2018.12.001>).

2. Orekhov B., Fischer F. The 91st Volume – How the Digitised Index for the Collected Works of Leo Tolstoy Adds A New Angle for Research // Digital Humanities 2018: Book of Abstracts / Libro de resúmenes. Mexico : Red de Humanidades Digitales A. C., 2018.

3. Peschio J., Pil'shchikov I., Vigurskii K. Academic Digital Libraries Russian Style: An Introduction to The Fundamental Digital Library of Russian Literature and Folklore // Slavic & East European Information Resources. 2005. VoReferencesl. 6. № 2-3. P. 45-63.

4. Gus'kov. S. N. Marginalii Y. G. Oksmana v sobranii sochineniy A. I. Gertsena // Russkaya literatura. 2020. № 4, (v pechati).

Вопросы цифровой интеграции научно-технических знаний архивов

Г. З. Залаев

РГАНТД

Аннотация. Рассматриваются основные аспекты интеграции в единое пространство знаний архивов и библиотек. Показано состояние фондов и развитие информационных технологий Российского государственного архива научно-технической документации (РГАНТД).

Проанализировано возможность создания единой системы научно-технических знаний на базе РГАНТД, ГПНТБ и ВИНТИ.

Ключевые слова: Российский государственный архив научно-технической документации, РГАНТД, цифровая трансформация, цифровизация, информационные технологии.

Issues of digital integration of scientific and technical knowledge of archives

G. Zalaev

RGANTD

Abstract. The main aspects of integration of archives and libraries into a single knowledge space are considered. It shows the state of funds and the development of information technologies in the Russian state archive of scientific and technical documentation (RGANTD).

The possibility of creating a unified system of scientific and technical knowledge on the basis of RGANTD, GPNTB and VINITI is analyzed.

Keyword: Russian state archive of scientific and technical documentation, RGANTD, digital transformation, digitalization, information technologies.

Архивы всегда представляли собой относительно самостоятельный и обособленный компонент советской и российской информационной инфраструктуры. Это связано, прежде всего, с определенной степенью закрытости архивов для широких кругов, которая складывалась годами,

и корни которой идут от того периода, когда архивная служба входила в состав НКВД и МВД (середина тридцатых — середина шестидесятых годов). Архивный фонд РФ содержит знания, накопленные столетиями, выраженные в документах, имеющие политическое, хозяйственное, научное, социально-культурное или историческое значение.

Особое место в системе федеральных архивов занимает Российский государственный архив научно-технической документации (РГАНТД) (до 1995 года Российский научно-исследовательский центр космической документации — РНИЦКД).

РГАНТД осуществляет государственное хранение научно-технической документации (отчеты НИР и ОКР, чертежи), фонодокументов (включая переговоры борт-земля), кино- и видеодокументов.

Источники комплектования — организации космической отрасли, личные архивы космонавтов и специалистов по ракетно-космической технике, воспоминаний деятелей науки и техники.

В результате реорганизации РГАНТД расширил профиль своего комплектования. В настоящий момент архив комплектуется документами из 20 отраслей народного хозяйства, включая документы предприятий Госкомоборонпрома, документы от научно-технических организаций гражданского и военного профиля, государственной и других форм собственности.

На государственном хранении в РГАНТД находится 692 360 единиц хранения, в том числе: научно-технической документации — 364 936 ед., документов личного происхождения — 40 054 ед., кинодокументов — 8 331 ед., фотодокументов — 105 829 ед., фонодокументов — 9 853 ед.

Источниками комплектования архива являются свыше 500 организаций. Архив активно комплектуется документами личных фондов.

На государственное хранение был принят 3D-документ — 3D-модель телевизионной башни, построенной по проекту Шухова.

Российский государственный архив научно-технической документации является ведущим архивом

по разработке и внедрению современных информационных технологий:

- развивает и поддерживает локальную сеть архива (более 140 рабочих станций);

- обеспечивает поддержку автоматизированных архивных технологий и баз данных архива и отраслевых ресурсов;

- проводит работы по оцифровке и загрузке в системы хранения данных документов на бумажной основе, а также аудиовизуальных документов из фондов РГАНТД;

- обеспечивает программно-техническую поддержку и информационную безопасность Интернет-центра РГАНТД, в состав которого входят 21 сервер, 7 систем хранения данных, коммуникационное оборудование.

На серверах Интернет-центра РГАНТД размещаются официальный сайт Росархива, портал «Архивы России», информационные ресурсы Росархива и федеральных архивов.

РГАНТД разработал и эксплуатирует АИПС с цифровыми копиями документов на платформе ПИК КАИСА.

В АИПС РГАНТД включены 12 информационно-справочных баз данных, интегрированных в единый программно-информационный комплекс.

В настоящее время РГАНТД проводит работы по загрузке информационных систем Росархива в Государственную единую облачную площадку (ГЕОП).

Таким образом, РГАНТД имеет базовую программно-техническую платформу для осуществления интеграции знаний с другими научными центрами.

Попытка интеграции РГАНТД с другими организациями предпринималась и ранее. Это создание Государственной автоматизированной системы научно-технической информации – ГАСНТИ, создание которой было начато в 1970-е гг. в СССР.

В ГАСНТИ входил целый ряд крупных и авторитетных организаций, таких как ГПНТБ, БЕН, ИНИОН, ВИНТИ и т. д.

Современное направление в развитии цифрового общества и цифровой экономики – цифровая трансформация, дает основу и требует по новому взглянуть на интеграцию знаний, которые накоплены в архивах, библиотеках, системе научной и научно-технической информации.

Современные информационные технологии, цифровизация позволяют создать Единую информационную сеть для архивов и библиотек. Стартовым этапом может быть создание «триады» РГАНТД – ГПНТБ – ВИНТИ.

Проблемы:

- 1) финансирование проекта,
- 2) принадлежность организаций к разным ведомствам,
- 3) программно-технические решения.

Единое информационное пространство архивов науки в контексте модели ИСАРАН

И. Н. Ильина

Архив РАН

Аннотация: В статье освещается история создания Информационной системы «Архивы РАН» (ИСАРАН), процесс формирования ее структуры и базового контента, эффект функционирования системы как интеграционной модели единого информационного пространства системы академических архивов. Представлены основные составляющие элементы объединенной базы данных, в том числе, электронный Центральный фондový каталог, а также программное обеспечение, разработанное для унифицированного участия архивов-партнеров в развитии ИСАРАН как корпоративного информационного ресурса. Информационная система охарактеризована как уникальная разработка, рассчитанная на особый и характерный для Академии наук вид архивных документов — научную документацию во всем ее разнообразии (управленческая и научно-организационная документация, результаты исследовательской деятельности во всех областях науки, кино-фото-аудио-документы, личные документы академиков, членов-корреспондентов, выдающихся ученых: рукописи статей, дневники, письма, фотографии; тематические коллекции документов и мн. др.), а также как инструмент целенаправленного поиска в море архивной информации, хранящейся в академических архивах. Констатируется актуальность собраний архивных документов, востребованность электронных баз данных архивов науки и цифровых документов научным сообществом. Анализируется уровень и проблемы взаимодействия научных учреждений, хранящих архивные комплексы, в свете современного состояния российской академической науки.

Ключевые слова: Российская академия наук, Архивы РАН, ИСАРАН, ЦФК, базы данных, архивные фонды.

Unified information space of science archives in the context of the ISA RAS model

I. N. Ilina

Archive RAS

Abstract: The article describes the history of creation of Information system «the archives of the Russian Academy of Sciences» (ISARN), the process of forming its structure and basic content, the effect of the functioning of the system integration model as a common information space systems academic archives. The main components of the unified database, including the electronic Central stock catalog, are presented, and the software developed for the unified participation of partner archives in the development of ISARAN as a corporate information resource is presented. The information system is described as a unique development designed to work with a special and characteristic type of archival documents for the Academy of Sciences – scientific documentation in all its diversity (management and scientific and organizational documentation, research results in all fields of science, film-photo-audio documents, personal documents of academicians, corresponding members, outstanding scientists: manuscripts of articles, diaries, letters, photographs; thematic collections of documents, etc. etc.), as well as a tool for purposeful search in the sea of archival information stored in academic archives. It States the relevance of collections of archival documents, the demand for electronic databases of archives of science and digital documents by the scientific community. The article analyzes the level and problems of interaction between scientific institutions that store archival complexes in the light of the current state of Russian academic science.

Keywords: Russian Academy of Sciences, Archives of the Russian Academy of Sciences, ISA RAS, CFK, databases, archive funds.

1. Введение

Идея единого информационного пространства архивов науки (академических архивных учреждений) возникла в Архиве РАН и начала реализовываться почти 20 лет назад. Собрания академических архивов, являясь частью Архивного фонда Российской Федерации и Архивного фонда РАН, в совокупности составляют специальную сферу научной документации, а потому взаимодействие архивов науки неизбежно и необходимо. Научное значение документальных комплексов ежедневно подтверждается массовым обращением к ним российских и зарубежных ученых. В качестве немаловажных факторов привлекательности для пользователей служат возможности удаленного доступа к архивам, расширенного автоматизированного поиска документов и ознакомления с их цифровыми образами. Инструментами обеспечения этих возможностей в архивах науки служат ло-

кальные базы данных и тематические информационные системы, способные, так или иначе, удовлетворить поисковые и исследовательские запросы специалистов.

Вопрос об информационном объединении ведомственной системы архивных учреждений Российской академии наук к началу 2000-х гг., что называется, витал в воздухе, однако стал решаться в стенах Архива РАН в одностороннем инициативном порядке. По замыслу, специальная информационная система должна быть способна обеспечить интеграцию всех многочисленных научных архивов учреждений РАН, архивов региональных научных центров и Отделений Академии наук в общую программную, структурную и функциональную среду. При этом принимались во внимание весьма скромные материальные возможности архивов-партнеров.

2. История проекта ИСАРАН

Информационная система «Архивы Российской академии наук» (ИСАРАН) была создана Архивом РАН в результате серии последовательных шагов в русле реализации концепции информатизации архивной деятельности в стране. [1]. В качестве перспективной цели рассматривалось развитие ИСАРАН как объединенной базы данных академических архивных учреждений, которые собирают, хранят и используют документальные собрания Архивного фонда РАН [2]. Первоначально ИСАРАН формировалась за счет создания тематических баз данных по документам Архива РАН, сплошной оцифровки описей фондов, сканирования документов и создания специального программного обеспечения (ПО). Осуществление этого спектра работ происходило в Архиве РАН, главным образом, за счет дополнительного (грантового) финансирования.

Первым в этом контексте стал проект «Применение информационных технологий в освоении источниковой базы истории науки (Разработка БД “Архив РАН” на платформе “клиент-сервер”» (РГНФ, проект № 04-03-12004в на 2004-2006 гг., рук. Г. А. Савина) [3]. Результатом работ явилось соз-

дание ядра многофункциональной базы данных (БД) «Архив РАН» как основы информационной системы, обслуживающей сотрудников Архива и историко-научные интересы потребителей через Интернет. Многофункциональная база данных

«Архив РАН» состояла из трех основных разделов, соответствующих направлениям деятельности Архива: учет и обеспечение сохранности документальных материалов (раздел «Учет»), комплектование новыми документальными массивами академических учреждений и материалами личного происхождения (раздел «Комплектование»), использование ведомственного архивного фонда в различных целях (раздел «Каталог»). В основу структуры базы данных и интерфейса ее разделов был положен Международный стандарт архивного описания документов (ISAD), который кроме принципа многоуровневой структуры данных, принятого во всем мире, закрепил органическую связь между архивной деятельностью и основами теории информатики на всех этапах: от создания информации до ее потребления.

Во многом благодаря проекту, в 2006 г. был создан сайт «Архивы РАН», зарегистрированный в качестве средства массовой информации (Свидетельство Росохранкультуры Эл ФС77-26549 от 13.12.2006), на странице которого была представлена многофункциональная БД с возможностью поиска неслужебной информации в режиме реального времени. Контент БД был в тот момент ограничен документальным собранием Архива РАН, но название сайта уже свидетельствовало о перспективе создания ведомственного корпоративного ресурса, а в последующем — банка данных по всему Архивному фонду Российской академии наук.

Далее пополнить контент многофункциональной БД позволили два проекта Архива РАН — «Космос — 2007» (Фонд Форда, проект № 1000-1553-4, рук. Савина Г. А.) и «Создание общедоступного информационного ресурса по документам гуманитарных учреждений Коммунистической академии ЦИК СССР» (рук. Загребаева В. Н., по Программе фундаментальных исследований Отделения

историко-филологических наук РАН «Власть и общество в истории») [3].

В 2009 г. две уникальные разработки Архива РАН – многофункциональная база данных «Учет и каталог» и программное обеспечение «Учет и каталог» – были зарегистрированы в Роспатенте (№ 2009620604 и № 2009617035 соответственно) [1].

Наращиванию контента ИСАРАН в течение ряда лет (2009–2015) способствовал проект «Отбор и сканирование архивных документов, хранящихся в Архиве РАН, для пополнения контента Электронной библиотеки “Научное наследие России”» (рук. к.и.н. Афиани В. Ю.) в рамках целевой программы Президиума РАН по созданию электронной библиотеки «Научное наследие России» (рук. академик Савин Г. И.). Исторические справки о лицах-фондообразователях Архива РАН, библиографические списки их работ, репрезентативные пакеты документов и фотографий ученых, размещенные в Электронной библиотеке, были одновременно присоединены к информационному массиву ИСАРАН и объединены гиперссылками с имеющимися в сети Интернет профильными ресурсами [4].

Решающее значение для ИСАРАН имела реализация проекта «Создание базы данных “Электронные описи Архива РАН”» (РГНФ, проект № 11-01-12045в, рук. к.и.н. Киселев М. Ю.). В 2011–2013 гг. были оцифрованы 1350 описей фондов, хранящихся в Архиве РАН в г. Москве, и в сети Интернет представлены все описи на документальные фонды Архива РАН, что существенно облегчило работу исследователей с архивными документами, особенно в удаленном режиме [5].

Принципиальным моментом формирования ИСАРАН как модели ведомственного информационного ресурса стал проект «Создание единого справочно-информационного банка данных системы архивов РАН (Центральный фондовый каталог)», осуществленный в 2013–2015 гг. (РГНФ, проект № 13-03-12010в, рук. к.и.н. Ильина И. Н.). Для реализации проекта были созданы специальные программные приложе-

ния к ИСАРАН, проведена проверка информации традиционных (бумажных) фондовых каталогов каждого академического архива, а академическим архивам предоставлена возможность самостоятельно формировать электронный научно-справочный аппарат к своим фондам и заниматься полноценной каталогизацией документов собственных архивных собраний с возможностью оцифровки материалов в любом объеме. Весь подготовленный в течение 2013–2015 гг. контент электронного ЦФК Архивного фонда РАН стал доступен потребителям историко-научной информации на сайте ИСАРАН, который пополняется в режиме реального времени и не имеет ограничений к просмотру [6].

Последним по времени в этой серии работ (2016–2017 гг.) стал проект Архива РАН «Практическая реализация концепции веб-портала “Мнемосина” на базе Информационной системы “Архивы РАН” (ИСАРАН)» (РФФИ, № 16-01-12002, рук. к.и.н. Злобин Е. В.). Создание портала в полной мере продемонстрировало интеграционный потенциал ИСАРАН, а включение в объединенную базу данных документальных собраний не только архивов, но и библиотек и музеев РАН, способствовало расширению единого информационного пространства до пределов всех «институтов памяти» Академии наук [7].

3. Программное обеспечение ИСАРАН

Начиная с 2013 г., в системе академических архивов стала распространяться web-версия программного обеспечения (ПО) «Учет и каталог» ИСАРАН с функцией ведения БД в удаленном доступе посредством сетевых технологий, позволяющая редактировать метаданные на всех четырех уровнях архивного описания: фонд — опись — дело — документ.

Специально разработанное в Архиве РАН ПО изначально было ориентировано на перспективную идею создания объединенной базы данных (ОБД), структурно совпадающей с ЦФК Архивного фонда РАН, обеспечивая ее интеграционными инструментами программного приложения «Учет и каталог». Ядро ИСАРАН — база данных «Учет

и каталог» — было построено на международных стандартах архивного описания, унифицированной структуре данных и являлось типовой платформой для всех архивных учреждений РАН. Регистрация ПО «Учет и каталог» в Роспатенте открыла возможность обеспечения архивов-партнеров по объединенной базе данных необходимым лицензированным программным продуктом на безвозмездной основе [3].

Приглашая академические архивы РАН к совместной работе, Архив РАН предложил модификацию программного обеспечения, подходящую конкретному архиву, с учетом степени его вовлеченности в общий проект и технических возможностей рабочих мест — от программного комплекса «ИСАРАН» в полной его комплектации (10 баз данных) до web-версии БД «Учет и каталог». Первые два уровня описания «Фонд» и «Опись» формируют комплекс данных, необходимый для ведения ЦФК Архивного фонда РАН (списки фондов, реестры описей, статистика). Вторые два уровня той же структуры данных «Дело» и «Документ» способствуют наполнению контента электронного каталога всеми участниками корпоративного информационного ресурса во всем разнообразии форматов файлов (в том числе и мультимедийных).

4. Специфика и миссия ИСАРАН

В итоге к настоящему времени ИСАРАН сформировалась как уникальная программно-информационная среда архивов науки с совершенствующейся структурой и идеологией, как разветвленная, постоянно развивающаяся система, регулярно пополняемая электронным контентом. Выполняя, в первую очередь, функцию учета документов, ИСАРАН реализуется как поисковый инструмент в море научной документации. Эта миссия системы является центральной в концепции корпоративной деятельности ведущих архивных учреждений РАН, территориально распространенных по всей стране. Сегодня ИСАРАН — это комплекс, управляемый программным модулем «Гайпер», состоящий из 10 баз данных, которые, автоматизируя основные архивные виды работ (комплектование, описание, учет, хранение и использование архивной информации), интег-

рируют сведения для научно-исследовательской деятельности и решения прикладных архивоведческих задач. В числе этих баз данных: «Учет и Каталог», «Персональный состав РАН», «История учреждений РАН», «Учреждения-фондообразователи», «Лица-фондообразователи», «Читальный зал и использование», «Каталог файлов», «Печатник», «Архивы» и «Сотрудники».

ИСАРАН позиционируется также как основа функционирования информационного ресурса широкого круга «институтов памяти» (библиотек, музеев и архивов академической сферы), что свидетельствует о неисчерпаемости ее возможностей [7, 8].

Web-версия ИСАРАН, размещенная в сети Интернет (www.isaran.ru), доступна для всех заинтересованных пользователей в целях их научно-исследовательской и практической работы.

Объединенная база данных (ОБД) ИСАРАН предоставляет пользователям возможность обращения к научно-справочному аппарату ресурса во всей его полноте (списки фондов, реестры описей, оцифрованные описи, справки и обзоры фондов и др.), а также к графическим образам документов, видео- и аудиофайлам. Электронный каталог ОБД дает возможность сквозного поиска информации по всему комплексу документов Российской академии наук, причем как по документам учреждений, так и личным фондам ученых. Взаимодействие архивных учреждений позволяет формировать тематически близкие массивы информации, необходимые для изучения исторического прошлого РАН, и обеспечивать потребности специалистов в научной документации родственных учреждений страны в одной предметной области [6]. С позиции перспективности развития интеграционных связей между ресурсами архивов-участников ИСАРАН наиболее многообещающими выглядят базы данных «Учет и каталог», «Персональный состав РАН», «История учреждений РАН» [10].

В контексте востребованности ОБД ИСАРАН показательна статистика посещаемости основных сайтов. В среднем от 60 тыс. до 90 тыс. исследователей и других потребителей

архивной информации в год обращаются к сайтам www.arran.ru и www.isaran.ru. Так, например, к началу октября 2020 г. эти сайты посетили около 62 тыс. и 52 тыс. человек соответственно.

5. Центральный фондовый каталог

Переход к созданию объединенной базы данных (ОБД) на основе ИСАРАН был невозможен без реформирования Центрального фондового каталога академических архивов, а именно – перевода его в электронный формат [9]. Электронный Центральный фондовый каталог (ЦФК) Архивного фонда РАН – это централизованный банк интегрированных метаданных информационных ресурсов документальных собраний институтов памяти (архивов, музеев и библиотек РАН) с широкими возможностями общего и отдельного поиска, фильтрации данных, составления учетно-отчетной документации (статистики, паспорта) и научно-справочного аппарата (списки фондов, фондовые каталоги, описи фондов, путеводители), формирования общего контента (электронного каталога) по широкому спектру тематики документальных собраний информационных хранилищ архивов – участников интегрированного ресурса [6].

Известно, что Архивный фонд РАН сложен и многообразен по своему составу. Он включает научную документацию с результатами исследовательской деятельности академических учреждений во всех областях науки; деловую, управленческую, научно-организационную документацию, отражающую многоплановую деятельность по руководству фундаментальными исследованиями в стране; фоно-фотокино-документы и электронную документацию; личные фонды академиков, членов-корреспондентов РАН и выдающихся ученых; тематические коллекции и многое другое. В Архивный фонд РАН входят такие всемирно известные документальные собрания, как фонды Отдела рукописей и Архива востоковедов Института восточных рукописей, собрания Пушкинского Дома, ценнейшие документы по истории русской и западноевропейской литературы Ин-

ститута мировой литературы и Архива А. М. Горького, уникальный комплекс документальных материалов Института археологии РАН, связанных с проведением археологических исследований на территории России и за рубежом. Постоянное хранение документов Архивного фонда Академии наук осуществляют научные архивы отделений и научных центров РАН.

Архивные фонды и коллекции учреждений РАН охватывают весь комплекс научных дисциплин, существующих на протяжении истории человечества.

Первоначально (2013–2015 гг.) ЦФК включал данные 46 архивных собраний учреждений РАН, находящихся в Московском, Санкт-Петербургском, Северо-Западном регионах, Поволжье, Урале, Сибири и Дальнем Востоке. Помимо Архива РАН и его Санкт-Петербургского филиала, к электронному ЦФК подключились архивы таких ведущих учреждений, как Институт мировой литературы им. Горького РАН, Институт археологии РАН, Институт российской истории РАН, Институт русской литературы (Пушкинский Дом), Кунсткамера, а также научные архивы Карельского, Кольского, Казанского, Уфимского, Дагестанского и Коми Научных центров РАН, другие архивы институтов и музеев РАН. Совместные действия участников проекта осуществлялись на основе единого ПО, разработанного и предоставленного им Архивом РАН безвозмездно. В ЦФК заложена возможность перманентного формирования контента, в течение длительного периода времени, и поддержание актуального состояния ведомственного архивного фонда на текущий момент времени. Сегодня в объединенный список ЦФК включены данные 77 архивов, 5536 фондов, 7724 описи. В электронном каталоге представлено около 270 тысяч записей.

Контент ЦФК РАН доступен для всех заинтересованных пользователей в режиме реального времени через глобальную сеть Интернет на специализированном сайте «ИСАРАН» (Информационная система «Архивы РАН»): <http://www>.

arran.ru/?q=ru/isaran) в разделе <http://isaran.ru/?q=ru/archives>

В «Положении об электронном ЦФК», разработанном в целях ведения централизованного государственного учета документов Архивного Фонда РАН в 2013 г., сформулировано понимание ЦФК как ведомственной информационной системы и модели разветвленной сети хранилищ ретроспективной информации. В нем отражены общие требования к формированию ЦФК, его составу и содержанию, зафиксированы права и обязанности участников электронного ЦФК.

Работа над созданием электронного ЦФК выявила общую заинтересованность архивных учреждений в подобном объединенном ресурсе, открытость ЦФК для постоянного расширения контента, его востребованность со стороны пользователей. Однако при этом практика информационного наполнения ЦФК оказалась не свободна от целого ряда проблем.

6. Портал «Mnemosyne» / «Мнемосина»

Портал «Мнемосина» (<http://arran.ru/?q=ru>) стал ожидаемым результатом интеграции «институтов памяти», хранящих документальные комплексы, и показателем долгосрочной перспективы расширения информационного архивного пространства. Объединив содержание трех сайтов (Архив РАН, ИСАРАН и Виртуальный музей), портал был назван Mnemosyne по имени богини памяти Мнемосины (др. греч. *Μνημοσύνη*), матери девяти муз древней Греции. Создание портала происходило в 2016–2017 гг. при финансовой поддержке РФФИ (проект № 16-01-12002 «Практическая реализация концепции веб-портала “Мнемосина” на базе Информационной системы “Архивы РАН” (ИСАРАН)», рук. к.и.н. Злобин Е. В.).

Перспективная идея и стратегия веб-портала «Мнемосина» заключается в поэтапном формировании единого структурированного информационного пространства для академической триады — архивов, библиотек и музеев —

в фокусе документального изучения наследия отечественной науки и культуры, истории Российской академии наук. [12].

На портале «Мнемосина» в единое информационное пространство интегрированы распределенные системы хранения и данные «институтов памяти» РАН (сайты и веб-страницы архивов, музеев и библиотек — хранилищ исторических источников). Они представлены в качестве описаний на сайте «Архивы РАН» в разделе «Институты памяти РАН» в подразделах: система архивов — система музеев — система библиотек — в сочетании с централизованной базой метаданных ЦФК на сайте «ИСАРАН», где в режиме онлайн фиксируется текущая статистика объектов описания (электронных записей) объединенной базы данных ИСАРАН.

Можно сказать, что порталом реализуется концепция единого информационного поля исторической памяти Академии наук: в нем интегрирован интеллектуальный мемориальный ресурс (источники и памятники по истории науки), важный для представления и популяризации, исследования и интерпретации истории науки. Выявленные в ходе мониторинга сведения о составе и содержании документальных материалов, хранящихся в музеях и библиотеках РАН, добавлены к созданному ранее электронному ЦФК ведомственного Архивного фонда РАН. Для исследователей это — удобная точка доступа к научно-справочному аппарату документальных собраний ведомственных «институтов памяти». Более того, единство хранящихся в «институтах памяти» материалов очевидно: их генетически связывает общее происхождение как результатов деятельности Академии наук на протяжении всей ее истории (в событиях, лицах, открытиях, институциональных реорганизациях).

Формирование портала позволило обозначить значительный круг новых учреждений-участников в работе над общим, тематически связанным контентом «Мнемосины», базирующимся на функциональных возможностях Информационной системы «Архивы РАН».

Составной частью портала является сайт «Виртуальный музей истории РАН» (<http://virtmus.arran.ru/ru/welcome>), где размещен мультимедийный контент раздела «Музеи

РАН». Он основывается на идее демонстрации главных научных открытий, изобретений и достижений выдающихся ученых, научных коллективов, учреждений и руководящих органов Академии наук, их реальной роли в истории страны и мирового сообщества; преимуществ академической формы организации научных исследований, ее места в формировании научного, технического, экономического и социокультурного потенциала России. Все это, на наш взгляд, должно способствовать созданию научно-обоснованного и целостного образа истории Российской академии наук.

Тематическая структура экспозиции формируется из сведений о событиях (создание Академии, переименования, реорганизации ее и руководящих органов, создание и реорганизация отделений, ведущих научных учреждений), деятельности, важнейших открытиях и изобретениях, портретной галереи выдающихся ученых — членов и сотрудников РАН, а также связях с зарубежными институциями науки в общем мировом процессе эволюции научных знаний. В экспозиции виртуального музея представлены ключевые, уникальные экспонаты, предметы, документы и изображения, раскрывающие место Академии наук в истории отечественной и мировой науки.

7. Современное состояние и проблемы развития

ИСАРАН

Архивный фонд Академии наук начал формироваться практически одновременно с созданием самой Академии. За почти 300-летнее существование Российской академии наук ее учреждения собрали богатейшие материалы по истории науки, культуры, искусства, а в научно-отраслевых архивах учреждений представлены документы разных эпох на многочисленных языках мира (в том числе древних), коллекции археологических экспедиций, тематические собрания, материалы по увековечению памяти лиц и событий. По этой причине создание единого справочно-информационного банка данных системы архивов Россий-

ской академии наук (Центральный фондový каталог) можно рассматривать в качестве серьезной попытки информационного охвата части единого целого, заполнить лауну в общем информационном пространстве, которая длительное время оставалась мало доступной большинству заинтересованных людей.

Однако в ходе создания объединенного ресурса первой обнаружилась собственно проблема поставки информации, в т. ч. полноты предоставляемых архивами-участниками данных. Не будем скрывать, что формирование электронного ЦФК было и в целом остается делом рук сотрудников Архива РАН в Москве. Предложенный Архивом РАН для всеобщего пользования алгоритм работы не был воспринят однозначно в ряде архивных учреждениях. В некоторых архивах по собственному усмотрению ограничили передаваемую информацию, не представив списки фондов и не открыв доступ к описям, что, безусловно, отрицательно сказывается на качестве и масштабах корпоративного ресурса. На наш взгляд, эта проблема — не только результат традиционной замкнутости некоторых архивных организаций, отсутствия мотивированного стремления вести дополнительную работу, но также следствие целого круга взаимосвязанных обстоятельств, в которых оказались академические архивные учреждения в наши дни.

Проблемы работы по поддержанию и развитию корпоративного ресурса архивных учреждений РАН совершенно одинаковы и повсеместны — это проблемы организационного, нормативно-правового, кадрового, материально-технического и финансового характера. При наличии программно-технических возможностей созданных Архивом РАН для формирования объединенного информационного ресурса на основе документальных собраний Академии наук, Архив РАН и его партнеры ограничены в своих действиях. Негативное воздействие вышеназванных причин на архивную сферу особенно ощутимо в последние годы: в условиях реформирования академической науки и ее неоднократного переподчинения система архивных учреждений РАН структурно и юридически «рассыпалась»,

ослабили внутренние организационные и деловые связи, нарушены единые технологические процессы, сказывается отсутствие общего руководства архивным делом в подведомственных организациях со стороны Министерства науки и высшего образования РФ. Существенную роль играют нехватка целевого финансирования и технического обеспечения информационной работы в архивах, возникновение кадровых «провалов» в части профессиональных кадров архивистов и IT-специалистов и др. В настоящее время работа над корпоративным архивным проектом, не будучи составным элементом государственного задания, ведется только во внеплановом порядке крайне узким кругом специалистов-энтузиастов, а пополнение информационного контента практически остановлено. Дальнейшее осуществление этой работы возможно только в русле восстановления и упорядочения ведомственной академической архивной системы.

Литература

1. ИСАРАН-web: <http://www.arran.ru/?q=ru/isaran> (Дата обращения 03.11.2020).
2. Савина Г. А. Интеграционные возможности Информационной системы «Архивы РАН» в создании корпоративного электронного ресурса ретроспективной и текущей информации // *Фундаментальная наука: проблемы изучения, сохранения и реставрации документального наследия. Материалы Межд. научн. конференции, Москва, Архив РАН, 4–7 июня 2013 г. М.: Архив РАН, 2013. С. 212–219.*
3. Загребаяева В. Н., Злобин Е. В., Савина Г. А. 10 лет Информационной системе «Архивы РАН» (ИСАРАН): итоги и перспективы. <http://arran.ru/?q=ru/node/746> (Дата обращения 03.11.2020).
4. Афиани В. Ю., Ильина И. Н. Интеграция электронных ресурсов Архива РАН в Электронную библиотеку «Научное наследие России» // *Информационное обеспечение науки: новые технологии. Сборник научных трудов. М., Научный мир, 2009. С. 81–88.*
5. Каленов Н. Е., Сотников А. Н., Ильина И. Н. Архивная информация в Электронной библиотеке «Научное наследие России» // *Фундаментальная наука: проблемы изучения, сохранения и реставрации документального наследия. Материалы Международной научной конференции. Москва, 4–7 июня 2013 г. М: Архив РАН, 2013. С. 25–35.*

6. Киселев М. Ю. Создание базы данных «Электронные описи Архива РАН» // *Фундаментальная наука: проблемы изучения, сохранения и реставрации документального наследия: Материалы Международной научной конференции. Москва, 4-7 июня 2013 г. М.: Архив РАН, 2013. С. 128-130.*
7. Загребаева В. Н., Злобин Е. В., Савина Г. А. База данных «Центральный фондовый каталог» архивов РАН-ФАНО // *Вестник архивиста. № 4. 2015. С. 282-292.*
8. Портал «Мнемосина»: <http://www.arran.ru/?q=ru>. (Дата обращения 03.11.2020).
9. Злобин Е. В., Загребаева В. Н., Савина Г. А. Портал «Мнемосина» – пример реализации единого информационного пространства // *Информационный бюллетень ассоциации История и компьютер. 2016. № 45. С. 208-210.*
10. ЦФК РАН: <http://isaran.ru/?q=ru/cfk>. (Дата обращения 03.11.2020).
11. Загребаева В. Н., Злобин Е. В., Савина Г. А. База данных «Персональный состав РАН» – структура и содержание // *Документация в информационном обществе: нормативно-методическое обеспечение управления документами. Доклады и сообщения на XXI Международной научно-практической конференции 18-19 ноября 2014 г. М., 2015. С. 416-420.*
12. Положение об электронном ЦФК: <http://isaran.ru/?q=ru/node/68> (дата обращения 03.11.2020).
13. Академические институты памяти на веб-портале «Мнемосина»: хранилища документальных собраний в едином информационном пространстве. <http://www.arran.ru/?q=ru/node/829> (дата обращения 03.11.2020).
14. Загребаева В. Н., Злобин Е. В., Савина Г. А. Информационный портал «Мнемосина»: история создания, структура, технологии / *Цифровая гуманитаристика: ресурсы, методы, исследования. Материалы Международной научной конференции. (Пермь, 16-18 мая 2017 г.): в 2-х ч. / Пермский государственный национальный исследовательский университет. Пермь, 2017. Ч. 2. С. 177-179.*

References

1. ISARAN-web: <http://www.arran.ru/?q=ru/isaran>. (03.11.2020).
2. Savina G. A. Integratsionnye vozmozhnosti Informatsionnoi sistemy «Arkhivy RAN» v sozdanii korporativnogo elektronnoho resursa retrospektivnoi i tekushchei informatsii // *Fundamentalnaia nauka: problemy izucheniia, sokhraneniia i restavratsii dokumentalnogo nasledii. Materialy Mezhd. nauchn. konferentsii, Moskva, Arkhiv RAN, 4-7 iyunia 2013 g. M.: Arkhiv RAN, 2013. S. 212-219.*

3. Zagrebaeva V. N., Zlobin E. V., Savina G. A. 10 let Informatsionnoi sisteme «Arkhivy RAN» (ISARAN): itogi i perspektivy. <http://arran.ru/?q=ru/node/746> (03.11.2020).
4. Afiani V. Iu., Ilina I. N. Integratsiia elektronnykh resursov Arkhiva RAN v Elektronnuu biblioteku «Nauchnoe nasledie Rossii» // Informatsionnoe obespechenie nauki: novye tekhnologii. Sbornik nauchnykh trudov. M., Nauchnyi mir, 2009. S. 81–88.
5. Kalenov N. E., Sotnikov A. N., Ilina I. N. Arkhivnaia informatsiia v Elektronnoi biblioteke «Nauchnoe nasledie Rossii» // Fundamentalnaia nauka: problemy izucheniia, sokhraneniia i restavratsii dokumentalnogo nasledii. Materialy Mezhdunarodnoi nauchnoi konferentsii. Moskva, 4–7 iunია 2013 g. M: Arkhiv RAN, 2013. S. 25–35.
6. Kiselev M. Iu. Sozdanie bazy dannykh «Elektronnye opisi Arkhiva RAN» // Fundamentalnaia nauka: problemy izucheniia, sokhraneniia i restavratsii dokumentalnogo nasledii: Materialy Mezhdunarodnoi nauchnoi konferentsii. Moskva, 4–7 iunია 2013 g. M.: Arkhiv RAN, 2013. S. 128–130.
7. Zagrebaeva V. N., Zlobin E. V., Savina G. A. Baza dannykh «Tsentralnyi fondovyi katalog» arkhivov RAN-FANO // Vestnik arkhivista. № 4. 2015. S. 282–292.
8. Portal «Mnemosina»: <http://www.arran.ru/?q=ru>. (03.11.2020).
9. Zlobin E. V., Zagrebaeva V. N., Savina G. A. Portal «Mnemosina» – primer realizatsii edinogo informatsionnogo prostranstva / Informatsionnyi biulleten assotsiatsii Istorii i kompiuter. 2016. № 45. S. 208–210.
10. TsFK RAN: <http://isaran.ru/?q=ru/cfk>. (03.11.2020).
11. Zagrebaeva V. N., Zlobin E. V., Savina G. A. Baza dannykh «Personalnyi sostav RAN» – struktura i sodержanie // Dokumentatsiia v informatsionnom obshchestve: normativno-metodicheskoe obespechenie upravleniia dokumentami. Doklady i soobshcheniia na XXI Mezhdunarodnoi nauchno-prakticheskoi konferentsii 18–19 noiabria 2014 g. M., 2015. S. 416–420.
12. Polozhenie ob elektronnom TsFK: <http://isaran.ru/?q=ru/node/68> (03.11.2020).
13. Akademicheskie instituty pamiati na veb-portale «Mnemosina»: khranilishcha dokumentalnykh sobranii v edinom informatsionnom prostranstve. <http://www.arran.ru/?q=ru/node/829> (03.11.2020).
14. Zagrebaeva V. N., Zlobin E. V., Savina G. A. Informatsionnyi portal «Mnemosina»: istoriia sozdaniia, struktura, tekhnologii / Tsifrovaia gumanitaristika: resursy, metody, issledovaniia. Materialy Mezhdunarodnoi nauchnoi konferentsii. (Perm, 16–18 maia 2017 g.): v 2-kh ch. / Permskii gosudarstvennyi natsionalnyi issledovatel'skii universitet. Perm, 2017. Ch. 2. S. 177–179.

**Информационная система
«История геологии и горного дела» как
элемент единого цифрового пространства
научных знаний в области истории науки**

И. Г. Малахова, И. П. Второв

ГИН РАН

Аннотация. Информационная система «История геологии и горного дела» открыла доступ к научной биографической, библиографической информации и документам. Её созданию предшествовал долгий путь сбора, обработки и оцифровки материалов. Опыт разработки и эксплуатации системы может быть полезен при создании единого цифрового рабочего пространства РАН, в которое система может быть интегрирована и стать моделью для других направлений изучения истории науки. Возможности системы показаны на примере истории Геологического института РАН.

Ключевые слова: история геологии, история горного дела, информационная система, научная база данных, единое цифровое пространство, Геологический институт РАН.

**Information System “History of geology and mining”
as a model of common digital workplace in the history
of science, and a practical experience**

I. G. Malakhova, I. P. Vtorov

Geological Institute RAS

Abstract. The Information System “History of Geology and Mining” has opened an access to biographical and bibliographic information, and scientific documents. A long time of collecting, processing and digitizing materials preceded the creation of the System. The experience in the development and operation of the Information System may be integrated and useful in creating of the Common Digital Workplace of the Russian Acade-

my of Sciences. The Information System could be a model for more fields of the history of science. The history of the Geological Institute of the Russian Academy of Sciences is an example of the System in action.

Keywords. History of geology, history of mining, information system, scientific database, Unified Digital Workplace of the Russian Academy of Sciences, Geological Institute of the Russian Academy of Sciences.

Создание базы данных по истории геологии и горного дела стало возможным благодаря концентрации информации в одной структурной ячейке системы научных подразделений РАН – в группе истории геологии Геологического института. Это подразделение было создано в 1949 г., более 40 лет работало под руководством лидера историков геологических наук Советского Союза члена-корреспондента РАН Владимира Владимировича Тихомирова (1915–1994), и в настоящее время группа истории геологии является структурным подразделением Геологического института РАН.

Первым шагом к созданию информационного продукта стало участие группы истории геологии в Программе Президиума РАН «Электронная библиотека Научное наследие России» (2008–2015). В рамках работ по этой программе отрабатывались подходы к формированию цифровых ресурсов, отражающих объекты различных типов, форматы метаданных, связи между объектами и т. п.

Опыт сотрудников МСЦ РАН и БЕН РАН позволил выработать концепцию онлайн информационной системы, которая, наряду с собственно ЭБ «Научное наследие России», была реализована в виде информационной системы «История геологии и горного дела».

Информационная система «История геологии и горного дела» (higeo.ginras.ru) – это база данных, содержащая биографические, библиографические сведения об учёных и организациях, связанных с развитием геологии и горного дела [1]. Информационная система была создана на базе сетевого настраиваемого программного комплекса «SciRus», разработанного в БЕН РАН [2, 3, 4]. На октябрь 2020 г. в информационной системе представлено 1114 персоналий: российские и зарубежные геологи и горные инженеры, в том

числе члены Российской академии наук, избранные за весь период ее существования с 1724 г. Внутренний интегрированный поиск по многим выделенным параметрам является удобным и основным функционалом системы. Массив данных концентрируется вокруг каждой персоналии учёного и включает:

- Краткие биографические данные, научные биографии.
- Организации, в которых работали учёные, членство в научных обществах.
- Направления и географию исследований.
- Списки научных трудов и публикации об учёных.
- Связанные документы (анкеты, рукописи, корреспонденция и пр.) в формате PDF.
- Портреты и другие изображения в формате JPG.

Основой для формализации массива информации стали собранная библиографическая информация, публикации историков науки Геологического института (ГИН) РАН и труды, выходявшие под их руководством и при непосредственном участии, а именно:

- Серия «Очерки по истории геологических знаний» (32 выпуска с 1953 г.).
- 50 томов издания Комиссии Геологической изученности СССР (1956–1991 гг.).
- 8 томов биобиблиографического словаря в рубрике «Материалы к истории геологии в СССР» (1965–1978 гг.).
- Три тома о членах Российской академии наук (русские академики и члены-корреспонденты; иностранные члены) [5, 6, 7].

Обращение к современным информационным технологиям было связано с необходимостью постоянно дополнять и актуализировать информацию, сделать её максимально доступной и удобной для поиска и научной работы, значительно снизить затраты на публикации. Первый опыт был получен во время работы по Программе Президиума РАН – Электронная библиотека «Научное наследие России» (e-heritage.ru), объединившей информационные ресурсы научных институтов, архивов и библиотек РАН. В настоящее

время Информационная система интегрирована с Электронной библиотекой ссылками из библиографий на биографические очерки и полнотекстовые публикации.

В 2015 г. Информационная система была предложена на конференции Федерального агентства научных организаций РАН в качестве модели единой сети данных по истории наук о Земле, для формирования соответствующего «Центра коллективного пользования» учёных [8]. С 2017 г. ресурс используют студенты 5 курса Геологического факультета МГУ им. М. В. Ломоносова в учебном курсе «История геологии».

Русская Википедия рассматривает информационную систему как авторитетный научный источник информации, поэтому на её страницы имеется более 500 ссылок и примечаний, созданы шаблоны по их интеграции, проводятся встречи с редакторами Википедии. Информационная система постоянно расширяет контент и является надёжным информационным ресурсом и инструментом для работы историков и специалистов по наукам о Земле.

Опыт создания и развития информационной системы «История геологии и горного дела» может быть использован в Российской академии наук как действующая модель общего информационного цифрового рабочего пространства по истории науки. Она может быть интегрирована в единое цифровое пространство РАН. Этому должно предшествовать формирование рабочей группы из заинтересованных историков науки различных направлений с целью координации наполнения контента биографической и библиографической информацией. Регламент и правила предоставления и размещения информации нужны для совместимости и унификации публикаций. Соответствующая комиссия по созданию единого цифрового ресурса РАН должна сформировать редколлегию, необходимую для работы информационного ресурса. Права на публикацию материалов должны быть только у квалифицированных членов цифровой редколлегии, которые руководствуются общими установками и целями.

Чем крупнее станет информационный ресурс РАН, тем больше в нём будет установлено взаимосвязей, которые бу-

дут уточнять, дополнять функционал и повышать информационную насыщенность и устойчивость системы.

Формирование единого цифрового рабочего пространства РАН и включение в него информационного ресурса по истории науки должно способствовать улучшению качества, ускорению и углублению изысканий, расширению базы авторитетных источников информации, повышению цитируемости отечественных авторов, проверяемости и исключению дублирования работ. Научные исследования должны использовать самые современные инструменты. Система должна быть надёжно защищена и иметь резервные копии. Возможно введение системы блокчейн для унификации и перепроверки библиографических описаний. По мере роста информационной системы работа с ней будет всё более целостной, системной, глубокой и надёжной.

Создание дублирующего англоязычного дублирующего интерфейса увеличило посещаемость системы за счёт поисковых запросов из зарубежных стран. Важную роль в охвате информации по истории геологии и горного дела сыграло международное сотрудничество.

Инициатива объединения историков геологии разных стран принадлежала В. В. Тихомирову. В 1967 г. он стал первым президентом Международной Комиссии по истории геологических наук (ИНИГЕО, INHIGEO – inhigeo.com), созданной под эгидой Международного союза геологических наук (IUGS) и Международного союза истории и философии науки (IUHPS). Историки геологии Геологического института РАН продолжают работу в ИНИГЕО.

За последние три года Информационную систему посетило почти 36 тысяч пользователей, причём среднемесячные визиты продолжают расти (*рис. 1*).

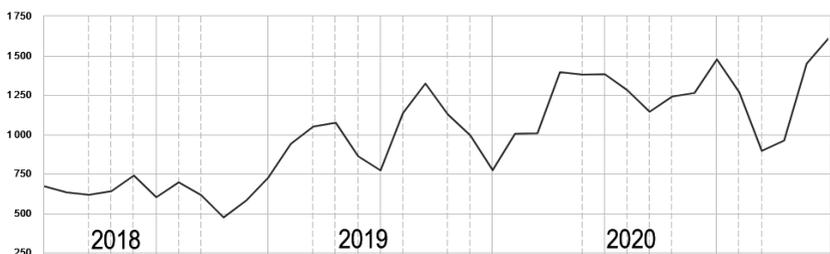


Рис. 1. Ежемесячная динамика пользования Информационной системой за последние 3 года (визиты в месяц, Яндекс Метрика на 3.11.2020)

По данным Яндекс Метрики основные пользователи были из России (93,4 %), а также из Украины (0,89 %), Казахстана (0,71 %), Белоруссии (0,61 %), США (0,58 %), Германии (0,46 %) и Армении (0,39 %), всего зафиксирована 91 страна. Запросы из России приходили в основном из Москвы и Московской области, из Санкт-Петербурга и Ленинградской области. Активность проявляли Свердловская, Новосибирская и Иркутская области и Республика Коми. Основными источниками интернет-трафика стали переходы из поисковых систем (67,3 %), переходы по ссылкам с сайтов (23,7 %) и прямые заходы (7,1 %). Обычно пользователи ищут конкретного автора или его публикации.

Топ-10 самых популярных по поисковым обращениям учёных: Э. Зюсс (Австрия), П. И. Бутов, И. Н. Стрижов, М. Д. Белонин, В. М. Севергин, А. Н. Криштофович, Л. Эли де Бомон (Франция), С. Г. Боч, В. К. Котульский и Л. С. Лейбензон.

Посетитель в среднем просматривает три страницы (глубина просмотра 3,44). Возрастные группы пользователей распределены достаточно равномерно: > 55 лет – 28 %, 18–24 года – 23 %, 45–54 года – 22 %, 25–34 года – 14 %, 35–44 года – 11 %. Доля планшетов и смартфонов при просмотре Системы увеличилась за 3 года с 8 до 14 %.

Информационная система «История геологии и горного дела» зарегистрирована в Роспатенте как элек-

тронный ресурс для сбора, хранения, представления и анализа информации в области наук о Земле (рис. 2).

Информация о Геологическом институте РАН в Информационной системе может служить наглядным примером использования научных данных.

Геологический институт АН СССР был создан в 1930 г. в Ленинграде — в геологической столице страны того времени, что было предопределено наличием организационной инфраструктуры, политикой Академии наук и кадровым потенциалом. В период масштабной реорганизации Академии наук СССР Институт переехал в Москву в 1934 г. [9].

90 лет ГИН — серьезный повод для проведения посвященных ему мероприятий и публикаций. Система предлагает широкий спектр материалов по истории ГИН РАН для заинтересованных лиц.



Рис. 2. Свидетельство о государственной регистрации базы данных № 2019622087 от 15 ноября 2019 г., правообладатель контента – ГИН РАН

За прошедшие десятилетия Институт менял названия – Геологический институт АН СССР, Институт геологических наук АН СССР, Геологический институт РАН. Информация содержится на всех соответствующих страницах в Системе и сопровождается фотографиями здания (рис. 3).

организация

Связанные персоналии:	[Просмотреть]
Название оригинальное / Original title:	Геологический институт РАН
Изображение организации / Photo:	
Краткие сведения / Brief info:	ГИН РАН
Тип организации / Activities:	Научные учреждения
История / History:	1930 - Геологический институт АН СССР (1934 - переезд из Ленинграда в Москву) 1938 - Институт геологических наук АН СССР (ИГН АН СССР) 1956 - Геологический институт АН СССР (ГИН АН СССР) 1991 - Геологический институт РАН (ГИН РАН) ИГН АН СССР - http://igeeo.ginras.ru/view-record.php?tbl=org&id=59 ГИН АН СССР - http://igeeo.ginras.ru/view-record.php?tbl=org&id=206
Членство / Members:	
Награды / Awards:	1969 — Орден Трудового Красного Знамени (за успехи в развитии геологической науки и подготовке высококвалифицированных научных кадров) 1972 — Юбилейный почётный знак в ознаменование 50-летия образования Союза ССР (за достижение выдающихся результатов во Всесоюзном социалистическом соревновании в ознаменование Пятидесятилетия образования СССР)
Лауреаты / Awards winners:	
Ссылки / Links:	

Рис. 3. Представление организации в Информационной системе на примере ГИН РАН

Первым директором ГИН АН СССР был академик В. А. Обручев (рис. 4).

Всего 16 ведущих ученых страны занимали этот пост. Информация о них и полные списки научных трудов размещены в Информационной системе (см. А. А. Борисяк, А. Д. Архангельский, А. Н. Заварицкий, И. Ф. Григорьев (дважды), А. Е. Ферсман, Д. С. Белянкин, М. И. Варенцов, Н. С. Шатский, А. В. Пейве, П. П. Тимофеев, А. Л. Книппер, Ю. Г. Леонов, М. Г. Леонов, М. А. Федонкин и К. Е. Дегтярёв).

В Информационной системе содержится краткая информация и библиография 220 научных сотрудников Института, работавших за все годы его истории.

Непременным дополнениям к персоналиям являются фотографии (портреты и групповые) — всего около 350 изображений.

Свет на историю Института, организацию исследований, многочисленные контакты и личные отношения проливают многочисленные документы научных сотрудников (около 250).

персоналия

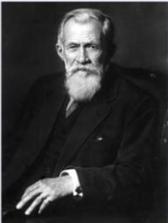
организация:	Императорская Санкт-Петербургская академия наук Московская горная академия Институт мерзлотоведения АН СССР Геологический институт АН СССР Коллекция по изучению четвертичного периода
Связанные документы:	[Дроссентера]
Связанные фотографии:	[Дроссентера]
ФИО / Surname, 1st name, 2nd name (in Russian):	Обручев Владимир Афанасьевич
Имя (русский язык) / 1st name (in Russian):	Владимир
Отчество (русский язык) / 2nd name (in Russian):	Афанасьевич
Фамилия (русский язык) / Surname (in Russian):	Обручев
Имя (английский язык) / 1st name:	Vladimir A.
Фамилия (английский язык) / Surname:	Obuchev;
Век / Century:	XIX - XX
Дата рождения / Date of birthday:	1863, 10 октября
Место рождения / Place of birthday:	с. Клепичино, Тверская губ., Российская империя
Дата смерти / Date of death:	1956, 19 июня
Место смерти / Place of death:	г. Москва, Россия
Ученое звание / Scientific degree:	действительный член
Область научных исследований / Fields of interest:	Геология / Общие вопросы, Геология / Тектоника, Геология / Геология полезных ископаемых, Геология / Инженерная геология и мерзлотоведение, Геология / Геология нефти и газа, Геология / История геологических наук
География исследований / Regions of explorations:	Сибирь, Дальний Восток, Восточная Азия, Центральная Азия
Членство в обществах / Scientific institutions membership:	
Биография / Biography:	
Библиография / Bibliography:	Основные научные труды В.А. Обручева <i>Обручев В.А. Пески и степи Закаспийской области // Изв. РГО. 1887. Т. 23. №. 2. С. 174-190. URL Отд. изд. СПб.: тип. А.С. Суворина. 1888. 17 с.</i> <i>Обручев В.А. Предварительный отчет о геологических исследованиях в Закаспийской области в 1886 г. // Изв. Тихомиров Р.Г., Нильсенвикова N.A. Centenary of the birth of V.A. Obuchev, soviet geologist, geographer, traveller and pioneer in permafrost research // Arctic list of North America. 1963. Vol. 16. N 4. P. 235-268 - part. Teil E. Obuchevsches geologische Forschungen im ostlichen Zentralasien // Petermanns Mit. 1894. Bd. 40. H. 12. S. 285-290 - Karte. Vladimir Afanasiev Obuchev (Obuchschew) // Leopoldina. 1961. Bd. 4.5. N 3. S. 94-97. Составитель И.Г. Мельцова</i>
Фотогалерея / Photo gallery:	
Автор записки: Date written last modified:	admin 2020-08-06 13:41:22

Рис. 4. Представление персоналии учёного на примере В. А. Обручева

Многофункциональный поиск внутри системы позволяет проследить развитие научных направлений и сконцентрироваться на различных регионах проведения геологических работ (рис. 5).

Информационная система

История геологии и горного дела

Поиск

персоналия / География исследований / Regions of explorations содержит Урал

Еще поле

Искать Сортировать результаты по полю персоналия / ФИО / Surname, 1st name, 2nd name (in Russian) персоналии

Найдено записей: 127 [в новом окне](#) [Вернуться в главное меню](#)

- [Ахметьев Михаил Алексеевич / Akhmetiev, Michael](#)
- [Баньковский Лев Владимирович / Bonikowski, Lev](#)
- [Безруков Пантелеймон Леонидович / Bezrukov, Panteleymon](#)
- [Белянкин Дмитрий Степанович / Belyankin, Dmitry](#)
- [Бетехтин Анатолий Георгиевич / Betekhtin, Anatoly](#)
- [Блохин Алексей Александрович / Blokhin, Alexey](#)
- [Болдырев Анатолий Кашионович / Boldyrev, Anatoly](#)
- [Бонштедт-Куплетская Эльза Максимилановна / Bonstedt-Kupletskaya, Elsa](#)
- [Борнеман-Старынкевич Ирина Дмитриевна / Borneman-Starynkevich, Irina](#)
- [Брянчаннинова Наталья Игоревна / Bryanchaninova, Natalia](#)
- [Буров Александр Петрович / Burov, Alexander](#)
- [Буртман Валентин Семенович / Burtman, Valentine](#)
- [Вагнер Пётр Иванович / Wagner, Peter](#)
- [Варлаков Александр Сергеевич / Varlakov, Aleksander](#)
- [Вахрушев Георгий Васильевич / Vakhrushev, Georgy](#)

Рис. 5. Пример поиска персоналий учёных, работавших на Урале

Одной из важнейших сторон работы группы истории геологии ГИН РАН является выполнение информационных запросов, поступающих из различных ведомств и учреждений России и от иностранных коллег. Материалы, содержащиеся в ИС, дополняются специально подготовленными аналитическими и информационными данными.

Возможности, преимущества и перспективы работы с Информационной системой и неоднократно докладывались и обсуждались на научных конференциях в России и за рубежом.

Литература

1. Каленов Н. Е., Малахова И. Г. Интегрированный общедоступный информационный ресурс «История геологии и горного дела» // Информационные ресурсы России. 2017. № 1. С. 19–23.
2. Сенько А. М. Информационная система SciRus: принципы построения и перспективы развития // Научный сервис в сети Интернет: технологии параллельного программирования. М., 2006. С. 58–59.
3. Якшин М. М. Web-интерфейс системы «Наука России» // Современные технологии в информационном обеспечении науки. М., 2003. С. 47–52.
4. Якшин М. М. Развитие платформы SciRus // Информационное обеспечение науки: новые технологии. М.: БЕН РАН, 2015. С. 203–207.
5. Соловьёв Ю. Я., Бессуднова З. А., Пржедецкая Л. Т. Отечественные действительные и почетные члены Российской академии наук. XVIII–XX вв. / Отв. ред. Ю. Я. Соловьёв. М.: Научный мир, 2000. 548 с.
6. Соловьёв Ю. Я., Хомизури Г. П., Бессуднова З. А. Отечественные члены-корреспонденты Российской академии наук XVIII – начала XXI века: геология и горные науки / Отв. ред. Ю. Я. Соловьёв. М.: Наука, 2007. 607 с.
7. Малахова И. Г., Бессуднова З. А., Хомизури Г. П., Минаева Е. Л. Иностранцы члены Российской академии наук. XVII–XXI вв.: Геология и горные науки / отв. ред. И. Г. Малахова. М.: ГЦ РАН, 2012. 504 с. URL: <http://elpub.wdcb.ru/ebooks/formemb.pdf> (дата обращения: 3.11.2020).
8. Малахова И. Г., Второв И. П. История геологических и горных наук // Информационные ресурсы Центры коллективного пользования и уникальные научные установки в организациях, подведомственных ФАНО России: [Москва. 20–21 окт. 2015.]: Тез. докладов. М.: ФАНО России, 2015. 2 с.
9. Геологический институт Академии наук СССР. 1931–1937. Архив РАН. Фонд 664. 55 ед. хр. <https://clck.ru/SGsHE> (дата обращения 07.11.2020).

References

1. Kalenov N. E., Malahova I. G. Integrirovannyj obshchedostupnyj informacionnyj resurs "Istoriya geologii i gornogo dela" // Informacionnyye resursy Rossii. 2017. № 1. S. 19–23.
2. Sen'ko A. M. Informacionnaya sistema SciRus: principy postroeniya i perspektivy razvitiya // Nauchnyj servis v seti Internet: tekhnologii parallel'nogo programmirovaniya. M., 2006. S. 58–59.

3. YAkshin M. M. Web-interfejs sistemy "Nauka Rossii" // *Sovremennye tekhnologii v informacionnom obespechenii nauki*. M., 2003. S. 47–52.
4. YAkshin M. M. Razvitie platformy SciRus // *Informacionnoe obespechenie nauki: novye tekhnologii*. M.: BEN RAN, 2015. S. 203–207.
5. Solov'yov Yu. Ya., Bessudnova Z. A., Przhedekaya L. T. Otechestvennye dejstvitel'nye i pochetnye chleny Rossijskoj akademii nauk. XVIII–XX vv. / *Otv. red. Yu. Ya. Solov'yov*. M.: Nauchnyj mir, 2000. 548 s.
6. Solov'yov Yu. Ya., Homizuri G. P., Bessudnova Z. A. Otechestvennye chleny-korrespondenty Rossijskoj akademii nauk XVIII – nachala XXI veka: geologiya i gornye nauki / *Otv. red. Yu. Ya. Solov'yov*. M.: Nauka, 2007. 607 s.
7. Malahova I. G., Bessudnova Z. A., Homizuri G. P., Minina E. L. Inostrannye chleny Rossijskoj akademii nauk. XVIII–XXI vv.: Geologiya i gornye nauki / *Otv. red. I. G. Malahova*. M.: GC RAN, 2012. 504 s. URL: <http://elpub.wdcb.ru/ebooks/formemb.pdf> (data obrashcheniya: 3.11.2020).
8. Malahova I. G., Vtorov I. P. Istoriya geologicheskikh i gornyh nauk // *Informacionnye resursy Centry kollektivnogo pol'zovaniya i unikal'nye nauchnye ustanovki v organizacijah, podvedomstvennyh FANO Rossii*: [Moskva. 20–21 okt. 2015.]: Tezisy dokladov. M.: FANO Rossii, 2015. 2 s.
9. Geologicheskij institut Akademii nauk SSSR. 1931–1937. Arhiv RAN. Fond 664. 55 ed. hr. <https://clck.ru/SGsHE> (data obrashcheniya 07.11.2020).

Развитие программного обеспечения электронной библиотеки «Научное наследие России»

К. П. Погорелко

МЦ РАН

Аннотация. Структура данных и программное обеспечение электронной библиотеки «Научное наследие России» создавалось в 2007 г. и в настоящее время не отвечает потребностям системы. В статье описаны решения, принятые при реализации нового варианта программного обеспечения. Эти решения затрагивают как организацию структуры базы данных, так и протоколы взаимодействия с системой. Особое внимание уделяется разработке средств поиска информации.

Ключевые слова: электронная библиотека, цифровая библиотека, программное обеспечение, поиск информации, asp.net core.

Development of the Software for the Electronic Library «Scientific Heritage of Russia»

К. P. Pogorelko

JSCC RAS

Abstract. The data structure and software of the Scientific Heritage of Russia electronic library was created in 2007 and currently does not meet the needs of the system. The article describes the decisions made when implementing a new version of the software. These decisions affect both the organization of the database structure and the protocols for interacting with the system. Particular attention is paid to the development of information retrieval tools.

Keywords: electronic library, digital library, software, information search, asp.net core.

Электронная библиотека (ЭБ) «Научное наследие России» начала создаваться в 2007 г. [1] и функционирует в полном объеме более 10 лет [2]. Библиотека пользуется постоянным спросом [3]. Распределение по годам количества запросов к электронным публикациям, включенным в ЭБ, приведено на *рис. 1*.

Программное обеспечение, поддерживающее функционирование библиотеки, несколько раз модифицировалось. Так, в 2013 г. представление полнотекстовых публикаций было выделено в отдельную подсистему [4]. Современное состояние библиотеки приводится в [5].

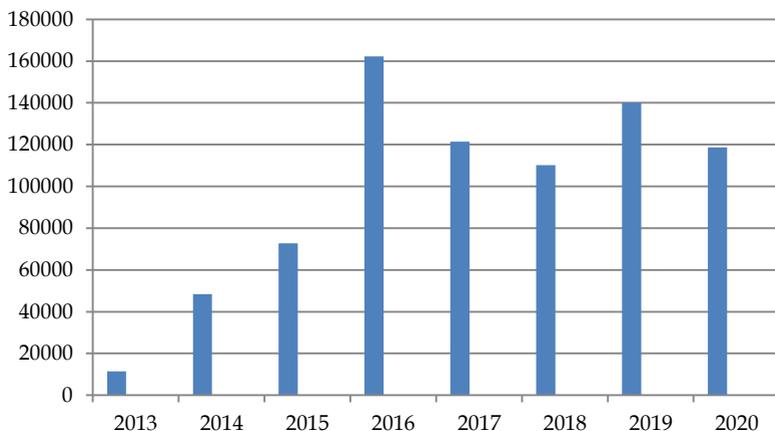


Рис. 1 Количество запросов на электронные публикации

В настоящее время назрела необходимость переработки программного обеспечения. Это обусловлено как возросшими требованиями к поисковым возможностям системы и ее быстродействию, так и к более полному удовлетворению требований, выдвигаемых технологическими процессами подготовки информации.

Поскольку средства разработки, на которых базируется существующая система, устарели, было принято решение о коренной переработке как программного обеспечения,

поддерживающего работу системы, так и ее внутренней структуры и интерфейсов.

Для реализации новой версии программного обеспечения была выбрана современная платформа asp.net core. Основной особенностью данной платформы является универсальность приложений, разработанных на ее базе. Эти приложения могут функционировать как на операционных системах Microsoft, так и на операционных системах семейства Unix. Кроме того, при развертывании веб-приложений, разработанных на базе этой платформы, они оформляются как исполняемый файл, что обеспечивает их высокую эффективность. Для хранения данных был выбран SQL сервер PostgreSQL, который имеет реализации для разных платформ и обеспечивает достаточно эффективное манипулирование данными.

Разрабатываемая система объединена в единый проект, состоящий из следующих частей: подсистема подготовки метаинформации по различным объектам, подсистема подготовки электронных публикаций и подсистема поиска и визуализации информации. Такое решение позволяет учитывать возможные изменения состава и структуры данных сразу во всем проекте. Кроме того, в зависимости от производительности серверов и нагрузки на систему, появляется возможность разворачивания системы как на одном сервере, так и разнесения составляющих проекта по разным серверам.

В предыдущей версии системы информация располагалась в нескольких базах данных, между которыми были организованы протоколы обмена. Помимо задержек с публикацией и увеличением вероятностей сбоя, это создавало определенные трудности в организации перекрестных ссылок между объектами, которые были размещены в разных базах. В новой версии системы все данные содержатся в единой базе, что обеспечивает единообразие их обработки, облегчает манипулирование данными и обеспечивает отсутствие задержек при их публикации.

Для обеспечения контроля качества вводимых данных принята двухуровневая система полномочий. На первом

уровне оператор, имеющий полномочия на ввод данных, осуществляет их ввод. По окончании ввода оператор устанавливает признак «ввод окончен». Выпускающий редактор, просматривая документы с признаком окончания ввода, проверяет введенные данные и, если данные введены с необходимым качеством, устанавливает признак «опубликовано». После того как документ получает этот статус, он становится доступным для поиска и показа в библиотеке. Оператор, осуществивший ввод документа, получившего статус «опубликовано», теряет право коррекции данного документа, если он не имеет прав выпускающего редактора.

В системе реализованы средства для поддержки процедур выявления и обработки ошибочных данных. Ошибки могут выявляться как в ручном режиме при просмотре данных, так и в автоматическом режиме, например, при проверке

корректности внешних ссылок. Все выявленные ошибки приписываются в виде списка к тому объекту, к которому они относятся. Оператор имеет возможность просматривать выявленные ошибки и, после их исправления, присваивать документу соответствующий статус.

В структуре базы данных предусмотрена возможность снабжать связи между метаданными различных объектов дополнительными атрибутами, конкретизирующими данную связь. Например, для связи объектов «персона» – «публикация» можно указать значение атрибута «автор», «редактор», «о нем» и т. д.

В новой версии системы существенно расширены ее поисковые возможности. Поисковый модуль системы позволяет обрабатывать запросы, состоящие из практически неограниченного количества термов, связанных между собой логическими операциями «И», «ИЛИ» или «И-НЕ». Каждый поисковый терм относится к тому или иному выбранному пользователем элементу метаданных (полно) объекта. Поле в зависимости от информации, которую оно содержит, относится к некоторому типу. В зависимости от того, к какому типу относится поле, пользователю предоставляется возможность задать те или иные условия для отбора его содержимого.

В системе предусмотрены следующие типы полей и соответствующие им поисковые условия:

1. Текстовое поле. К этому типу относятся поля, содержащие относительно короткую текстовую информацию. Для этого типа имеется возможность задать фрагменты текста с уточнением «содержится», «начинается с» и «равно» (по умолчанию используется условие «содержится»).

2. Полнотекстовое поле. К этому типу относятся поля, содержащие значительную по объему текстовую информацию, например, биографию ученого. Для такого типа полей имеется возможность задать словоформы, которые должны присутствовать в поисковом поле. Для организации поиска применяются средства полнотекстового поиска PostgreSQL.

3. Числовое поле. Для полей такого типа имеется возможность задать условие «больше», «равно» или «меньше» и числовое значение, с которым необходимо производить сравнение. В ряде случаев значения числовых полей в базе данных могут быть не определены (год рождения персоны, год издания книги и др.). Система допускает ввод в эти поля нечисловых значений типа «середина 17-го века» и т. п. При задании условий на числовые поля записи с неопределенными значениями никогда не попадут в результирующую выдачу. Чтобы иметь возможность обеспечить полноту поиска, для числовых полей добавлена опция «Добавить документы с неустановленным значением». При задании этой опции выбираются как документы, поля которых отвечают заданному условию, так и документы, у которых значение поля не является числовым.

4. Кодификатор. К этому типу относятся поля, значение которых ограничено определенным перечнем, например, рубрика ГРНТИ или тип публикации. Для задания условий на поле этого типа пользователю выдается соответствующий кодификатор и предоставляется возможность отметить нужные значения.

5. Дата. К этому типу относятся поля, значение которых является датой, например, дата поступления в библиотеку. Для задания условий на поле этого типа поль-

зователю обеспечивается возможность задать необходимую дату и условие «больше», «равно» или «меньше».

Для запросов с большим объемом полученных результатов предусмотрена возможность постраничной выдачи. При этом поиск в базе осуществляется один раз и результат выполнения запроса запоминается. Это позволяет, при пролистывании страниц, не осуществлять повторный поиск, а оперативно выдавать данные очередной страницы.

Перечисленные выше возможности позволяют формировать запрос к полям, относящимся к определенному элементу метаданных. Исключением является возможность поиска публикаций по фамилии автора. В новой системе реализована попытка предоставления пользователю возможности универсального поиска, при которой можно задавать условия не только к определенному типу метаданных, но и к метаданным, связанным с этим типом, и, в свою очередь, к метаданным, связанным со связанными метаданными и т. д. Решение этой задачи потребовало разработки как интуитивно понятного построителя поискового запроса, так и стратегии формирования запроса к базе данных которая, с одной стороны, обеспечивала бы простую алгоритмизацию и, с другой стороны, достаточную эффективность сформированного запроса. Обе задачи в новой системе решены. На *рис. 2* приведен скриншот с примером запроса на поиск соавторов Эйлера.

Ищем у которых

Найдено персон 4

- Шриттер Иван Михайлович [1740, 10 (21) октября]
- Фусс Павел Николаевич [1798, 21 мая (1 июня)]
- Гольдбах Христиан [1690, 18 марта]
- Фусс Николай Иванович [1755, 29 января]

Рис. 2. Пример выполнения запроса к универсальной системе поиска

Новая версия системы в настоящее время находится в стадии опытной эксплуатации. Текущая версия ЭБ «Научное наследие России» продолжает функционировать и доступна по адресу <http://e-heritage.ru>. Для перехода на новую систему реализована отдельная программа переноса данных из баз старой системы в базу данных новой системы. После доработки и окончательного тестирования новой версии она заменит предыдущую без каких-либо перерывов в работе.

Обрабатываемые при реализации системы подходы могут быть положены в основу базового программного обеспечения для поддержки различных компонентов единого цифрового пространства научных знаний [6, 7].

Работа выполнена в МСЦ РАН в рамках государственного задания № 0580–2021–0014.

Литература

1. Каленов Н. Е., Савин Г. И., Сотников А. Н. Технология создания электронной библиотеки «Научное наследие России» // Научная книга. 2007. Т. 1–4. С. 170–173.
2. Каленов Н. Е., Савин Г. И., Серебряков В. А., Сотников А. Н. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты, системы и алгоритмы. 2012. Т. 4. № 100. С. 30–40.
3. Погорелко К. П. Динамика использования электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2017. С. 192–200.
4. Погорелко К. П. Новая система презентации электронных книг в системе «Научное наследие России» // Информационное обеспечение науки: новые технологии. Сборник научных трудов / Каленов Н. Е. (ред.). М.: БЕН РАН, 2013. С. 32–35.
5. Каленов Н. Е., Кириллов С. А., Соболевская И. Н., Сотников А. Н. Современное состояние электронной библиотеки «Научное наследие России» // Труды НИИСИ РАН. Математическое и компьютерное моделирование сложных систем: теоретические и прикладные аспекты. 2018. Т. 8. № 6. С. 166–169.
6. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О Едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.
7. Антопольский А. Б. и др. Принципы построения и структура Единого цифрового пространства научных знаний (ЕЦПЗ) // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2020. № 4. С. 9–17.

References

1. Kalenov N. E., Savin G. I., Sotnikov A. N. Technology of creating an electronic library «Scientific heritage of Russia» // Nauchnaya kniga. 2007. T. 1–4. P. 170–173.
2. Kalenov N. E., Savin G. I., Serebryakov V. A., Sotnikov A. N. Principles of construction and formation of the electronic library «Scientific heritage of Russia» // Software products, systems and algorithms. 2012. T. 4. № 100. P. 30–40.
3. Pogorelko K. P. Dynamics of the use of the electronic library «Scientific heritage of Russia» // Information support of science: new technologies: Collection of scientific papers. M.: LNS RAS, 2017. P. 192–200.
4. Pogorelko K. P. New system of presentation of e-books in the system «Scientific heritage of Russia» // Information support of science: new

technologies. Collection of scientific papers / Kalenov N. E. (ed.). M.: LNS RAS, 2013. P. 32–35.

5. Kalenov N. E., Kirillov S. A., Sobolevskaya I. N., Sotnikov A. N. The current state of the electronic library «Scientific heritage of Russia» // Proceedings of NIISI RAS. Mathematical and computer modeling of complex systems: theoretical and applied aspects. 2018. T. 8. № 6. P. 166–169.

6. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. On the Unified Digital Space of Scientific Knowledge // Bulletin of the Russian Academy of Sciences. 2019. T. 89. № 7. P. 728–735.

7. Antopol'skij A. B. et al. Principles of construction and structure of the Unified digital space of scientific knowledge (UDSSK) // Scientific and technical information. Series 1: Organization and methodology of information work. 2020. № 4. P. 9–17.

Автоматизированная система формирования и поддержки базы данных результатов научной деятельности сотрудников академических организаций

С. А. Власова

МСЦ РАН

Аннотация. В статье описывается автоматизированная система создания и поддержки базы данных научных трудов сотрудников академических учреждений, разработанная специалистами МСЦ РАН. Информационная база системы содержит данные о следующих объектах: авторах, связанных с ними организациях (местах их работы), публикациях на аналитическом и монографическом уровнях, источниках (изданиях на сводном уровне — журналах, сборниках), докладах, сделанных на научных конференциях, симпозиумах, семинарах. В системе реализован административный модуль, предназначенный для ввода и редактирования данных. Пользовательский модуль системы представляет собой специальный поисковый аппарат, осуществляющий поиск информации о публикациях, источниках, докладах, мероприятиях, авторах путем обработки поисковых запросов. Отличительной особенностью системы является введенное понятие «эквивалентных» объектов. Эквивалентными считаются объекты, представленные в системе различными метаданными, но относящимися к одной физической сущности. Такими объектами являются «персоны», соответствующие одному автору с различными написаниями фамилии в библиографических описаниях публикаций; организации, имеющие различные варианты названий; статьи, опубликованные без изменений на различных языках.

Ключевые слова: научные труды, библиографические описания, автоматизированная система, база данных, поисковые запросы.

Automated system for creating and supporting a database of research results of academic organizations' employees

S. A. Vlasova

JSCC RAS

Abstract. The article describes the automated system for creating and maintaining a database of scientific works of academic institution's employees, developed by specialists of the Joint Supercomputer Center RAS. The system's information base contains data about objects: the authors, related organizations (places of their work), publications at the analytical and monographic levels, sources (publications at the summary level – journals, collections), reports made at scientific conferences, symposia, seminars. The system has an administrative module designed to enter and edit data. The user's module of the system is a special search engine that searches for information about publications, sources, reports, events, authors by processing search queries. A distinctive feature of the system is the introduced concept of «equivalent» objects. Such objects are «persons» corresponding to the same author with different spellings of the last name in the bibliographic descriptions of publications; organizations with different versions of names; articles which are published without changes in different languages.

Keywords: scientific works, bibliographic descriptions, automated system, database, search queries.

В настоящее время перед каждой научной организацией стоит задача учета научных трудов ее сотрудников (публикаций, докладов, авторских свидетельств и т. п.). В организациях Российской академии наук выполнение этой задачи обеспечивали информационно-библиотечные подразделения, которые вели картотеки трудов сотрудников, содержащие данные об их научных трудах. С развитием вычислительной техники на смену традиционным картотекам пришли информационно-аналитические системы, поддерживающие базы данных (БД) трудов сотрудников и обеспечивающие к ним доступ пользователей через Интернет [1–6].

В 2017 г. в Библиотеке по естественным наукам РАН была разработана и внедрена в техническую эксплуатацию типовая автоматизированная система формирования и поддержки базы данных публикаций [7]. Реализация системы была основана на технологии Microsoft ASP.NET 4 на платформе Microsoft .NET Framework в среде разработки Microsoft Visual Studio 2018 с использованием языка программирования С#. БД системы поддерживаются Microsoft SQL Server.

В 2019–2020 гг. система поддержки базы данных научных трудов получила свое дальнейшее развитие в МСЦ РАН. Система обеспечивает создание и поддержку следующих взаимосвязанных объектов:

- публикации на аналитическом и монографическом уровнях;
- источники (издания на сводном уровне, в которых опубликованы статьи);
- доклады;
- научные мероприятия, на которых сделаны доклады;
- персоны (авторы публикаций и докладов);
- организации (места работы персон).

Система состоит из двух модулей: административного и пользовательского.

В административном модуле (<http://dirsmc.ru/bd/adm.aspx>) осуществляются следующие процессы:

- ✓ ввод и редактирование данных об операторах, работающих с системой;
- ✓ ввод новых записей публикаций, докладов, мероприятий, персон, организаций;
- ✓ редактирование метаданных всех объектов;
- ✓ поиск и просмотр зарегистрированных в системе объектов;
- ✓ создание групп эквивалентных записей.

Между объектами в системе реализованы связи типа «эквивалентные записи». Связанные таким образом объекты система воспринимает как одинаковые. Необходимость ввода эквивалентных персон обусловлена тем, что написание фа-

мии и имени одного автора в разных библиографических описаниях может отличаться. Например, в англоязычных публикациях персоны «Каленов Н. Е.» можно встретить следующие написания данного автора: «N. Kalenov», «N. E. Kalenov», «Nikolay Kalenov». Эквивалентность публикаций возникает тогда, когда в базе данных отражена статья на языке оригинала и ее полностью переведенные на другой язык версии. Связь эквивалентности для организаций устанавливается, когда организация меняет свое название (например, ИМГ АН СССР и ИМГ РАН).

Рассмотрим процесс ввода в систему публикаций и докладов. Ввод данных новой публикации начинается с ввода ее авторов в том порядке, который представлен в публикации. Для каждого автора проверяется его наличие в системе, в случае отсутствия происходит процесс регистрации новой персоны. В систему вводятся данные автора (фамилия и инициалы, дополнительная информация), и он привязывается к организации, которая либо уже существует в системе, либо выполняется ее регистрации.

После окончания ввода авторов система предоставит форму для ввода метаданных публикации: название публикации, вид издания (статья, монография), источник, год издания, том, номер, страницы, адрес полного текста, идентификаторы во внешних базах данных (см. *рис. 1*). При вводе названия публикации система показывает уже зарегистрированные публикации, в которых находит совпадения в названиях, по первым словам, вводимого названия публикации и которые имеют тех же авторов (см. *рис. 1*). Для привязки к вводимой публикации источника (издания, где опубликована статья), его нужно найти по фрагментам названия, а в случае отсутствия – зарегистрировать в системе (ввести название и дополнительную информацию). После окончания ввода всех необходимых метаданных публикация будет зарегистрирована в системе.

Ввод новой публикации

Власова С.А.
Каленов Н.Е.

Название публикации

Власова С.А., Каленов Н.Е. Опыт автоматизации технологических процессов МБА // Информационное обеспечение науки: новые технологии: Сб. науч. тр. / Каленов Н.Е. (ред). - М.: Научный мир, 2009. - 342 с. , 2009. - С. 208-217.

Вид публикации ▼

Источник

Год

Том

Номер

Страницы

Язык публикации ▼

Идентификатор во внешних базах данных

Адрес полного текста

Рис. 1. Регистрация публикации

Регистрация нового доклада так же, как и ввод новой публикации начинается с ввода авторов. Затем система предоставляет форму для ввода метаданных доклада: название доклада, вид доклада (пленарный, секционный, стендовый, приглашенный), мероприятие, дополнительная информация. К докладу нужно привязать мероприятие, на котором был сделан доклад. По фрагментам названия мероприятия определяется его наличие в системе. В случае его отсутствия предоставляется форма для ввода следующих метаданных: название мероприятия, вид мероприятия (конференция, семинар, симпозиум, совещание), место проведения мероприятия (город и страна), даты проведения мероприятия, ссылка на сайт мероприятия, ссылка на материалы (см. *рис. 2*).

Ввод нового мероприятия

Название мероприятия: 8-я Международная научно-практическая конференция "Научное издание международного уровня – 2019: стратегия и тактика управления и развития"

Вид мероприятия: конференция

Место проведения (страна): Россия

Место проведения (город): Москва

Даты мероприятия: год 2019 с 24 . 04 по 26 . 04 (число.месяц)

Ссылка на сайт мероприятия: https://conf.rasep.ru/WCSP/WCSP2019

Ссылка на материалы:

Дополнительная информация:

Рис. 2. Регистрация мероприятия

Следует отметить, что метаданные организаций в системе представлены в виде иерархической структуры: организация может включать подразделения, в которых есть отделы, которые, в свою очередь, могут включать лаборатории и т. д. Администратор системы вводит название организации, затем названия ее подразделений, далее к каждому подразделению привязывает названия его отделов и т. д. На рис. 3 показан интерфейс для редактирования организаций на примере организации «Научно-исследовательский институт системных исследований (НИИСИ РАН)». Здесь можно корректировать названия организации и подразделений, добавлять (или удалять) подразделения на любом уровне.

ОРГАНИЗАЦИИ

[Научно-исследовательский институт системных исследований \(НИИСИ РАН\)](#)

[Межведомственный суперкомпьютерный центр Российской академии наук \(МСЦ РАН\)](#)

[Отдел информационных ресурсов и систем \(ОИРС\)](#)
[Санкт-Петербургское отделение \(СПО\)](#)

[Новое подразделение](#) [Редактировать](#)

[Выход](#)

Рис. 3. Административный блок. Редактирование организации

Пользовательский блок системы (<http://dirsmc.ru/bd/>) был создан с учетом опыта предыдущих разработок автора [8, 9]. Он представляет собой поисковый аппарат, осуществляющий поиск информации путем обработки поисковых запросов любой сложности. Для формирования запроса в поисковую форму вводятся термины (слова без знаков препинания) в одну, две или три строки (см. *рис. 4*). В терминах возможно использование правого усечения (символ *). По умолчанию слова связываются оператором «И», который можно поменять на «ИЛИ». Строки по умолчанию связаны оператором «И», который можно поменять на «ИЛИ» и «И НЕ».

Для каждой строки введенных терминов в поисковой форме выбирается наименование области (поля), в которой должен быть осуществлен поиск:

- фамилия персоны,
- название публикации,
- название журнала / сборника,
- название доклада,
- мероприятие (название, страна, город),
- организация,
- идентификатор во внешних базах данных.

БД научных трудов сотрудников МСЦ РАН

Фамилия персоны И

И

Название публикации И

И

Название доклада И

Год: от по

Показывать по

Сортировка в

Найдено записей: 6

[Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека "Научное наследие России" как интегратор научной информации // Информационные системы и процессы: Сб. науч. трудов/ под ред. проф. В.М. Тютюнника, 2016. - № 15. - С. 21-29.](#)

[Каленов Н.Е., Савин Г.И., Соболевская И.Н., Сотников А.Н. Цифровые музейные коллекции и представление объектов естественно-научного музейного хранения в электронной библиотеке "Научное наследие России" // Научные основы и практика реализации цифровых проектов в сфере культуры и образования.\(Электронная библиотека\), 2016. - С. 33-46.](#)

[Каленов Н.Е., Савин Г.И., Серебряков В.А., Сотников А.Н. Принципы построения и формирования электронной библиотеки "Научное наследие России" // Программные продукты и системы, 2012. - Т. 4. - № 100. - С. 30-40.](#)

[Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека "Научное наследие России" как составляющая интеграционных процессов // Вестник Библиотечной Ассамблеи Евразии, 2011. - Т. 3. - С. 52-55.](#)

[Каленов Н.Е., Савин Г.И., Сотников А.Н. Принципы построения и технологической поддержки электронной библиотеки "Научное наследие России" // IT-Технологии и электронные ресурсы. Межрегиональная науч.- практ. конф. 27.09.- 01.10.2010 г. Респ. Хакасия, г. Абакан, 2010. \(электронный ресурс на CD-ROM\), 2010.](#)

[Антопольский А.Б., Савин Г.И., Сотников А.Н. Электронная библиотека "Научное наследие России": новый этап развития // Информационные ресурсы России, 2010. - № 2 \(114\). - С. 2-6.](#)

Рис. 4. Поисковая форма системы. Пример запроса на поиск публикаций

По умолчанию в поисковой форме в первой строке выбрано поле «Фамилия персоны», во второй — «Название публикации», в третьей — «Название доклада» (см. рис. 4). Поисковый запрос можно ограничить годами публикаций (мероприятий), выбрав необходимые года из выпадающих списков «Год: от ... по ...». Результаты поиска могут быть отсортированы по году или по алфавиту описаний найденных записей в прямом или обратном порядке. По умолчанию сортировка производится по году в обратном порядке (вначале выдаются публикации (мероприятия) текущего года). В поисковой форме системы имеется возможность настройки выдачи найденной информации, выбрав необходимую опцию из выпадающего списка «Показывать»:

- публикации и доклады,

- публикации,
- доклады,
- журналы / сборники,
- мероприятия,
- персоны,
- организации.

По умолчанию в поисковой форме выбрана опция «Публикации и доклады». Результат выполнения поисковых запросов выдается на экран порциями, размер которых задается в выпадающем списке «Показывать ... по ...» (по умолчанию — 20 документов на странице).

В том случае, если все поисковые поля формы оставить пустые, после проведения поиска (нажать на кнопку «Поиск») система выдаст все зарегистрированные записи в соответствии с выбранной опцией «Показывать».

На *рис. 4* приведен пример запроса на поиск совместных публикаций авторов «Савин» и «Сотников» за 2010–2016 гг., названия которых содержат термины «научное» и «наследие». По этому запросу система находит 6 публикаций, которые выдаются в виде стандартных библиографических описаний. В описаниях публикаций авторы и названия журналов (сборников) являются активными ссылками. Название публикации будет являться активной ссылкой, в том случае если метаданные публикации содержат адрес на ее полный текст.

Ссылка на название источника (журнала, сборника) позволяет получить список всех статей, зарегистрированных в системе и опубликованных в данном источнике. При переходе по ссылке от фамилии автора пользователь получит следующую информацию: место работы (название организации, подразделение, отдел) персоны, все эквивалентные записи персоны, дополнительные данные (идентификаторы в различных БД), а также полный список всех публикаций данного автора (см. *рис. 5*).

Ссылка от названия организации (отделения, отдела) позволяет перейти на статьи всех персон, относящихся к данной организации (отделения, отдела).

Соболевская И.Н.

Научно-исследовательский институт (НИИСИ РАН). / Межведомственный суперкомпьютерный центр Российской академии наук (МСЦ РАН). / Отдел информационных ресурсов и систем (ОИРС)

Публикации: 26

Эквивалентные записи:
I. Sobolevskaia
I.N. Sobolevskaia
Irina Sobolevskaia
Sobolevskaia I.N.
И.Н. Соболевская

Идентификаторы в БД Scopus: 57204961984 РИНЦ: 970897

1 2

- N. Kalenov, I. Sobolevskaia, A. Sotnikov, S. Kirillov The use of 3D visualization technology web-collections for the formation of virtual exhibitions // CEUR Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019), Novorossiysk-Abrau, Russia, September 23-28, 2019., 2020. - Vol. 2543. - P. 382-390.
- И.Е. Каленов, С.А. Кириллов, И.Н. Соболевская, А.Н. Сотников Визуализация цифровых 3D-объектов при формировании виртуальных выставок // Электронные библиотеки, 2020. - Т. 23, - № 3. - С. 418-432.
- И.Е. Каленов, И.Н. Соболевская, А.Н. Сотников Формализация процессов формирования пользовательских коллекций в цифровом пространстве научных знаний // Электронные библиотеки, 2020. - Т. 23, - № 3. - С. 433-450.
- N. Kalenov, I. Sobolevskaia, A. Sotnikov 1. Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // CEUR Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019), Novorossiysk-Abrau, Russia, September 23-28, 2019., 2020. - Vol. 2543. - P. 391-398.
- Nikolay Kalenov, Irina Sobolevskaia, Alexander Sotnikov Hierarchical Representation of Information Objects in a Digital Library Environment // Communications in Computer and Information Science, 2019. - Vol. 1093. - P. 93-104.

Рис. 5. Пример полученной информации о персоне

Система обеспечивает возможность выгрузки необходимых пользователю библиографических описаний публикаций в текстовый файл или в структурированный файл формата CSV. Для этого пользователю нужно отметить интересующие его публикации и воспользоваться кнопкой «Выгрузка записей» внизу страницы (см. рис. 4).

Помимо рассмотренного выше поиска публикаций в пользовательском блоке системы аналогичным образом реализован поиск и просмотр докладов. На рис. 6 представлен запрос на поиск докладов, названия которых содержат термин «3D», сделанных в 2018–2019 гг. сотрудниками МСЦ РАН.

В результате проведенного поиска система выдает список записей найденных докладов. Каждая запись содержит

перечень авторов доклада, название доклада и описание мероприятия (название, даты мероприятия, место проведения мероприятия). Фамилии авторов представляют собой активные ссылки, переход по которым обеспечит выдачу информации об авторе и всех докладов данного автора, зарегистрированных в системе. Ссылка от названия мероприятия обеспечивает переход на сайт мероприятия. Название доклада будет являться активной ссылкой, в том случае если метаданные доклада содержат адрес на публикацию доклада или его презентацию.

БД научных трудов сотрудников МСЦ РАН

Организация: МСЦ РАН

И: И

Название публикации: И

И: И

Название доклада: 3D И

Год: от 2018 по 2019

Показывать: Доклады по 20

Сортировка: Год в обратном порядке

Поиск Сброс

Найдено записей: 3

- I.N. Sobolevskaya, A.N. Soltnikov Principles of 3D Web-collections Visualization. 3rd International Conference on Computer-Human Interaction Research and Applications - CHIRA 2019. 20.09 - 21.09.2019 г., Австрия, г. Вена
- Сотников А.Н., Соболевская И.Н., Кириллов С.А., Каленов Н.Е. Применение технологии визуализации 3D WEB-коллекций для формирования виртуальных выставок. XX Всероссийская научная конференция "Научный сервис в сети Интернет". 17.09 - 22.09.2018 г., Россия, г. Новороссийск
- Сотников А.Н., Соболевская И.Н., Кириллов С.А., Чередниченко И.Н. Анализ методов визуализации 3D web-коллекций. XXII научно-практический семинар «Информационное обеспечение науки: новые технологии». 15.10 - 19.10.2018 г., Россия, г. Кунгур

Выгрузка записей: Текст

Рис. 6. Пример запроса на поиск докладов

Так же, как и в случае с публикациями, система позволяет выгружать выбранные записи докладов в текстовом формате и в формате CSV.

Представленная автоматизированная система поддержки базы данных научных трудов в настоящее время функционирует в технологическом режиме в МСЦ РАН.

В системе зарегистрированы 325 персон из 46 организаций; 739 статей, опубликованные в 348 изданиях. Кроме того, в этом году начался ввод данных о сделанных докладах на научных мероприятиях. Было зарегистрировано 24 мероприятия и 38 докладов, сделанных в 2018–2020 гг.

Работа выполнена в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН в рамках государственного задания № 0580-2021-0014.

Литература

1. Захарова С. С., Гуреева Ю. А. Научные публикации: от картотеки трудов до библиографических профилей // Библиосфера. 2017. Т. 2. С. 85–89.
2. Мазов Н. А., Гуреев В. Н. Создание базы данных трудов сотрудников организации как основы для достоверного библиометрического анализа // Труды ГПНТБ СО РАН. 2015. № 9. С. 103–109.
3. Василенко Е. А., Панфилов В. И., Жуков Д. Ю., Сивуха Д. В. Разработка наукометрической базы данных публикационной активности ученых РХТУ им. Д. И. Менделеева // Научно-техническая информация. Сер. 1. 2013. № 11. С. 26–30.
4. Столяров Р. А., Чугреев В. Л. Автоматизированная система учета результатов интеллектуальной деятельности в научной организации // Вопросы территориального развития. 2015. № 6 (26). С. 4.
5. Кочергина Т. А., Исхакова Л. Д., Каленов Н. Е., Якшин М. М. База данных «Труды сотрудников НЦВО РАН» как часть информационно-поисковой системы «Волоконная оптика» // Системы и средства информатики. 2013. Т. 23. № 2. С. 284–296.
6. Левченко О. И., Соловьев А. В. Формирование базы данных публикаций сотрудников Института физики твердого тела РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Каленов Н. Е., Цветкова В. А. (ред.). М.: БЕН РАН, 2015. С. 215–221.
7. Власова С. А. Автоматизированная система поддержки корпоративной базы данных научных публикаций // Программные продукты, системы и алгоритмы. 2018. Вып. 2. С. 42–46.
8. Власова С. А., Каленов Н. Е. Информатика в академической библиотеке // Системы и средства информатики, 2016. Т. 26. № 3. С. 162–178.
9. Власова С. А., Каленов Н. Е. Интернет-каталог Библиотеки по естественным наукам Российской академии наук как специальная информационно-поисковая система, ориентированная

на квалифицированного пользователя // Системы и средства информатики, 2019. Т. 29. № 1. С. 86–95.

References

1. Zakharova S. S., Gureeva Yu. A. Nauchnye publikatsii: ot kartoteki trudov do bibliograficheskikh profiley // Bibliosfera. 2017. Vol. 2. P. 85–89.

2. Mazov N. A., Gureev V. N. Sozdanie bazy dannykh trudov sotrudnikov organizatsii kak osnovy dlya dostovernogo bibliometricheskogo analiza // Trudy GPNTB SO RAN. 2015. № 9. P. 103–109.

3. Vasilenko E. A., Panfilov V. I., Zhukov D. Yu., Sivukha D. V. Razrabotka nauko-metricheskoy bazy dannykh publikatsionnoy aktivnosti uchenykh RKhTU im. D. I. Mendeleeva // Nauchno-tehnicheskaya informatsiya. Ser. 1. 2013. № 11. P. 26–30.

4. Stolyarov R. A., Chugreev V. L. Avtomatizirovannaya sistema ucheta rezul'tatov intellektual'noy deyatel'nosti v nauchnoy organizatsii // Voprosy territorial'nogo razvitiya. 2015. № 6 (26). P. 4.

5. Kochergina T. A., Iskhakova L. D., Kalenov N. E., Yakshin M. M. Baza dannykh «Trudy sotrudnikov NTsVO RAN» kak chast' informatsionno-poiskovoy sistemy «Volokonnaya optika» // Sistemy i sredstva informatiki. 2013. Vol. 23, № 2. P. 284–296.

6. Levchenko O. I., Solov'ev A. V. Formirovanie bazy dannykh publikatsiy sotrudnikov Instituta fiziki tverdogo tela RAN // Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov / Kalenov N. E., Tsvetkova V. A. (red.). M.: BEN RAN, 2015. P. 215–221.

7. Vlasova S. A. Avtomatizirovannaya sistema podderzhki korporativnoy bazy dannykh nauchnykh publikatsiy // Programmnye produkty, sistemy i algoritmy. 2018. Iss. 2. P. 42–46.

8. Vlasova S. A., Kalenov N. E. Informatika v akademicheskoy biblioteke // Sistemy i sredstva informatiki, 2016. Vol. 26, № 3. P. 162–178.

9. Vlasova S. A., Kalenov N. E. Internet-katalog Biblioteki po estestvennym naukam Rossiyskoy akademii nauk kak spetsial'naya informatsionno-poiskovaya sistema, orientirovannaya na kvalifitsirovannogo pol'zovatelya // Sistemy i sredstva informatiki, 2019. Vol. 29. № 1. P. 86–95.

Некоторые аспекты представления 3D-объектов в едином цифровом пространстве научных знаний

И. Н. Соболевская

МЦЦ РАН

Аннотация. Рассматриваются аспекты формирования и представления мультимедийных объектов в едином цифровом пространстве научных знаний (ЕЦПНЗ). Описаны приемы использования новых технологических решений для переноса изображений физических объектов в цифровое пространство. Приведен пример технологии представления виртуальных 3D-моделей в интернет-пространстве. Описан принцип формирования научной виртуальной выставки в ЕЦПНЗ.

Ключевые слова: научная информация, структура пространства знаний, 3D-модель, музейный объект, виртуальная выставка контент.

Some aspects of 3D-objects presentation in a common digital space of scientific knowledge

I. N. Sobolevskaya

JSCC RAS

Abstract. The paper deals with the issues of multimedia objects presentation in a common digital space of scientific knowledge. Examples of using new technological solutions for transferring images of physical objects into virtual space are given. The technology for representing digital 3D-models in the environment of a common digital space of scientific knowledge is proposed. The principle of scientific virtual exhibitions formation in the environment of a common digital space of scientific knowledge is considered.

Keywords: scientific information, structure of knowledge space, 3D-model, museum object, virtual exhibition content.

Введение

На сегодня существует несколько крупных агрегаторов информационных ресурсов, предоставляющих доступ не только к отдельным цифровым копиям печатных изданий, но и позволяющих формировать и просматривать тематические коллекции или виртуальные выставки. На сайтах таких сервисов представлены переведенные в цифровую форму монографии, художественные альбомы, ноты, картографические материалы, рукописи, периодические издания и др. [1] Активно развиваются научные электронные ресурсы, предоставляющие широкому кругу пользователей такие сервисы, как:

- поиск по различным научным дисциплинам и источникам в специальных базах данных;
- поиск по различным научным дисциплинам и источникам в полнотекстовых электронных изданиях крупных научных издательств мира;
- доступ к найденным информационным ресурсам;
- использование приложений, превращающих электронные библиотеки в исследовательский полигон (виртуальная обсерватория, виртуальная химическая или биологическая лаборатория и т. п.);
- сохранность научной информации;
- интеграция цифровых образов информационных объектов различной природы;
- обеспечение возможностей научного сотрудничества в региональном и международном масштабах.

Методы представления мультимедийных объектов в ЕЦПНЗ

При создании единого цифрового пространства научных знаний (ЕЦПНЗ) необходимо решать задачи как формирования контента ЕЦПНЗ, так и его визуализации для научных, образовательных и познавательных целей [2]. Визуализация контента ЕЦПНЗ позволит эффективно предоставлять пользователю информацию о динамических про-

цессах и объемных объектах (реконструкция исторических памятников, моделирование физических и технологических процессов, визуализация решений уравнений механики или химических реакций), а также предоставлять такие «научно-популярные» сервисы, как экскурсии по музеям, научным лабораториям, обсерваториям и т. д.

Для представления мультимедийных объектов, инструментарий формирования ЕЦПНЗ должен обеспечивать следующие основные возможности:

- детализация всего изображения (или его части) с сохранением качества изображения;
- получение справочной (или пояснительной информации) об объекте;
- воспроизведения аудиосопровождения, соответствующего статичному или динамичному визуальному ряду;
- воспроизведения видеоряда с использованием функции «стоп-кадра»;
- создания пользовательских коллекций;
- использования игровых компонентов с информационными составляющими.

Методы представления объектов библиотечного и архивного хранения в цифровой форме, а также предоставления доступа к ним, на сегодняшний день разработаны достаточно хорошо. Однако представление мультимедийных объектов в ЕЦПНЗ предполагает не только создание средств для описания этих объектов и обеспечения доступа к ним, но и формирование средств интеграции и погружения мультимедийных объектов в цифровую среду. Кроме того, для музейных объектов необходимо формирование качественных, с точки зрения зрительного восприятия, цифровых образов их 3D-моделей. Информационная среда, содержащая мультимедийные объекты наряду с объектами библиотечного и архивного хранения, позволит создавать междисциплинарные мультимедийные цифровые коллекции и формировать виртуальные выставки.

Для представления цифровых 3D-моделей в среде ЕЦПНЗ целесообразно использовать, так называемую, техно-

логию интерактивной анимации. Эта технология основана на программной смене (прокручивании) фиксированного набора видов объекта (кадров) с помощью специализированных интерактивных программ отображения, имитирующих смену точки взгляда на исходный объект и, позволяющих масштабировать как все изображение, так и его фрагмент. Для создания такой интерактивной анимации необходим набор заранее снятых сцен, которые будут использоваться как отдельные кадры экспозиции (рис. 1) [3].



Рис. 1. Набор заранее снятых сцен, которые будут использоваться как отдельные кадры экспозиции

Структура виртуальной выставки

Виртуальные выставки, как информационный ресурс, предоставляющий пользователям разнородную информацию (цифровые копии печатной продукции, архивных документов, мультимедийные материалы и т. п.), объединенную по определенным признакам, позволяют эффективно решать задачи представления информационных ресурсов,

в том числе, цифровых музейных коллекций, интегрированных средствами ЕЦПНЗ [4, 5].

Несмотря на то, что каждая виртуальная выставка уникальна по своему наполнению, при формировании таких выставок в среде ЕЦПНЗ можно выделить следующие основные типы разделов:

- основной тематический раздел;
- интерактивный раздел;
- биографический раздел;
- раздел видеоматериалов;
- раздел фотодокументов;
- библиотека;
- раздел «Коллекция 3D-объектов»;
- раздел отзывов;
- контакты.

Основной тематический раздел. В данном разделе содержится подборка материалов, раскрывающих основную тему виртуальной выставки. Как правило, это статьи (в том числе авторские, написанные специально для раскрытия тематики выставки), подборки редких фотоматериалов и изображений, выдержки из публицистических и периодических изданий (рис. 2).



Михаил Михайлович Герасимов (1907 – 1970)

КТО

Михаил Михайлович Герасимов — замечательный советский учёный, археолог и реставратор, антрополог и скульптор. Он автор метода пластической портретной реконструкции, лауреат Сталинской премии (1950), доктор исторических наук (1956), Заслуженный деятель науки РСФСР (1969).

ЧЕМ ЗНАМЕНИТ

Открыл и исследовал одну из интереснейших по находкам каменного и костяного инвентаря позднелеолитических стоянок Сибири - [Мальта](#). На стыке своих научных интересов разработал [метод](#) восстановления внешнего облика людей по костным остаткам. Основал [Лабораторию пластической реконструкции](#) при Институте этнографии и антропологии РАН. Создал около 200 скульптурных реконструкций облика древних людей и более 20 портретов исторических деятелей.

Рис. 2. Основной тематический раздел (на примере виртуальной выставки «Портреты по скелетам»)

Интерактивный раздел. Управляющие элементы интерфейса раздела должны дать возможность пользователю взаимодействовать с элементами виртуальной экспозиции. Задачей данного раздела является повышение заинтересованности пользователя (посетителя виртуальной выставки), переход от пассивного восприятия информации к активному пониманию коллекции. Основными элементами интерактивного раздела являются научные викторины, интеллектуальные игры, 3D-анимации (рис. 3).



Рис. 3. Интерактивный раздел (на примере виртуальной выставки «Сад жизни»)

Биографический раздел. Этот раздел содержит биографические справки и портреты ученых (рис. 4).

[Главная](#) [Биография](#) [Литература](#) [Труды М.М. Герасимова](#) [Олицетворяя прошлое](#) [Каталог работ](#) [Документы эпохи](#)
о проекте М.М. Герасимова о М.М. Герасимове биографии метода реконструкции 26 моделей и атрибуции эпохи

БИОГРАФИЯ

М.М. Герасимова

БИОГРАФИЯ



Михаил Михайлович родился 15 сентября 1907 года в Петербурге, в семье врача. Вскоре после рождения сына семья переехала в Иркутск. Именно там определилось научное призвание Михаила Михайловича. С ранних лет его увлекала антропология и археология. В 11 лет он впервые участвовал в раскопках стоянки древнего человека на Верхоненской Горе, с 15 лет работал в археологических экспедициях под руководством профессора этнологии Иркутского университета Бернгарда Эдуардовича Петри (1884–1937).

В 1926 году восемнадцатилетний Михаил стал сотрудником Иркутского краеведческого музея, а через год ему доверили заведование археологическим отделом. К этому времени относятся его первая научная публикация о раскопках палеолитического местонахождения у Переселенческого пункта в Иркутске (1926) и первые опыты в реконструкции облика первобытного человека (1927). В 1932 году М. М. Герасимов переехал в Ленинград. Работу в Институте истории материальной культуры (ИИМК) он совмещал с работой реставратором в Государственном Эрмитаже, не оставляя при этом и археологических изысканий (Мальта: 1934 и 1936–1937). В начале 1941 года в Ташкенте по инициативе Комитета Навои была создана научная комиссия для

Рис. 4. Биографический раздел (на примере виртуальной выставки «Портреты по скелетам»)

Раздел видеоматериалов. Видео материалы в этом разделе представлены документальными фильмами, архивными видеоматериалами, и/или научно-популярными фильмами. Просмотр видеоматериалов возможен в различных режимах. Также реализуются все необходимые элементы управления для просмотра видео. Кроме материалов, предоставленных создателями виртуальной выставки, в разделе кинодокументов формируются ссылки на видеоматериалы, находящиеся в свободном доступе в сети интернет, если таковые имеются (рис. 5).

Российский государственный архив кинофотодокументов
(РГАКФД)

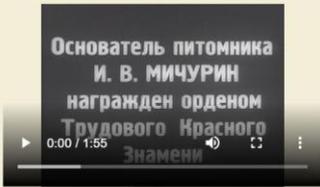
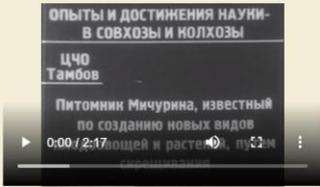
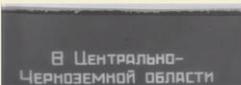
№	Название РГАК	Краткое описание	Год выпуска	Длительность
1	Совкиножурнал № 69\248.		1929	1 мин. 54 сек.
2	Социалистическая деревня №24.		1931	2 мин. 24 сек.
				

Рис. 5. Раздел видеоматериалов (на примере виртуальной выставки «Сад жизни»)

Раздел фотодокументов. Раздел, как правило содержит уникальные фотодокументы, предоставленные участниками проекта по формированию виртуальной выставки (рис. 6).

Электронная библиотека. В разделе представлены публикации по заданной тематике из фондов различных электронных библиотек. Дополнительно представляются ссылки на издания, находящиеся в свободном доступе в сети интернет. Для изданий, которые по каким-либо причинам еще не оцифрованы, предоставляются библиографические списки, позволяющие пользователям коллекции найти и прочитать эти книги в научных библиотеках (рис. 7).



Рис. 6. Раздел фотодокументов (на примере виртуальной выставки «Портреты по скелетам»)

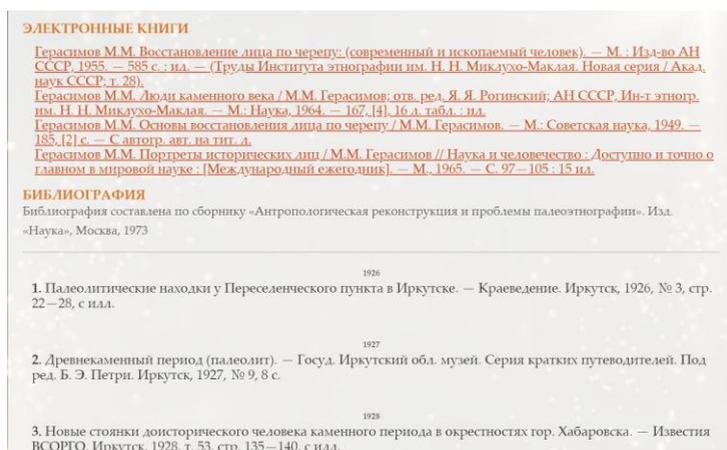


Рис. 7. Электронная библиотека (на примере виртуальной выставки «Портреты по скелетам»)

Раздел «Коллекция 3D». Раздел представляет собой галерею, состоящую из 3D-макетов оцифрованных архивных или музейных предметов. Предметы оцифрованы таким образом, чтобы пользователь междисциплинарной коллекции смог детально рассмотреть предмет во всех ракурсах, «поворачи-

вая» объект (рис. 8). «Вращение» возможно благодаря разработанным средствам представления 3D-объектов.

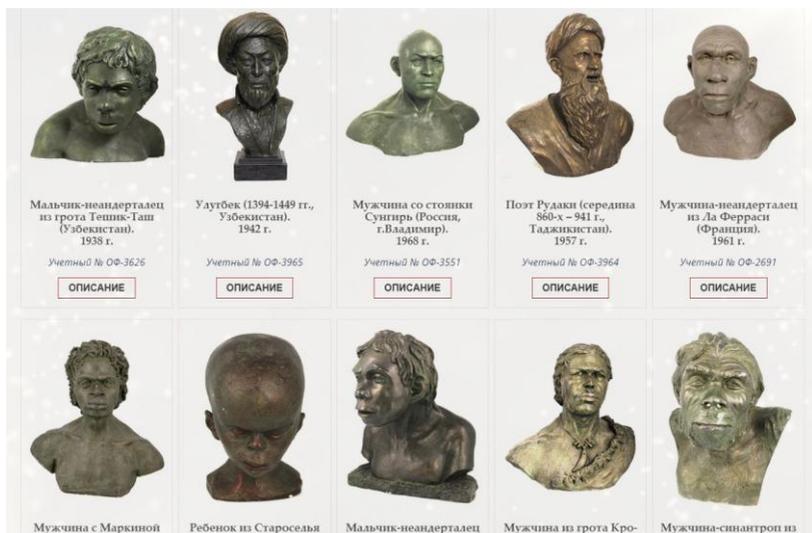


Рис. 8. Коллекция 3D (на примере виртуальной выставки «Портреты по скелетам»)

Раздел отзывов. Данный раздел создается для обратной связи с посетителями виртуальной выставки. Здесь можно (после обязательной регистрации) обменяться мнениями, выступить с различными сообщениями по тематике выставки.

Раздел «контакты». В данном разделе указываются контакты администратора виртуальной выставки для связи с ним.

Заключение

Описанная схема формирования виртуальных тематических выставок реализована в цифровых проектах, интегрированных с электронной библиотекой «Научное наследие России» [6]. Эти выставки посвящены 160-летию со дня рождения И. В. Мичурина и развитию генетики в СССР (<http://>

vim.benran.ru/) и научному наследию М. М. Герасимова (<http://acadlib.ru/>).

Исследования в области формирования ЕЦПНЗ проводятся в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН в рамках госзадания 0580-2021-0016.

Литература

1. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук, 2019. Т. 89. № 7. С. 728–735.
2. Антопольский А. Б. и др. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.
3. Сотников А. Н., Соболевская И. Н., Кириллов С. А., Каленов Н. Е. Применение технологии визуализации 3D веб-коллекций для формирования виртуальных выставок // Труды XXII Всероссийской научной конференции «Научный сервис в сети интернет». 2019. С. 621–626.
4. Munster S. Digital Heritage as a Scholarly Field-Topics, Researchers, and Perspectives from a Bibliometric Point of View // Acm journal on computing and cultural heritage. 2019. 12(3), № 12.
5. Wuttke U., Spiecker C., Neuroth H. PARTHENOS – A Digital Research Infrastructure for eHumanities and eHeritage // Bibliothek forschung und praxis. 43(1). 2019. P. 11–20.
6. Sobolevskaya I. N., Sotnikov A. N. Principles of 3D Web-collections Visu-alization. // Proceedings of the 3rd Inter-national Conference on Computer-Human Interaction Research and Appli-cation. SCITEPRESS. 2019.

References

1. Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N. O edinom cifrovom prostranstve nauchnyh znaniy // Vestnik Rossijskoj akademii, 2019. T. 89. № 7. S. 728–735.
2. Antopol'skij A. B. i dr. Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy (ECPNZ) // Nauchno-tehnicheskaya informaciya. Ser. 1. 2020. № 4. S. 9–17.
3. Sotnikov A. N., Sobolevskaya I. N., Kirillov S. A., Kalenov N. E. Primenenie tekhnologii vizualizacii 3D web-kollekcij dlya formirovaniya virtual'nyh vystavok // Trudy XXI Vserossijskoj nauchnoj konferencii «Nvuchnyj servis v seti internet. 2019. S. 621–626.

4. Munster S. Digital Heritage as a Scholarly Field-Topics, Researchers, and Perspectives from a Bibliometric Point of View // *Acm journal on computing and cultural heritage*. 2019. 12(3), № 12.
5. Wuttke U., Spiecker C., Neuroth H. PARTHENOS – A Digital Research Infrastructure for eHumanities and eHeritage // *Bibliothek forschung und praxis*. 43(1). 2019. P. 11–20.
6. Sobolevskaya I. N., Sotnikov A. N. Principles of 3D Web-collections Visu-alization. // *Proceedings of the 3rd Inter-national Conference on Computer-Human Interaction Research and Appli-cation*. SCITEPRESS. 2019.

Формирование цифровых ресурсов для электронной библиотеки «Научное наследие России»

С. А. Кириллов

МСЦ РАН

Аннотация. В статье описаны методы формирования и представления цифровых образов объектов различных типов (печатных изданий, архивных материалов, предметов музейного хранения, аудио-видеоматериалов) для электронной библиотеки «Научное наследие России».

Ключевые слова: информационное сопровождение науки, научная информация, контент.

Formation of digital resources for the electronic library «Scientific Heritage of Russia»

S. A. Kirillov

JSCC RAS

Abstract: The article describes methods for the formation and presentation of digital images of objects of various types (printed publications, archival materials, museum items, audio-video materials) for the electronic library «Scientific Heritage of Russia».

Keywords: information support for science, scientific knowledge preservation, popularizing science.

Электронная библиотека «Научное наследие России» (ЭБ ННР) представляет собой типовое решение электронной библиотеки, позволяющей интегрировать помимо электронных версий печатной продукции информацию другого рода, в частности: архивные материалы, предметы музейного хра-

нения, представленные в виде: графики, фотографий, 3D-моделей музейных предметов, аудио- и видео материалов.

Электронная библиотека «Научное наследие России» представляет собой распределенную информационную систему, объединяющую множество дистанционно удаленных участников, работающих по единым правилам.

Базовыми компонентами ЭБ ННР являются типовые центры подготовки цифрового контента, размещенные у держателей информационных фондов (в архивах, библиотеках, музеях, научных институтах) в различных регионах России.

Основная задача таких центров — обеспечить формирование информационных ресурсов следующих видов:

- цифровых копий печатных и рукописных документов библиотечного, архивного и музейного хранения;
- цифровых моделей музейных предметов;
- цифровых копий изобразительных и аудио-видео материалов.

Разработанная архитектура типового центра формирования цифрового контента предполагает выполнение следующих технологических задач:

- подбор контента;
- описание контента (формирование метаданных);
- оцифровка;
- обработка оцифрованного контента;
- верстка электронного издания (медиа-объекта);
- публикация электронного издания на Интернет-портале.

Технология формирования цифрового контента

В рамках проекта по созданию ЭБ ННР разработано типовое решение построения электронной библиотеки. Сотрудниками МСЦ РАН, ВЦ РАН и БЕН РАН разработаны специальные прикладные программные средства [1–8], на базе которых реализован технологический аппаратно-программный комплекс подготовки электронных изданий, состоящий из следующих подсистем.

Подсистема диспетчеризации. Данная подсистема разработана для организации ведения учета электронных изданий, предназначенных для формирования контента ЭБ ННР. Подсистема обеспечивает исключение возможности дублирования при изготовлении электронных изданий, планирование их подготовки и ряд сервисных функций, связанных с организацией работ по формированию цифровых объектов.

Подсистема описания контента. Эта подсистема включает специальную программную оболочку и базу данных, в которую вводятся метаданные, описывающие информационные объекты ЭБ ННР. Подсистема располагается на специальном технологическом сервере; доступ к ее содержимому возможен только авторизованным пользователям, обладающим правами ввода и (или) корректировки данных.

Подсистема оцифровки. Подсистема предназначена для изготовления электронных копий объектов библиотечного, архивного и музейного хранения, которые были выбраны в качестве контента для включения в ЭБ ННР. Для создания электронной библиотеки информация (например, страницы книги) переводится в цифровой вид с использованием различных типов сканеров, отличающихся разрешающей способностью, возможностью сканирования различного типа изображений и другими свойствами. В связи с этим в проекте сформировались следующие основные типы комплексов оцифровки:

- Комплекс черно-белого сканирования на основе планетарного сканера MINOLTA PS7000. Сканер Minolta PS7000 оцифровывает книги формата А5–А3 с разрешением 600 dpi.

- Комплекс сканирования на основе цветного планетарного сканера ПланСкан С2-ЦА-600. Этот комплекс сканирования пришел на смену устаревшего комплекса на основе сканера Minolta PS7000. Оператор сканера ПланСкан С2-ЦА-600 имеет возможность сканировать любые типы публикаций до формата А2 (в цветном, сером и черно-белом режиме).

- Комплекс планетарного цветного сканирования ПланСкан А2-VC-B. Комплекс планетарного цветного сканирования ПланСкан А2-VC-B по характеристикам аналогичен ПланСкан С2-ЦА, однако он более компактен, оборудован встроенной светодиодной системой подсветки. Основной особенностью данного комплекса является возможность сканировать ветхие и плотно сшитые издания.

- Комплекс сканирования на основе сканера Пауэрскан Д14000 А0-20/25. Основной особенностью данного комплекса является возможность сканировать материалы особо большого формата (вплоть до размера А0), с высокой разрешающей способностью, например, газеты, карты, плакаты, картины.

- Комплекс сканирования на основе сканера микрофиш Kodak ABR 3000 DSV. Данный комплекс служит для оцифровки материалов, хранящихся на микрофишах и микроплёнке.

- Комплекс 3D-оцифровки на основе поворотной платформы Resam T-50, управляющей программы 3D-Maker и цифрового фотоаппарата Canon EOS600D. Комплекс позволяет выполнять в автоматическом режиме 3D-съёмку музейных предметов высотой до 150 см и весом до 50 кг.

Применение указанных комплексов оцифровки позволяет выполнять максимально точный перевод публикаций в цифровую форму.

Подсистема формирования электронных изданий. Эта подсистема состоит из набора программных продуктов, предназначенных для создания конечного продукта – электронных копий объектов библиотечного, архивного и музейного хранения. В рамках подсистемы используется ряд сервисных программ.

Программа NNMetaDis предназначена для формирования электронных книг. Она позволяет выполнять сборку книги из оцифрованных страниц, вводить библиографические описания изданий и их оглавления.

При формировании 3D-моделей предметов музейного хранения применяются программы Autodesk 3dsMax,

3DMaker, а также набор Java-скриптов собственной разработки.

Для представления фотографий, портретов, карт, картин и т. п. применяется набор макрокоманд собственной разработки в среде Adobe PhotoShop.

Для представления аудио и видео-ресурсов используется набор редакторов, кодеков и конверторов свободного и открытого программного обеспечения.

Подсистема контроля и хранения контента. Подсистема включает методы контрольных операций: проверку качества сканирования книги, наличия страниц, порядка их размещения, правильности и полноты метаданных и т. д. Система включает методы хранения электронных изданий в специальной архивной системе.

Функционирование ЭБ как общедоступной поисковой системы обеспечивает сервер интернет-портала. Пользовательский интерфейс ЭБ ННР обеспечивает многоаспектный поиск информации, ее визуализацию, навигацию по связанным объектам Библиотеки. В рамках ЭБ ННР предусмотрен ряд сервисов, предусматривающих интеграцию с информационными ресурсами других организаций.

Разрабатываемые в проекте ЭБ ННР методы формирования и представления цифровых образов различных типов объектов (печатных изданий, архивных материалов, предметов музейного хранения, аудио- и видеоматериалов) содействуют развитию электронной информационно-библиотечной среды: помогают совершенствовать электронное описание документов, приемы и навыки оцифровки, расширяют поисковые возможности системы. Создаваемый фонд оцифрованных документов и справочно-библиографический аппарат вносят существенный вклад в формирование единого цифрового пространства научных знаний.

Исследования в области формирования ЕЦПНЗ проводятся в МСЦ РАН — филиале ФГУ ФНЦ НИИСИ РАН в рамках госзадания 0580–2021–0014.

Литература.

1. Каленов Н. Е., Савин Г. И., Серебряков В. А., Сотников А. Н. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы, 2012. Т. 4, № 100. С. 30–40.
2. Погорелко К. П. Эволюция программного обеспечения системы подготовки материалов для электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: сборник научных трудов. М.: Научный Мир, 2011. С. 260–263.
3. Якшин М. М. Платформа SciRus – основа технологического комплекса электронной библиотеки «Научное наследие России» // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: XVI Всероссийская научная конференция RCDL-2014. Труды конференции. М., 2014. С. 362–368.
4. Якшин М. М. Развитие платформы SciRus // Информационное обеспечение науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2015. С. 203–207.
5. Погорелко К. П. Система экспорта данных электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2015. С. 200–202.
6. Сотников А. Н., Кириллов С. А. Технология подготовки электронных изданий для электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2015. С. 178–190.
7. Забровская И. Е., Кириллов С. А., Кондратьева Е. А., Пругло О. А., Сотников А. Н. Вопросы формирования фондов электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сборник научных трудов. М.: БЕН РАН, 2017. С. 184–191.
8. Каленов Н. Е., Кириллов С. А., Соболевская И. Н., Сотников А. Н. Современное состояние электронной библиотеки «Научное наследие России» // Труды НИИСИ РАН. Математическое и компьютерное моделирование сложных систем: теоретические и прикладные аспекты, 2018. Т. 8. № 6. С. 166–169.

References

1. Kalenov N. E., Savin G. I., Serebryakov V. A., Sotnikov A. N. Principy` postroeniya i formirovaniya e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Programmy`e produkty` i sistemy`, 2012. T. 4, № 100. S. 30–40.

2. Pogorelko K. P. E`volyuciya programmnoo obespecheniya sistemy` podgotovki materialov dlya e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Informacionnoe obespechenie nauki: novy`e texnologii: sbornik nauchny`x trudov. M.: Nauchny`j Mir, 2011. S. 260–263.
3. Yakshin M. M. Platforma SciRus — osnova texnologicheskogo kompleksa e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // E`lektronny`e biblioteki: perspektivny`e metody` i texnologii, e`lektronny`e kollekcii: XVI Vserossiyskaya nauchnaya konferenciya RCDL-2014. Trudy` konferencii. M., 2014. S. 362–368.
4. Yakshin M. M. Razvitie platformy` SciRus // Informacionnoe obespechenie nauki: novy`e texnologii: Sbornik nauchny`x trudov. M.: BEN RAN, 2015. S. 203–207.
5. Pogorelko K. P. Sistema e`ksporta danny`x e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Informacionnoe obespechenie nauki: novy`e texnologii: Sbornik nauchny`x trudov. M.: BEN RAN, 2015. S. 200–202.
6. Sotnikov A. N., Kirillov S. A. Texnologiya podgotovki e`lektronny`x izdaniy dlya e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Informacionnoe obespechenie nauki: novy`e texnologii: Sbornik nauchny`x trudov. M.: BEN RAN, 2015. S. 178–190.
7. Zabrovskaya I. E., Kirillov S. A., Kondrat`eva E. A., Pruglo O. A., Sotnikov A. N. Voprosy` formirovaniya fondov e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Informacionnoe obespechenie nauki: novy`e texnologii: Sbornik nauchny`x trudov. M.: BEN RAN, 2017. S. 184–191.
8. Kalenov N. E., Kirillov S. A., Sobolevskaya I. N., Sotnikov A. N. Sovremennoe sostoyanie e`lektronnoj biblioteki «Nauchnoe nasledie Rossii» // Trudy` NIISI RAN. Matematicheskoe i komp`yuterno modelirovanie slozhny`x sistem: teoreticheskie i prikladny`e aspekty`, 2018. T. 8. № 6. S. 166–169.

Разработка тезаурусов для поиска и навигации в цифровом пространстве знаний

Н. Н. Шабурова

ИФП СО РАН

В. Н. Белоозеров

ВИНИТИ РАН

Аннотация. Рассмотрены принципы разработки тезаурусов, установления тезаурусных отношений и дефинитивных связей, а также решения ключевых задач, стоявших перед разработчиками. Представлены результаты создания классификационно-тезаурусной сети смысловых связей терминов и тематических рубрик, используемых при создании информационных ресурсов по физике и электронике. Материал может быть предложен как подход к созданию семантического инструментария для навигации по единому пространству знаний.

Ключевые слова: онтология, тезаурус, ключевые слова, классификационные системы, единое пространство знаний, родовидовые связи, дефинитивная связь.

Development of thesaurus for search and navigation in the digital knowledge space

N. N. Shaburova

Institute for Applied Physics SB RAS

V. N. Beloozerov

VINITI RAS

Abstract. The principles of thesaurus development, establishment of thesaurus and definitive relations, as well as of the solution of key problems faced by the developers, are considered. The results of creating the classification-thesaurus network of semantic links of terms and thematic

headings used in the field of developing information resources in physics and electronics are presented. The material can be proposed as an approach to the creation of semantic tools for navigating the common space of knowledge.

Keywords: ontology, thesaurus, keywords, classification systems, unified knowledge space, generic relationships, definitive relationships.

1. Введение

Создание единого пространства знаний состоит в объединении информационных ресурсов общими средствами поиска и извлечения знаний. Универсальным средством для этого является естественный язык, который и служит людям в этом качестве в условиях, когда источники знания представлены в воспринимаемой человеком форме и в обозримом количестве. Развитие компьютерных средств хранения знаний требует использования компьютерных языков поиска и интерпретации компьютерных данных, доступных в едином пространстве. Перспективы компьютерного анализа смысла **полного** текста остаются проблематичными. Поэтому целью нашей работы стало исследование возможности и путей использования **метаданных** публикаций, которые сами уже являются продуктом интеллектуальной интерпретации документа с целью облегчить поиск содержащихся в публикации знаний. Такими метаданными являются ключевые слова и классификационные индексы, которыми снабжаются практически все документы в информационном пространстве науки и техники. Этим наш подход отличается от направления развития современных поисковых машин, которые ориентированы на «свободную лексику» и игнорируют труд индексаторов научных публикаций, результаты которого уже вложены в имеющиеся информационные ресурсы.

Предположение, что цифровое пространство будет включать разнородные автономные источники знаний, ставит задачу согласования средств, с помощью которых систематизированы сведения в разнородных ресурсах [1]. Однако сейчас отсутствует единство и универсальность этих средств описания тематики. Разные источники информации описы-

вают свои документы разными тематическими рубрикаторами, а многие вообще не пользуются тематической рубрикацией, а индексируют документы ключевыми словами, которые отнюдь не всегда позволяют определить тематику работы. Эта ситуация может быть преодолена установлением сети смысловых связей рубрик используемых классификационных систем при привязке к ним ключевых слов, индексирующих документы данной рубрики и также связанных сетью смысловых отношений.

Иными словами, для реализации в среде связанных компьютерных ресурсов единого пространства знаний по определённой научной области требуется создать классификационно-тезаурусную сеть смысловых связей терминов и тематических рубрик, которые были использованы при создании информационных ресурсов данной области. Примером интеграции в одном инструменте различных средств доступа к содержанию ресурсов может служить успешно действующая в Национальной библиотеке по медицине США Единая медицинская лингвистическая система, включающая «Метатезаурус» [2], который объединил лексику около двухсот классификаторов, справочников, словарей и номенклатур по частным медицинским проблемам. Для физики и электроники мы начали осуществлять такую же идею с 2008 г. [3]

2. Тезаурус тематических рубрик

В 2009 г. была создана первая версия Тезауруса тематических рубрик по физике полупроводников (ТРФП) [4], который включал рубрики следующих классификаций:

- Универсальная десятичная классификация (УДК).
- Библиотечно-библиографическая классификация (ББК).
- Государственный рубрикатор научно-технической информации (ГРНТИ).
- Рубрикатор ВИНТИ.

– Классификационная схема по физике и астрономии Американского физического общества (Physics and Astronomy Classification Scheme – PACS).

ТРФП в составе около 1600 дескрипторов был выложен на сайте научной библиотеки Института физики полупроводников им. А. В. Ржанова СО РАН (<http://lib.isp.nsc.ru/library/links.php>) и депонирован в ВИНТИ [5]. В табл. 1 показаны образцы статей тезауруса. Каждому дескриптору приписаны коды пяти классификаций и даны обычные тезаурусные связи дескрипторов.

Таблица 1

**Образцы статей Тезауруса тематических рубрик
по физике полупроводников**

антисегнетоэлектрики = диэлектрики, не являющиеся сегнетоэлектриками, но обладающие определенной спецификой электрических свойств. Основной признак антисегнетоэлектрика – наличие структурного фазового перехода, сопровождающегося значительной аномалией диэлектрической проницаемости

*ББК В379.331.5:331.7

ВИНТИ 291.19.35

ГРНТИ 29.19.35

УДК 537.226.4

В: сегнетоэлектричество

антиферромагнетизм полупроводников = одно из магнитных состояний вещества, отличающееся тем, что элементарные (атомные) магнитики соседних частиц вещества ориентированы навстречу друг другу (антипараллельно), и поэтому намагниченность тела в целом очень мала. Этим антиферромагнетизм отличается от ферромагнетизма, при котором одинаковая ориентация элементарных магнетиков приводит к высокой намагниченности тела.

ББК В379.233.4

*УДК 537.611.45:537.622.5:621.315.59

%УДК 537.311.322.04:537.611.45

В: антиферромагнетизм – теория

В: антиферромагнетики и слабый ферромагнетизм

В: магнитные свойства полупроводников

антиферромагнетизм – теория ББК В373.35 УДК 537.611.45.01 В: магнетизм – теория Н: антиферромагнетизм полупроводников
антиферромагнетики См: антиферромагнитные материалы
антиферромагнетики и слабый ферромагнетизм ББК В373.34/35 ВИНТИ 291.19.43 ГРНТИ 29.19.43 УДК 548:537.611.45/.46 В: магнитные свойства твёрдых тел Н: антиферромагнетизм полупроводников Н: антиферромагнитные материалы

Продолжение табл. 1

антиферромагнитные материалы УДК 537.622.5 С: антиферромагнетики В: антиферромагнетики и слабый ферромагнетизм В: магнитные материалы В: магнитные свойства материалов
арсенид галлия / полупроводниковые свойства/ *ББК Г125.315:Г123.31-2:з843.3 *УДК 546.681'19:621.315.59 В: галлий и его соединения – полупроводниковые свойства Н: полупроводниковые соединения
атомная физика УДК 539.18 Н: свойства атомов
атомносиловая микроскопия в спинтронике ВИНТИ: 291.19.36.23.27 В: микроскопия В: экспериментальные методы спинтроники

В дальнейшем тематика физики полупроводников была дополнена терминами смежных областей физики, электроники и нанотехнологий, а модель тезауруса была предложена для построения онтологий других предметных

областей [6]. В настоящее время реализуется именно эта модель как широкая система классификаций и ключевых слов для различных областей знания.

Постановка задачи и начальный этап работ по созданию базы данных ключевых слов ТЕРМИН в которой реализована семантическая сеть смысловых связей терминов, описана в статьях [7, 8].

В рамках этой концепции разработаны тезаурусы «Физика» и «Электроника» для областей знания, определяемых разделами ГРНТИ: 29 *Физика* и 47 *Электроника. Радиотехника* [9]. В основе тезаурусов лежат ключевые слова, указанные индексаторами ВИНТИ и БЕН РАН как наиболее важные для каждой подрубрики данных разделов ГРНТИ. Каждое ключевое слово сопровождалось определением соответствующего понятия, почерпнутым, как правило, из авторитетных словарей и руководств. Словник был дополнен терминами, выделенными из наименований рубрик, входящих в данные разделы ГРНТИ, а также из связанных по смыслу рубрик других классификаций – ББК, УДК, PACS, Международной патентной классификации. Лексика других рассмотренных классификаций (Scopus, Web of Science, OECD Fields of Science, Номенклатура ВАК) вполне укладывается в пределы собранного массива терминов. Термины из классификаций также были снабжены определениями.

Объём лексики обоих тезаурусов составляет около 1200 словарных единиц (примерно 700 по физике и 500 по электронике).

Выбор терминов из используемых классификаций и назначенных документам ключевых слов обеспечивает привязку нашей системы к реальным признакам описания тематики документов, имеющих в пространстве поиска.

3. Онтологическая интерпретация тезаурусных отношений

При установлении смысловых связей терминов мы исходили из того, что тезаурусы будут составлять компо-

нент формальной онтологии, в которой они должны отражать родовидовые связи объектов, прагматические отношения которых должна отражать система ассоциативных связей разного рода, которые в свою очередь могут составлять отдельный тезаурус связей с открытым составом членов. В силу этого мы устанавливали иерархическую связь «выше — ниже» только между понятиями одной категории, которые соотносятся как род и вид, т. е. множество денотатов одного термина является подмножеством другого. Остальные случаи смысловой связи обозначались ассоциативным отношением терминов. Этим тезаурусы «Физика» и «Электроника» базы данных ТЕРМИН отличаются от обычных информационно-поисковых тезаурусов (включая ТРФП), в которых основанием для иерархической связи является соотношение не денотатов понятий, а стоящих за терминами множеств документов. То же самое справедливо и для отношения эквивалентности терминов. Термины эквивалентны тогда и только тогда, когда они обозначают тождественные множества объектов. Отношение тождества терминов означает именно тождество денотатов, а не тождество массивов документов про них. Если между двумя понятиями А и Б установлено онтологическое отношение «А тождественно Б» или «А включает Б», то это отношение будет справедливым и в поисковом смысле, но не наоборот.

Различие между двумя интерпретациями отношений можно пояснить следующим примером. В поисковом смысле часто антонимы можно рассматривать как эквиваленты; например все статьи о *неустойчивости плазмы* релевантны запросу об *устойчивости плазмы*, поскольку критерии этих явлений совпадают, и их определения равно соответствуют и тому, и другому понятию. Но в онтологическом смысле эти понятия исключают друг друга и не могут рассматриваться как эквивалентные или пересекающиеся.

Что же касается «ассоциативного» отношения пересечения $A \times B$, то в онтологическом смысле оно понимается как наличие у денотатов общих атрибутов или как смежность соответствующих реалий, что вполне сходится с его понима-

нием в поисковом смысле как пересечение массивов релевантных документов.

4. Интерпретация дефинитивных связей

Базовый массив связей терминов в системе ТЕРМИН был получен автоматически на основе «дефинитивного» поиска соответствий (употребление одного термина в составе дефиниции другого). Такой поиск не даёт возможности квалифицировать найденную связь по категориям видов тезаурусных отношений. Уточнение вида автоматически установленной связи по категориям «совпадение — вхождение — пересечение» осуществлялась интеллектуальным рассмотрением вручную. При этом существенную долю дефинитивных связей (примерно четверть) пришлось исключить как установленные из-за формального совпадения слов с семантически не связанными смыслом (например, когда термины были приписаны к понятию «*время*» из-за того, что в их определениях встретилось выражение «*в то время как*»).

Но и среди оставленных действительных связей не все целесообразно использовать при обычном документальном поиске. Так, например, понятия «рога» и «корова» явно связаны, и это может быть отражено в их определениях. Но использовать документы о рогах как релевантные для поиска документов об коровах (и наоборот) вряд ли целесообразно в общем случае. Однако, если в поисковой системе будет реализован специфические модальности поиска «особенности внешнего вида» или «средства обороны», то связь таких терминов будет востребована. Поэтому мы такие связи не ликвидировали, а оставляли как особую категорию «слабых пересечений» в качестве кандидатов на установление специфических режимов поиска, учитывающих прагматических отношения объектов онтологической реальности.

5. Задачи и решения

Интерес представляет не только содержание тезаурусов, но также и принципы решения ключевых задач, стоявших перед разработчиками.

ПЕРВОЕ. Определение состава лексики тезаурусов. Поскольку тезаурусы должны направлять навигацию и поиск по соответствию смысла документов в разных ресурсах, лексика должна включать обозначения этих смыслов — ключевые термины и рубрики тематических классификаций. Так что, как было отмечено выше, в тезаурусы включены все рубрики классификаций взаимодействующих ресурсов и/или выделенные из их наименований термины.

Введение в тезаурусы всего объёма ключевых терминов, приписанных документам в поисковых массивах, технически нереально; поэтому наша процедура состояла в том, что действующим индексаторам ВИНТИ и БЕН РАН было предложено заиндексировать наиболее существенными (ключевыми) терминами все рубрики ГРНТИ.

Эти два слоя лексики — из классификаций и из практики индексирования — и составили словарь тезаурусов. Для каждой тематической области это оказался вполне обозримый набор нескольких сот терминов, с одной стороны доступный для экспертной обработки, а с другой стороны, покрывающий смысловое поле области знания «с точностью до ГРНТИ» — общесетевого набора тематических категорий.

ВТОРОЕ. Установление семантики терминов. Одновременно с выявлением терминов им были приписаны определения, главным образом заимствованные из авторитетных источников (энциклопедий и словарей).

ТРЕТЬЕ. Поиск семантически связанных терминов. Здесь была применена полуавтоматическая процедура. Компьютерная программа во всём массиве выделила пары терминов, определения которых ссылаются друг на друга. Если определение одного термина использует другой термин, то эти термины семантически связаны. Экспертный просмотр позволил удалить те пары терминов, связь которых

была установлена по случайному употреблению слов в нетерминологическом смысле.

ЧЕТВЁРТОЕ. Определение качества (характера) смысловой связи терминов. Эта задача потребовала переосмысления обычно применяемых отношений между дескрипторами тезаурусов. Наша лексика включает дескрипторы двух категорий — ключевые термины и тематические рубрики. Что касается рубрик, то они связаны отношениями тех классификационных систем, из которых они извлечены. Эти иерархические отношения фиксируют связь рубрик по включению друг в друга множеств документов, входящих в эти рубрики в соответствующем информационном ресурсе. Это вполне соответствует связям дескрипторов в информационно-поисковых тезаурусах, где дескриптор рассматривают как множество индексированных им документов. Но за большинством отдельных терминов никакие документы не стоят, они обозначают не документы, а понятия об объектах действительности. Соответственно этой природе между терминами могут быть установлены отношения включения и пересечения объёмов (экстенсионалов) и содержания (интенсионалов) понятий. Строгое указание именно этих онтологических отношений позволяет рассматривать наши тезаурусы как онтологии и использовать их для навигации по связям вещей, а не документов о вещах.

Согласно этому представлению путём экспертного анализа алгоритмически установленных связей терминов среди них выделены строго родовидовые связи (по включению интенсионалов и экстенсионалов) и ассоциативные связи (частичное пересечение экстенсионалов и интенсионалов).

Таким образом, в словарях «Физика» и «Электроника» базы данных ТЕРМИН были устанавливаются следующие виды связей терминов, показанные на *табл. 2*.

Таблица 2

Смысловые связи дескрипторов тезауруса

<i>Знак</i>	<i>Названия</i>	<i>Значение</i>
$A = B$	понятия (термины, рубрики) совпадают	Термины А и Б обозначают тождественные множества реалий (синонимы)
$A \gg B$	понятие А шире, включает Б	Термин А обозначает множество реалий, в которое включено множество реалий, обозначаемых термином Б, при чём объёмы этих множеств соизмеримы.
$A \ll B$	понятие А уже, входит в Б	Термин А обозначает множество реалий, включённое во множество реалий, обозначаемых термином Б, при чём объёмы этих множеств соизмеримы.
$A > < B$	понятия пересекаются	Множества реалий, обозначаемые терминами А и Б пересекаются в существенной части.
$A > B$	рубрика А шире, включает Б	Рубрика, соответствующая понятию А, включает в себя информацию о понятии Б
$A < B$	рубрика А уже, входит в Б	Рубрика, соответствующая понятию А, содержит часть информации о понятии Б
$A - B$	термины дефинитивно связаны	Реалии, обозначаемые терминами А и Б, связаны прагматическими связями, но их множества, вероятно, не пересекаются, относясь к различным онтологическим категориям.

Пятая задача. **Установление парадигматических связей.** Она относится к числу перспективных. В тезаурусы следует ввести кроме вышеописанных логических отношений прагматические связи объектов действительности, выражаемых дескрипторами тезаурусов. Это такие связи как:

- часть — целое;
- причина — следствие;
- событие — место;
- событие — время;

- множество — экземпляр;
- свойство — носитель свойства;
- ... и многие другие, более частные.

В идеале предполагается разработка самостоятельного тезауруса отношений объектов действительности, который в совокупности с тезаурусами дескрипторов объектов и тезаурусом тематических рубрик будет представлять полноценную онтологию информационного пространства НТИ, пригодную для реализации задач информационного поиска, смысловой навигации и других функций искусственного интеллекта. Но это предстоит осуществить по мере дальнейшего развития работ.

Работа выполняется при поддержке РФФИ (проект 20-07-00103)

Литература

1. Калёнов Н. Е., Сотников А. Н. Архитектура Единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 5–8.
2. Metathesaurus / U. S. National Library of Medicine : [Электронный ресурс]. URL: https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html
3. Белоозеров В. Н., Шабурова Н. Н. Классификационные системы как средство поиска информации по физике полупроводников // Библиосфера. 2008. № 3. С. 34–42.
4. Белоозеров В. Н., Шабурова Н. Н. Сопоставительный тезаурус классификационных систем по физике полупроводников // Информационное обеспечение науки: новые технологии : Сборник научных трудов / Н. Е. Калёнов (ред). М.: Научный Мир, 2009. С. 311–322.
5. Белоозеров В. Н., Шабурова Н. Н. Тезаурус тематических рубрик по физике полупроводников // Депонировано в ВИНТИ 2013-12-24, № 379-В2013.
6. Белоозеров В. Н., Шабурова Н. Н. Тезаурус библиографических классификаций как модель интеграции информационных ресурсов // Международная конф. [27–28 окт. 2011, Москва] «Информационное общество: Состояние и тенденции межгосуд. обмена научно-технической информацией в СНГ». М: ВИНТИ, 2011. С. 8–9.
7. Антопольский А. Б. [и др.]. Разработка онтологии информационного пространства знаний на основе дефинитивных связей // Научно-техническая информация. Серия 1. Организация и методика информационной работы. 2017. № 11. С. 19–24.

8. Antopol'skii A. B. [et al.]. The development of a semantic network of keywords based on definitive relationships // Scientific and Technical Information Processing. 2017. 44(4). P. 261–265.

9. Белоозеров В. Н., Шабурова Н. Н. О разработке классификационно-тезаурусной онтологии для предметной области физики и радиоэлектроники // Информационное обеспечение науки: новые технологии : Сб. науч. трудов / Ред. П. П. Трескова. Екатеринбург, 2018. С. 75–87.

References

1. Kalenov N. E., Sotnikov A. N. Architecture of the common digital space of scientific knowledge // Informacionnie resursi Rossii. 2020. № 5. P. 5–8.

2. Metathesaurus / U. S. National Library of Medicine : [Электронный ресурс]. URL: https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html

3. Beloozerov V. N., Shaburova N. N. Classification systems as means of searching for information on semiconductor physics // Bibli-osfera. 2008. № 3. P. 34–42.

4. Beloozerov V. N., Shaburova N. N. Comparative thesaurus of classification systems in semiconductor physics // Informacionnoe obespechenie nauki: novwe texnologii : Sbornik nauchnyx trudov / Kalenov N. E. (ed.). Moscow: Nauchnyj mir, 2009. P. 311–322.

5. Beloozerov V. N., Shaburova N. N. The thesaurus of subject headings in physics of semiconductors // Deposited in VINITI 2013–12–24, № 379-B2013.

6. Beloozerov V. N., Shaburova N. N. Thesaurus of bibliographic classifications as a model for integrating information resources // Mezhdunarodnaya konferenciya : Informacionnoe obshhestvo : Sostoyanie i tendencii mezghosudarstvennogo obmena nauchno-texnicheskoy informacij v SNG. Moscow: VINITI, 2011. P. 8–9.

7. Antopol'skij A. B. [et al.]. Development of an ontology of the information space of knowledge based on definitive links // Nauchno-texnicheskaya informaciya. Seriya 1. Organizaciya i metodika informacionnoj rabotw, 2017. № 11. P. 19–24.

8. Antopol'skii A. B. [et al.]. The development of a semantic network of keywords based on definitive relationships // Scientific and Technical Information Processing. 2017. 44(4). P. 261–265.

9. Beloozerov V. N., Shaburova N. N. On the development of the classification-thesaurus ontology for the subject area of physics and radio electronics // Informacionnoe obespechenie nauki: novwe texnologii : Sbornik nauchnyx trudov / Ed. P. P. Treskova. Ekaterinburg, 2018. P. 75–87.

Российские публикации по информатике в базе данных «Web of Science»

А. В. Глушановский

БЕН РАН

Аннотация. Рассматривается современное состояние и эволюция за десять последних лет потока российских публикаций, отраженных в БД Web of Science (WOS), по тематической области «Информатика» (соответствующая тематическая область WOS – Computer Science). Рассмотрены возможности библиометрических методов анализа массива научных публикаций и причины их применения при анализе российской науки.

Показано, что данный поток рос с 2010 г. по 2019 г. темпом, опережающим среднемировой. В результате Россия переместилась за этот период с 34-ого на 15-ое место среди стран мира, публикующих подобную информацию, что, однако, не вполне соответствует требованиям национального проекта «Наука», требующего вхождения в мировую пятерку стран по основным научным направлениям.

Проанализирован состав массива российских публикаций 2019 г. по типам входящих в него документов. Установлено, что основная масса документов относится к типам: «публикация в журнале» и «публикация в сборнике трудов конференции». Определено количество журналов, публикующих материалы российских ученых, их распределение по квартилям WOS и количество организаций, представивших эти материалы. Аналогично определено количество конференций, материалы которых представлены в массиве, и количество представивших их организаций. Отдельно определено количество организаций РАН и число представленных ими документов обоих типов. Отмечено, что в массиве журнальных публикаций доля публикаций РАН существенно выше, чем в части публикаций трудов конференций.

Установлено, что в рассматриваемом массиве число документов типа «труды конференций», примерно, в три раза превышает число документов типа «публикации в журнале». Отмечено, что это является нетипичным для массивов публикаций большинства других стран мира. Показано, что уровень цитирования трудов конференций существенно ниже, чем статей в журналах, что может говорить о более низком интересе к ним со стороны мирового научного сообщества. Показано, что «коэффициент качества» публикационного массива журнальных статей (рассчитанный согласно методике, утвержденной Министерством высшего образования и науки РФ в 2020 г.) существенно выше, чем

тот же коэффициент для массива публикаций трудов конференций, несмотря на существенно большее число документов в последнем, что говорит о некотором снижении общего качества массива (в терминах упомянутой выше методики) в связи с «перекосом» в область документов типа «труды конференций», что требует более внимательного рассмотрения в дальнейшем.

Ключевые слова: информатика, Web of Science, библиометрический анализ, мировой поток публикаций, статьи из журналов, труды конференций, качество массива публикаций, цитирование.

Russian Informatics Publications in «Web of Science» Data Base

A. V. Glushanovskiy

LNS RAS

Abstract. This article describes a Current State and Evolution of Russian's Publication for Computer Science Flow in Web of Science (WOS) Data Base. The Reasons for Bibliometric Methods Using for Russian Scientific Publications Analyze and it's Opportunities are shown.

The presented Research shows the Increasing Volume of Russian Computer Science Publications for 2010 to 2019, which outstrips the Increasing of the World Publications Flow. This allows Russia to rise from 34 to 15 Position in World Countries Classification in this Scientific Publication Area. But this Position is not enough for the National Project «Science» Goal Achievement — minimum fifth Position in this Classification.

The List of Document's Types of analyzed Document's Set was determined. It concludes two main Types: Journals Articles and Conferences Proceeding Papers and additionally a little quantity of Documents some another Types. For two main Types the Lists of Journals, Conferences and Organizations (Authors from which) were established. All Journals have been dissipated to WOS Journals Quartiles.

The Citation of published Documents was calculated separately for every from tow main Document Types. In was shown, that the Citing of the Journals Articles is more in Times that Conference Proceedings Papers. That Fact shows the very less Interest to Proceeding Conference Papers then to Journal Articles from the Scientific Society.

For two main Document Types Sets have been calculated «Quality Coefficient» according the Russian Science Ministry Methodic. It shows that this Coefficient for Journals Articles is much more than for Conference Proceedings despite of less Quantity of the Journal's Article Set. This Fact shows, that the special Attention must be focused to this Problem in Future.

Keywords: informatics, Web of Science, world publication's flow, publication's set, journal's article, conference proceedings, bibliometric analyzes.

1. Введение

Применение библиометрических методов для оценки научной деятельности с самого начала и до настоящего времени является предметом дискуссий как, собственно в научной среде, так и в среде управления наукой. Однако в последнее время эти методы находят все более широкое применение в работе научных фондов и органов управления и финансирования науки различных стран мира. Опыт использования их в некоторых странах рассматриваются, например, в публикации директора ГПНТБ СО РАН А. Е. Гуськова с соавторами [1]; а также в работе социологов М. А. Юревича и Д. С. Еркиной [2].

С 2009 г. подобные методы начали применяться и органами управления наукой нашей страны для оценки научной деятельности в России. Эволюция их внедрения в российской науке достаточно подробно прослежена в упомянутой выше статье [1], а также в публикации сотрудников БЕН РАН В. А. Цветковой и Ю. В. Мохначевой [3]. По общему мнению самих специалистов в области библиометрии [1, 3, 4] библиометрические методы, *при правильном их использовании грамотными специалистами*, в определенной степени, отображают объективные реально существующие тенденции развития отечественной и мировой науки. В целом, большинство специалистов приходит к выводу, что, при всей дискусионности, иного метода объективной оценки научных публикаций, независимой от пристрастий экспертов и редакций журналов, пока не найдено.

Так или иначе, в принятом на государственном уровне национальном проекте «Наука», определяющем в значительной мере деятельность научных учреждений РФ на данный момент, выдвинуто требование о вхождении Российской Федерации к 2024 г. в пятерку ведущих стран мира по числу научных статей в приоритетных для нее об-

ластях науки [5]. Это вхождение оценивается, в основном, по вхождению российских публикаций в источники, включаемые в БД Web of Science (WOS) и Scopus. При этом, в отчетности научных организаций (согласно последней версии методики расчета «Комплексного балла публикационной результативности», утвержденной заместителем Министра науки и высшего образования РФ С. В. Люлиным 25.08.2020 г.) с наивысшими показателями учитываются публикации, отраженные в БД WOS, затем (с существенно более низкими коэффициентами) — публикации, отраженные в БД Scopus, и затем — все остальные.

В отчетных показателях в весьма малой степени учитываются публикации в значительной части российских научных журналов (за исключением входящих в БД WOS и Scopus, а также в перечень журналов ВАК). Из учета публикаций, в частности, исключена БД «Российский индекс научного цитирования» (РИНЦ), которая, при всех ее недостатках, фиксирует значительную часть русскоязычных научных публикаций. Таким образом, установленный Минобрнауки порядок, фактически, оставляет «за бортом» основную массу публикаций в российских научных журналах, не включенных в упомянутые БД.

В связи с данным подходом в настоящей статье рассматривается ситуация с российскими публикациями по информатике (как безусловно приоритетной для РФ области науки), отражаемыми в БД WOS — наиболее «ценной» согласно упомянутой методике Минобрнауки.

Так как БД WOS является англоязычной, следует вначале уточнить терминологию, Термин «Информатика» в русском языке имеет несколько значений. На сайте «Википедия» [6] указывается:

Считается, что под терминами «informatics» в европейских странах и «информатика» в русском языке понимается направление, именуемое в английском языке «computer science». К другому направлению, посвящённому изучению структуры и общих свойств объективной (научной) информации, иногда называемому

документалистикой (документальной информатикой) или автоматическим анализом документов, близок термин «*information science*».

В рамках данной статьи мы будем далее рассматривать то направление информатики, которое в английском языке и, соответственно, в БД WOS определяется термином «*computer science*».

2. Российские публикации по тематике «*computer science*» в БД WOS

Тематически БД WOS подразделяется на более, чем 200 тематических разделов, именуемых «*Subject Categories*» – предметные категории. Следует заметить, что в этой БД предметная категория присваивается журналу, в целом, а не каждой конкретной статье. Один и тот же журнал, в силу тематической направленности, может принадлежать к двум и более предметным категориям.

Согласно русскоязычной версии справочника этой БД (https://images.webofknowledge.com/WOKRS535R100/help/r_u_RU/WOS/hp_subject_category_terms_tasca.html), к тематике «*computer science*» относятся семь предметных категорий:

Таблица 1

Предметные категории Computer Science

Subject Category
Computer Science Interdisciplinary Applications
Computer Science Theory Methods
Computer Science Information Systems
Computer Science Artificial Intelligence
Computer Science Software Engineering
Computer Science Cybernetics
Computer Science Hardware Architecture

Как и почти по всем предметным категориям WOS, суммарный мировой поток публикаций по этим категориям за десять последних лет существенно увеличился. Эти увеличения отражено в *табл. 2*.

Таблица 2

Мировой поток публикаций — Computer Science

Год	Кол-во публикаций	Прирост к предыдущему (%)
2010	118561	
2013	132458	12 %
2015	186153	41 %
2017	193928	4 %
2019	190357	-2 %

Из таблицы 2 следует, что наиболее значительный прирост публикаций по данной тематике пришелся на период 2013–2015 гг., после чего количество публикаций, примерно, стабилизировалось.

За пять последних лет (2015–2019 гг.) вклад в этот поток внесли 198 стран мира. В 2019 г. 177 стран.

Далее рассмотрим вклад российских ученых в эту часть общемирового потока публикаций. В табл. 3 представлен поток российских публикаций по тематике Computer Science за 2010–2019 годы.

Следует пояснить, что у каждой публикации может быть несколько соавторов (в том числе – из разных стран), а с другой стороны, один и тот же автор может указать несколько аффилиаций. В связи с этим под российской публикацией в дальнейшем понимается любая публикация, в которой хотя бы одна аффилиация одного автора является российской.

Таблица 3

Российский поток публикаций — Computer Science

Год	Кол-во публикаций	Прирост к предыдущему (%)
2010	796	
2013	928	17 %
2015	2213	138 %
2017	3871	75 %

2019	3521	-9 %
------	------	------

Согласно *табл. 3*, максимальный прирост российских публикаций также пришелся (следуя мировым тенденциям) на 2013–2015 гг., но темп этого прироста был значительно выше среднемирового. Если за период 2010–2019 гг. мировой поток возрос в 1,6 раза, то российский за тот же период – в 4,4 раза, и этот рост продолжался и в период 2015–2017 гг. В то же время и некоторое снижение объема публикаций в период 2017–2019 гг. в России оказалось более ощутимым, чем в мире, в целом.

В соответствии с этим темпом роста, опережающим среднемировой, доля России возрастала, и Россия перемещалась на все более высокое место среди стран, участвовавших в этом потоке.

Таблица 4

**Изменение позиции и доли участия России
в мировом потоке публикаций.**

Год	Процент публикаций РФ в мировом потоке	Место в списке стран
2010	0,70 %	34
2013	0,70 %	33
2015	1,20 %	23
2016	1,60 %	17
2017	2,00 %	14
2018	1,83 %	15
2019	1,82 %	15

В результате, за десять лет Россия поднялась в списке стран на 19 мест и почти втрое увеличила свою долю в общем потоке, хотя эта доля (как и место в списке) к 2017 г. стабилизировалась и по-прежнему остается достаточно малой. Сам список первой десятки стран в течение десяти лет остается более или менее стабильным (за исключением появления

в 2013 г. среди лидеров и сразу на достаточно высокой позиции публикационного потока Индии).

Таблица 5

Первая десятка стран по объему потока публикаций.

Место	2010	2013	2015	2017	2019
1	КНР	КНР	КНР	КНР	КНР
2	США	США	США	США	США
3	ФРГ	Индия	Индия	Индия	Индия
4	Великобритания	ФРГ	ФРГ	Великобритания	Великобритания
5	Франция	Великобритания	Великобритания	ФРГ	ФРГ
6	Япония	Франция	Франция	Япония	Япония
7	Испания	Япония	Япония	Франция	Франция
8	Канада	Испания	Италия	Италия	Канада
9	Италия	Канада	Канада	Канада	Италия
10	Тайвань	Италия	Испания	Испания	Респ. Корея

Здесь и далее под потоком публикаций Великобритании понимается суммарный поток Англии, Шотландии, Уэльса и Северной Ирландии, учитываемых в БД WOS по отдельности.

Далее сосредоточимся на более подробном анализе публикаций 2019-го года – последнего полного года, учтенного на данный момент в БД WOS. Объем публикаций и доли первых пятнадцати стран в потоке 2019 г. представлены в табл. 6.

Рассмотрим более подробно состав российских публикаций этого года по типам входящих документов. Подавляющее большинство рассматриваемого массива составляют два типа документов – статьи в журналах и публикации в трудах конференций. При этом небольшая часть (27 публикаций) из трудов конференций опубликована в виде статей в журналах, а подавляющее большинство – в отдельных сборниках трудов конференций. Дополнительно к этим двум подмассивам в WOS присутствуют также

(в весьма незначительном объеме) отдельные главы из книг и обзоры. В БД WOS присутствуют также в небольшом количестве редакционные материалы и поправки к статьям, но эти публикации далее не учитывались (что и объясняет некоторое различие в данных России в табл. 3 и 6), так как они, как правило, не содержат новых научных результатов.

Таблица 6

Доля первых 15-ти стран в мировом потоке. 2019 г.

Страна	Колич-во публикаций	Процент в мировом потоке
Peoples Resp. China	56 046	29,932
Usa	36 237	19,353
India	12 839	6,857
United Kingdom	11 141	4,960
Germany	9245	4,937
Japan	7239	3,866
France	6904	3,687
Canada	6537	3,491
Italy	6084	3,249
South Korea	6038	3,225
Australia	5665	3,025
Spain	5583	2,982
Brazil	4099	2,189
Taiwan	3477	1,857
Russia	3456	1,822

Для большинства других стран также характерно подавляющее преобладание этих же двух типов документов (статьи в журналах и сборниках трудов конференций), но процентное соотношение этих типов в российском массиве заметно отличается от массивов других стран.

Таблица 7

Процентное распределение основных типов документов

Страна	Процент статей	Процент трудов
--------	----------------	----------------

	из журналов	конференций
КНР	63,2	36,2
США	40	59,1
Индия	53,3	45
Бразилия	46,4	51,7
Тайвань	53,7	45,7
Россия	23,8	73,2

Для сравнения взяты первые три страны общего списка и две страны (Бразилия и Тайвань) соседствующие в этом списке с Россией.

Нетрудно видеть, что в российском массиве существует значительное преобладание публикаций из сборников конференций (73,2 %), в отличие от большинства стран, где эти типы документов наличествуют в близких по соотношению долях. Некоторым исключением является массив публикаций США, но и там этот «перекосяк» не является столь заметным.

Далее рассмотрим более подробно документы двух упомянутых основных типов.

В массиве российских публикаций 2019 г. присутствуют 823 статьи из 209-ти журналов. В БД WOS одним из важных библиометрических параметров журнала является квартиль — принадлежность журнала к одной из четырех групп журналов определенной тематической категории, ранжированных по убыванию уровня цитируемости его публикаций (импакт-фактор). В понятиях БД WOS квартиль является показателем научного уровня журнала. С этой позиции 209 журналов, в которых опубликованы статьи российских авторов, распределяются по квартилям (Q1 — самый высокий и далее до Q4) следующим образом.

Таблица 8

Распределение журналов по квартилям

Квартиль	Журналов	Процент	Статей	Процент
----------	----------	---------	--------	---------

Q1	59	28,20 %	240	29,2 %
Q2	49	23,40 %	105	12,8 %
Q3	26	12,40 %	66	8,0 %
Q4	29	13,90 %	214	26,0 %
Без кв-ля	46	22,00 %	198	24,0 %
Всего	209	100 %	823	100,0 %

Следует отметить, что распределение статей по журналам является весьма неравномерным. Первые двенадцать журналов (среди них семь журналов высшего (первого) квартиля), содержащие десять и более российских публикаций каждый, дают в сумме 450 статей (более 50 % всех публикаций). Все 823 статьи представляют (с учетом организаций всех российских и зарубежных соавторов) 1079 организаций. Среди них 17-ю организациями РАН представлено 234 статьи (более четверти общего количества статей).

В рассматриваемом массиве представлена также 2531 публикация из трудов 338-ми научных конференций. Эти публикации (с учетом организаций всех соавторов) представлены участниками из 1308-ми организаций. Среди них – 181 публикация, представленная 45-ю организациями РАН. Таким образом, в массиве журнальных публикаций доля публикаций РАН существенно выше, чем в части публикаций трудов конференций.

Определенный перекос в сторону преобладания трудов конференций над журнальными статьями наблюдается не только применительно к российским публикациям в области Computer Science. В одной из предыдущих работ [7] автором отмечалась похожая ситуация в области российских публикаций по разделу «Физика». Этот момент отмечался и в работах других специалистов. Так, в работе [8] группа авторов из ГПНТБ СО РАН указывает на эту особенность, как на одну из стратегий для увеличения представительности страны и ее организаций в мировых базах данных, но отмечает, что *«...такая стратегия положительно влияет на качество конференций и их трудов, на развитие международно-*

го сотрудничества, но не может быть достоверным индикатором повышения уровня научных исследований». Авторы публикации [9] в журнале *Scientometrics* также отмечают, что вопрос о том, является ли это позитивной тенденцией, остается дискуссионным.

При этом авторы работы [9] обращают внимание на низкий уровень цитирования таких документов (особенно вне России). Разный уровень цитирования журнальных публикаций и публикаций трудов конференций (что, возможно, отражает меньший интерес к ним со стороны научного сообщества) может быть оценен на базе рассматриваемого нами массива российских публикаций в области *Computer Science*. Автором было определено (средствами БД *WOS*) цитирование упомянутых выше 823-х журнальных публикаций и 2531 публикации трудов конференций. Результаты представлены в табл. 9.

Таблица 9

Цитирование (по типам документов)

Массив документов	Число док-тов	Число цитирований	Число цитирований (без самоцитирования)	Среднее цитирование.
	2647	623	580	0,24
Цитирование трудов конференций РАН	187	58	54	0,31
Цитирование ж-лов	821	1812	1753	2,21
Цитирование ж-лов РАН	234	364	337	1,48

Из рассмотрения таблицы 9 следует, что и в нашем случае среднее цитирование статей журналов почти в восемь раз выше, чем материалов конференций (для публикаций РАН этот показатель близок к пяти), что подтверждает высказанное ранее предположение.

И наконец, заметим, что упомянутая выше методика расчета «Комплексного показателя публикационной резуль-

тативности» [10], утвержденная Минобрнауки, имеет в своей основе «коэффициент качества (“научного уровня”) статьи/журнала», опирающийся (для документов из БД WOS) на ту же систему квартилей. При определении качества публикации по этой методике каждому документу присваивается поправочный коэффициент по следующему правилу:

Q1	20
Q2	10
Q3	5
Q4	2,5
Без кв-ля	1

Согласно данным *таб. 8*, 823 журнальные публикации дадут суммарный показатель качества:

$$240 \times 20 + 105 \times 10 + 66 \times 5 + 214 \times 2,5 + 198 = 13\,093$$

В то же время, 2531 публикация трудов конференций (публикации, не имеющие квартиля) дадут суммарный показатель качества всего 2351 (примерно, в пять раз меньше). То есть и с этой точки зрения, принятой Минобрнауки, наличие относительно большего массива трудов конференций и потоке публикаций не слишком положительно сказывается на качестве массива.

3. Выводы

Российские публикации по тематике Computer Science в БД WOS, увеличиваясь в количественном отношении в течение десяти лет с темпом, опережающим среднемировой, переместились за этот период вверх на 19 мест в мировом потоке, но пока занимают в нем лишь 15-ое место. Это пока достаточно далеко от требуемого национальным проектом «Наука» входа в первую мировую пятерку.

Поток российских публикаций по типам документов состоит, в основном, из статей в журналах и материалов конференций, причем материалы конференций по объему

в 2019 г., приблизительно, втрое превышают количество статей из журналов.

Статьи, публикуемые в журналах, в основном, размещаются в журналах достаточно высокого (по критериям WOS) научного уровня.

В то же время рост потока материалов конференций не всегда способствует общему повышению научного качества массива российских публикаций, в целом.

Наращивание массива публикаций за счет материалов конференций также не вполне отвечает росту качества массива, рассчитываемого согласно критериям, утвержденным Министерством науки и высшего образования РФ.

В связи с последними двумя пунктами научным организациям при формировании потока своих публикаций, возможно, следует обратить дополнительное внимание на его видовой состав.

Литература

1. Гуськов А. Е., Косяков Д. В., Селиванова И. В. Методика оценки результативности научных организаций // Вестник Российской академии наук. 2018. № 5. С. 430–443.

2. Юревич М. А., Еркина Д. С. «Публикационное ралли»: прямая угроза или новые возможности для научного сообщества? // Социология науки и технологий. 2017. Т. 8, № 2. С. 104–117.

3. Цветкова В. А., Мохначева Ю. В. Научная среда и публикационная активность: риски библиометрических оценок // Культура: теория и практика: электрон. науч. журн. 2020. № 2. С. 11–18.

4. Лазарев В. С. Власть библиометрических иллюзий над ленивыми, профанация плодотворных идей и проклятье «парабиблиометрической» оценки науки // Научный редактор и издатель. 2019. Т. 4, № 1/2. С. 12–20.

5. Паспорт национального проекта «Наука». Москва, 2018. 34 с. <http://static.government.ru/media/files/vCAoi8zEXRVsuy2Yk7D8hvQbpbUSwO8y.pdf> (дата обращения 09.11.2020).

6. Википедия. <https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0> (дата обращения 09.11.2020).

7. Глушановский А. В. Библиометрический анализ качества массива российских публикаций в области физики из БД Web of Science Core Collection // Библиосфера, 2020. № 2. С. 49–60. DOI: 10.20913/1815-3186-2020-2-49-60.

8. Гуськов А. Е., Косяков Д. В., Селиванова И. В. Стратегии повышения убликационной активности университетов – участников Проекта 5-100 // Научные и технические библиотеки. 2017. № 12. С. 5–18.

9. Moed HF, Markusova V and Akoev M (2018) Trends in Russian research output indexed in Scopus and Web of Science. *Scientometrics* 116: 1153–1180.

10. Методика расчета качественного показателя государственного задания «Комплексный балл публикационной результативности» для научных организаций, подведомственных Министерству науки и высшего образования Российской Федерации, на 2020 год/ – <https://m.minobrнауки.gov.ru/upload/iblock/6ce/6ce0e88ac94831159db8927cece01f9a.pdf> (дата обращения 28.11.2020).

References

11. Guskov A. E., Kosyakov D. V. and Selivanova I. V. Methodology to assess the effectiveness of scientific organizations // *Vestnik Rossiiskoi akademii nauk* / 2018. № 5: 430–443. (In Russ.).

12. Yurevich M. A. and Erkina D. S. “Publication rally” a direct threat or new opportunities for the scientific community // *Sociologiya nauki i tekhnologii* / 2018. Т. 8 № 2: 104–117. (In Russ.).

13. Tsvetkova V. A. and Mokhnacheva Yu. V. The scientific environment and publication activity: risks of bibliometric estimates. // *Kul'tura: teoriya i praktika elektron. nauch. zhurn.* 2020. № 2: 11–18. (In Russ.).

14. Lazarev V. S. The power of bibliometric illusions over the lazy, of the science “parabibliometric” assessment. *Nauchnyi profanation of fruitful ideas and the curse redactor i izdatel' 4 (1/2): 12–20.* (In Russ.).

15. Passport of the national project “Science” [Pasport natsional'nogo proekta “Nauka”]. Moscow. <http://static.government.ru/media/files/vCAoi8zEXRVsuy2Yk7D8hvQbpbUSwO8y.pdf>.

16. Wikipedia – <https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0>.

17. Bibliometric analysis of Russian publications' quality in physical area, included to the Web of Science Core Collection Data Base. // *Bibliosphere.* 2020. № 2. 49–59.

18. Guskov AE, Kosyakov D. V., Selivanova I. V. Strategies to increase the publication activity of universities participating the Project 5-100 // *Nauchnye i tekhnicheskie biblioteki* 12: 5–18.

19. Moed HF, Markusova V and Akoev M (2018) Trends in Russian research output indexed in Scopus and Web of Science. *Scientometrics* 116: 1153–1180.

20. О корректировке государственного задания с учетом методики расчета комплексного балла публикационной результативности. Письмо Миннауки от 14.01.2020 г. <http://docs.cntd.ru/document/564894817> (09.11.2020).

Классификационные индексы документов электронных каталогов естественнонаучных и технических библиотек: аспектный анализ

А. Н. Сысоев, В. Н. Белоозеров

ВИНИТИ РАН

Аннотация. Электронные каталоги библиотек, применяющие универсальные иерархические классификации и осуществляющие углублённую систематизацию и многоаспектное индексирование входящих изданий обладают базой для проведения аспектного анализа документального фонда. Его использование расширяет возможности библиометрических методов арсенала библиотек и позволяет проводить аналитические исследования профильного библиотечного фонда, в том числе оценивать объёмы и интенсивность информационных потоков, включая междисциплинарность отдельных направлений научных исследований. Применение отдельных методов аспектного анализа рассмотрено на примере корпуса данных БЕИ РАН.

Ключевые слова: междисциплинарность, многоаспектное индексирование, иерархические классификации, аспектный анализ.

Classification indexes of documents of electronic catalogues of natural and technical libraries: aspect analysis

Sysoev A. N., Beloozerov V. N.

VINITI RAS

Annotation. Electronic catalogues of libraries that use universal hierarchical classifications and perform in-depth systematization and multi-aspect indexing of incoming publications have a basis for conducting an aspect analysis of the documentary fund. Its use expands the possibilities of bibliometric methods from the library arsenal and allows conducting analytical research of the specialized library fund, including evaluating the vol-

ume and intensity of information flows, including the interdisciplinarity of individual research areas. The application of certain methods of aspect analysis is considered on the example of the data corpus of the BEN RAS.

Key words: interdisciplinarity, multi-dimensional indexing, hierarchical classification, aspect-analysis.

Состояние проблемы

Уход библиотечного сообщества от вопросов совершенствования используемых международных классификаций знания к вопросам адаптации упрощённых рубрикаторов к проблеме однозначного установления научных связей поднимает «вечные» вопросы: о дисциплинарности и соподчинённости наук, свойственные более задачам создания классификаций чем их совершенствования [1, 2].

Вместе с тем, базируясь на данных каталогов Elsevier SciVal [3], Elsevier Scopus [4], Thomson Reuters WoS [5] или ELibrary РИНЦ [6], исследователи опираются, главным образом, на мнение издателей и на компетенцию составителей каталогов, которые иногда слишком вольно позиционируют издания в той или иной области знания. Причём важнейшей методологической особенностью как Scopus, так и Clarivate Analytics (до 2016 года Thomson Reuters) служит известный факт, что тематику отдельных публикаций эти базы не определяют, а присваивают ее в соответствии с тематикой журнала, в котором располагается публикация. Причиной этому использование упрощённых классификаций и очевидная сложность предмета, которая доступна только либо профильному специалисту, либо библиотекарю-предметнику который каждодневно дирижирует потоком входящих научных и технических изданий. Но ни издатели, ни ритейлеры на рынке информации зачастую не в состоянии достичь необходимой детальности и точности в описании естественно-научной литературы, размещаемой в создаваемых ими электронных каталогах и вынуждены идти на известные упрощения. Тем более, что рубрикаторы, используемые для этой цели, не превышают объёма в несколько сотен рубрик, и не идут в сравнение с той же УДК, количество наименований в которой достигает поч-

ти 8000 тысяч рубрик. Не озабоченный вопросами освоения международных классификаций, для непонятной ему цели межбиблиотечной коммуникации, информационный ритейл оставляет эти проблемы, предпочитая достигать коммуникации с потребителем информации на базе упрощённых классификаций науки, минуя библиотеки. Упрощённый подход к знанию не может не приводить к курьёзам при тематическом поиске, а оценки междисциплинарности попросту выглядят невозможными.

Лишённые этих недостатков библиотечные электронные каталоги, разрабатываемые квалифицированными систематизаторами-предметниками, сегодня остаются без внимания. Их уверенно оттесняют с рынка информации, поскольку к конкурентной борьбе с маститыми ритейлерами они явно не готовы как с финансовой, так и с организационной стороны. Хотя, руководствуясь методиками индексирования и располагая опытом систематизации научных изданий в знакомой ему области, библиотекарь-предметник нередко вынужден оспаривать тематическую локализацию отдельных изданий, которые известными электронными каталогами, отнесены не в свою научную нишу.

Если для издательств вопросы глубины применяемых классификаций значения не имеют, то для информационного ритейла, создающего обширные электронные каталоги, проблема крайне актуальна и имеет очевидное решение: использовать при тематической рубрикации документов известные международные универсальные классификации знаний, а при отсутствии собственного опыта, привлекать опыт ведущих естественнонаучных и технических библиотек.

Аспектный анализ: постановка задачи, методология и приёмы

Аспектный анализ следует относить к арсеналу методов наукометрии. Он направлен на выявление динамики связей наук и установление количественной меры или силы междисциплинарных связей.

Философия аспектного анализа построена на предпосылке, что пространство Универсума знаний является непрерывным, и никакая наука в нём не может существовать в изолированном состоянии: в каждой его точке можно наблюдать взаимосвязанность с другими областями знания. В контексте этих представлений феномен междисциплинарности является фундаментальным свойством знания. Междисциплинарность присутствует всегда и служит мерой интереса конкретной науки к окружающему её миру знаний и, наоборот, окружающего мира к этой науке, и зависит только от состояния науки в данный момент времени и качества исходных данных. Универсальные иерархические классификации знания УДК, ДКД, ББК и другие являются «картами знаний», благодаря которым становится возможным охватить пространство Универсума знаний в его доступном состоянии [7].

Для естественнонаучных библиотек свойственна углублённая систематизация и многоаспектное индексирование документов составными индексами (рис. 1), которые, включают коды двух или более классов, представляющих разные аспекты темы. Электронные каталоги библиотек, в которых осуществляется углублённая систематизация и многоаспектное индексирование входящих изданий служат информационной базой, предоставляющей данные для аспектного анализа. Процесс создания составного индекса и применения аспектных ссылок подробно изложен в литературе по библиотечному делу и регламентирован соответствующими стандартами [8].

Анализируя издание и определяя его тему, систематизатор, в первую очередь, решает вопрос о степени сложности документа и принимает решение, какой индекс следует применить – простой, сложный или составной. Далеко не всякий документ получает составной индекс. Так, в открытом электронном каталоге БЕН РАН, применяющей в систематизации УДК, по состоянию на конец 2019 г. использовано 374957 индексов систематизации, из которых 296568 не используют аспекты, а 78389 индексов составлены с использованием

приёмов многоаспектного индексирования. Примеры библиографических записей с многоаспектными индексами УДК:

102154; **619:616.98:578.825.1**; Герпесвирусные болезни животных. / Закутский Н. И., Хухоров И. Ю., Жестерев В. И. и др.; Всеросс. НИИ ветеринарии, вирусологии и микробиологии. Владимир; Покров : Фолиант, 2003. 282 с. Библиография : с. 255–275. ISBN 5–94210–021–7.

103063; **614.876:621.039.74:004.9:519.816**; Трахтенгерц Э. А. Компьютерная поддержка управления ликвидацией последствий радиационного воздействия. / Трахтенгерц Э. А., Шершаков В. М., Камаев Д. А.; Рос. АН. Ин-т пробл. упр. им. В. А. Трапезникова и др. М. : СИНТЕГ, 2004. 456 с. : ил. (Системы и проблемы управления). Библиография : с. 440–455. ISBN 5–89638–071–2.

109696; **611.013:573.2:681.3.06:510.5**; Cytowski Jerzy, Gielecki Jerzy. An application of genetic algorithm to edge detection of blood vessels. Warszawa : Inst. of computer science Pol. acad. of sciences, 1995. 24 с. : ил. (Pr. IPI PAN, ISSN 0138–0648; 794). Парал. загл. сер. англ. Рез. англ., пол. Библиография : с. 23–24. Print. as a ms.; Применение генетического алгоритма для обнаружения границ кровеносных сосудов.

1405; **615:001.4:811.124(07)**; Шиканова Татьяна Аркадьевна. Латинский язык и основы фармацевтической терминологии : Учебное пособие для мед. вузов. / Шиканова Т. А.; Сиб. гос. мед. ун-т. Каф. лат. яз. и мед. терминологии. Томск : Печат. мануфактура, 2004. 128 с. ISBN 5–94476–043–5.

144084; **612.6:575.17:004.9**; Handbook on analyzing human genetic data : computational approaches and software. / Lin Shili, Zhao Hongyu ed. Heidelberg etc. : Springer, cop. 2010. XIV, 333 с. : ил., табл. Библиография в конце ст. указ.: с. 331–333. ISBN 978–3–540–69263–8; Руководство по анализу генетических данных человека: методы вычислений и программное обеспечение.

77131; **614.876:539.1.074:002.1:004.65**; Радиация и риск : Состояние базы данных РГМДР на 1 декабря 1997 г. Спец. вып. 1999. 141 с. ил. Рез. работ англ. Библиография в конце.

Многоаспектное индексирование относится к приёмам углублённого анализа документов и осуществляется систематизаторами-предметниками естественнонаучных библиотек на этапе обработки поступающих изданий. При индексировании систематизатор прямо указывает комплементарные области знаний, к которым адресуется документ. Крайний слева индекс отмечает главную тему документа и его поло-

жение в классификации общепринятых областей знания, тогда как последующие индексы, отделённые то него двоеточием, характеризуют аспекты знания, к которым он адресуется. При интерпретации индекса главную тему следует соотносить с каждым аспектом, и тогда перед интерпретирующим в полной мере раскрывается содержание издания. Таким образом, картина аспектных ссылок, адресованных к смежным и комплементарным наукам и дисциплинам оказывается как бы скрытой внутри составного индекса. Для воссоздания этой картины нужно извлечь эти ссылки, не теряя при этом связь с темой индекса [9].

При этом необходимо помнить, что картина связей всякой науки составляется, с одной стороны, ссылками науки на своё окружение, а с другой — ссылками её окружения на эту науку. Обязательный учёт обоих факторов даёт возможность увидеть картину связей целиком и избавляет исследователя от случайных заблуждений по поводу междисциплинарности отдельных научных направлений.

Данные для представления связей наук черпаются непосредственно из составных индексов систематизации электронного каталога, и при анализе они используются в двух направлениях.

Первое направление опирается на оценку количества индексов, использующих аспектные ссылки на интересующие области знания. Такая оценка даёт представление о существовании устойчивых связей между отдельными областями знания. Результат такой оценки указывает на необходимые условия для существования мультидисциплинарности: существование аспектов, адресованных другим дисциплинам, и позволяет оценить относительную силу междисциплинарных воздействий.

Второе направление использует подсчёт количества самих аспектных ссылок, указывающих на связи с определёнными областями знания, которые скрыты в индексе систематизации. Эта оценка выявляет достаточные условия, однозначно говорящие о присутствии или отсутствии связей между науками и дисциплинами

в количественном выражении. В качестве показателя выступает количество накопленных аспектных ссылок.

Универсальные иерархические библиотечно-библиографические классификации, такие как УДК, ББК, ДКД и другие, подчинены законам уровневых систем, и подходы в аспектном анализе разделяются на уровневые и междууровневые. Первый предполагает учёт аспектных ссылок на каком-либо одном уровне. Второй — учёт характера междууровневых связей; в этом случае важно понимать, что общая картина связей наук складывается из суммы ссылок анализируемой науки на комплементарные области знания и ссылок этих областей на анализируемую науку.

В электронных каталогах естественнонаучных библиотек наиболее важную роль в процессе формирования всякой междисциплинарной связи играет тема, под которой собираются аспекты. Она определяет оттенки знания, которое необходимо ей для своего раскрытия. Прослеживание максимумов аспектных ссылок, накопленных под определенной темой, позволяет говорить о междисциплинарных интересах отдельных научных направлений. Максимальные количества накопленных аспектных ссылок, собранных под определённой темой и на определённом уровне иерархии, свидетельствует о доминанте научного интереса к данному направлению.

От теории к практике

Рассмотрим пример осуществления аспектного анализа на реальном корпусе данных каталога БЕН РАН, использующего УДК. Этот корпус представлен генеральной совокупностью данных, заключенной в 224748 объектах систематизации (документах), которые накоплены в открытом электронном каталоге БЕН РАН за период около 30 лет, актуализированном на конец 2019 года. Он отражен в 374957 индексах, сформированных в режиме углублённой систематизации (*табл. 1*).

**Характеристика корпуса данных систематизации
каталога БЕН РАН**

Оценочные параметры	по классам УДК									
	0	1	2	3	5	6	7	8	9	Всего
Составных индексов	12210	1093	59	3339	36739	22806	232	846	1065	78389
Другие индексы	30262	7838	313	12504	151473	85957	1220	2511	4490	296568
Всего индексов										374957
Всего индексов скрытых в аспектах										81853
Объектов систематизации										224748

Видовой состав документальной базы: монографии – 128255 единиц, многотомные издания – 18671 единиц, выпуски сериальных изданий – 49820 единиц, выпуски материалов конференций – 5762 единиц, выпуски журнальных изданий – 10859 единиц, авторефераты диссертаций – 4817 единиц, диссертации – 2 единицы, препринты – 2038 единиц. Нематериальные единицы учёта в каталоге представлены сводными уровнями, обеспечивающими внутреннюю группировку изданий – 4921 индексов систематизации.

Обобщенные результаты систематизации представлены в виде распределения по девяти областям знания нулевого уровня иерархической классификации УДК (рис. 1). На диаграмме представлены абсолютные значения количества индексов систематизации. Всего – 374957 индексов. Черным цветом представлены индексы, содержащие аспекты; серым – не содержащие аспекты.

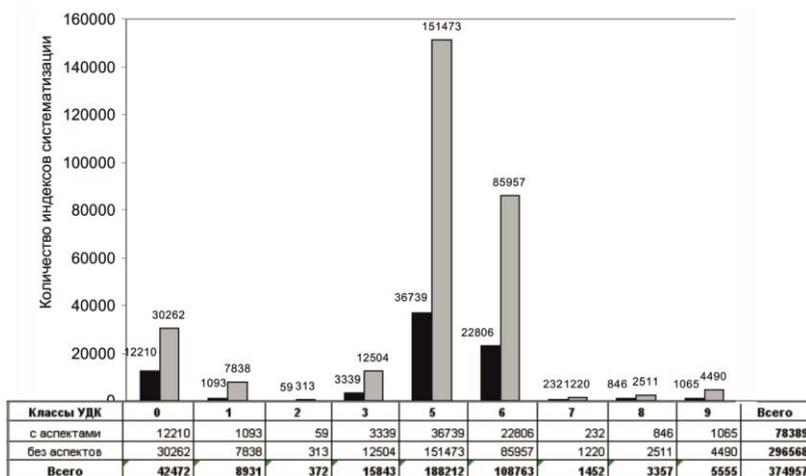


Рис. 1. Представление корпуса данных в виде распределения индексов систематизации

Распределение отражает профильный характер документов, накопленных в библиотеке, где максимумы отмечают их принадлежность к областям знания: 0 — Знание в целом; 3 — Экономика; 5 — Математика и естественные науки и 6 — прикладные науки. Такая картина верифицирует данные корпуса естественнонаучных документов, свидетельствует о правильной ориентации работы библиотеки в целом, соответствии фонда её профилю и, в сущности, является типической для научно-технических и естественно-научных библиотек.

Распределение составных индексов в представленных областях знания (тёмное) при углублённой систематизации отражает количество накопленных составных индексов в каждом из классов и отмечает классы, отвечающие необходимому условию существования междисциплинарности. Кроме того на рис. 1 отражена профильность каталога библиотеки, которая концентрируется в классах: 0, 3, 5 и 6.

Обобщая оценку распределения, можно заметить, что среди научных исследований, завершившихся публикацией документа, представленного в библиотечном фонде, меж-

дисциплинарные исследования составляют около 20 %. Эту оценку, определённую как доля составных индексов среди документов некоторой области знания, можно использовать в качестве показателя общего уровня междисциплинарной активности. Любопытно, что она располагается в диапазоне оценок известной эмпирической закономерности, выявленной Паретто.

Матрица перекрёстных аспектных ссылок нулевого уровня УДК в каталоге БЕН РАН (табл. 2) даёт представление о распределении аспектных ссылок по темам документов. В качестве элементов этой матрицы выступает абсолютное количество аспектных ссылок между различными областями знаний. Главная диагональ матрицы суть количество аспектных ссылок внутри данной области. Она отражает погруженность области знаний в решение своих собственных задач, например класс 2 «Религия. Богословие» аспектных ссылок на себя в каталоге не содержит, а класс 5 «Естественные науки» из общего числа ссылок — 34916 имеет две трети — 23465 ссылок на себя. Примечательно, что выделяются классы 0, 5, 6, имеющие максимумы накопленных аспектных ссылок на главной диагонали, такая картина гносеологически мотивирована и говорит о гармоничности развития естественно-научной сферы знания. Интересно отметить, что вклад в общий объём информации со стороны прикладной сферы (класс 6) и со стороны сферы абстрактных исследований (класс 0) почти одинаков.

Таблица 2

**Распределение аспектных ссылок по классам
верхнего (нулевого) уровня УДК**

Количество аспектных ссылок по классам											
Темы документов по классам	Классы	0	1	2	3	5	6	7	8	9	Всего
	0	8387	268	6	877	1844	1369	215	117	125	13208
	1	133	415	20	194	121	133	18	48	28	1110
	2	9	14	0	8	6	1	4	11	8	61
	3	1146	270	11	791	506	579	11	59	81	3454
	5	6361	757	10	1366	23465	5526	55	20	672	38232
	6	4152	319	4	986	8454	9505	125	25	83	23653
	7	83	8	5	7	47	71	9	1	7	238
	8	163	86	9	96	137	292	6	41	32	862
	9	391	14	6	63	336	96	17	32	80	1035
Всего	20825	2151	71	4388	34916	17572	460	354	1116	81853	

Более наглядным для отображения связей наук и научных направлений является представление информационных потоков, относящихся к конкретному классу в пределах нулевого уровня иерархии. На таких схемах (рис. 2 – Стрелками показаны информационные потоки между областями знаний) проступают закономерности, которые не заметны в матричном отображении. Например, у экономики (класс 3) характерно преобладание числа обращений к классу 5 над обращениями из класса 5. Это свидетельствует о поиске ею приложения своего социально-экономического фактора в естественнонаучной сфере.

Вместе с тем, надо отметить, что предложенная визуализация информационных потоков выполнена в «плоском» представлении, поскольку данные целиком расположены на одном иерархическом уровне. Тогда как сопоставление данных различных иерархических уровней, очевидно, потребует объёмного и многоуровневого отображения для углубления представлений о связях наук.

КЛАССЫ	5	
	темы	аспекты
0	1844	6361
1	121	757
2	6	10
3	506	1366
5	23465	23465
6	8454	5526
7	47	55
8	137	20
9	336	672

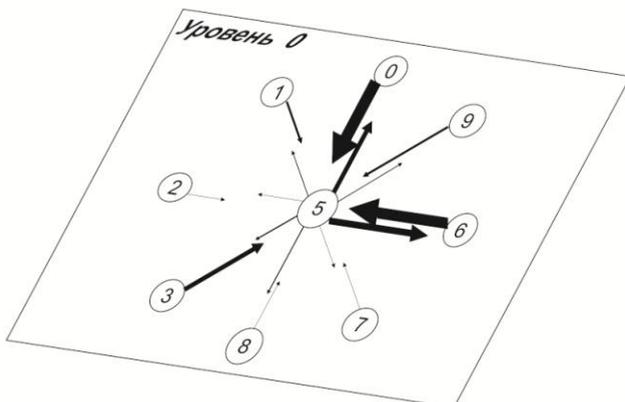


Рис. 2. Взаимодействие областей знания

Не менее выразительно выглядит матрица перекрёстных ссылок на первом уровне класса 5, являющегося профильным исследуемому каталогу. На рис. 4 отражено распределение общего количества индексов систематизации каталога БЕН РАН для класса 5 «Математика и естественные науки». Этот класс представлен в анализируемом каталоге 188 212-мя индексами систематизации, из которых 151 473 не содержат аспектов, а 36 739 индексов являются составными и содержат аспекты. На рис. 3 приведено распределение количества аспектных (черный цвет) и безаспектных (серый) индексов по десяти областям знания на первом уровне УДК класса 5.

Анализ распределения рассматриваемой группы индексов выделяет класс 56 «Палеонтология», систематизация которого содержит почти в два раза больше составных (аспектных) индексов, чем простых (безаспектных). При этом значение соответствующего элемента главной диагонали матрицы (табл. 3) равно 1, что свидетельствует о практически полном отсутствии интереса к внутренним проблемам. Наблюдаемый эффект характеризует идеальную междисциплинарность: научную дисциплину, занятую проблемами смежных наук. По тематике своих прямых исследований она

привлекает все науки, кроме Астрономии и Физики, а в своих ссылках опирается на Науки о Земле и Биологию.

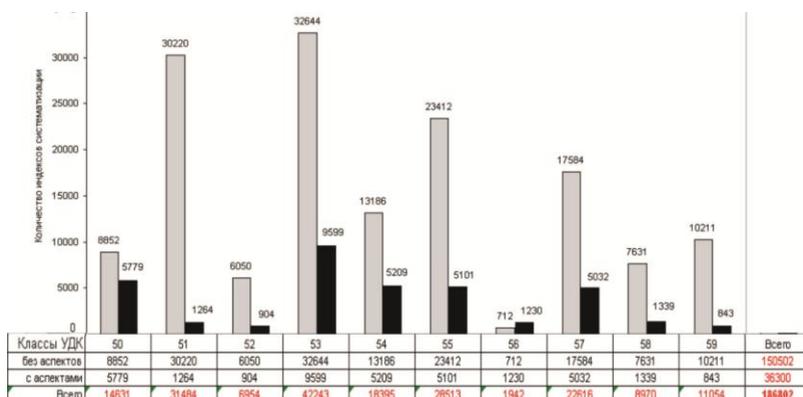


Рис. 3. Распределение индексов систематизации по областям знания в классе 5 УДК

Рассмотрение матрицы в табл. 3 позволяет отметить самоуглублённость Физики, Химии, Биологии и Наук о Земле, которые имеют максимальное количество накопленных аспектных ссылок на главной диагонали, что свидетельствует об исследовательской активности, направленной на решение внутренних задач. Документы по тематике класса 57 «Биология» обнаруживают через аспектные ссылки тенденцию к междисциплинарным связям практически по всем направлениям естественнонаучных исследований, кроме Палеонтологии и Астрономии.

Таблица 3

Матрица перекрёстных аспектных ссылок первого уровня.

		Количество аспектных ссылок по классам										
Темы документов по классам	Классы	50	51	52	53	54	55	56	57	58	59	Всего
	50	1014	27	78	45	411	223	2	436	450	418	3104
	51	2	436	7	287	5	12	0	6	0	0	755
	52	37	36	134	32	4	210	1	10	3	0	467
	53	0	2570	8	3407	297	10	0	5	1	1	6299
	54	115	173	8	351	3078	189	0	97	7	1	4019
	55	279	149	294	205	492	2005	26	109	16	1	3576
	56	6	173	0	0	3	819	1	316	92	31	1441
	57	95	186	46	356	663	123	21	620	537	466	3113
	58	28	6	19	4	46	29	0	297	95	5	529
59	15	2	4	6	5	15	1	205	13	60	326	
Всего	1591	3758	598	4693	5004	3635	52	2101	1214	983	23629	

Переходя к междууровневым сравнениям, отметим, что, в отличие от уровневых оценок, междууровневые оценки требуют рассмотрения как прямых (табл. 4), так и обратных (табл. 5) матриц перекрестных ссылок. Структура связей такой междууровневой оценки показана на рис. 4.

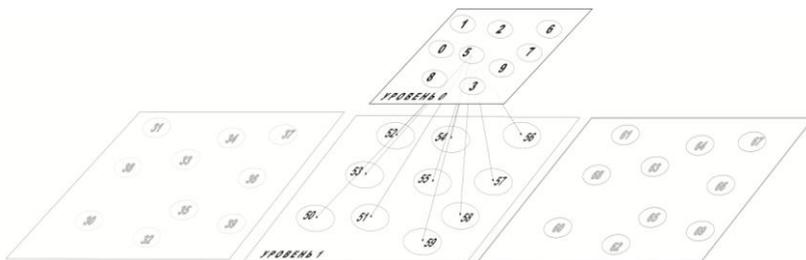


Рис. 4. Структура связей для междууровневых оценок

Прямая матрица манипулирует объёмом аспектных ссылок, заключённых на вышестоящем уровне по отношению к нижестоящему. Так, для класса 5 существует 22420 аспектных ссылок, которые адресованы к нижестоящим классам от 50 до 59.

Таблица 4

Прямая матрица перекрёстных аспектных ссылок с нулевого уровня на первый

Количество аспектных ссылок по классам												
Темы документов по классам	Классы	50	51	52	53	54	55	56	57	58	59	Всего
	0	53	1152	18	290	66	96	8	68	3	15	1769
	1	21	48	5	12	8	18	0	15	1	1	129
	2	0	5	0	0	0	0	0	0	1	0	6
	3	87	193	9	71	31	47	2	30	3	7	480
	5	1591	3589	597	4695	4018	3578	52	2101	1214	985	22420
	6	582	591	323	1814	2061	512	4	1820	546	199	8452
	7	13	4	3	16	0	2	1	2	3	2	46
	8	20	24	3	16	17	16	0	15	6	13	130
	9	62	21	23	8	9	153	5	44	3	1	329
	Всего	2429	5627	981	6922	6210	4422	72	4095	1780	1223	33761

Таблица 5

**Обратная матрица перекрёстных аспектных ссылок
с первого уровня на нулевой**

Количество аспектных ссылок по классам											
Темы документов по классам	Классы	0	1	2	3	5	6	7	8	9	Всего
	50	542	24	3	626	3106	1605	23	0	311	6240
	51	139	54	0	267	755	70	4	2	3	1294
	52	305	1	1	21	467	79	3	1	25	903
	53	2090	402	0	127	6299	934	3	2	32	9889
	54	513	16	0	57	4018	785	6	2	11	5408
	55	799	18	1	129	3578	478	9	7	211	5230
	56	84	2	0	3	1268	10	0	0	5	1372
	57	792	95	0	88	3113	1027	4	2	34	5155
	58	438	8	4	13	529	387	2	2	14	1397
	59	371	3	0	7	326	149	0	1	16	873
	Всего	6073	623	9	1338	23459	5524	54	19	662	37761

Задачи междууровневых оценок — отразить динамику потребности на информацию, связанную с конкретной наукой или научным направлением; она решается выявлением классификационного индекса, наиболее часто фигурирующего в аспектных ссылках. В этой связи при аспектном ана-

лизе удобно использовать сквозные трассировки максимумов накопленных ссылок. Такой приём удобен, когда мы хотим выяснить, на каком разделе знания более всего концентрировался интерес анализируемой области науки. Покажем это на примере Медицины. С этой целью подготовим выборку систематизированных документов по теме 61 «Медицинские науки», содержащих аспектные ссылки на все прочие классы. На верхнем (нулевом) уровне максимум накопленных аспектных ссылок располагается в области класса 6 «Прикладные науки» и содержит 9963 аспектные ссылки (рис. 5). Далее подготовим выборку индексов по первому уровню класса 6 и увидим, что максимум накопленных аспектных ссылок располагается в классе 62 в объёме 3093 ссылок. Теперь производим подсчёт ссылок для второго уровня класса 62, где максимум встречен в классе 621 в объёме 1618 ссылок. И переходя последовательно по уровням, мы достигаем четвёртого уровня, на котором причина интереса проступает достаточно отчётливо и не требует комментариев (рис. 5):

<621.37 Радиотехника. Техника электромагнитных колебаний>

<621.38 Электроника. Фотоэлектроника. Электронные лампы. Трубки. Рентгенотехника>

<621.39 Электросвязь. Радиосвязь. Техника и аппаратура для передачи изображений. Телеуправление>

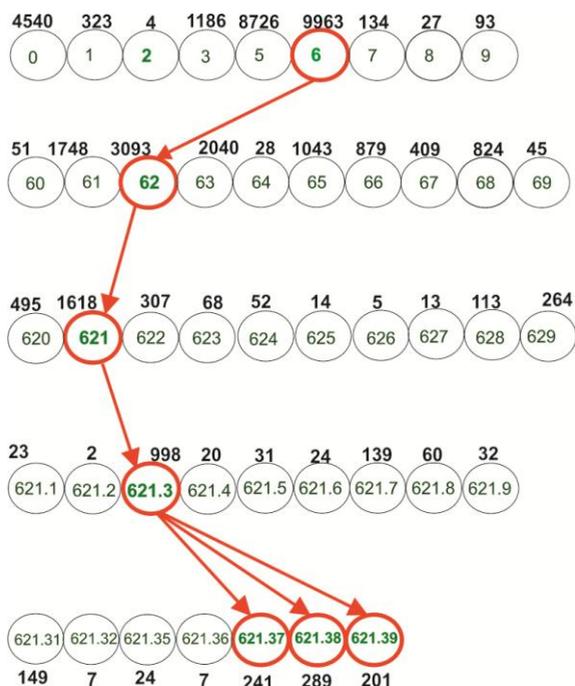


Рис. 5. Схема последовательного анализа максимумов накопленных аспектных ссылок

Выводы

Таким образом, аспектный анализ может служить достаточно чувствительным методом библиометрических исследований и при этом не требует упрощения представлений о структуре научного знания. Опираясь на широко распространённый в библиотеках на классификационный универсум он доступен в любой библиотеке, использующей универсальные иерархические библиотечно-библиографические классификации такие как УДК, ББК, ГРНТИ и другие. Применение перечисленных приёмов анализа с использованием таблиц перекрёстных ссылок в среде иерархически организованных данных позволяет делать выводы о связях научных дисциплин, направлениях и динамике развития наук, состоянии профильного библиотечного фонда.

Литература

1. Leydesdorff L. Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance? / *Scientometrics* (2018) 116:2113–2121; [\(0123456789\),-volV\)](https://doi.org/10.1007/s11192-018-2810-y(0123456789),-volV)(0123456789),-volV).
2. Leydesdorff L., Wagner C. S., Bornmann L. Diversity measurement: Steps towards the measurement of interdisciplinarity? / *Journal of Informetrics*. 2019. Т.13. № 3. pp. 904–905.
3. Elsevier SciVal // <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/scival>
4. Elsevier Scopus // <https://www.scopus.com>
5. Thomson Reuters WoS // <http://webofknowledge.com>
6. ELibrary РИНЦ // http://elibrary.ru/project_risc.asp
7. Белоозеров В. Н., Астахова Т. С., Антошкова О. А., Смирнова О. В., Сукиасян Э. Р. Методическое пособие по Универсальной Десятичной Классификации. М.: ВИНТИ, 2007. 151 с.
8. СИБИД ГОСТ Р 7.0.66–2010 (ИСО 5963:1985) Индексирование документов. Общие требования к координатному индексированию = ISO 5963:1985 Documentation – Methods for examining documents, determining their subjects and selecting indexing terms (MOD). / Издание официальное / М. : ИПК Издательство стандартов. 2011. 16 с.
9. Сысоев А. Н., Белоозеров В. Н. Аспектный анализ классификационных индексов документов библиотечного фонда // *Научно-техническая информация, Сер. 1*. 2020. № 8. С. 18-26.

Принципы обеспечения повторного использования данных и их применение в инфраструктурах исследовательских данных

Н. А. Скворцов

ФИЦ «Информатика и управление» РАН

Аннотация. Развитие технологий сопровождения данных, обеспечивающего их повторное использование в научном сообществе весьма актуально. С этой целью были предложены принципы поддержки обнаруживаемых, доступных, интероперабельных и повторно используемых данных, известные под аббревиатурой FAIR. Инфраструктуры исследовательских данных разрабатываются в соответствии с этими принципами. Однако принципы FAIR-данных интерпретируются по-разному, и подходы к обеспечению интероперабельности и повторного использования исследовательских данных могут существенно отличаться. Кроме прочего, эти принципы декларируют обеспечение понимания данных как человеком, так и машиной, и в этом смысле признаются рассчитанными на перспективу. В статье приведён обзор решений для обеспечения повторного использования данных в инфраструктурах данных и исследован потенциал концептуального моделирования, позволяющего семантически связывать данные методы и другие ресурсы, классифицировать и идентифицировать их в предметной области, интегрировать и повторно использовать, существенно повышать производительность исследований и автоматизацию при работе с неоднородными данными.

Ключевые слова: инфраструктуры исследовательских данных, повторное использование данных, принципы FAIR-данных.

The principles for ensuring data reuse and applying them at research data infrastructures

N. A. Skvortsov

FRC "Computer Science and Control" RAS

Abstract. Enhancement of data curation technologies supporting their reuse in the research community is highly relevant. For this purpose, the principles of findable, accessible, interoperable, and reusable data known

by FAIR abbreviation were proposed. Research data infrastructures are developed following these principles. However, the FAIR data principles are interpreted differently, and approaches to ensuring interoperability and reuse of research data may differ significantly. Among other problems, these principles declare understanding of data by both humans and machines and are recognized as prospective in this sense. The paper reviews solutions for data reuse in data infrastructures and explores the potential of conceptual modeling that allows to semantically link data, methods and other resources, classify and identify them in the domain, integrate and reuse them. It significantly increases the productivity of research and automation of heterogeneous data processing.

Keywords: research data infrastructures, data reuse, FAIR data principles.

1. Введение

Существенная часть времени в большинстве исследовательских проектов и по сей день тратится на выявление семантики используемых данных из различных источников и решение проблем их неоднородности. Однако при нынешних объёмах и разнообразии данных в науке и количестве направлений их необходимого анализа уже нет возможности решать эти проблемы вручную. Поэтому сегодня в центре внимания находится развитие технологий сопровождения и сохранения данных, обеспечивающих их интероперабельность и повторное использование в научном сообществе.

Создаются глобальные инфраструктуры исследовательских данных, которые направлены на обеспечение доступности научных данных. В них неоднородность данных также остаётся велика и разрешается, в основном, силами исследователей при решении научных задач. А сервисы инфраструктур данных тем временем обеспечивают исследователей необходимой информацией для преодоления неоднородности данных. Тем не менее, научное сообщество ведёт активные исследования для того, чтобы минимизировать неоднородность данных средствами современных разрабатываемых инфраструктур данных.

Данная работа посвящена обзору подходов, применяемых в инфраструктурах данных для обеспечения повторного

использования данных с учётом общепринятых и декларируемых в научном сообществе принципов его обеспечения. Также исследован потенциал концептуального моделирования для решения обозначенной проблемы.

В следующем разделе приводятся принципы обеспечения повторного использования исследовательских данных, их возможные интерпретации и соответствующие направления в исследованиях. В разделе 3 представлен обзор различных решений в инфраструктурах данных в соответствии с этими принципами, преимущества и недостатки таких решений. В разделе 4 обсуждаются возможности, которые может привнести в инфраструктуру данных концептуальное моделирование на основе формальных спецификаций предметных областей исследования.

2. Принципы обеспечения повторного использования данных

В ходе исследований очищаются, обогащаются и классифицируются исходные данные наблюдений или результатов предшествующих исследований, реализуются научные методы и при их применении генерируются данные, содержащие новые знания об объектах исследования. Все эти результаты полезны, если они могут быть воспроизведены и повторно используются в дальнейших исследованиях, над другими исходными данными, другими исследовательскими группами. Поэтому в дисциплине сопровождения данных (data curation), нацеленной на управление данными в течение их жизненного цикла, обеспечивающее их оптимальное долгосрочное использование, важной частью стали обеспечение доступности и повторного использования данных и методов, а также воспроизводимости результатов исследований. И на протяжении ряда лет эти проблемы являются предметом интенсивных исследований.

С учётом этих целей формировались требования к способам длительного хранения и обеспечения доступа к данным. Декларировалось, что данные важно описывать метаданными для определения их семантики

и происхождения для обеспечения поиска и выбора данных, релевантных проводимым исследованиям. Также данные должны быть связаны со средствами их обработки и анализа, методами предметной области исследования. Хорошим представителем реализации этих требований стал, в частности, проект WF4Ever [1], направленный на воспроизводимость результатов исследований. В центре этого проекта находятся повторно используемые научные потоки работ. Формируются контейнеры, называемые исследовательскими объектами (research object), в которых спецификации потоков работ сопровождаются данными, метаданными, документацией, ссылками, правилами доступа. Поддерживаются сообщества исследователей, проекты, которым предоставляется доступ к исследовательским объектам, возможность публикации и поиска потоков работ.

Позже на основании накопленного в научном сообществе опыта достижения интероперабельности и повторного использования данных и на основании обсуждения в научном сообществе (в первую очередь, в сообществе FORCE11 [2], нацеленном на электронные публикации на основе семантических технологий) были предложены принципы управления данными, известные под аббревиатурой FAIR [3]. Данные должны быть однозначно идентифицированы, семантически определены с помощью общих словарей и онтологий, сопровождаются информацией о происхождении, должны иметь четкие правила доступа, соответствовать известным протоколам, стандартам или иметь известные отображения в них. Свойства обнаруживаемости (возможности поиска), доступности и интероперабельности (возможности совместного функционирования в системе) являются условиями для основного целевого свойства данных — возможности их повторного использования (см. табл. 1). Те же требования предъявляются к метаданным, методам, процессам, связанным с данными, также являющимся разновидностями данных.

Актуальность декларируемых целей, запоминающаяся и интуитивно понятная аббревиатура, простота формулировки принципов, универсальность и абстрактность без навязывания определённых реализаций быстро сделали их популярными, и они были поддержаны на самом высоком

уровне. Принципы FAIR-данных уже сыграли и продолжают играть большую роль в активизации и формировании направления исследований повторного использования данных. Однако неформальное определение этих принципов также послужило причиной появления совершенно различных их толкований, иногда противоречивых или ограничивающих их цели. В результате, дальнейшие исследования стали существенно разнонаправлены:

1. Появились предложения дальнейшей детализации требований к данным, в том числе, детализированные списки, расписывающие и усложняющие каждый их принцип, новые красивые аббревиатуры, вносящие дополнительные неформальные принципы для различных областей и целей.

Таблица 1

Принципы FAIR-данных

<p>Findable – обнаруживаемые</p> <ul style="list-style-type: none"> • Постоянные уникальные идентификаторы • Обширное описание метаданными • Метаданные включают идентификаторы данных • Регистрация или индексирование для поиска 	<p>Accessible – доступные</p> <ul style="list-style-type: none"> • Стандартный протокол • Извлечение по идентификаторам • Возможна авторизация и аутентификация • Метаданные доступны, даже когда данные больше не доступны
<p>Interoperable – интероперабельные</p> <ul style="list-style-type: none"> • Использование формального языка представления знаний • Использование словарей, отвечающих FAIR • Ссылки на другие данные 	<p>Reusable – повторно используемые</p> <ul style="list-style-type: none"> • Определение множества атрибутов применимости • Лицензирование • Детальное описание происхождения • Поддержка стандартов предметных сообществ

2. Принципы FAIR часто рассматриваются не как задающие направление, но как требования, с точки зрения следования которым можно оценивать существующие и новые базы данных и репозитории научных данных.

В результате, предложены метрики оценки и даже системы сертификации баз данных и репозиторий на соответствие принципам FAIR-данных.

3. Подбираются технологии и разрабатываются интерфейсы, которые объявляются обеспечивающими реализацию сформулированных принципов.

5. FAIR-данные примеряются к различным прикладным сферам существенно по-разному, исходя из их предваряющего опыта работы с данными и обычно без больших изменений в подходах.

4. Делаются выводы о следствиях принципов FAIR-данных для различных аспектов управления данными, в частности, цитирования данных, планов управления данными, инфраструктур данных. В частности, в крупных инфраструктурах исследовательских данных эти принципы стали неотъемлемой основой развития, что рассматривается в данной работе подробнее.

Из-за широкого поля интерпретаций принципов FAIR-данных их авторам приходилось иногда прилагать усилия, чтобы вернуть им изначальный замысел [4]. В этом плане важно отметить некоторые ориентиры. FAIR-данными могут являться практически любые ресурсы, связанные с данными. Метаданные, словари и онтологии, интерфейсы и протоколы, реализации методов, потоки работ и программы, результаты интеграции данных и методов, цитирования, документы и другие разновидности данных могут рассматриваться с точки зрения объектов повторного использования и могли бы отвечать свойствам FAIR. Таким образом, речь не только и не столько о публикации наборов данных.

Ещё одним важным аспектом принципов FAIR-данных является декларируемая ими возможность управления данными машиной (machine-actionability) помимо человека. Таким образом, приветствуется переход от обработки машиной предопределённых наборов метаданных к обеспечению корректной интерпретации данных и метаданных, с которыми машина ранее не работала. В том числе, для этого декларируется важность использования формальных языков пред-

ставления знаний, позволяющих производить вывод, доказывать непротиворечивость описаний для проверки соответствия семантики данных.

В реальности, в условиях неоднородности данных действия над ними, производимые по инициативе машины, исходя из семантики данных, на сегодня достижимы с трудом. Даже автоматизация на основе директив человека может быть нетривиальной задачей. Если обеспечение поиска по метаданным и доступности данных могут быть реализованы без затруднений различными подходами, то принцип обеспечения интероперабельности данных встречает проблемы, связанные, в первую очередь, с определением семантики данных и необходимостью работать с неоднородными данными.

Поэтому принципы FAIR-данных признаются рассчитанными на перспективу. Невозможно сказать, что определённый проект выполняет принципы FAIR полностью или не выполняет их вовсе. Однако стремление к их реализации задаёт направление в сторону разрешения неоднородности данных, автоматизации обработки, анализа данных, научных исследований и повторного использования данных и ресурсов различной природы.

Конкретные реализации принципами FAIR-данных не продиктованы и не предполагаются как требования. Тем не менее, есть определённые технологии, помогающие приблизиться к их выполнению.

В частности, самими авторами принципов FAIR-данных предлагался набор технологий [5], основанных на языке RDF в качестве основополагающей модели данных. Используются контейнеры данных (Linked Data Platform – LDP), средства отображения данных из разных моделей (RDF Mapping Language – RML) и средства извлечения данных совместно с метаданными (Triple Pattern Fragments – TPF). Нередко принципы FAIR-данных в целом связывают с технологиями связанных открытых данных (Linked Open Data – LOD). На самом же деле, такого ограничения нет, и оно непродуктивно.

Распространённым подходом к реализации принципов FAIR-данных стал программный продукт FAIR Data Point (FDP) [6], обеспечивающий создание, хранение и обслуживание метаданных. Он позволяет публиковать метаданные цифровых объектов и обнаруживать информацию (метаданные) о доступных цифровых объектах. Метаданные должны соответствовать схемам, определенным сообществами предметных областей. Такие репозитории и инфраструктуры исследовательских данных как EUDAT [7], Zenodo [8], BioPortal [9] использовали FDP для реализации возможностей повторного использования данных и различных ресурсов и интеграции с другими инфраструктурами данных, обеспечивающими поиск ресурсов по метаданным.

3. Следование принципам FAIR-данных в инфраструктурах исследовательских данных

В инфраструктурах исследовательских данных, призванных поддерживать доступность исследовательских данных, упростить доступ к ним и сопровождать их средствами обработки и анализа, принципы FAIR-данных стали ведущими для выбора направления развития технологий. Ведь, в первую очередь, инфраструктуры данных предоставляют услуги хранения метаданных о существующих наборах научных данных, поиска по ним и решения научных задач над предоставляемыми данными. Поэтому в инфраструктурах данных концепции FAIR упоминаются в основополагающих документах, декларируется их поддержка, обсуждаются подходы к следованию данным принципам. На сегодня концепции инфраструктур данных остаются неустойчивыми и активно обсуждаемыми. В связи с различными интерпретациями принципов FAIR-данных и с тяготением к привычным средствам работы с данными в определённых областях подходы к обеспечению интероперабельности и повторного использования исследовательских данных могут существенно отличаться.

Основу обеспечения возможности поиска ресурсов в инфраструктурах данных составляет сбор метаданных и поиск по ним. Снабжение данных метаданными для семантически значимого поиска чаще всего реализуется на основе определённых наборов общих параметров. Используются текстовые определения, ключевые слова, ссылки на термины словарей, редко типизированные значения. Для разных предметных областей могут выбираться собственные наборы атрибутов, помимо универсальных. Однако такой подход не учитывает специфические свойства разных типов объектов реального мира, данные о которых необходимо найти. Метаданные описывают обычно наборы данных в целом и редко полностью описывают состав и структуру баз данных для поиска релевантных фрагментов.

В инфраструктурах данных организуются реестры ресурсов, возвращающие идентификаторы релевантных данных в качестве ответов на поисковые запросы по метаданным. В основном, поиск по метаданным производится для дальнейшего ручного интегрирования ресурсов и решения задач с их привлечением. Автоматизированного поиска с последующим повторным использованием ресурсов машиной не предлагается, хотя в некоторых инфраструктурах сообщества исследователей и стремятся добиться этого. Это связано, в первую очередь, с неформальностью самих метаданных. В качестве идентификаторов ресурсов чаще всего используются стандарты DOI и URI. Первый гарантирует постоянство идентификаторов, но менее гибкий и интуитивно понятный. Глобальная идентификация с прямым доступом к необходимым фрагментам данных обсуждается редко, чаще осуществляется доступ к ресурсу в целом. Практически не реализуется идентификация данных, хранимых в подмножестве атрибутов, данных об определённых объектах реального мира, выбор срезов данных по условиям и тому подобные подходы к идентификации данных. Затрагиваются альтернативные способы идентификации, которые связывали бы данные, относящиеся к одним и тем же сущностям реального мира для возможности подмены релевантных ресурсов.

Обеспечение доступности данных не вызывает затруднений при использовании стандартных протоколов сети Интернет, в том числе, с механизмами разграничения доступа. Архивы метаданных без актуальных доступных данных обычно не поддерживаются. Технологии долговременного хранения данных разной природы, обеспечивающие одновременно эффективный и произвольный доступ к данным и метаданным, в рамках обсуждения FAIR-данных затрагиваются редко, обычно только при обсуждении архитектур инфраструктур данных.

Как уже говорилось ранее, основной проблемой обеспечения интероперабельности данных и, как следствие, реализации всей концепции FAIR-данных становится ручной подход к интеграции неоднородных данных. Автоматизация интеграции найденных релевантных данных и ресурсов возможна только при определённых условиях в перспективе, о чём будет упомянуто в следующем разделе.

Унификация интерфейсов, высокоуровневых протоколов и регламентов доступа к данным, применяемая в инфраструктурах данных, решает, в основном, проблему согласования транспортной функции, но не решает проблемы неоднородности самих передаваемых данных, либо решает её только для определённых типов объектов предметных областей. В качестве решения проблемы неоднородности моделей данных часто принимается решение поддерживать простые форматы данных общего назначения, такие как CSV и JSON. Однако и в этом случае основная причина неоднородности — семантические различия в данных — не учитывается вовсе, а оставляется на откуп пользователям данных для ручной работы. При создании инфраструктур данных практически игнорируются проблемы преобразования между моделями данных [10]. Стандартизацию представления данных, учитывающую их семантику, в разных предметных областях принято основывать на разных принципах. Это может быть стандартизация набора атрибутов с определённой семантикой предметной области и формата значений для них для того, чтобы любые записи в предметной области были однородны. Под-

держка форматов и стандартов представления данных, используемых в определённых предметных областях, позволяет работать с чётко определённой семантикой данных, известной и принимаемой сообществами этих областей, но при этом могут быть сильно ограничены разновидности сущностей, выразимых их посредством в предметной области.

Выше упоминалось, что нередко концепцию FAIR-данных связывают с технологиями LOD на основе языка RDF. Это происходит, должно быть, оттого что в списке принципов есть схожие на требования LOD. В частности, это требование о создании в данных ссылок на другие релевантные данные, которое, возможно, позаимствовано в качестве одного из принципов FAIR-данных. Однако в инфраструктурах данных этот принцип реализуется связью данных, используемых совместно, в частности, посредством создания контейнеров связанных ресурсов. В инфраструктурах данных часто используется идея создания контейнеров, содержащих данные, метаданные, ссылки на внешние ресурсы, программы, документацию и другие ресурсы, используемые совместно с данными, необходимые для их обработки и анализа.

Для обеспечения автоматизации обработки данных и управления данными машиной чаще всего используются подходы, которые действительно реализуют возможности автономной работы машины, но фактически являются предопределёнными человеком. Среди таких подходов спецификация потоков работ для поэтапной обработки данных, инициирование действий на основе событий, создание планов управления данными для упреждающего определения таких действий как публикация данных, удаление неактуальных данных и других. Однако в рамках реализации принципов FAIR-данных есть стремление реализовать автоматизацию на основе определения семантики данных. Такова, например, инициатива интернета FAIR-данных и услуг (Internet FAIR Data and Services – IFDS) [11].

Ключевыми принципами для достижения интероперабельности неоднородных данных являются использование формальных языков представления знаний и словарей. Каза-

лось бы, для реализации этих принципов естественным было бы использование языков формальных онтологий и автоматического вывода над ними. Однако онтологии часто даже не упоминаются в некоторых инфраструктурах данных, либо используются только для прямых ссылок на понятия. О том, какие задачи для обеспечения повторного использования данных могут решаться с помощью онтологий, подробнее сказано в следующем разделе.

Необходимо привести примеры разных подходов к решению проблем разной семантики данных. Так, наиболее распространённая инфраструктура астрономических данных CDS [12] решает проблемы неоднородности данных на основе требования к публикуемым астрономическим каталогам использовать одни и те же стандартизованные атрибуты с определённой семантикой при всякой возможности. Используя пересечение атрибутов в разных каталогах, инфраструктура предоставляет специализированные интерфейсы и языки запросов, которые могут работать с несколькими каталогами одновременно. На деле оказывается, что и в эти атрибуты часто вкладывается различная семантика, не описанная авторами каталогов. Метаданные о структуре каталогов имеют текстовый, частично структурированный вид. Семантика атрибутов описывается дескрипторами UCD, принимаемыми сообществом, но допускающими неоднозначность. Сервисы DADI [12] предоставляют доступ к существующим репозиториям астрономических данных, включая CDS, и соответствующим услугам. Они декларируют следование принципам FAIR-данных, однако не привносят существенного развития в сторону семантических подходов. Вероятно, авторы считают достаточными существующие подходы. Практика же показывает, что при работе с астрономическими данными на основе существующих услуг семантическая неоднородность данных остаётся существенной, и решение проблем их неоднородности требует существенных усилий практически в каждом проекте.

В области биомедицины общепризнанным ресурсом является BioPortal [9]. Он издавна объединяет множество онтологий, известных в данной области, описывающих разные

аспекты затрагиваемых предметных областей и пересекающиеся друг с другом. С некоторых пор осознана необходимость согласования этих онтологий. Несмотря на то, что основная масса онтологий до сих пор остаётся не затронутой согласованием, и они просто доступны в хранилище BioPortal, тем не менее, в процессе решения научных задач в предметной области вырисовался костяк онтологий, которые претерпели изменения ради согласования друг с другом. В их основе лежит методология описания различных задач, процессов, экспериментов в биомедицине, как лабораторных, так и информационных. В биомедицине широко используется аннотирование разного рода ресурсов в терминах этих онтологий. Предлагаемые сервисы упрощают составление семантических аннотаций. Причём аннотирование допускает не просто ссылки на понятия, но и выражения в их терминах, что позволяет наиболее точно описывать смысл аннотируемых сущностей. По самим онтологиям можно производить поиск релевантных понятий. Таким образом, он является хорошей основой для создания сервисов, обеспечивающих интероперабельность данных.

Обеспечение повторного использования данных в исследовательских инфраструктурах, в первую очередь, находится в зависимости от обеспечения поиска, доступности и интероперабельности данных, рассмотренных выше. Помимо этого, в соответствии с принципами FAIR-данных необходимо поддерживать стандарты сообществ предметных областей, и целесообразно оценивать атрибуты, описывающие нефункциональные свойства данных. Метаданные, необходимые для проверки возможности повторного использования, а не для поиска данных, могут включать описание происхождения данных и лицензионные условия использования.

На деле данные редко снабжаются метаданными об их происхождении, и в инфраструктурах данных их непросто выявить. Поддержка стандартов предметной области, безусловно, осуществляется. Однако они слабо интегрированы в другие средства поддержки FAIR-данных,

и добавляют ручной работы по их интегрированию и адаптации потребителями данных.

Можно сделать общие выводы о состоянии реализации принципов FAIR-данных и, соответственно, обеспечения повторного использования данных в исследовательских инфраструктурах. Очевиден разрыв между подходами к доступу к данным, превалирующими в конкретных дисциплинах. Традиционные способы интеграции данных в них влияют наибольшим образом на решения инфраструктур. Решение проблем неоднородности данных и интеграции релевантных средств их анализа занимает значительную часть усилий пользователей, и инфраструктуры данных, в основном, предоставляют необходимую для этого информацию вместе с данными. Подходы к обработке и анализу данных, управляемым машиной, исходя из их семантики, фактически не реализованы. Однако всем этим проблемам в исследованиях инфраструктур данных уделяется пристальное внимание, что позволяет надеяться на скорое качественное изменение реализации повторного использования исследовательских данных.

Сегодня разрабатывается инфраструктура исследовательских данных Европейское облако открытой науки (European Open Science Cloud – EOSC) [13]. Эта инициатива наибольшим образом связана с реализацией принципов FAIR-данных, так как авторы концепции FAIR, в том числе, занимаются исследованиями для этой инфраструктуры. Поэтому необходимо рассмотреть решения, которые предлагаются в этом проекте. В основе реализации услуг EOSC во многом используются услуги существующих инфраструктур данных, и облако призвано объединить ресурсы, доступные во многих инфраструктурах данных. В первую очередь, интегрируются услуги инфраструктур данных EGI [8] и EUDAT [7] в рамках проекта EOSC-Hub [14]. Они составляют средства регистрации ресурсов, описания метаданных и аннотирования ресурсов, поиска по метаданным и доступа к ресурсам, преобразования, публикации и долговременного хранения данных и другие.

Помимо реализованных услуг в основу EOSC вкладываются решения, которые развивались для реализации принципов FAIR-данных. В исследовательских проектах над данными заблаговременно разрабатываются планы управления данными (DMP) [15] на всё время реализации проектов, что позволяет автоматизировать управление данными проекта. Для связывания данных с релевантными им ресурсами разной природы используется концепция цифровых объектов (Digital Objects – DO) [16]. Это контейнеры, доступные по глобальным идентификаторам, агрегирующие данные, метаданные, информацию о происхождении данных, используемые стандарты и форматы, программный код, лицензии, идентификаторы и ссылки на внешние ресурсы. Для хранения и манипулирования этими видами ресурсов создаются реестры. Для управления цифровыми объектами разработан специальный протокол DOIP [17], определяющий набор операции манипулирования ими. Для автоматизации предлагается разделение на услуги для исследователей и услуги для машинных агентов. Поддерживается движимая типом автоматическая обработка данных. В связи с инфраструктурой данных EOSC развивается инициатива интернета FAIR-данных и услуг IFDS [11]. Она рассматривает соединение данных с инструментами и вычислениями, которые соответствуют им семантически. Таким образом, в инфраструктурах обработка данных может производиться автоматически релевантными инструментами, исходя из семантики самих данных. Такие исследования являются прямым следствием и заслугой руководящих принципов FAIR-данных.

4. Потенциал концептуального моделирования для реализации принципов повторного использования данных в исследовательских инфраструктурах

Как показано выше, наибольшие затруднения для повторного использования данных в исследовательских инфра-

структурах вызывают разрешение неоднородности данных и возможность автоматизации их обработки, связанные с разницей в семантике данных. Поэтому использование семантических технологий видится естественным направлением исследований и развития в данной области, и избегание определения семантики данных вряд ли решит насущные проблемы их повторного использования. Среди подходов к определению семантики данных, в первую очередь, используются средства онтологического и концептуального моделирования. Они могут стать основой решения проблем неоднородности данных и автоматизации обработки и анализа данных. На самом деле, возможность и целесообразность использования формальных концептуальных описаний данных и автоматизации на основе концептуальных подходов в инфраструктурах данных декларируются и исследуются, однако они используются ограниченно, либо избегаются в пользу слабо автоматизируемых оценочных подходов на основе мер близости. Использование методов машинного обучения для связывания релевантных ресурсов также является оценочным подходом к установлению семантики данных и, скорее, применимо для рекомендаций пользователю или эксперту, нежели для потенциальной автоматизации повторного использования ресурсов при решении научных задач.

Для применения семантических подходов в полной мере необходимо формальное описание знаний предметных областей исследования: онтологии предметных областей, предложение общих схем данных для наиболее часто используемых описаний в выбранных общих моделях данных. В инфраструктурах данных с исследовательскими сообществами, работающими в определённых предметных областях, должны быть связаны формальные онтологии затрагиваемых ими областей.

Описание семантики ресурсов предметной области, включая релевантные данные, методы и другие ресурсы, требует их семантического аннотирования в терминах онтологий предметных областей. Семантические аннотации ресурсов являются основными метаданными, связывающими

идентификаторы ресурсов с понятиями предметной области, описанной онтологиями. По этим понятийным метаданным производится семантический поиск данных и ресурсов в исследовательских инфраструктурах. Для основных разделов знаний также составляются концептуальные схемы со структурами для разных типов объектов и спецификациями их поведения и применяемых к ним методов и преобразований или вычислений.

Метаданные сохраняются в реестрах, построенных на основе онтологий. Такие реестры являются хорошо классифицированными коллекциями ресурсов, повторно используемых сообществом. Сообщество может собирать коллекции как наборов данных, так и реализованных методов, потоков работ, программ, описание интерфейсов которых в терминах онтологий позволяет семантически связать данные с другими релевантными ресурсами, в том числе научными методами. При регистрации данные и реализации методов семантически (на основе семантических аннотаций) интегрируются с концептуальными схемами данных в сообществах с учётом также преобразования моделей данных. А после интеграции данные и ресурсы могут многократно использоваться повторно в предметной области без дополнительной интеграции. Таким образом, регистрация данных или других ресурсов, включающая их интеграцию в предметную область сообщества, производится единожды.

Задачи в инфраструктуре данных могут решаться на основе найденных релевантных зарегистрированных ресурсов с данными, методами, программами. Решение задач с использованием общих схем в сообществах позволяет повторно использовать найденные данные и ресурсы с использованием результатов их интеграции, проведённой ранее при регистрации. При необходимости публикации результатов решения задач, включая новые данные и разработанные методы, они регистрируются в реестрах на основе семантического аннотирования в терминах онтологий сообщества и становятся доступны для повторного использования сообществом. То же проделывается и с вновь ис-

пользованными внешними ресурсами, которые ранее отсутствовали в реестрах.

Формальное описание онтологий и стандартных концептуальных схем предметных областей для такого сценария работы инфраструктур данных требует усилий. Однако, во-первых, в большинстве областей исследований с интенсивным использованием данных есть наработки в этом направлении, которые можно использовать и развивать. Во-вторых, сообщества, работающие в определённых предметных областях, весьма и весьма заинтересованы в этой работе. Они могут не только быть пользователями инфраструктур данных, но и коллаборативно разрабатывать и согласовывать спецификации предметных областей в своих сообществах, а затем требовать приверженности этим спецификациям от входящих в сообщество членов. Это, кстати, соответствует принципу поддержки стандартов предметных сообществ. В-третьих, однажды согласованные спецификации могут повторно использоваться многократно и претерпевать изменения и развитие с фиксацией версий, что минимизирует дальнейшие усилия по определению онтологий, схем и интеграции ресурсов.

Применение онтологий и семантического аннотирования вполне соответствует принципам обширного описания данных метаданными. Ведь достаточно подробный анализ существенных свойств большинства объектов предметной области фактически приводит к концептуализации предметной области, а при формальном подходе — к созданию онтологий предметной области.

Формальность онтологии позволяет качественно классифицировать данные практически о любых объектах предметной области и обеспечивает автоматизацию поиска с учётом семантики данных. Онтологии также позволяют описывать данные одновременно с нескольких ракурсов, используя аннотации в терминах разных онтологий или выражения на основе онтологических понятий в аннотациях. Поддержка реестров метаданных на основе онтологий и поиска в них на основе механизмов вывода и языков запро-

сов к онтологиям реализует принципы обнаруживаемости данных.

С другой стороны, применение онтологий и концептуальных схем соответствует принципам использования формальных языков представления знаний и словарей. Использование формальных концептуальных моделей является основой для возможности автоматического вывода. В частности, онтологии в моделях, выразимых на языке онтологий OWL, поддерживают автоматическое решение задач выполнимости формул и включения между их множественными интерпретациями, столь важных для проверки семантического соответствия между разными описаниями ресурсов и поиска ресурсов, релевантных запросу в терминах онтологии, а значит, для решения проблем интероперабельности данных и их обнаруживаемости. Таким образом, автоматический вывод над семантическими спецификациями может стать залогом к обеспечению управления данными машиной в инфраструктурах данных. При этом подробное описание данных, семантики методов и элементов их интерфейсов в терминах онтологий, а также интеграция схем гарантирует корректность и автоматического применения методов к данным. В качестве формальных языков представления знаний могут использоваться разные диалекты языка OWL. Однако это не единственная возможность. Например, стандарты SQL с их расширениями могут быть достаточно формальным основанием для управления данными.

Создание концептуальных схем данных соответствует принципам поддержки стандартов сообществ. Концептуальные схемы предметных областей, форматы данных, высокоуровневые протоколы, принятые в предметных сообществах, стандарты и регламенты работы с данными, безусловно, должны быть использованы для унификации представления данных и установления семантики данных. Концептуальные схемы предметных областей могут определять не только структуры и ограничения для представления данных предметной области, но и возможное манипулирование данными,

поведение описываемых объектов, методы, связанные с определёнными типами сущностей.

Метаданные в терминах онтологий призваны не только для описания данных с точки зрения семантики предметной области для их классификации и поиска. Часть метаданных необходима для выражения других, нефункциональных аспектов описания данных, таких как авторство, актуальность, полнота, качество данных, лицензионные условия, необходимых для корректного повторного использования данных. Поддержка сообществом специализированных онтологий, отражающих нефункциональные характеристики, используемые в предметной области, соответствует принципу описания атрибутов применимости данных. Такими онтологиями могут являться, в частности, онтология качества данных, онтология происхождения данных, и данные могут описываться с их точек зрения тем же способом, который используется для описания семантики данных в предметной области сообщества. Ссылки данных на другие данные также предпочтительно выстраивать на отношениях, семантически определённых в предметной области с помощью онтологий. Для поддержки метаданных, когда ресурс более не доступен, необходимо снабжать их явной информацией об актуальности данных среди других метаданных происхождения. Также и лицензии должны быть четко понятны для человека и интерпретируемы машиной, и их описание может быть реализовано посредством специальных онтологических описаний.

5. Заключение

В работе рассмотрено состояние выполнения принципов FAIR-данных в инфраструктурах исследовательских данных. Возможность и целесообразность использования формальных описаний данных и автоматизации на основе концептуальных подходов в инфраструктурах данных декларируются и исследуются, однако они редко используются в достаточной мере. В работе исследован потенциал концептуального моделирования, позволяющего семантически опи-

сывать, классифицировать и накапливать используемые в сообществах данные, избавляться от неоднородности данных в сообществе. Подход позволяет искать и семантически корректно повторно использовать интегрированные релевантные задачам данные, методы и другие ресурсы вплоть до автоматизации работы с неоднородными данными, что существенно повышает производительность исследований.

Работа выполнена при поддержке ЦКП «Информатика» ФИЦ ИУ РАН [18], Российского фонда фундаментальных исследований, проекты 19-07-01198, 18-29-22096, 18-07-01434.

Литература

1. Belhajjame, K. et al. Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica2012). 2012. P. 1-12.
2. Improving Future Research Communication and e-Scholarship. Bournea, P. et al (eds.) The Future of Research Communications and e-Scholarship. 2011. <https://www.force11.org/>
3. Wilkinson, M. D., et al.: The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3, 160018. (2016). <https://doi.org/10.1038/sdata.2016.18>
4. Mons, B., et al. Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science Cloud. Information Services & Use, 37(1), p. 49-56. 2017. <https://doi.org/10.3233/isu-170824>
5. Wilkinson, M. D., et al. Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science 3, e110. 2017. <https://doi.org/10.7717/peerj-cs.110>
6. FAIR Data Point design specification. 2019. <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec>
7. The EUDAT Collaborative Data Infrastructure. <https://www.eudat.eu/>
8. EGI Foundation. <https://www.egi.eu/>
9. Noy, N. et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research, Volume 37, Issue suppl_2, 1 July 2009, Pages W170-W173, <https://doi.org/10.1093/nar/gkp440>
10. Stupnikov S., Kalinichenko L. (2019) Extensible Unifying Data Model Design for Data In-tegration in FAIR Data Infrastructures. In: Manolopoulos Y., Stupnikov S. (eds) Data Ana-lytics and Management in Data Intensive Domains. DAMDID/RCDL 2018. Communica-tions

in Computer and Information Science, vol 1003, P. 17–36. Springer, Cham. doi.org/10.1007/978-3-030-23584-0_2

11. Mons B. FAIR Science for Social Machines: Let's Share Metadata Knowlets in the Internet of FAIR Data and Services. Data Intelligence. Vol. 1, № 1. 2019. P. 22–42. https://doi.org/10.1162/dint_a_00002

12. Genova, F. Data as a research infrastructure CDS, the Virtual Observatory, astronomy, and beyond. EPJ Web of Conferences, vol. 186, 01001–2018. <https://doi.org/10.1051/epjconf/201818601001>

13. European Open Science Cloud. <https://www.eosc-portal.eu/>

14. EOSC-Hub. <https://eosc-hub.eu/>

15. Guidelines on Data Management in Horizon 2020. Directorate-General for Research & Innovation, EC. 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

16. Wittenburg, P.: From persistent identifiers to digital objects to make data science more efficient. Data Intelligence. Vol. 1, No 1. 2019. P. 6–21. https://doi.org/10.1162/dint_a_00004

17. Digital Object Interface Protocol Specification. Ver. 2.0. DONA Foundation. 2018. https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf

18. Положение о ЦКП «Информатика». <http://www.frccsc.ru/ckp>

References

1. Belhajjame, K. et al. Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica2012). 2012. P. 1–12.

2. Improving Future Research Communication and e-Scholarship. Bourne, P. et al (eds.) The Future of Research Communications and e-Scholarship. 2011. <https://www.force11.org/>

3. Wilkinson, M. D., et al.: The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3, 160018. (2016). <https://doi.org/10.1038/sdata.2016.18>

4. Mons, B., et al. Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science Cloud. Information Services & Use, 37(1), p. 49–56. 2017. <https://doi.org/10.3233/isu-170824>

5. Wilkinson, M. D., et al. Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science 3, e110. 2017. <https://doi.org/10.7717/peerj-cs.110>

6. FAIR Data Point design specification. 2019. <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec>

7. The EUDAT Collaborative Data Infrastructure. <https://www.eudat.eu/>
8. EGI Foundation. <https://www.egi.eu/>
9. Noy, N. et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, Volume 37, Issue suppl_2, 1 July 2009, Pages W170–W173, <https://doi.org/10.1093/nar/gkp440>
10. Stupnikov S., Kalinichenko L. (2019) Extensible Unifying Data Model Design for Data In-tegration in FAIR Data Infrastructures. In: Manolopoulos Y., Stupnikov S. (eds) *Data Ana-lytics and Management in Data Intensive Domains. DAMDID/RCDL 2018. Communica-tions in Computer and Information Science*, vol 1003, P. 17–36. Springer, Cham. doi.org/10.1007/978-3-030-23584-0_2
11. Mons B. FAIR Science for Social Machines: Let's Share Metadata Knowlets in the Internet of FAIR Data and Services. *Data Intelligence*. Vol. 1, № 1. 2019. P. 22–42. https://doi.org/10.1162/dint_a_00002
12. Genova, F. Data as a research infrastructure CDS, the Virtual Observatory, astronomy, and beyond. *EPJ Web of Conferences*, vol. 186, 01001–2018. <https://doi.org/10.1051/epjconf/201818601001>
13. EOSC-Hub. <https://eosc-hub.eu/>
14. European Open Science Cloud. <https://www.eosc-portal.eu/>
15. Guidelines on Data Management in Horizon 2020. Directorate-General for Research & Innovation, EC. 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
16. Wittenburg, P.: From persistent identifiers to digital objects to make data science more efficient. *Data Intelligence*. Vol. 1, No 1. 2019. P. 6–21. https://doi.org/10.1162/dint_a_00004
17. Digital Object Interface Protocol Specification. Ver. 2.0. DONA Foundation. 2018. https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf
18. Regulations of CKP “Informatics”. <http://www.frccsc.ru/ckp>

**Выбор оптимального количества
процессорных ядер для алгоритма
многократной маркировки перколяционных
кластеров на высокопроизводительных
вычислительных системах**

С. Ю. Лапшина

МСЦ РАН

Аннотация. В статье рассматривается выбор оптимального количества запрашиваемых процессорных ядер для запуска алгоритма многократной маркировки перколяционных кластеров на современных суперкомпьютерных системах, установленных в Межведомственном суперкомпьютерном центре Российской академии наук.

Алгоритм может быть использован в любой области в качестве инструмента дифференцирования кластеров решетки большого размера, т. к. ему на вход подаются данные в формате, независимом от приложения. В МСЦ РАН этот инструмент был использован для изучения задачи распространения эпидемий, для чего была разработана соответствующая мультиагентная модель.

В ходе имитационных экспериментов применялся усовершенствованный для применения на многопроцессорной системе вариант алгоритма многократной маркировки перколяционных кластеров Хошена — Копельмана, связанный с механизмом линковки меток, который также может быть использован в любой области в качестве инструмента дифференцирования кластеров решетки большого размера.

Ключевые слова: мультиагентное моделирование, перколяционный кластер, механизм линковки меток, высокопроизводительные вычислительные системы, процессорные ядра.

The Optimal Processor Cores' Number choice for the Parallel Cluster Multiple Labeling Technique on High-Performance Computing Systems

S. Y. Lapshina

JSCC RAS

Abstract. The article is about the research of a optimum number of processor cores for launching the Parallel Cluster Multiple Labeling Technique on modern supercomputer systems installed in the JSCC RAS.

This technique may be used in any field as a tool for differentiating large lattice clusters, because it is given input in a format independent of the application. At the JSCC RAS, this tool was used to study the problem of the spread of epidemics, for which an appropriate multiagent model was developed.

In the course of imitation experiments, a variant of the Parallel Cluster Multiple Labeling Technique for percolation Hoshen-Kopelman clusters related to the tag linking mechanism, which can also be used in any area as a tool for differentiating large-size lattice clusters, was used to be improved on a multiprocessor system.

Key words: multi-agent simulation, percolation's cluster, parallel cluster multiple labeling technique, high-performance computing systems, processor cores.

Ситуация с распространением нового коронавируса COVID-19 показала, что в настоящее время инфекционные болезни остаются одной из ведущих причин преждевременной смерти людей. 30 января 2020 г. на заседании комитета по чрезвычайным ситуациям Всемирной организации здравоохранения (ВОЗ) вспышка COVID-19 была признана чрезвычайной ситуацией в области общественного здравоохранения, имеющей международное значение. 11 марта ВОЗ объявила эпидемию коронавируса пандемией.

На 14 июля 2020 года подтверждено 12 964 809 случаев заражения по всему миру, 570 288 летальных исходов [1]. В Китае — 85 623 случая, 4 649 летальных исходов. В США — 3 286 063 случая заражения, 134 704 летальных исхода. В Европе большинство случаев пришлось на Российскую Фе-

дерацию — 739 947 случаев заражения, 11 614 летальных исходов; Великобританию — 290 137 случаев заражения, 44 830 летальных исходов; Испанию — 255 953 случаев заражения, 28 406 летальных исходов.

При анализе распространения данной эпидемии (и эпидемий прошлых лет) замечена тенденция, что в определенный момент возникает скачкообразный переход в распространении эпидемии.

При моделировании процессов распространения массовых эпидемий и пандемий часто возникает эффект того, что незначительные изменения значений одного или нескольких параметров (например, вероятности инфицирования отдельных представителей) могут привести к скачкообразному изменению поведения всей популяции (болезнь из локальной и неопасной переходит в стадию широкомасштабной пандемии). Одним из способов исследования подобных эффектов является изучение формирования и роста *перколяционных кластеров*.

Исследования свойств перколяционных кластеров целесообразно сопровождать активным использованием разнообразных технологий компьютерного моделирования, в том числе и технологии мультиагентной имитации [2–4].

Принципиальной особенностью мультиагентной имитации по сравнению с другими технологиями (агрегатной, дискретно-событийной и т. д.) является возможность выявления и регистрации свойств поведения сложной системы под воздействием сугубо индивидуалистического характера поведения её представителей. Уточнение полученных результатов с целью совершенствования исходной модели порой требует проведения большого количества ресурсоёмких имитационных экспериментов (главным образом на многопроцессорных вычислительных архитектурах), использования специализированных алгоритмов распараллеливания и подбора оптимальных значений входных параметров [5].

В мультиагентной модели распространения эпидемий, разработанной в МСЦ РАН как часть программно-аппаратного комплекса BIOCLUST [6], применяется много-

процессорный алгоритм многократной маркировки перколяционных кластеров (ММПК) [7].

Важным моментом работы параллельного алгоритма ММПК является правильный подбор количества процессорных ядер, на которых будет производиться обработка исходной решетки.

В ходе работы алгоритма она загружается в оперативную память узла и, принимая во внимание ее большой размер, было бы логично проводить распараллеливание процесса ее обработки на большое количество частей.

Но с другой стороны в ходе работы алгоритма нужно проводить обмен данными между пограничными ячейками частей исходной решетки. И если таких частей будет слишком много, то время обмена данными может превысить отведенный лимит времени на обработку задания.

Проводя запуск алгоритма на суперкомпьютере, необходимо найти баланс между увеличением количества запрашиваемых вычислительных ядер и задержками, связанными с обменом данными между пограничными ячейками.

Для исследования оптимального количества запрашиваемых процессорных ядер для запуска алгоритма программа маркировки кластеров Load запускалась в МСЦ с входным параметром вероятности p от 0.01 до 1 с шагом в 0.01 при постоянных значениях модельного времени $t = 30$ дней на 48–304 процессорных ядрах на следующих суперкомпьютерах: МВС-10П ОП, МВС-10П МП2, МВС-10П Торнадо, МВС-100К.

МВС-10П ОП предоставляется пользователям Центра в режиме коллективного доступа к трём разделам: Haswell, Broadwell и Skylake:

- Haswell (42 вычислительных модуля на базе процессоров Intel Xeon E5-2697 v3, 128 ГБ оперативной памяти на модуль, пиковая производительность модуля — 1.1648 TFLOPS, 1176 ядер в разделе);

- Broadwell (136 вычислительных модулей на базе процессоров Intel Xeon E5-2697 v4, 128 ГБ оперативной памяти

на модуль, пиковая производительность модуля – 1.3312 TFLOPS, 4352 ядра в разделе);

- Skylake (58 вычислительных модулей на базе процессоров Intel Xeon Gold 6154, 192 ГБ оперативной памяти на модуль, пиковая производительность модуля – 3.456 TFLOPS, 2088 ядер в разделе).

Общим для установок на МВС-10П ОП является использование в качестве коммуникационной среды низколатентной сети Intel Omni-Path.

МВС-10П МП2 KNL – суперкомпьютер из 38 вычислительных модулей на базе процессоров Intel Xeon Phi 7290, 96 ГБ оперативной памяти на модуль, пиковая производительность модуля – 3.456 TFLOPS, 2736 ядер в системе.

МВС-10П Торнадо – суперкомпьютер из 207 вычислительных модулей, каждый модуль имеет в своем составе 2 процессора Xeon E5-2690, 64 ГБ оперативной памяти, два сопроцессора Intel Xeon Phi 7110X, пиковая производительность модуля – 371.2 GFLOPS, 3312 ядер в системе.

МВС-100К – суперкомпьютер из 110 вычислительных модулей на базе процессоров Intel Xeon E5450, 8 ГБ оперативной памяти на модуль, пиковая производительность модуля – 96 GFLOPS, 880 ядер в системе.

На МВС-10П ОП среднее время расчета составило:

- раздел Haswell – 360 сек;
- раздел Broadwell – 376 сек;
- раздел Skylake – 417 сек.

Минимальное время запуска:

- раздел Haswell – 322 сек на 128 ядрах;
- раздел Broadwell – 361 сек на 208 ядрах;
- раздел Skylake – 371 сек на 128 ядрах.

Среднее время расчета на МВС-10П МП2 KNL составило 1201 сек (почти в три раза больше, чем на любом из разделов МВС-10П ОП), минимальное время запуска – 1172 сек на 128 ядрах.

Среднее время расчета на МВС-10П Торнадо составило 263 сек (примерно на 25 % меньше, чем на любом из разделов МВС-10П ОП), минимальное время запуска — 235 сек на 160 ядрах.

Среднее время расчета на МВС-100К составило 570 сек (примерно на 50 % больше, чем на разделах МВС-10П ОП), минимальное время запуска — 480 сек на 128 ядрах.

На *рис. 1* показан сводный график зависимости времени работы программы Load от количества запрашиваемых процессорных ядер на основных системах МСЦ РАН. Минимальное время расчета показывает МВС-10П Торнадо. Для большинства суперкомпьютеров минимальное время счета достигается при использовании 128–208 ядер.

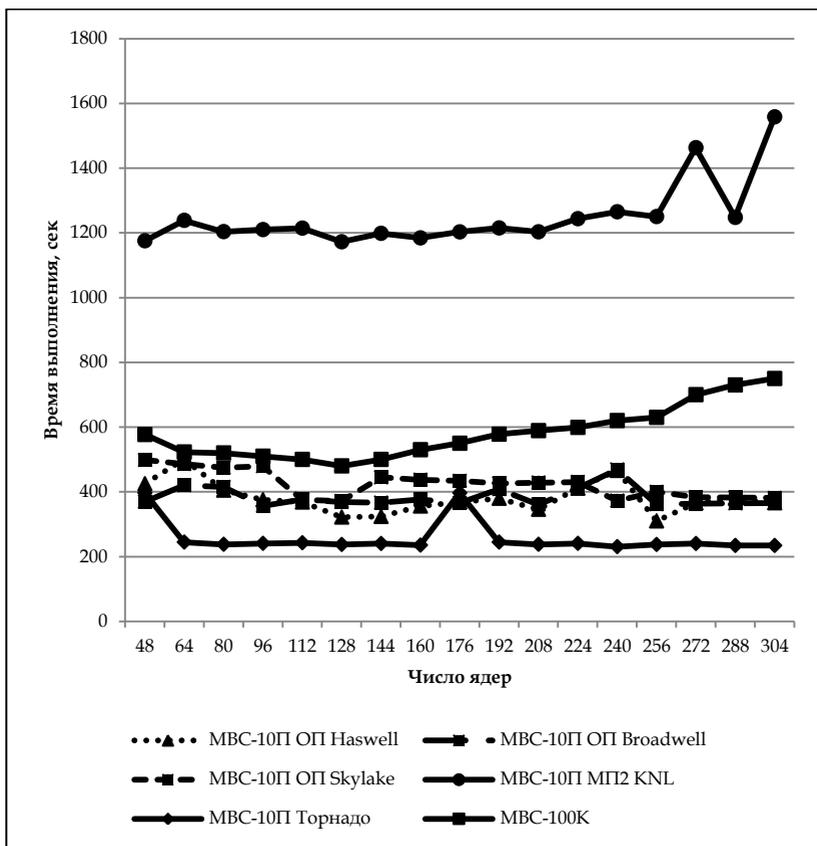


Рис. 1. Сводный график зависимости времени расчета от количества запрашиваемых процессорных ядер

Работа выполнена в МСЦ РАН в рамках государственного задания № 0580-2021-0016 и в рамках проекта РФФИ № 19-07-00861.

Расчеты проводились на высокопроизводительных вычислительных системах MVC-10П МП2 KNL, MVC-10П ОП, MVC 10П Торнадо, MVC-100К в Межведомственном суперкомпьютерном центре Российской академии наук.

Литература

1. World Health Organization Homepage <https://www.who.int/>, last accessed 2020/07/15
2. Калихман Р. С., Шебеко Ю. А. Моделирование роста перколяционных кластеров на компьютерах с сильно выраженной параллельной архитектурой // Сборник научных трудов РАН «Вычислительные технологии». Том 4. № 10. Сибирское отделение РАН. 1995 г.
3. Кондратьев М. А., Ивановский Р. И., Цыбалова Л. М. Применение агентного подхода к имитационному моделированию процесса распространения заболевания // Научно-технические ведомости СПбГУ/ Наука и образование. 2010. Т. 2. № 2(100). С. 189-195.
4. Тарасевич Ю. Ю. Перколяция: теория, приложения, алгоритмы. Учебное пособие. М.: УРСС, 2002. 64 с.
5. Утакаева И. Х. Имитационное моделирование распространения эпидемий на основе агентного подхода // Научный журнал КубГАУ. 2016. № 121(07) URL: <http://ej.kubagro.ru/2016/07/pdf/85.pdf> (дата обращения 09.10.2020).
6. S. Yu. Lapshina. High-Performance Computations in Multi-agent Simulation Problems of Percolation Cluster's Behavior // Lobachevskii Journal of mathematics. Volume 40, Issue 3 (2019), p. 341-348.
7. S. Yu. Lapshina. Parallel Cluster Multiple Labeling Technique // Lobachevskii Journal of Mathematics, Volume 40 Issue 5 (2019), p. 555-561.

References

1. World Health Organization Homepage <https://www.who.int/>, last accessed 2020/07/15
2. Kalihman R. S., Shebeko Yu A. Modelirovanie rosta perkolyacionnyh klasterov na kompyuterah s silno vyrazhennoj parallelnoj arhitekturoj // Sbornik nauchnyh trudov RAN Vychislitelnye tekhnologii. Tom 4. № 10. Sibirskoe otdelenie RAN. 1995 g.
3. Kondratev M. A., Ivanovskij R. I., Cybalova L. M. Primenenie agentnogo podhoda k imitacionnomu modelirovaniyu processa rasprostraneniya zabolevaniya // Nauchno-tekhnicheskie vedomosti. SPbGU. Nauka i obrazovanie. 2010. T 2. № 2(100). S 189-195.
4. Tarasevich Yu Yu. Perkolyaciya teoriya prilozheniya algoritmy // Uchebnoe posobie. M. URSS 2002. 64 s.
5. Utakaeva I. H. Imitacionnoe modelirovanie rasprostraneniya ehpidemij na osnove agentnogo podhoda // Nauchnyj zhurnal KubGAU. 2016. 121(07) URL: <http://ej.kubagro.ru/2016/07/pdf/85.pdf> (data obrashcheniya 09.10.2020).

6. S. Yu. Lapshina. High-Performance Computations in Multi-agent Simulation Problems of Percolation Cluster's Behavior // Lobachevskii Journal of mathematics. Volume 40, Issue 3 (2019), p. 341-348
7. S. Yu. Lapshina. Parallel Cluster Multiple Labeling Technique // Lobachevskii Journal of Mathematics, Volume 40 Issue 5 (2019), p. 555-561

Национальная исследовательская компьютерная сеть: инфраструктурно- сервисная основа единого цифрового пространства научных знаний

А. Г. Абрамов, А. В. Евсеев

Санкт-Петербургское отделение МСЦ РАН

А. А. Гончар

МСЦ РАН

Аннотация. В работе приведены общие сведения об архитектуре национальной исследовательской компьютерной сети нового поколения (НИКС), эксплуатируемой и развиваемой МСЦ РАН по заданию Минобрнауки России в качестве администратора и оператора с 2019 года. Обсуждены ключевые функции и направления функционирования НИКС сегодня и в ближайшей перспективе развития, некоторые актуальные задачи сферы науки и образования страны, требующие привлечения имеющихся инфраструктурных и сервисных возможностей сети, а также ее поступательной комплексной модернизации. Приведена информация о текущем состоянии в отношении географического и целевого пользовательского охвата НИКС внутри страны, сформированном заделе и потенциальном расширении направлений международного сотрудничества с зарубежными национальными научно-образовательными сетями и сетевыми консорциумами в интересах сферы науки и образования России. Обозначены предоставляемые НИКС возможности для совершенствования единого цифрового пространства научных знаний, повышения уровня доступности, информационной безопасности и сетевой защиты формирующих его ресурсов и сервисов. Представлены функционирующие на базе инфраструктуры НИКС отдельные специализированные научные сервисы, сервисы информационных технологий и сервисы коллективного пользования, построенные, в том числе на перспективных и массово эксплуатируемых технологиях федеративной аутентификации и развиваемые во взаимодействии с глобальным научно-образовательным сообществом.

Ключевые слова: национальная исследовательская компьютерная сеть нового поколения, НИКС, научно-образовательная сеть, ин-

фраструктурно-сервисная платформа, научные сервисы, федеративная аутентификация, AAI, RUNNetAAI, eduGAIN.

National Research Computer Network: infrastructure and service basis for a Common digital space of scientific knowledge

A. G. Abramov, A. V. Evseev

SPb branch of JSCC RAS

A. A. Gonchar

JSCC RAS

Abstract. The paper provides general information on the architecture of the new generation National Research Computer Network (NIKS), operated and developed by JSCC RAS on the assignment of the Ministry of Science and Higher Education of the Russian Federation as an administrator and operator since 2019. The key functions and directions of functioning of NIKS today and in the short term development, some urgent tasks of the sphere of science and education of the country, requiring the engaging of the existing infrastructure and service capabilities of the network, as well as its progressive comprehensive modernization. Information is provided on the current state of NIKS with the geographical and target user coverage within the country, the groundwork formed and the potential expansion of areas of international cooperation with foreign national research and education networks and network consortia in the interests of the sphere of science and education of Russia. The opportunities provided by NIKS for improvement of a Common digital space of scientific knowledge, increasing the level of accessibility, information security and network protection of its resources and services are denoted. Individual specialized scientific services, information technology services and services for collective use functioning on the basis of the NIKS infrastructure, built, including on promising and massively exploited technologies of federated authentication and developed in cooperation with the global research and education community are presented.

Keywords: new generation national research computer network, NIKS, research and education network, infrastructure and service platform, scientific services, federated authentication, AAI, RUNNetAAI, eduGAIN.

1. Введение

Решение весьма актуальной задачи по формированию базирующегося на передовых технологиях, высоко востребованного и массово доступного единого цифрового пространства научных знаний (ЕЦПНЗ) [1, 2] в значительной мере увязано с опережающим развитием глобальной отраслевой телекоммуникационной сети, эксплуатируемой и совершенствуемой в интересах сферы науки и образования, выступающей в качестве высокопроизводительной инфраструктурно-сервисной платформы для хранения, анализа и обмена научно-технической информацией при осуществлении научно-исследовательской деятельности, реализации совместных проектов с предоставлением надежного повсеместного доступа к цифровым научным и образовательным ресурсам в доверенном и защищенном сетевом периметре.

Соответствующие функции традиционно возлагаются на существующие в большинстве развитых стран мира национальные научно-образовательные сети (National Research and Education Network, NREN), которые отвечают также за обеспечение высокоскоростной сетевой связности научных и образовательных организаций своих стран, межсетевое взаимодействие с зарубежными NREN и сетевыми консорциумами с повышенными требованиями к качеству сервиса, доступ целевых пользователей в глобальное ИКТ-пространство, в публичные сети, развитие и предоставление пользователям специализированных научных сервисов, сервисов информационных технологий и коллективного пользования [3–5]. В связи с этим достаточно упомянуть хорошо известные в профессиональной среде научно-образовательные сетевые консорциумы GÉANT (Европа), NORDUnet (Скандинавские страны), национальные научно-образовательные сети Internet2 (США), DFN (Германия), SURFnet (Нидерланды), RENATER (Франция), GARR (Италия) и др.

Стоит заметить, что пользователями ведущих NREN, помимо научных и образовательных организаций разного уровня и ведомственной принадлежности, являются также

учреждения культуры и искусства, медицины и здравоохранения, включая библиотеки, музеи, архивы, медицинские научные и исследовательские институты и центры, больницы, а также издательства научной и учебной литературы как значимые источники данных для ЕЦПНЗ. Согласно сложившейся мировой практике функционирования и развития NREN, востребованные научные и образовательные ресурсы и сервисы преимущественно размещаются в инфраструктуре таких отраслевых сетей и в отдельных случаях доступны пользователям только при непосредственном подключении к ним.

В нашей стране основные функции NREN сегодня выполняет Национальная исследовательская компьютерная сеть нового поколения (НИКС, <https://www.niks.su>), созданная в 2019 г. по заданию Минобрнауки России в результате интеграции независимо эксплуатировавшихся на протяжении 25 лет специализированных сетей – федеральной университетской сети RUNNet и сети Российской академии наук RASNet [6–8]. В соответствии с поручением Минобрнауки России функции администратора и оператора НИКС на основании государственного задания выполняет Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН (МСЦ РАН).

2. НИКС: общие сведения и особенности архитектуры

В инфраструктурном отношении НИКС представляет собой глобальную географически распределенную гетерогенную сеть передачи данных федерального масштаба с присутствием по состоянию на 2020 г. в 34 субъектах страны. НИКС напрямую предоставляет телекоммуникационные услуги более 150 государственным организациям высшего образования и науки, в большинстве своем подведомственным Минобрнауки России [7, 8]. В обмене данными между собой и с зарубежными организациями (подключенными к своим локальным NREN) с использованием инфраструк-

турных возможностей НИКС по приблизительным оценкам участвуют более 300 профильных российских организаций.

К сети подключено 16 ведущих российских суперкомпьютерных центров (СКЦ) науки и образования, более 150 центров коллективного пользования научным оборудованием (ЦКП), более 100 уникальных научных установок (УНУ), с обеспечением безбарьерного доступа пользователей к таким объектам научной инфраструктуры коллективного пользования (при наличии принципиальной возможности).

На основе глобальной доверенной среды передачи данных реализуется сетевая связность участников научного, научно-технического и инновационного взаимодействия, в том числе федеральных органов управления наукой и образованием, научных организаций и образовательных организаций высшего образования, создаваемых в рамках Национального проекта «Наука» научно-образовательных центров мирового уровня (НОЦ), научных центров мирового уровня (НЦМУ), центров компетенции научно-технологической инициативы (ЦК НТИ).

В отношении текущего уровня технологического развития можно отметить, что магистральная инфраструктура НИКС простирается от Амстердама до Хабаровска, при этом типичная пропускная способность каналов связи внутри России составляет от 1 до 10 Гбит/с. Сеть имеет два независимых подключения к зарубежным NREN через инфраструктуру международных консорциумов GÉANT и NORDUnet с пропускной способностью по 10 Гбит/с [4–6].

Выстроенное на базе НИКС (и ее сетей-предшественниц) многолетнее устойчивое взаимодействие с европейскими межгосударственными консорциумами научно-образовательных сетей, исторически воспринимающих ее в качестве российской NREN, предоставляет российским организациям-пользователям возможности полноправного участия в международных исследовательских проектах, обеспечивает инфраструктурную поддержку совместных исследований, доступ через реализованную связность к исследовательским установкам класса «мегасайенс», вос-

требуемым ресурсам и сервисам NREN, предоставляет возможности для трансфера перспективных ИКТ.

В экосистеме НИКС расположены потенциальные источники цифрового контента для ЕЦПНЗ, включая многочисленные российские научные и образовательные веб-ресурсы и сервисы, отраслевые государственные информационные системы федерального и регионального уровней, платформы дистанционного обучения, реестры онлайн-курсов образовательных организаций, проблемно-ориентированные порталы, электронные библиотеки и архивы, тематические и междисциплинарные цифровые коллекции научных данных и др.

Одной из первоочередных задач в текущих реалиях является предоставление повсеместного защищенного доступа учащихся и преподавателей к интенсивно развиваемым и внедряемым в системе образования платформам онлайн-обучения, в том числе в рамках реализуемого в настоящее время глобального интегрирующего проекта ГИС «Современная цифровая образовательная среда» (СЦОС). НИКС должна отводиться первоочередная роль в части предоставления инфраструктурной платформы для организации доступа в защищенном сегменте к подобного рода объектам критической информационной инфраструктуры сферы образования и науки, федеральным и государственным информационным системам на основе сертифицированных решений с применением технологий VPN и современных протоколов шифрования сетевого трафика.

Архитектуру платформы НИКС, которую целесообразно рассматривать в качестве инфраструктурно-сервисной основы единой цифровой среды сферы науки и образования страны (*рис. 1*), можно представить, таким образом, в разрезе следующих ее ключевых компонентов и соответствующих им задач, ориентированных, в первую очередь, на удовлетворение постоянно растущих потребностей целевых пользователей и способствующих интеграции на базе сети потенциала российского научно-образовательного общества [7, 8]:

- сетевая связность участников со специальными требованиями, сервисы уровня инфраструктуры;
- доверенная среда передачи данных, безбарьерный доступ к объектам научной инфраструктуры коллективного пользования, цифровому научно-образовательному контенту;
- инфраструктура для цифровых платформ (ЦП) Минобрнауки России, отраслевых прикладных ИКТ-платформ, государственных информационных систем;
- содействие международному сотрудничеству в сфере образования, науки и технологий;
- информационная безопасность и федеративная аутентификация;
- базовые сетевые сервисы, сервисы телематических служб, сервисы информационных технологий.



Рис. 1. НИКС как инфраструктурно-сервисная основа единой цифровой среды сферы науки и образования

НИКС в рамках запланированной на ближайшие годы комплексной модернизации инфраструктуры и дальнейшей географической экспансии [7, 8] нацелена на подключение максимального числа ведущих научных и образовательных организаций страны, основных участников научного, науч-

но-технического и инновационного взаимодействия, размещение в своей доверенной и защищенной среде востребованных объектов научной инфраструктуры коллективного пользования в первую очередь отраслевых СКЦ (во взаимодействии с Национальной суперкомпьютерной инфраструктурой, в концепции создания и обеспечения функционирования которой закреплены соответствующие роли и функции НИКС).

3. Сервисная платформа НИКС

НИКС использует для решения собственных задач при эксплуатации сети, предоставляет своим пользователям, принимает участие во внедрении и развитии как «традиционных» сетевых, так и специализированных «нишевых» сервисов для образования и науки, которые можно классифицировать и подразделить следующим образом [9]:

- инфраструктурные сервисы;
- базовые сетевые сервисы и сервисы телематических служб;
- сервисы мониторинга и управления инфраструктурой;
- общие сервисы информационных технологий;
- специализированные сервисы для науки и образования;
- профессиональные сервисы.

В отношении построения отвечающего современным требованиям ЕЦПНЗ, объединяющего разнородные и расположенные в распределенной сетевой среде научные и образовательные ресурсы, особая роль, как представляется, должна быть отведена сервисам информационной и сетевой безопасности, сбора и анализа статистики использования ресурсов, а также федеративной аутентификации, авторизации и контроля доступа.

Технологии федеративной аутентификации (Authentication and Authorization Infrastructure, AAI) широко интегрированы в публичное Интернет-пространство и используются для бесшовного перехода между массово востребованными

ресурсами коллективного пользования общего назначения. Применительно к сфере науки и образования эти технологии решают задачу создания и обеспечения функционирования безбарьерной доверенной среды для предоставления студентам, преподавателям, научным сотрудникам, исследователям возможностей повсеместного доступа к научной и образовательной информации (в том числе с разными правами и в отсутствие жесткой «привязки» к IP-адресам конкретной организации) [10–14].

Соответствующие методики и технологии управления доступом развиты и успешно эксплуатируются в рамках удостоверяющих федераций, представляющих собой объединения сервис-провайдеров и провайдеров идентификации, присоединившихся к общему регламенту в целях безопасного обмена информацией о пользователях (при взаимодействии с системой управления учетными записями) и ресурсах и предоставления доступа к последним в соответствии с концепцией «единого окна» (Single Sign On, SSO) [10, 11]. При этом задействуются единые учетные данные пользователя с возможностью сквозной аутентификации на разных входящих в удостоверяющую федерацию ресурсах, которая производится однократно на узле идентификации «домашней» организации (с которой пользователь аффилирован).

Примерами предоставляемых крупными сервис-провайдерами ресурсов с возможностями федеративного доступа являются базы данных научного цитирования Scopus и Web of Science, полнотекстовые базы научной литературы и периодики Elsevier, Springer, EBSCO, Cambridge University Press, Wiley & Sons, IEEE и др., отдельные цифровые коллекции с результатами уникальных тематических научных исследований.

Специалистами НИКС создана национальная удостоверяющая федерация RUNNetAAI [13, 14], которая с 2018 г. участвует в международном проекте eduGAIN (<https://edugain.org>) – интерфедерации, объединяющей 70 национальных федераций по всему миру с 4 тыс. узлами провайдеров идентификации и более 3 тыс. узлами сервис-провайдеров [3]. В рамках RUNNetAAI реализован

и функционирует в интересах российских организаций-участниц федеративный доступ к Scopus и Web of Science, к системе «Антиплагиат» и ряду других ресурсов.

Следует заметить, что технологии ААИ позволяют, при налаживании обмена необходимой информацией между входящими в удостоверяющую федерацию провайдерами, формировать достоверную статистику использования ресурсов, к примеру, в рамках проекта национальной подписки на научные информационные ресурсы, обязанности оператора которой с 2020 г. возложены на РФФИ (вопросы участия фонда в RUNNetAAI находятся в стадии проработки).

В качестве других примеров сервисов информационных технологий и специализированных прикладных сервисов, в том числе с федеративным доступом, эксплуатируемых и развиваемых в настоящее время или планируемых к реализации на базе НИКС в ближайшей перспективе, можно указать на [6, 7, 9]:

— сервис международного роуминга в Wi-Fi сетях для научно-образовательного сообщества с едиными учетными данными и аутентификацией на стороне «домашней» организации в рамках проекта eduroam (<https://eduroam.ru>) [12, 13, 15];

— сервис безопасного доступа к НИКС и локальным сетям пользователей из публичных сетей на базе свободно распространяемого программного обеспечения (с учетом опыта международного проекта eduVPN, <https://eduvpn.org>);

— облачные сервисы коллективного пользования и совместной работы на базе свободно распространяемого программного обеспечения;

— облачные сервисы распределенной работы с научными данными (архивное и временное хранение, виртуализация, обмен данными, обработка и анализ);

— сервисы коммуникаций в режиме реального времени для проведения конференций, семинаров, персонального профессионального общения, научные социальные сети (ве-

бинары, ВКС, стриминговые трансляции, участие в международных проектах eduMEET и др.);

— сервисы визуализации и проведения численных расчетов — «научные блокноты» (Science Notebook as a Service, SNaaS).

Приведенный список не претендует на полноту — так, здесь оставлены без внимания перспективные (с учетом профильной деятельности и компетенций сотрудников МСЦ РАН) прикладные сервисы высокопроизводительных вычислений и искусственного интеллекта [16, 17], сервисы генерации и распространения цифровых сертификатов, разработки и тестирования на базе инфраструктуры НИКС новых ИКТ, сервисы в контексте набирающей популярность технологии блокчейн применительно к оптимизации издательского процесса, рецензирования научных статей, грантовых проектов, прикладной наукометрии и т. д.

Работа выполнена в МСЦ РАН в рамках государственного задания № 0580-2021-0014 и в рамках проекта РФФИ № 20-07-00773.

Литература

1. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук, 2019. Т. 89. № 7. С. 728-735.
2. Савин Г. И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России, 2020. № 5. С. 3-5.
3. GEANT Compendium of National Research and Education Networks in Europe — 2019 Edition. <https://compendium.geant.org>
4. Абрамов А. Г., Евсеев А. В. RUNNet как национальная научно-образовательная сеть России: цели, основные задачи, телекоммуникационная инфраструктура и сервисы // Информатизация образования и науки, 2018. № 4(40). С. 3-15.
5. Abramov A. G., Evseev A. V. RUNNet: infrastructural and service basis of the national research and education network of the Russian Federation // Proc. VIII Int. Conf. «Distributed Computing and Grid-technologies in Science and Education» (GRID 2018), Dubna, September 10-14, 2018. P. 52-57.
6. Абрамов А. Г., Евсеев А. В. Концептуальные аспекты создания в Российской Федерации национальной исследовательской компьютерной сети нового поколения // Информационные технологии, 2019. Т. 25. № 12. С. 724-733.

7. Концепция функционирования и развития Национальной исследовательской компьютерной сети нового поколения на 2020–2024 годы (проект). М.: МСЦ РАН, 2020. 47 с.
8. Абрамов А. Г., Гончар А. А., Евсеев А. В., Савин Г. И., Шабанов Б. М. Национальная исследовательская компьютерная сеть нового поколения: текущее состояние и концепция развития // Информационные технологии, 2020 (направлено в редакцию).
9. Абрамов А. Г. Сервисная платформа национальной исследовательской компьютерной сети нового поколения в контексте перспектив развития // Информатизация образования и науки, 2020. № 4(48) (принято к публикации).
10. Chadwick D. W. Federated Identity Management // In: Foundations of security analysis and design V. Springer Berlin Heidelberg. 2009. P. 96–120.
11. Bertino E., Takahashi K. Identity Management: Concepts, Technologies, and Systems. Artech House, 2011. 198 p.
12. Овсянников А. П., Савин Г. И., Шабанов Б. М. Удостоверяющие федерации научно-образовательных сетей // Программные продукты и системы, 2012. № 4. С. 3–7.
13. Абрамов А. Г., Васильев И. В., Порхачёв В. А. Развитие инфраструктуры аутентификации и авторизации для удостоверяющей федерации в рамках проектов eduGAIN и eduroam на базе сети RUNNet // ИТНОУ: Информационные технологии в науке, образовании и управлении, 2017. № 4. С. 56–64.
14. Абрамов А. Г., Васильев И. В., Порхачёв В. А. Принципы функционирования и управления удостоверяющей федерацией RUNNetAAI в рамках интерфедеративного взаимодействия с проектом eduGAIN // Информатизация образования и науки, 2019. № 2(42). С. 40–47.
15. Абрамов А. Г., Васильев И. В., Морин Ю. Н., Овсянников А. П., Порхачёв В. А. Вопросы совершенствования российского сегмента сервиса роуминга в беспроводных сетях eduroam в условиях интеграции научно-образовательных сетей RUNNet и RASNet // Труды НИИСИ РАН, 2019. Т. 9. № 6. С. 67–76.
16. Савин Г. И., Шабанов Б. М., Баранов А. В., Овсянников А. П., Гончар А. А. Об использовании федеральной научной телекоммуникационной инфраструктуры для суперкомпьютерных вычислений // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика, 2020. Т. 9. № 1. С. 20–35.
17. Шабанов Б. М., Овсянников А. П., Баранов А. В., Лещев С. А., Долгов Б. В., Дербьшев Д. Ю. Проект распределенной сети суперкомпьютерных центров коллективного пользования // Программные системы: теория и приложения, 2017. Т. 8. № 4(35). С. 245–262.

References

1. Antopolskii A. B., Kalenov N. E., Serebriakov V. A., Sotnikov A. N. O edinom tsifrovom prostranstve nauchnykh znaniy // Vestnik Rossiiskoi akademii nauk, 2019. T. 89. № 7. P. 728–735.
2. Savin G. I. Edinoe tsifrovoe prostranstvo nauchnykh znaniy: tseli i zadachi // Informatsionnye resursy Rossii, 2020. № 5. P. 3–5.
3. GÉANT Compendium of National Research and Education Networks in Europe – 2019 Edition. <https://compendium.geant.org>.
4. Abramov A. G., Evseev A. V. RUNNet kak natsionalnaia nauchno-obrazovatelnaia set Rossii: tseli, osnovnye zadachi, telekommunikatsionnaia infrastruktura i servisy // Informatizatsiia obrazovaniia i nauki, 2018. № 4(40). P. 3–15.
5. Abramov A. G., Evseev A. V. RUNNet: infrastructural and service basis of the national research and education network of the Russian Federation // Proc. VIII Int. Conf. «Distributed Computing and Grid-technologies in Science and Education» (GRID 2018), Dubna, September 10–14, 2018. P. 52–57.
6. Abramov A. G., Evseev A. V. Kontseptualnye aspekty sozdaniia v Rossiiskoi Federatsii natsionalnoi issledovatel'skoi kompiuternoi seti novogo pokoleniia // Informatsionnye tekhnologii, 2019. T. 25. № 12. P. 724–733.
7. Kontseptsiia funktsionirovaniia i razvitiia Natsionalnoi issledovatel'skoi kompiuternoi seti novogo pokoleniia na 2020–2024 gody (proekt). M.: MSTs RAN, 2020. 47 p.
8. Abramov A. G., Gonchar A. A., Evseev A. V., Savin G. I., Shabanov B. M. Natsionalnaia issledovatel'skaia kompiuternaia set novogo pokoleniia: tekushchee sostoianie i kontseptsiia razvitiia // Informatsionnye tekhnologii, 2020 (napravleno v redaktsiiu).
9. Abramov A. G. Servisnaia platforma natsionalnoi issledovatel'skoi kompiuternoi seti novogo pokoleniia v kontekste perspektiv razvitiia // Informatizatsiia obrazovaniia i nauki, 2020. № 4(48) (priniato k publikatsii).
10. Chadwick D. W. Federated Identity Management // In: Foundations of security analysis and design V. Springer Berlin Heidelberg. 2009. P. 96–120.
11. Bertino E., Takahashi K. Identity Management: Concepts, Technologies, and Systems. Artech House, 2011. 198 p.
12. Ovsyannikov A. P., Savin G. I., Shabanov B. M. Udostoveriaiushchie federatsii nauchno-obrazovatelnykh setei // Programmnye produkty i sistemy. 2012. № 4. P. 3–7.
13. Abramov A. G., Vasilyev I. V., Porkhachev V. A. Razvitie infrastruktury autentifikatsii i avtorizatsii dlia udostoveriaiushchei federatsii v ramkakh projektov eduGAIN i eduroam na baze seti RUNNet // ITNOU: Informatsionnye tekhnologii v nauke, obrazovanii i upravlenii, 2017. № 4. P. 56–64.

14. Abramov A. G., Vasilyev I. V., Porkhachev V. A. Printsipy funktsionirovaniia i upravleniia udostoveriaiushechei federatsiei RUNNetAAI v ramkakh interfederativnogo vzaimodeistviia s proektom eduGAIN // Informatizatsiia obrazovaniia i nauki, 2019. № 2(42). P. 40–47.

15. Abramov A. G., Vasilyev I. V., Morin Yu.N., Ovsyannikov A. P., Porkhachev V. A. Voprosy sovershenstvovaniia rossiiskogo segmenta servisa rouminga v besprovodnykh setiakh eduroam v usloviakh integratsii nauchno-obrazovatelnykh setei RUNNet i RASNet // Trudy NIISI RAN, 2019. T. 9. № 6. P. 67–76.

16. Savin G. I., Shabanov B. M., Baranov A. V., Ovsyannikov A. P., Gonchar A. A. Ob ispolzovanii federalnoi nauchnoi telekommunikatsionnoi infrastruktury dlia superkompiuternykh vychislenii // Vestnik Iuzhno-Uralskogo gosudarstvennogo universiteta. Serii: Vychislitelnaia matematika i informatika, 2020. T. 9. № 1. P. 20–35.

17. Shabanov B. M., Ovsyannikov A. P., Baranov A. V., Leshchev S. A., Dolgov B. V., Derbyshev D.Yu. Proekt raspredelennoi seti superkompiuternykh tsentrov kollektivnogo polzovaniia // Programmnye sistemy: teoriia i prilozheniia, 2017. T. 8. № 4(35). P. 245–262.

Презентация цифрового пространства знаний через сайт научной библиотеки

Т. Н. Харыбина, И. А. Митрошин, Е. В.

Бескаравайная

БЕН РАН

Аннотация. В статье представлена информация о сайте библиотеки, рассмотрены традиционные библиографические сервисы – оригинальные библиографические продукты, сложные электронные ресурсы – базы данных, сочетающие элементы библиографических, справочных, полнотекстовых материалов. Представлены данные об информационных продуктах научных библиотек, как звене единого научного информационного пространства. Представленная информация характеризует деятельность академических библиотек во внешней среде, которая раскрывает изменения, связанные с развитием информационных технологий: использование ресурсов открытого доступа и аналитики данных при обработке научной информации. Использование портала библиотеки облегчает ученым планирование, разработку, проведение и публикацию будущих исследований. Авторами выделены приоритеты в развитии портала и определены планы на будущее.

Ключевые слова: единое цифровое пространство, информационные технологии, научные библиотеки; цифровые библиотеки, электронные ресурсы, инновационные услуги, популяризация и продвижение цифровой библиотеки.

Presentation of the digital space of knowledge through the scientific library website

T. N. Kharybina, E. V. Beskaravaynaya, I. A. Mitroshin

LNS RAS

Abstract. The article reflects information on library websites, which is publicly available. Traditional bibliographic indexes are reviewed – original bibliographic products, complex electronic resources, databases that combine elements of bibliographic, reference, full-text materials. Data on information products of scientific libraries, as an element of a single information space. Briefly characterize the activities of academic libraries

in the external environment, reveal changes in the digital environment and social requirements associated with the development of IT: the use of open access technologies and data analytics in the processing of scientific data. In addition, the portal can make it easier for scientists to plan, design, conduct and publish future research. The authors highlighted development priorities and identified plans for the future.

Keywords: common digital space; information technology, digital libraries; electronic resources; innovative services; popularization, promotion, library site.

Первым шагом на пути создания единого информационного пространства, стала принятая в 1995 г. концепция «...формирования и развития России и соответствующих государственных информационных ресурсов (ЕРЭПЗ)» [1], подразумевающая оцифровку книжных, архивных и музейных фондов и создание электронных библиотек по разным отраслям знания и сферам творческой деятельности. На сегодняшний день уже сложилось четкое понимание организации этих процессов на основании практики, анализа, наполнения и поддержке контента [2–4]. Через призму имеющегося опыта мы приходим к пониманию, что процесс создания Единого информационного пространства (ЕИП) является многоуровневым, требующим технических, организационно-технологических и человеческих ресурсов.

Именно научная библиотека представляет собой платформу, максимально приближенную к потребностям пользователей, где имеются все необходимые элементы по взаимодействию с традиционным (печатным) информационным пространством, а теперь еще и с цифровым. На протяжении 20 лет в Центральной библиотеке (ЦБП ПНЦ РАН) Пущинского научного центра РАН, являющейся отделом Библиотеки по естественным наукам (БЕН) РАН разрабатывается комплексная система информационно-библиотечного обслуживания научных исследований, включающая все виды информационно-библиотечных услуг. Важным элементом в реализации данной системы стало внедрение автоматизации в деятельность БЕН и ЦБП: создание собственного сайта и баз данных собственной генерации. В 2001 г. сотрудники ЦБП получили свой первый грант Рос-

сийского фонда фундаментальных исследований (РФФИ) на «Создание комплексной автоматизированной информационной системы поддержки фундаментальной науки в ПНЦ РАН» [5–6]. В настоящее время особое место в данной работе занимает портал по физико-химической биологии, обеспечивающий получение исчерпывающей библиографической, реферативной, полнотекстовой, патентной и фактографической информации из ведущих отечественных и мировых информационных ресурсов в дистанционном режиме. В основу его создания легли принципы развития единого информационного пространства, включающие:

- хранение знаний (опубликованной в прошедшие годы научной информации);
- предоставление к ним свободного доступа к хранящимся знаниям;
- оперативное отражение новой информации;
- обеспечение многофункционального информационного сопровождения научных исследований;
- аналитическую работу.

Фонды централизованной библиотечной системы (ЦБС) ЦБП, в которую входит собственно центральная библиотека и её 7 филиалов в академических институтах ПНЦ РАН, насчитывают 770 000 экземпляров. Сводные электронные каталоги обеспечивают свободный доступ к размещенной в них информации из любой точки мира, а заказ материалов возможен по абонементу Центральной библиотеки посредством удаленного просмотра каталогов в рамках системы межбиблиотечного абонемента (МБА).

Накопление информационных ресурсов на сайте ЦБП происходит не только за счет приращения библиотечных фондов, но и за счет ресурсов собственной генерации — баз данных, отражающих наиболее достоверную и ценную в научном отношении информацию по тематике исследований обслуживаемых пользователей. В процесс её отбора вовлечены библиотекари и учёные, что исключает проникновение информационного шума. Эти ресурсы удобны для текущего и ретроспективного поиска и служат надёжным

подспорьем в исследовательской работе научных сотрудников ПНЦ. Кроме того, проблемно-ориентированные БД, созданные в ЦБП, рассматриваются как объект информационного анализа для выявления тенденций развития отдельных научных направлений ПНЦ и их прогнозирования.

Для сохранения и представления научного потенциала НИИ ПНЦ РАН в информационном пространстве в библиотеке созданы и поддерживаются БД трудов сотрудников ПНЦ РАН; БД патентов; БД диссертаций по физико-химической биологии, обновляющиеся по мере поступления новой информации.

Еще одно важное направление при формировании структуры портала ЦБП – это оперативное отражение изменений в области физико-химической биологии через предоставление информации о новых поступлениях материалов в фонды ЦБС ПНЦ и электронных тематических выставках.

В настоящее время важнейшим направлением в деятельности Российской академии наук является координация усилий по научному просвещению, распространению и пропаганде научных знаний с привлечением сил и возможностей научно-исследовательских институтов, библиотек и других учреждений

В рамках этого направления, начиная с 2013 года, сотрудники ЦБП занимались масштабным проектом «Научные Школы академического исследовательского Центра и их вклад в отечественную фундаментальную науку» на примере Пущинского научного Центра РАН. На основании опубликованных материалов, архивных источников и библиометрического анализа научных трудов, была собрана информация о 30 научных Школах 9 институтов ПНЦ РАН. Были проанализированы и данные о публикационной и патентной активности; цитируемости; о полученных грантах, премиях и других наградах; сведения о международном сотрудничестве; о работе с научной молодежью; признании достижений Школ за рубежом, перспективах развития [7]. Результаты исследования размещены на сайте библиотеки и получили высокую оценку администрации Центра, как «...информационный ресурс популяриза-

ции науки, служащий ориентиром в осмыслении сложной и многоплановой истории академических научных Школ».

Другим эффективным проектом ЦБП стала база персональных бывших сотрудников ПНЦ РАН, проживающих ныне за рубежом. При поддержке гранта РФФИ (проект № 16-06-00297-а — «Российская научная диаспора академического исследовательского центра: её вклад, роль и место в российской науке»), нами был собран материал, содержащий сведения о сфере научной деятельности учёного, времени отъезда, месте жительства, области научных интересов, публикационной активности, цитируемости [8]. Визуализация этого исследования через портал призвана служить налаживанию контактов и организации совместных исследований между представителями зарубежных научных диаспор и учёными России, привлечению ученых-соотечественников к научно-профессиональным контактам.

Сопровождение научных исследований со стороны библиотеки включает создание электронных путеводителей по интернет-ресурсам. Данный вид деятельности является важным инновационным направлением в работе библиотек, способным облегчить труд пользователей глобальной сети [9]. Более того, в условиях, когда документ библиотекой не приобретается и не сохраняется на ее территории, отражение его в навигационно-поисковом аппарате становится важнейшим элементом в обеспечении доступности для пользователей библиотеки и может рассматриваться, как включение в пространство знаний отдельного документа [10].

На сайте ЦБП мы предприняли попытку создания подобного рода электронных путеводителей. Для упрощения поиска необходимой информации ресурсы сгруппированы по предметным направлениям в области физико-химической биологии, как имеющиеся в доступе, так и с возможным доступом в будущем. Раздел портала «Ресурсы по физико-химической биологии» содержит информацию о различных материалах со ссылками, описанием и вариантами доступа к ним.

Библиометрические исследования проводятся в ЦБП с начала 1980-х гг. За прошедший период полученные ре-

зультаты нашли применение как для обеспечения технологических процессов ЦБП, так и в качестве услуги библиотеки по предоставлению информационно-аналитических данных различным категориям пользователей [11–12]. Таким образом, еще одним направлением развития библиотечного портала является визуализация библиометрического анализа публикационной активности на основании данных из наиболее авторитетных и исчерпывающих информационных зарубежных (Web of Science, Scopus, MEDLINE) и отечественных (РИНЦ) ресурсов. Результаты аналитической деятельности институтов ПНЦ обновляются на сайте 1–2 раза в год и служат основой формирования предложений по оптимизации научной деятельности. Так, например, анализ быстроцитируемых работ отражает заинтересованность научной общественности тематикой статьи и задаёт направление предстоящих исследований института. Анализ цитирования отдельными организациями или учеными позволяет выявить возможных научных партнеров, заложить основы будущих совместных мероприятий. Изучение публикаций с точки зрения грантовой поддержки позволяет выявить тематики, получающие наибольшее финансирование.

В 2019 году, в рамках аналитической деятельности, сотрудниками библиотеки был проведен анализ научного потенциала трех НИИ ПНЦ РАН и собственно Федерального исследовательского центра (ФИЦ) «Пушкинский научный центр биологических исследований Российской академии наук». Результаты исследования публикационной активности учёных ПНЦ представлены на сайте в виде кратких содержательных справок о количестве публикаций, цитировании, зарегистрированных патентах, основных научных направлениях, полученных грантах, российском и международном сотрудничестве (<https://cnbp.ru/informatsiya/bibliometricheskij-analiz.html>).

Для постоянного информирования посетителей сайта библиотеки поддерживаются разделы с новостями (собственными и из других библиотек, издательств и пр.) и выставками (электронный вариант выставок, проводимых в ЦБП и её филиалах).

Исследование современных программных средств и изменившихся потребностей пользователей привело к решению об оптимизации в 2019 году существующего портала. Основной объем информации был перенесен со старого сайта и дополнен уже в новом формате. Кроме того, сейчас на сайте появился раздел с методическими материалами, где рассматриваются разработки ведущих специалистов в информационно-библиотечной отрасли, материалы и рекомендации издательств, юридических центров. Отдельно стоит отметить новую возможность пользователей сайта быстро связаться по интересующим его вопросам с сотрудниками библиотеки. Для этого на странице портала ЦБП теперь можно оставить заявку в специальной форме и отправить ее на электронную почту ответственному сотруднику данного сектора. Наиболее востребованными на практике оказываются услуги по оказанию консультативной, методической и информационной помощи молодым учёным, подготовка и проведение совместных конференций в области физико-химической биологии, формирование и ведение банков данных по биологической тематике, библиометрический анализ и патентные исследования [13]. Формы обратной связи, имеющиеся на портале, значительно повышают качество и скорость обслуживания посетителей. Так, с введением в эксплуатацию сайта библиотеки появились запросы не только от сотрудников обслуживаемых организаций, но и от посетителей сайта из других регионов РФ (Самарская область, Уральский федеральный округ и др.) и ряда других государств (Латвия, Румыния и др.). Это говорит о популярности современного портала ЦБП и его востребованности у различных групп пользователей.

В планах библиотеки по развитию библиотечно-информационного портала к концу 2020 г. стоит создание личного кабинета (ЛК) ученого, включающего информационное сопровождение научных исследований в режиме избирательного распространения информации (ИРИ). Планируется ввести интерактивные сервисы, обеспечивающие корректировку запросов по заданному научному направлению, и форму обратной связи с возможностью оценки

со стороны пользователей качества предоставляемых ресурсов.

Имея опыт в создании специализированного сайта и оценку его комфортности со стороны пользователей, сотрудники библиотеки видят наиболее важными ориентирами при дальнейшей разработке портала включение его в Единое Цифровое Пространство Научных Знаний и обеспечение следующих функций:

- избирательное наполнение сайта (контент),
- индексация сайта и работа по его продвижению в поисковых системах,
- размещение ссылок на сайт на информационных порталах,
- создание страниц в социальных сетях,
- информационные email-рассылки,
- обеспечение обратной связи с пользователями,
- регулярное выполнение аналитики сайта.

Согласно данным статистики, уже сегодня мы можем говорить о том, что с введением в эксплуатацию новой версии сайта было отмечено резкое увеличение количества его посетителей. Нами исследовались статистические данные посещаемости, собранные при помощи счетчика Hotlog, и полученные при помощи Яндекс-метрики. На текущий момент количество уникальных посетителей увеличилось примерно в 3 раза в сравнении с прошлой версией сайта.

В заключение хочется отметить, что с развитием единого информационного пространства развивающемуся portalу необходим интерфейс, соответствующий современным запросам и пожеланиям своих пользователей, отвечающий критериям целесообразности выбора актуальных ресурсов, скорости наполнения и обновления научной информации.

Литература

1. Концепция формирования и развития единого информационного пространства России и соответствующих государственных информационных ресурсов // Межотрасл. информ. служба. М., 1995. № 3. С. 3-8.

2. Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.
3. Шрайберг Я. Л., Цветкова В. А., Маршак Б. И. Особенности разработки и реализации крупной информационной системы национального масштаба в сфере образования и науки // Научно-техническая информация. Сер. 1. Орг. и методика инф. работы. 2014. № 11. С. 16–21.
4. Маркина Ю. В., Денисова Т. В. Формирование системы информационного обеспечения инновационного развития экономики региона // Вестник Южно-Уральского государственного университета. Серия: Экономика и менеджмент. 2017. Т. 11. № 2. С. 22–27.
5. Мохначева Ю. В., Слащева Н. А., Харыбина Т. Н., Кудеярова А. Ю. Определение информационной значимости электронных ресурсов для научных исследований // XI Конференция представителей региональных научно-образовательных сетей Relarn-2004, 30.05–06.06.2004 г. : тезисы докл. 2004. С. 55–57.
6. Слащева Н. А. Создание комплексной системы информационно-библиографического обслуживания специалистов Научного центра на базе современных технологий (На примере Пущинского научного центра РАН) : дис. ... канд. пед. наук : 05.25.03. Москва, 1999. 125 с.
7. Бескаравайная Е. В., Харыбина Т. Н., Мохначева Ю. В., Слащева Н. А. Изучение становления и развития научных школ в Пущинском научном центре РАН (ПНЦ РАН) // Информационное Обеспечение Науки. Новые Технологии. 2012. С. 110–120.
8. Мохначева Ю. В., Бескаравайная Е. В., Харыбина Т. Н. Представленность и профессиональная успешность научных диаспор Пущинского научного центра РАН за рубежом // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2016. № 12. С. 9–33.
9. Н. И. Гендина, Н. И. Колкова. Библиотека в едином информационном пространстве: необходимость создания электронных путеводителей по интернет-ресурсам. // Научные и технические библиотеки. 2018. № 7. С. 43–59.
10. Каленов Н. Е., Харыбина Т. Н. Отделение БЕН РАН в Пущинском научном центре как звено корпоративной системы информационного сопровождения науки // Вестник Российской академии наук. 2019. Т. 89, № 3. С. 255–259.
11. Мохначева Ю. В., Харыбина Т. Н. Использование метода цитат-анализа для оптимизации научных исследований (на примере библиотеки академического НИИ) // Библиотекосведение. 2009. № 2. С. 125–128.
12. Мохначева Ю. В., Харыбина Т. Н. Сравнительная оценка научной продуктивности исследовательских учреждений РАН и сектора

российской высшей школы по некоторым библиометрическим индикаторам (2000–2009 гг.) // Библиосфера. 2011. № 3. С. 57–64.

13. Митрошин И. А. Патентное информирование в научных библиотеках // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Н. Е. Каленов, В. А. Цветкова (ред.). 2017. С. 130–141.

References

1. Kontseptsija formirovanija i razvitija edinogo informatsionnogo prostranstva Rossii i sootvetstvujuschikh gosudarstvennykh informatsionnykh resursov // Mezhotrasl. inform. sluzhba. M., 1995. № 3. S. 3–8.

2. Antopol'skij A. B., Kalenov N. E., Serebrjakov V. A., Sotnikov A. N. O edinom tsifrovom prostranstve nauchnykh znanij // Vestnik Rossijskoj akademii nauk. 2019. T. 89. № 7. S. 728–735.

3. SHrajberg Я. L., TSvetkova V. A., Marshak B. I. Osobennosti razrabotki i realizatsii krupnoj informatsionnoj sistemy natsional'nogo masshtaba v sfere obrazovanija i nauki // Nauchno-tehnicheskaja informatsija. Ser. 1. Org. i metodika inf. raboty. 2014. № 11. S. 16–21.

4. Markina JU. V., Denisova T. V. Formirovanie sistemy informatsionnogo obespechenija innovatsionnogo razvitija ekonomiki regiona // Vestnik JUzhno-Ural'skogo gosudarstvennogo universiteta. Serija: Ekonomika i menedzhment. 2017. T. 11. № 2. S. 22–27.

5. Mokhnacheva JU. V., Slasheva N. A., KHarybina T. N., Kudejarova A. JU. Opredelenie informatsionnoj znachimosti elektronnykh resursov dlja nauchnykh issledovanij // XI Konferentsija predstavitelej regional'nykh nauchno-obrazovatel'nykh setej Relarn-2004, 30.05–06.06.2004 g. : tezisy dokl. 2004. S. 55–57.

6. Slasheva N. A. Sozdanie kompleksnoj sistemy informatsionno-bibliograficheskogo obsluzhivaniya spetsialistov Nauchnogo tsentra na baze sovremennykh tekhnologij (Na primere Puschinskogo nauchnogo tsentra RAN) : dis. ... kand. ped. nauk : 05.25.03. Moskva, 1999. 125 s.

7. Beskaravajnaja Elena Vjacheslavovna, KHarybina Tat'jana Nikolaevna, Mokhnacheva Julija Valer'evna, Slasheva Natal'ja Anatol'evna Izuchenie stanovlenija i razvitija nauchnykh shkol v Puschinskom nauchnom tsentre RAN (PNTS RAN) // Informatsionnoe Obespechenie Nauki. Novye Tekhnologii. 2012. S. 110–120.

8. Mokhnacheva JU. V., Beskaravajnaja E. V., KHarybina T. N. Predstavlennost' i professional'naja uspeshnost' nauchnykh diaspor Puschinskogo nauchnogo tsentra RAN za rubezhom // Nauchno-tehnicheskaja informatsija. Serija 1: Organizatsija i metodika informatsionnoj raboty. 2016. № 12. S. 9–33.

9. N. I. Gendina, N. I. Kolkova. Biblioteka v edinom informatsionnom prostranstve: neobkhodimost' sozdaniya elektronnykh putevoditelej po internet-resursam. // Nauchnye i tekhnicheskie biblioteki. 2018. № 7. S. 43–59.

10. Kalenov N. E., KHarybina T. N. Otdelenie BEN RAN v Puschinskom nauchnom tsentre kak zveno korporativnoj sistemy informatsionnogo soprovozhdenija nauki // Vestnik Rossijskoj akademii nauk. 2019. T. 89. № 3. S. 255–259.

11. Mokhnacheva JU. V., KHarybina T. N. Ispol'zovanie metoda tsitat-analiza dlja optimizatsii nauchnykh issledovanij (na primere biblioteki akademicheskogo NII) // Bibliotekovedenie. 2009. № 2. S. 125–128.

12. Mokhnacheva JU. V., KHarybina T. N. Sravnitel'naja otsenka nauchnoj produktivnosti issledovatel'skikh uchrezhdenij RAN i sektora rossijskoj vysshej shkoly po nekotorym bibliometricheskim indikatoram (2000–2009 gg.) // Bibliosfera. 2011. № 3. S. 57–64.

13. Mitroshin I. A. Patentnoe informirovanie v nauchnykh bibliotekakh // Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov / N. E. Kalenov, V. A. TSvetkova (red.). 2017. S. 130–141.

Портал БЕН РАН как источник научных знаний

Т. Н. Соловьева

БЕН РАН

Аннотация. В статье рассматривается наполнение и структура сайта БЕН РАН и, в первую очередь, те особенности, которые определяют сайт как портал БЕН РАН. Рассматривается новая версия сайта, созданная специально для портала БЕН РАН. Подчеркнуты его особенности и отличия от сайтов научных библиотек. Рассмотрены разделы сайта: Каталоги (сводные интернет-каталоги периодических и неперіодических изданий), научные семинары (Семинар «Информационное обеспечение науки: новые технологии» и научно-практические обучающие семинары), «Метауказатели» (Астрономия, Биология, Математика, Механика, Наука о Земле, Нанотехнология, Физика, Химия, Экология и Общенаучная информация в Интернет). А также рассмотрен указатель «Серіальные издания», который содержит описания и ссылки на сайты, поддерживающие электронные периодические издания по основным разделам естественных наук. Рассмотрен типовой сайт, созданный для научных библиотек, а также рассмотрены тематические сайты-порталы библиотек, входящих в сеть БЕН РАН. По каждому разделу в статье дается подробное описание данного ресурса, отражающее особенности, характерные для него именно как раздела портала БЕН РАН.

Ключевые слова: сайт, портал, электронный каталог, указатель, БЕН РАН.

LNS RAS portal as a source of scientific knowledge

T. N. Solovyeva

LNS RAS

Abstract. The article discusses the content and structure of the site of the LNS RAS and, first of all, the features that define the site as a portal of the LNS RAS. A new version of the site created specifically for the LNS RAS portal is being considered. Its features and differences from the sites of scientific libraries are emphasized. The following sections of the site are considered: Catalogs (consolidated online catalogs of periodical and non-periodical publications), Scientific seminars (Seminar «Information support of science: new technologies» and Scientific and practical training seminars),

«Meta-indicators» (Astronomy, Biology, Mathematics, Mechanics, Earth science, Nanotechnology, Physics, Chemistry, Ecology, and General Scientific information on the Internet). The index «serial publications» is also considered, which contains descriptions and links to sites that support electronic periodicals on the main sections of natural sciences. A typical site created for scientific libraries is considered, as well as thematic sites — portals of libraries that are part of the LNS RAS network. For each section, the article provides a detailed description of this resource, reflecting the features that are characteristic of it as a section of the LNS RAS portal.

Keywords: website, portal, electronic catalog, index, LNS RAS.

С самого начала работы с Интернет (1996 г.) Библиотека по естественным наукам (БЕН) РАН ставила своей задачей предоставить своим пользователям не только традиционные для библиотек ресурсы (фонды), но все многообразие информации в области естественных наук, имеющейся в сети. Так в начале 2000-х годов Библиотекой был реализован проект при поддержке РФФИ (грант № 03-07-90055) о создании на базе сайта БЕН РАН портала «Естественные науки в Интернет» [1]. Целью этого проекта является создание Интернет-ресурса, объединяющего различную информацию в области естественных наук и предоставляющего исследователю возможность получить доступ к интересующим его сведениям из одной точки — с сайта БЕН РАН. Для портала была создана новая версия сайта [2, 4], которая используется и в настоящее время. Для сайта принят функциональный стиль оформления с минимумом графических элементов. Наполнение сайта многообразной информацией в области естественных наук требует четкой структуризации как сайта в целом, так и информационного поля главной страницы, чтобы обеспечить тем самым оптимальную навигацию по сайту. Рубрики меню, представленные на главной странице, объединены в лаконичную структуру, состоящую из нескольких основных разделов: «Электронные ресурсы БЕН РАН», «Внешние ресурсы» и блок объявлений. На основе этой структуры строится информационное пространство главной страницы. Использование современных технологий Web-дизайна при формировании страниц сайта и в первую очередь главной страницы позволяют экономить

визуальное пространство страниц, для размещения необходимой информации.

Исходя из основной задачи, стоящей перед Библиотекой (информационное обеспечение научных исследований), сайт БЕН РАН ориентирован, главным образом, на пользователей из академической среды, а именно на научных работников институтов и научных центров, библиотеки которых входят в сеть БЕН РАН.

БЕН РАН является центральной библиотекой Централизованной библиотечной системой (ЦБС БЕН РАН) [3] в состав которой входят более 50-ти библиотек (список представлен на сайте Библиотеки (<http://www.benran.ru>) в разделе «Отделы в институтах РАН») академических институтов и научных центров Москвы и Подмосковья. Список библиотек, представленный на сайте в этом разделе, снабжен ссылками на сайты этих библиотек.

Для научных библиотек (НБ), входящих в сеть БЕН РАН, ее специалистами был разработан типовой сайт НБ [5], входящий в структуру Библиотеки. Типовой сайт ориентирован на удовлетворение информационных потребностей определенного научного коллектива – сотрудников научного учреждения, обслуживаемого конкретной библиотекой. Основная функция сайта – позволить пользователю виртуально посещать библиотеку, как в информационных, так и в справочных целях. Т.е. сайт НБ выполняет в первую очередь функцию так называемой электронной доски объявлений и позволяет удаленно обращаться в библиотеку как с конкретными информационными, так и справочными запросами. Основные элементы наполнения типового сайта НБ группируются в два блока: сведения о библиотеке, информация о ресурсах библиотеки.

Часть библиотек ЦБС предпочла типовому сайту сайт, разработанный своими специалистами со своей уникальной структурой. Некоторые из них по разнообразию и количеству ресурсов, представленных на этом сайте, по сути, выполняют роль портала в определенной области естественных наук (в соответствии с научной тематикой обслуживаемого института). К таким сайтам, к примеру, можно отнести сайт «Цен-

тральная библиотека в Пушкинском научном центре» [6] и сайт «Библиотека в Математическом институте им. В. А. Стеклова РАН».

Таким образом, с сайта БЕН РАН по ссылкам можно перейти на сайт любой из библиотек ЦБС и получить либо справочную информацию (типовой сайт НБ), либо углубленную информацию в определенной области естественных наук (тематические сайты порталы). Такой подход демонстрирует основной принцип портала БЕН РАН — возможность получить доступ к интересующим его сведениям из одной точки — с сайта БЕН РАН.

По принципу доступа из одной точки организованы многие ресурсы, представленные на сайте БЕН РАН. Одними из таких ресурсов являются сводные каталоги ЦБС БЕН РАН.

БЕН РАН, обладающая фондом естественно-научной литературы, поддерживает несколько сводных каталогов ЦБС БЕН РАН (каталоги периодических и непериодических изданий). БЕН РАН всегда придавала большое значение задачам по созданию сводных электронных каталогов и отображению их на своем сайте, таким образом, чтобы обеспечить максимальное удобство работы с ними в удаленном режиме для пользователей сети Интернет.

Хотя фонд БЕН РАН включает как периодические, так и непериодические издания, но, как показывают опросы, неоднократно проводившиеся БЕН РАН среди сотрудников академических институтов, обслуживаемых ЦБС БЕН РАН, именно научные зарубежные журналы являются для них самым важным и необходимым источником информации.

В соответствии с этой особенностью информационных потребностей абонентов Библиотеки основную часть фонда БЕН РАН составляют научные журналы, и, соответственно, большая часть из них — зарубежные журналы. Поэтому и при создании электронных каталогов первоочередное внимание уделялось сводному Интернет-каталогу журналов (ИКЖ) [7].

На данный момент ИКЖ БЕН РАН отражает около 12000 названий журналов.

Из-за отсутствия финансирования комплектования академических библиотек, печатная часть журнального фонда, практически, «заморожена». Благодаря государственной программе национальной электронной подписки виртуальная часть журнального фонда БЕН РАН регулярно пополняется и включает на данный момент большой список зарубежных журналов ведущих издательств мира и поэтому остается и в настоящее время актуальной для пользователей Библиотеки. Доступ к этому фонду пользователи могут получить как в центральной библиотеке, так и через библиотеки ЦБС БЕН РАН.

Благодаря использованию уникальной технологии, разработанной специалистами БЕН РАН, в ИКЖ отражены не только журналы, выписанные непосредственно Библиотекой, но и выборочная часть журналов национальной электронной подписки, наиболее соответствующая естественно-научному профилю БЕН РАН (порядка 3000 названий зарубежных журналов). Кроме того, упомянутой технологией предусмотрено формирование отсылок со страниц журналов ИКЖ на соответствующие страницы журналов на сайтах издательств. Таким путем организован доступ из библиотек ЦБС БЕН РАН через ИКЖ непосредственно к электронным полнотекстовым версиям журналов на сайтах издающих организаций, что является безусловным удобством для пользователей ЦБС БЕН РАН. Пользователи из любой точки Сети могут также воспользоваться ссылками ИКЖ и получать информацию о журналах на сайтах издателей в пределах установленных на этих сайтах ограничений.

Ссылками также обладает и часть журналов ИКЖ, не входящих в национальную электронную подписку. По таким ссылкам можно получить доступ к архивной информации (статьи прошлых лет), что особенно существенно для пользователей с учетом того, что некоторые журналы открывают свободный доступ к полным текстам через несколько лет после выхода номера. Таким образом, учитывая доступ по программе национальной электронной подписки, расширяется диапазон доступной пользователям полнотекстовой информации.

Если сравнивать предоставление аналогичной информации на сайтах большинства научных библиотек, то, как правило, в электронных каталогах отражается печатный фонд, а электронные журналы, доступные пользователям предоставляются в виде списков, расположенных на сайтах библиотек отдельно, что, по нашему мнению, менее удобно в работе с журналами.

Вторым по значимости каталогом на сайте БЕН РАН является сводный интернет-каталог книг и продолжающихся изданий [8], который содержит расширенные библиографические описания изданий (монографии, сборники статей, материалы конференций, труды организаций и научных обществ), поступающих в ЦБС БЕН РАН. В каталоге также отражены специальные выпуски научных журналов, посвященные конкретным событиям или персонам. Каталог позволяет осуществлять поиск по запросам повышенной сложности по различным элементам данных с логическими связками. Предусмотрена возможность поиска информации из фонда конкретной библиотеки, выбранной из списка. Часть изданий каталога снабжена ссылками на электронные версии. Это касается изданий, которые либо приобретены БЕН РАН у правообладателей, либо легитимно размещены в свободном доступе в Интернет.

Как показывают аналитические данные, собранные разработчиками каталога и опубликованные в источнике [8], запросы в каталоге книг и продолжающихся изданий поступают с различных IP-адресов и охватывают всю Россию. Кроме того, к каталогу БЕН РАН обращаются пользователи из стран «ближнего и дальнего зарубежья». Результаты анализа показывают, что интернет-каталог книг и продолжающихся изданий БЕН РАН достаточно широко востребован и не только пользователями академической среды, на которую он был ориентирован.

На сайте БЕН РАН, кроме представленных выше каталогов, имеется Имидж-каталог отечественных книг и Имидж-каталог зарубежных книг. Все Интернет-каталоги БЕН РАН имеют русскоязычную и англоязычную версии. Кроме того, на сайте имеются ссылки и на другие Интернет каталоги ве-

дущих научных библиотек России и мира. Они объединены в списки, представленные на странице сайта «Ресурсы центральных библиотек». На данной странице представлены списки: «Ресурсы центральных библиотек России», «Каталоги зарубежных библиотек», «Списки Интернет-каталогов библиотек мира».

«Метауказатели»

Одним из относительно новых, возникших с появлением Интернета, видов деятельности Библиотеки является помощь своим пользователям в навигации по ресурсам Сети. Применительно к БЕН РАН, это поиск сведений о научных ресурсах в области естественных наук, представленных в сети Интернет, и создание соответствующих указателей этих ресурсов. Такие указатели отражены на сайте в разделе меню «Естественные науки в Интернет» [9]. Этот раздел содержит список «метауказателей» по основным направлениям естественных наук. Мы рассматриваем этот ресурс как указатели, ссылающиеся, в свою очередь, на указатели ресурсов, созданные ведущими специалистами в основных областях естественных наук (в научных институтах, университетах, библиотеках по всему миру). Использование такой технологии с одной стороны существенно расширяет количество ссылок, доступных в итоге пользователю (каждая ссылка раскрывает новый перечень ссылок), а с другой стороны, до определенной степени гарантирует «авторитетность» указателя (каждая страница, открываемая по ссылке из него, составлена компетентными специалистами соответствующей области науки).

На сегодняшний день на сайте поддерживаются 10 указателей по основным областям естественных наук (Астрономия, Биология, Математика, Механика, Наука о Земле, Нанотехнология, Физика, Химия, Экология и Общенаучная информация в Интернет).

В систему указателей входит также и указатель «Серийные издания». Этот указатель содержит аннотированные ссылки на сайты, поддерживающие информацию

об электронных периодических изданиях по основным разделам естественных наук.

Научные семинары

К важным функциям научных библиотек, как научно-исследовательских организаций (БЕН РАН имеет именно такой статус) является проведение семинаров и других мероприятий, для обсуждения результатов их научной деятельности. С самого начала своей деятельности Библиотека регулярно проводит научные семинары, результаты которых отражены на сайте в разделе «Семинары БЕН РАН», в котором представлена деятельность двух семинаров, проводимых Библиотекой. Основным научным семинаром БЕН РАН является научно-практический семинар «Информационное обеспечение науки: новые технологии» [10].

Семинар «Информационное обеспечение науки: новые технологии» (далее Семинар) был задуман в 1985 году сотрудниками БЕН РАН как площадка для обмена мнениями библиотекарей, информационных работников и программистов по вопросам применения средств вычислительной техники для совершенствования процессов информационного обеспечения научных исследований и библиотечной технологии.

Семинар проводился Библиотекой регулярно один раз в два года, начиная с 1985 года по 2009 год, а, начиная с 2010 года каждый год (совместно с Центральной библиотекой УрО РАН).

В работе семинара, в разные годы, принимали участие специалисты из крупнейших библиотек России (РГБ, ГПНТБ, БЕН РАН, БАН, ГПНТБ СО РАН, ЦНБ УрО РАН, ГЦНМБ, ЦНСХБ и др.).

Если говорить о «географии», представленной на Семинаре, то в нем в разные годы участвовали представители практически всех регионов бывшего Советского Союза, где проводились и проводятся серьезные научные исследования. Это – все региональные отделения РАН, научные центры Санкт-Петербурга, Подмосковья, Поволжья, Карельский и Кольский; специалисты из Белоруссии, Грузии, Литвы, Латвии, Казахстана, Молдавии, Украины, Эстонии.

В Семинаре принимали участие представители Болгарии, Великобритании, Германии, Голландии, КНР, США, Югославии.

Материалы семинаров регулярно публиковались в специальных сборниках и размещены на сайте в электронном виде. За 33 года существования Семинара (в 2019 году работа Семинара приостановлена) в сборниках, отражающих его материалы, было опубликовано около 700 статей. Для Семинара специалистами БЕН РАН было разработано приложение «Электронная библиотека», которое включает в себя материалы 22-х семинаров. В электронной библиотеке Семинара, содержатся данные об авторах статей, месте их работы, библиографические описания статей и сборников, а также полные тексты статей и презентации докладов. Все материалы Семинара находятся в открытом доступе.

Кроме основного семинара, Библиотека регулярно проводит Научно-практические обучающие семинары, материалы которых также отражены в разделе «Семинары БЕН РАН» [11].

БЕН РАН на протяжении всего существования, уделяет значительное внимание профессиональной подготовке, как библиотечных специалистов, так и учёных — пользователей Библиотеки, избрав для этого форму очных научных семинаров и в Web-семинаров. Это помогает пользователям ориентироваться в современном информационном мире, в поиске научной информации.

В новых условиях потребность в информационно-обучающих семинарах, как для работников библиотек, так и для пользователей, представляемых библиотеками ресурсов существенно увеличилась. Стремительное развитие информационно-коммуникационных технологий в полной мере затронуло библиотечные процессы и сервисы. Библиотеки вынуждены работать с новыми электронными информационными ресурсами, включая электронные издания, базы и банки данных, электронные каталоги, электронные библиотеки и пр. Поэтому на сайте Библиотеки представлены не только ресурсы, доступ к которым требует определенных навыков пользователей, но и материалы обучающих семинаров, содержащие инструкции по доступу к научным ресур-

сам. Благодаря такому взаимодействию на сайте библиотечные сотрудники и пользователи академической среды имеют не только доступ к ресурсам, но и возможность получить ответы на свои вопросы, касающиеся практической работы с информационными системами и ресурсами.

С 2016 года научно-информационные семинары в БЕН РАН стали регулярными. Презентации прошедших семинаров можно посмотреть на сайте Библиотеки, в разделе «Научно-информационные семинары» (размещены с разрешения авторов). Научно-информационные семинары проводятся с приглашением ведущих отечественных и зарубежных специалистов. Семинары становятся популярными не только среди специалистов отделов БЕН РАН, но и специалистов других библиотек. Каждый семинар посещает до 60 человек. В 2019 году было проведено анкетирование участников семинаров. В ходе анкетирования участниками семинаров была отмечена их большая значимость не только для сотрудников библиотеки, но и для научных сотрудников, которые активно пользуются информационными ресурсами.

Кроме семинаров Библиотека проводит большую работу по привлечению своих пользователей к обучающим вебинарам ведущих фирм естественно-научного направления (Web of Science, World Scientific, Annual Reviews, Springer, и др.). Информация о каждом предстоящем вебинаре размещается на сайте Библиотеки в разделе «Объявления» и в социальных сетях.

Такая образовательная модель как вебинары [12] стала популярна в последнее время. Особенно эта форма образования приемлема для пользователей, работающих в дистанционном режиме.

Таким образом, Портал БЕН РАН, не только раскрывает для пользователей научные ресурсы в области естественных наук (как своей библиотеки, так и соответствующие по тематике и научному уровню ресурсы мировой сети), но и представляет им обучающие материалы по доступу и использованию широкого спектра мировых научных ресурсов по естественнонаучной тематике.

Литература

1. Власова С. А., Глушановский А. В., Каллистратова О. Д., Соловьева Т. Н. Портал «Естественные науки в ИНТЕРНЕТ» // Современные технологии в информационном обеспечении науки: Сборник научных трудов / Каленов Н. Е. (ред.). М.: Научный мир, 2003. С. 24–29.
2. Соловьева Т. Н. Сайт БЕН РАН-WEB портал в области естественных наук // Мир библиотеки, 2008. Т. 4. С. 7–10.
3. Каленов Н. Е. Концепция развития Централизованной библиотечной системы Библиотеки по естественным наукам Российской академии наук (ЦБС БЕН РАН) // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Каленов Н. Е., Цветкова В. А. (ред.). М.: БЕН РАН, 2015. С. 7–17.
4. Соловьева Т. Н. Новый сайт БЕН РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Каленов Н. Е. (ред.). М.: БЕН РАН, 2013. С. 148–155.
5. Ивановский А. А. Типовой сайт научной библиотеки, входящей в структуру БЕН РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / отв. редактор Трескова П. П.; сост. Оганова О. А., Уласовец М. А. Екатеринбург: ООО «УИПЦ», 2014. С. 175–177.
6. Митрошин И. А., Мохначева Ю. В., Харьбина Т. Н. Развитие тематического портала по физико-химической биологии // Скворцовские чтения. Библиотечное дело-2015: документно-информационные коммуникации и библиотеки в пространстве культуры, образования, науки. Материалы двадцатой международной научной конференции (22–23 апреля 2015 г.), 2015. Ч. III. С. 61–66.
7. Погорелко К. П. Новый вариант журнального каталога ЦБС БЕН РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Н. Е. Каленов, В. А. Цветкова (ред.). М.: БЕН РАН, 2017. С. 178–183.
8. Власова С. А., Каленов Н. Е. Новые поисковые возможности и востребованность каталога книг и продолжающихся изданий БЕН РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / отв. редактор П. П. Трескова; сост. О. А. Оганова. Екатеринбург, 2016. С. 171–178.
9. Глушановский А. В. Динамика изменения указателя интернет-ресурсов по естественным наукам Библиотеки по естественным наукам РАН // Клиентоориентированный подход в информационном обслуживании : материалы XIV Всероссийской научно-практической конференции «Электронные ресурсы библиотек, музеев, архивов» — «Информационное обслуживание в век электронных коммуникаций», 30 октября — 1 ноября 2019 г., Санкт-Петербург : сборник материалов / ЦГПБ им. В. В. Маяковского ; ред.-сост. И. Е. Прозоров ; пер. на англ. яз. и сост. А. В. Соколова; ред. А. В. Хромова. Санкт-Петербург : ЦГПБ им. В. В. Маяковского, 2019. С. 235–241.

10. Каленов Н. Е. Семинар «Информационное обеспечение науки: новые технологии» и история библиотечной автоматизации // Библиотекосведение, 2018. Т. 67. № 1. С. 49–55.

11. Соловьева Т. Н., Цветкова В. А. Научно-информационные семинары как вид деятельности БЕН РАН // Информационное обеспечение науки: новые технологии: Сборник научных трудов / Н. Е. Каленов, В. А. Цветкова (ред.). М.: БЕН РАН, 2017. С. 285–291.

12. Калашникова Г. В. О подходах к вопросам непрерывного образования в библиотечно-информационной сфере (на примере БЕН РАН) // Культура: теория и практика (Электронный журнал: <http://theoryofculture.ru/>), 2020. № 2 (35).

References

1. Vlasova S. A., Glushanovsky A. V., Kallistratova O. D., Solovyeva T. N. Portal «Natural Sciences on the INTERNET» // Modern technologies in information support of science: Collection of scientific papers / Kalenov N. E. (ed). M.: Scientific world, 2003. P. 24–29.

2. Solovyeva T. N. Site of the LNS RAS — WEB-portal in the field of natural Sciences // Mir Biblioteka, 2008. Vol. 4. P. 7–10.

3. Kalenov N. E. The concept of development of the Centralized library system of The library for natural Sciences of the Russian Academy of Sciences (CLS LNS RAS) // Information support of science: new technologies: Collection of scientific papers / Kalenov N. E., Tsvetkova V. A. (ed.). M.: BEN RAS, 2015. P. 7–17.

4. Solovyeva T. N. New site of LNS RAS // Information support of science: new technologies: Collection of scientific papers / Kalenov N. E. (ed.). Moscow: BEN RAS, 2013. Pp. 148–155.

5. Ivanovskiy A. A. Typical site of the scientific library that is part of the LNS RAS structure // Information support of science: new technologies: Collection of scientific papers / editor Treskova P. P.; comp. Oganova O. A., Ulasovets M. A. Yekaterinburg: UIPC LLC, 2014. Pp. 175–177.

6. Mitroshin I. A., Mokhnacheva Y. V., Harybina T. N. Development of a thematic portal on physical and chemical biology // Skvortsovskie readings. Librarianship-2015: document and information communications and libraries in the space of culture, education, and science. The proceedings of the twentieth international conference (22–23 April 2015), 2015. Part III. S. 61–66.

7. Pogorelko K. P. New version of the journals opac of CLS LNS RAS // Information support of science: new technology: Collection of scientific works / N. E. Kalenov, V. A. Tsvetkov (ed.). M.: LNS, 2017. S. 178–183.

8. Vlasova S. A., Kalenov N. E. New search capabilities and demand for the catalog of books and ongoing publications of the LNS RAS // Infor-

mation support of science: new technologies: Collection of scientific works / editor P. P. Treskova; comp. O. A. Oganova. Yekaterinburg, 2016. P. 171-178.

9. Glushanovsky A. V. Dynamics of index of Internet resources in the natural sciences Library for natural sciences of RAS // Client-Oriented approach in information service : materials of the XIV all-Russian scientific and practical conference «Electronic resources of libraries, museums, archives» – «Information service in the age of electronic communications», October 30 – November 1, 2019, Saint Petersburg : collection of materials / Mayakovsky Central STATE library of science ; ed.-comp. ed. by A. V. Khromov. Saint Petersburg : tsgpb named after V. V. Mayakovsky, 2019. P. 235-241.

10. Kalenov N. E. Seminar «Information support of science: new technologies» and the history of library automation // Library science, 2018. Vol. 67. № 1. P. 49-55.

11. Solovyeva T. N., Tsvetkova V. A. Scientific and informational seminars as a type of activity of LNS RAS // Information support of science: new technologies: Collection of scientific papers / N. E. Kalenov, V. A. Tsvetkova (ed.). M.: LNS RAS, 2017. Pp. 285-291.

12. Kalashnikova G. V. On approaches to continuing education in the library and information sphere (on the example of LNS RAS) // Culture: theory and practice (Electronic journal: <http://theoryofculture.ru/>), 2020. № 2 (35).

Процессы импортозамещения и конверсии: проблемы информационного обеспечения

О. В. Сютюренко, Е. Ю. Дмитриева

ВИНИТИ РАН

Аннотация. В рамках сопоставительного анализа по отраслям промышленности приводятся данные, иллюстрирующие актуальность проблемы импортозамещения и конверсии для реиндустриализации российской экономики на новой технологической основе. Показан комплекс приоритетных задач, проблемно-ориентированного характера, по информационному обеспечению процессов импортозамещения и конверсии военного производства. Рассматриваются факторы, препятствующие решению задач информационного обеспечения, а также экономические проблемы, детерминирующие развитие процессов импортозамещения и конверсии в российской экономике.

Ключевые слова: импортозамещение, информационное обеспечение, конверсия, экспорт, промышленное производство, жизненный цикл продукции, емкость рынка, технологический обмен, технико-экономическая информация, платежеспособный спрос.

Import substitution and conversion processes: problems of information support

E. Yu. Dmitrieva, O. V. Suintuirenko

VINITI RAS

Abstract. The article gives data of the problem of import substitution and conversion to the re-industrialization of the Russian economy on a new technological basis as a part of a comparative analysis by industry. A set of priority tasks, problem-oriented nature, on information support of processes of import substitution and conversion of military production is shown. Factors impeding the solution of information support tasks, as well as economic problems determining the development of import substitution and conversion processes in the Russian economy, are considered.

Keywords: Import substitution, information support, conversion, export, industrial production, product life cycle, market capacity, technological exchange, technical and economic information, solvent demand.

1. Актуальность проблемы

На современном этапе развития экономики России вопрос импортозамещения стал актуальным, в том числе в геоэкономической войне, в которой страна оказалась в роли одной из противоборствующих сторон. Замещение импорта продуктами отечественного производства было озвучено, как одна из основных задач, ещё в декабре 2015 года в послании Президента РФ Федеральному собранию [1]. Импортозамещение выступает одним из направлений перехода к новой модели развития экономики страны на перспективной технологической основе реиндустриализации, позволяет повысить ее экономическую безопасность. Импортозамещение направлено на рост экономики РФ за счет перехода от производства простых товаров (сырья, материалов) к наукоемкой и высокотехнологичной продукции и является целенаправленной промышленной политикой. При ее реализации одним из основных факторов развития экономики страны являются доходы от экспорта товаров. То есть, создаваемая на импортозамещающих производствах продукция должна быть ориентирована как на внутренний рынок, так на внешний. Актуальность задачи иллюстрируют следующие данные сопоставительного анализа по отраслям промышленности (в долях импортной продукции): станкостроение 90 %; тяжёлое машиностроение 60–80 %; лёгкая промышленность 70–90 %; радиоэлектронная промышленность 80–90 %; фармацевтическая, медицинская промышленность 70–80 %; машиностроение для пищевой промышленности 60–80 % [2,3].

В сфере информационных технологий: операционные системы >96 %; СУБД ~91 %; средства виртуализации ~99 %; коммуникационное ПО ~93 %; промышленно-технологическое ПО (АСУ ТП, PLM) >75 %; порталные решения >70 %; офисное ПО: >95 %; почтовые системы ~95 %;

системы управления деятельностью >60 %; специализированное ПО для решения отраслевых задач >70 %. В целом (в ценовом выражении) импорт товаров в Россию в 2018 году составил 238,2 млрд долларов [4].

Необходимо отметить, что экспорт высокотехнологичной продукции важнее и выгоднее для экономики страны, чем, например, экспорт топливных ресурсов или металлопродукции. Увеличение объема производства высокотехнологичной продукции обходится несоизмеримо дешевле, чем увеличение производства ресурсов и полуфабрикатов. Если развитым странам, экспортирующим высокотехнологичную продукцию, для увеличения объемов экспорта на 1 млрд дол. достаточны капиталовложения в объеме 500 млн дол., то России, ориентированной в основном пока на экспорт сырья, для достижения того же объема потребуется 2–4 млрд дол. [5].

В [6] выделяется три типа импортозамещения: а) инновационное, связанное с повышением качества и/или снижением производственных издержек и цены продукции; б) принудительное, часто сопровождаемое потерей качества, ростом цены, увеличением сроков реализации инновационных проектов; в) упреждающее, связанное с возможным ухудшением экономических взаимосвязей с другими странами в будущем.

Для оценки уровня импортозамещения (по отраслям) может использоваться целый ряд показателей.

1. Снижение отношения G/N , где G – стоимостное выражение импорта, а N – объем предложения в стране, представляющий собой сумму импорта и внутреннего производства. Этот показатель не всегда говорит об успешном импортозамещении, поскольку он может снижаться, если объём производства в отрасли падает, но величина импорта снижается еще быстрее. Тем не менее, на практике этот показатель используется достаточно часто.

2. Измерение доли импорта по отношению к общему объему предложения лежит в основе еще одного показателя: $(W_0 - W_1)S_1$, где W_0 и W_1 – отношение величины импорта к объему предложения в базисном и текущем году соответст-

венно, а S_1 — объем предложения в текущем году (импорт + отечественное производство).

3. Еще один показатель, свидетельствующий об импортозамещении, — степень самообеспеченности, которая рассчитывается, как отношение внутреннего производства к общей величине предложения (импорт + производство) в процентах, или $D/(R+D)$. Рост степени самообеспеченности может служить показателем импортозамещения.

К положительным последствиям импортозамещения как многофакторного процесса относятся следующее:

- рост ВВП и ускорение темпов роста экономики;
- повышение занятости населения за счет новых рабочих мест, как на основном импортозамещающем производстве, так и на вспомогательных предприятиях;
- диверсификация экономики и развитие новых производственных компетенций;
- улучшение платежного баланса за счет снижения затрат валюты на импорт;
- возможность регулирования импортозамещающих производств в национальных системах стандартов и качества.

В целом, качественными и объективными показателями успешности политики импортозамещения являются более высокие темпы экономического роста, стабильные цены, повышение потребления на душу населения, снижение безработицы, повышение производительности факторов производства и улучшение платежного баланса.

Задачи информационного обеспечения процессов импортозамещения актуальны для научно-промышленных объединений и предприятий, которые реализуют или планируют внедрение корпоративной программы импортозамещения, для предприятий планирующих расширение рынков сбыта, в том числе выход на зарубежные рынки.

2. Задачи информационного обеспечения процессов импортозамещения

Пул задач информационного обеспечения процессов импортозамещения включает в себя «традиционные» задачи информационной поддержки работ по фазам и этапам жизненного цикла продукции [7, 8], а также комплекс задач {А, В, С, D, E, F, G} проблемно-ориентированного характера, который более детально рассматривается ниже:

А. Подготовка и анализ данных по результатам целевых исследований российских и зарубежных рынков для отечественной продукции (и услуг). Здесь могут быть использованы результаты анализа данных таможенной статистики внешней торговли РФ и взаимной торговли стран ЕАЭС, сервисы открытой информационной системы «База данных таможенной статистики внешней торговли РФ», а также иные онлайн сервисы анализа данных таможенной статистики партнерских организаций системы торгово-промышленных палат РФ.

В. Проведение исследований и подготовка информационно-аналитических данных по определению потенциально перспективных товарных ниш. Данные включают в себя: а) результаты исследования характеристик рынка Y; б) анализ распределение долей рынка между фирмами-производителями.

С. Комплексный сопоставительный анализ рынка и выбор товарной ниши. Подготовка проекта («дорожной карты») информационной поддержки импортозамещения.

Д. Для лучшей ориентации на рынке и оптимальной корректировки политики ценообразования и продвижения — подготовка данных по результатам многоаспектного информационно-экономического анализа основных конкурентов.

Е. Информационная оценка реальной и потенциальной емкости рынка (внутреннего и внешнего) — позволяет оценить перспективы на данном рынке и избежать неоправданных рисков и потерь. Оценка

потенциала импортозамещения (в том числе экспортно-ориентированного) является основой для формирования рекомендаций по выбору общей стратегии импортозамещения и участия в программах господдержки.

Ф. Подготовка прогнозно-аналитической информации по своей доле на внутреннем и внешнем рынках (по временным фазам) — с учетом технико-экономического потенциала компании.

Г. Информационное моделирование и анализ спроса на разрабатываемую продукцию (по ключевым группам потребителей) с целью: а) установить конкурентоспособные цены на новую продукцию; б) подготовить основу для внесения изменений в сам продукт; в) оптимизировать каналы продвижения и рекламную стратегию.

Н. Информационный мониторинг каналов сбыта в целях оптимизации продвижения продукта до конечного потребителя.

Проблемы, препятствующие решению задач информационного обеспечения процессов импортозамещения: 1. Доступ к информации о заказах со стороны промышленных предприятий, особенно оборонно-промышленного комплекса (ОПК), ограничен для широкого круга российских поставщиков 2. Отсутствует доступная и качественная информация о технологических и производственных возможностях российских поставщиков, способных обеспечить поставки (материалов, деталей, узлов, агрегатов, приборов и пр.) в соответствии с требуемым качеством, в установленные сроки и приемлемой стоимостью. 3. Низкое качество подготовки технико-экономических обоснований, бизнес-планов и финансовых моделей инициаторами импортозамещающих проектов из-за недостаточной квалификации кадров. 4. «Узким» местом информационного обеспечения работ и этапов жизненного цикла является информационное обеспечение процесса оформления прав интеллектуальной собственности и владения (в том числе патентов и заявок). 5. Низкий уровень информационной поддержки взаимодействия ключевых аудиторий на этапах исследования, разработки, трансфера технологий. 6. Незаработанность методологии комплексного

информационного обеспечения работ фазы трансфера технологий (здесь объем внешней научно-технической информации обычно не превышает 10–15 %, а остальной объем — это внутрикорпоративная информация и информация не научно-технического характера). 7. Существенное ослабление роли и потенциала отечественных информационных центров, а также профессиональных возможностей отраслевых научно-технических институтов на фоне слабого развития современных инжиниринговых компаний.

3. Задачи информационного обеспечения процессов конверсии

Задачи информационного обеспечения процессов конверсии предприятий оборонного промышленного комплекса (ОПК) во многом аналогичны задачам информационного обеспечения процессов импортозамещения. Целями проведения конверсии являются:

- обеспечение технического перевооружения промышленности на основе эффективного использования научно-технического и производственного потенциала организаций оборонной промышленности для производства высокотехнологичной и конкурентоспособной на внутреннем и внешнем рынках продукции (работ, услуг);
- уменьшение государственных расходов и нагрузки на федеральный бюджет за счет снижения (оптимизации) оборонного заказа.

Вопреки сложившимся у нас представлениям, тенденции развития научно-производственного потенциала в экономически развитых странах таковы, что ОПК совершенствуется и модернизируется, главным образом, за счет инноваций, переходящих из гражданского сектора промышленности, где спрос и конкуренция в области товаров и услуг массового потребления (а также их значительно меньший жизненный цикл по сравнению с военной техникой) лучше стимулируют прогресс в технологическом развитии. В качестве примера непрерывности и стремительности роста технического уровня продукции широкого применения

и спроса можно привести разработку и изготовление микрочипов, производительность которых, в соответствии с пока выполняющейся закономерностью, каждые полтора года удваивается, а цена вдвое сокращается. Именно массовое гражданское производство в условиях рыночной экономики располагает необходимыми финансовыми ресурсами для создания научно-технических заделов и распространения основной массы нововведений в другие сферы экономики, в том числе и оборонного производства [9].

При этом отсутствие опыта работы военных предприятий в условиях рынка является значительной проблемой для конверсионного перехода. Необходима многоаспектная комплексная подготовка данного процесса. Можно выделить внешние и внутренние факторы успеха конверсии. К внешним факторам относятся: наличие госпрограммы, финансовая, информационная, организационная и нормативно-правовая поддержка государства. Внутренние факторы классифицированы по функциональному признаку на группы: управление, финансы, персонал, технологии, производство, маркетинг и сервис, логистика, взаимодействие с внешней средой.

Основные целевые задачи информационного обеспечения процессов конверсии включают в себя:

А. Предоставление информации по компаниям, фирмам, корпорациям, направлениям их работ и основных видов выпускаемой продукции (номенклатура и объемы), о финансовом состоянии, деловых связях, сделках, логистике, таможенной статистике и правилам таможенного регулирования, о результатах НИОКР, выполненных за счет госбюджета, по инжиниринговым, консалтинговым, сервисным услугам и т. д.

В. Подготовку информационно-аналитических данных: о состоянии рыночных сегментов (товары и капиталы); тенденциях развития; уровне платежеспособного спроса; демографических и климатических изменениях.

С. Предоставление данных: по результатам государственного целеполагания (приоритетами развития, планами,

прогнозами); о государственных (региональных) целевых научно-технических программах; о госзаказах.

Д. Организацию информационной поддержки процесса выбора новой продукции и путей наиболее экономичного перехода к производству новых товаров, который, с учетом эффекта *технологической однородности изделий* прежней и конверсионной продукции, обеспечит минимизацию суммарных затрат на конверсию. Наличие в различных изделиях технологически однородных изделий позволяет концентрировать их производство и тем самым повышать серийность и снижать себестоимость.

Е. Подготовку (с целью снижения неопределенности и риска принятия ошибочных решений) информационно-аналитических данных по результатам многофакторной оценки реализуемости конверсионного проекта с учетом детерминирующих факторов (финансирование, максимально допустимый временной интервал переходного периода, научно-технический потенциал).

Ф. Обеспечение информационной поддержки модернизации технологической базы, включая создание современной системы обеспечения непрерывной информационной поддержки основных бизнес-процессов предприятия, таких как конструкторско-технологическая подготовка, электронный документооборот, планирование производства и процессов его подготовки, оперативное управление производством, контроль производственного процесса, управление качеством и сопровождение выпущенной продукции. Такая поддержка реализуется в рамках CALS-технологии (Continuous Acquisition and Life cycle Support – непрерывная информационная поддержка жизненного цикла продукта) [10]. Достаточно эффективно могут быть использованы отечественные программные комплексы, например система CADLib Модель [11], платформа «1С Предприятие 8» (фактически функциональный аналог СУБД Oracle).

Г. Обеспечение информационной поддержки профессиональной подготовки и переподготовки руководящих кадров высшего и среднего звена предприятия.

Н. Информационную поддержку взаимодействия в области технологического обмена между гражданскими и военными производствами для использования имеющихся разработок при создании продукции двойного и гражданского назначения.

Следует отметить, что при решении конверсионных задач возрастает использование технико-экономической, экономической и технико-технологической информации. Поиск и восприятие этой информации в цифровом пространстве становится все более важной и сложной задачей. При этом существующие методы многокритериальной оценки инноваций, процессов трансфера технологий, оценки научно-технического уровня выпускаемой и разрабатываемой продукции, решения задач импортозамещения и конверсии недостаточно эффективны, т. к. не позволяют обрабатывать все возрастающие объемы цифровой научно-технической и технико-экономической информации. Перспективным направлением является использование технологии Больших Данных (Big Data), в том числе для реализации эконометрических моделей полных жизненных циклов наукоемкой продукции, сложных технических комплексов и систем. В определенной мере технологии Big Data это ответ на качественно новые задачи в промышленности и науке в структуре цифровой экономики.

4. Заключение

В заключение кратко остановимся на факторах, детерминирующих развитие процессов импортозамещения и конверсии:

1. Несоответствие возможностей существующей национальной информационной инфраструктуры современным требованиям новой российской экономической институциональной среды. Основная системная проблема – темпы развития и потенциал существующей Государственной системы научно-технической информации не позволяют в полной мере удовлетворять растущий спрос и расширяющийся спектр информационных потребностей пользователей

из инновационно-промышленной (и научно-образовательной) сферы народного хозяйства [12].

2. Явная структурно-функциональная недостаточность «промежуточного слоя», существующего между фундаментальной и прикладной наукой, и промышленностью, необходимого для создания инновационных продуктов и трансфера технологий. Такое положение объясняется целым рядом причин финансового, конъюнктурно-экономического, социального и технологического характера. В советский период «промежуточный слой» состоял из отраслевых прикладных НИИ и проектных организаций. В постсоветский период этот «промежуточный слой» практически деградировал, по отдельным направлениям он необратимо деформировался и фактически утратил имевшийся научно-технический потенциал.

3. В большинстве российских компаний отсутствуют отделы (а также методические материалы и обученные специалисты) научно-технической информации, маркетинга и интеллектуальной собственности, предназначенные для выполнения взаимосвязанного комплекса работ инновационного характера в рамках процессов импортозамещения и конверсии.

4. Существенной проблемой является недостаток квалифицированных кадров в целом ряде производств. Негативным фактором является, так называемая, утечка мозгов — процесс массовой эмиграции, при которой из страны по экономическим или иным причинам уезжали учёные, специалисты (в том числе IT-сферы), квалифицированные рабочие.

5. Наряду с методами экономического стимулирования распространения нововведений существенную роль в формировании спроса на новшества играет, так называемая, «несущая способность» среды, под которой понимается платежеспособный спрос на инновационный продукт или технологию со стороны производственного сектора или потребительского рынка, который и обеспечивает необходимую финансовую и материальную поддержку распространения новшеств. Бедность населения является в большей

степени экономической, а не социальной категорией, так как *детерминирует* рост платежеспособного спроса и, тем самым, рост инвестиционной активности, рост объемов производства, рост внутреннего рынка и экономики в целом. Наличие платежеспособного спроса на продукцию промышленных компаний является критически важным обстоятельством, обеспечивающим возможности повышения рентабельности и инновационной активности. Отсутствие уверенных перспектив сбыта продукции существенно подрывает стимулы к реализации любых стратегий развития процессов импортозамещения и конверсии России.

Литература

1. Послание Президента Российской Федерации Федеральному собранию Российской Федерации, 2016. URL: <http://www.kremlin.ru/acts/bank/41550>
2. Калинина В. В. Кластерный анализ состояния промышленности регионов РФ. // Экономика в промышленности. 2016. № 3. Июль – Сентябрь. С. 259–269. URL: <https://ecoprom.misis.ru/jour/article/view/548/511>
3. Кирбитова С. В., Кожина Н. А. Проблемы импортозамещения в России // Таможенная политика России на Дальнем Востоке. 2019. № 1(86). С. 61–72.
4. Внешняя торговля России в 2018 году. URL: <https://russian-trade.com/reports-and-reviews/2019-02/vneshnyaya-torgovlya-rossii-v-2018-godu/>
5. Новые возможности по активизации импортозамещения высокотехнологичной продукции. URL: <https://importozameshenie.ru/novye-vozmozhnosti-po-aktivizacii-importozamesheniya-vysokotekhnologichnoj-produkcii/>
6. Фальцман В. Импортозамещение в ТЭК и ОПК // Вопросы экономики. 2015. № 1. С. 116–124.
7. Слонтюренко О. В. Информационное обеспечение: факторы развития, управление, эффективность // НИИ. Сер. 2. Информационные процессы и системы. 2016. № 6. С. 7–15.
8. Слонтюренко О. В., Гиляревский Р. С. Задачи информационного обеспечения инновационного развития экономики и роль инжиниринга // Научно-техническая информация, Сер. 1. 2017. № 5. С. 5–14.

9. Багриновский К. А., Бендииков М. А., Хрусталеv Е. Ю. Современные методы управления технологическим развитием. М.: «Российская политическая энциклопедия», 2001. 272 с.

10. CALS-технологии. URL: <http://www.bestreferat.ru/referat-236518.html>

11. Информационная поддержка процесса и инфраструктуры производства. URL: <https://sapr.ru/article/24525>

12. Стютиренко О. В., Дмитриева Е. Ю. Государственная система научно-технической информации в структуре задач цифровой экономики // Научно-техническая информация. Сер. 1. 2019. № 9. С. 1–11.

References

1. Poslaniye Prezidenta Rossiiskoi Federatsii sobraniyu Rossiiskoi Federatsii, 2016. URL : <http://www.kremlin.ru/acts/bank/41550>

2. Kalinina V. V. Klasternyi analiz sostoyaniya promyshlennosti regionov RF // Ekonomika v promyshlennosti. 2016. N 3. Iul'-sentyabr. S. 259–269. URL: <https://ecoprom.misis.ru/jour/article/view/548/511>

3. Kirbitova S. V., Kozhina N. A. Problemy importozamesheniya v Rossii // Tamozhennaya politika Rossii na Dalnem Vostoke. 2019. N 1(86). С. 61–72.

4. Vneshniaya trgovlia Rossii v 2018 godu. URL: <https://russian-trade.com/reports-and-reviews/2019-02/vneshnyaya-torgovlya-rossii-v-2018-godu/>

5. Novye vozmozhnosti po aktivizatsii importozamesheniya vysokotekhnologichnoi produktsii. URL: <https://importozameshenie.ru/novye-vozmozhnosti-po-aktivizatsii-importozamesheniya-vysokotekhnologichnoj-produkcii/>

6. Faltzman V. Importozameshenie v TEK I OPK // Voprosy ekonomiki. 2015. N 1. S. 116–124.

7. Syntyurenko O. V. Informatsionnoe obespechenie: factory razvitiya, upravlenie, effektivnost // NTI. Ser. 2. Informatsionnye protsessy i sistemy. 2016. N 6. С.-7–15.

8. Syntyurenko O. V., Gilyarevskii R. S. Zadachi informatsionnogo obespecheniya innovatsionnogo razvitiya ekonomiki I rol' inginiringa // Nauchno-tekhnicheskaya informatsia. Ser. 1. 2017. N 5. S. 5–14.

9. Bagrinovskii K. A., Bendikov M. A., Hrustalev E. Yu. Sovremennyye metody upravleniya tekhnologicheskim razvitiem. М.: Rossiiskaya politicheskaya enziklopedia, 2001. 272 s.

10. CALS-tehnologii. URL: <http://www.bestreferat.ru/referat-236518.html>

11. Informatsionnaya podderzhka protsessu i infrastruktury proizvodstva. URL: <http://www.bestreferat.ru/referat-236518.html>

12. Syuntyurenko O. V., Dmitrieva E. Yu. Gosudarstvennaya sistema nauchno-technicheskoi informatsii v structure zadach tsifrovoi ekonomiki // Nauchno-technicheskaya informatsia. Ser. 1. 2019. № 9. S. 1-11.

Разработка единой системы наукометрического анализа на примере УрО РАН

Ю. Д. Прокофьева

ЦНБ УрО РАН

Аннотация. В статье на примере Уральского отделения Российской академии наук (УрО РАН) поднимается проблема научных коммуникаций между организациями, и, как следствие, отсутствие единой системы проведения наукометрических исследований. Для руководства в качестве дополнительного инструмента оценки деятельности организации предлагается использование информационно-аналитического продукта. Предлагаемый продукт информационно-аналитической деятельности – результат анализа, проведённого на основании «Методики оценки развития междисциплинарных исследований и деятельности научных организаций», разработанной в ЦНБ УрО РАН. В работе исследованы достоинства и недостатки методики, факторы, влияющие на результат, выявленные в ходе анализа организаций УрФО. Предложены две модели наукометрического анализа деятельности организаций, сформулированы условия для их реализации. Одним из важнейших условий реализации моделей названо наличие в научных организациях научных библиотек или специалистов, способных проводить наукометрические исследования.

Ключевые слова: наукометрический анализ, методика оценки, научная библиотека, научная коммуникация, наукометрия, библиометрия.

Development of a unified system of scientometric analysis on the example of the Ural Branch of the Russian Academy of Sciences

Yu. D. Prokofieva

C SL UrB RAS

Abstract. The article on the example of the Ural Branch of the Russian Academy of Sciences (Ural Branch of the Russian Academy of Sciences) raises the problem of scientific communications between organizations, and,

as a result, the absence of a unified system for conducting scientometric research. The existing system for evaluating the effectiveness of scientific organizations and the state of scientometric research in the organizations of the Ural Branch of the Russian Academy of Sciences are considered. As an additional tool for assessing the activities of an organization for management, the use of an information and analytical product is proposed. The proposed product of information and analytical activities is the result of the analysis carried out on the basis of the «Methodology for assessing the development of interdisciplinary research and the activities of scientific organizations», developed at the Central Scientific Library of the Ural Branch of the Russian Academy of Sciences. The paper presents the advantages and disadvantages of the methodology, the factors influencing the result, revealed during the analysis of the organizations of the Ural Federal District. Two models of scientometric analysis of organizations' activities are proposed, conditions for their implementation are formulated. One of the most important conditions for the implementation of the models is the presence in scientific organizations of scientific libraries or specialists capable of conducting scientometric research.

Keywords: scientometric analysis, assessment methodology, scientific library, scientific communication, scientometrics, bibliometrics.

Введение

С целью повышения качества публикаций российских учёных и сохранения темпов роста их количества, в России ежегодно совершенствуется система оценки результативности научных организаций и всё больше внимания уделяется качественным критериям. Но количественные показатели результативности остаются в приоритете. Значение баллов и коэффициентов считаются объективными показателями, так как они позволяют Министерству науки и высшего образования Российской Федерации оценить эффективность работы подведомственной организации, а также отдельного учёного.

Помимо установленных показателей, руководству организации для принятия решений, выработки стратегии развития, выбора способов поощрения сотрудников и поддержки научных исследований необходимо обладать детальной аналитически переработанной информацией.

Преимущественно источниками качественной информации для руководства служат отчёты сотрудников, отделов,

сводные отчёты организации и годовой отчёт отделения (в случае Российской академии наук). Отчеты о научно-исследовательской работе, как правило, содержат краткую фактическую информацию об исследовании, контекст исследования, обоснование актуальности и уровня развития темы. Но недостаток таких документов – отсутствие видения развития организации на фоне других организаций и в контексте общего развития науки в стране и в мире.

С целью создания более полного представления о деятельности организации в Центральной научной библиотеке Уральского отделения РАН разработана «Методика оценки развития междисциплинарных исследований и деятельности научных организаций» (далее Методика). Особенность данной Методики в том, что она основана на сведениях из наукометрических баз данных [11, 19].

В ходе исследования выявлена проблема – отсутствие коммуникации между организациями по вопросам сбора, учёта и обработки информации о деятельности научных организаций УрО РАН, представленной в наукометрических базах данных.

Цель исследования – разработка модели единой системы наукометрического анализа научных организаций УрО РАН на основании разработанной Методики.

Объектом исследования является система оценки эффективности деятельности научных организаций и учёных.

Предмет исследования – наукометрический анализ, как инструмент оценки эффективности деятельности научных организаций и учёных УрО РАН.

Задачи исследования:

1. Рассмотреть существующую систему оценки эффективности научных организаций.

2. Оценить сегодняшнее состояние наукометрических исследований в организациях УрО РАН.

3. Представить опыт разработки «Методики оценки развития междисциплинарных исследований и деятельности научных организаций». Оценить результаты проведённого анализа на основании Методики и выявить достоинства, не-

достатки и факторы, влияющие на результаты использования Методики.

4. Разработать модель единой системы наукометрического анализа УрО РАН.

Единая система оценки на основании информации из наукометрических баз данных позволит получать актуальную, своевременную информацию, основанную на общепринятых критериях. Как следствие, это даст возможность сформировать общее видение развития отдельных институтов и Уральского отделения в целом.

Вопросами наукометрических исследований и библиометрических методов анализа в России наиболее активно занимаются Лаврик О. Л., Плешакова М. А., Калюжная Т. А., Цветкова В. А., Мохначева Ю. В., Харыбина Т. Н., Бескаравайная Е. В., Москалева О. В., Акоев М. А., Земсков А. И. и другие. Одними из первых этой проблемой на Урале занялись сотрудники ЦНБ УРО РАН [7, 17, 18, 21, 22].

В контексте проблематики данного исследования публикации практически отсутствуют. Среди работ российских авторов нет исследований, посвящённых единой системе сбора и учёта наукометрических данных на примере организаций типа Российской академии наук. В некоторых публикациях присутствует описание идей и опыта, которые близки по замыслу отдельным этапам будущей системы. Например, в ряде публикаций в качестве инструмента для сбора и анализа наукометрических данных предлагается использовать готовые программные продукты, имеющие название CRIS системы (Current Research Information Systems) [1, 6, 20, 23].

Проблема научной коммуникации и взаимодействия между организациями рассмотрена в статье А. Б. Антопольского [2].

Концепция единой точки доступа или единой системы сбора информации о показателях встречается в публикациях российских и зарубежных авторов [6, 13, 29, 31].

Важность взаимодействия работников библиотек со специалистами научных организаций упоминается во многих российских и зарубежных работах [3, 5, 8, 9, 31, 32].

Примерами эффективной системы научной коммуникации являются университеты. Прогрессивный опыт взаимодействия академического сообщества, руководства университетов и специалистов по наукометрии реализован за счёт создания специализированных отделов и центров [14, 15, 16, 25, 26, 27]. Стоит отметить, что не всегда эти отделы создаются на базе библиотеки.

Большинство зарубежных публикаций отражают опыт преимущественно университетов. Примеры таких стран, как Австрия, Швеция, Дания, Хорватия, Австралия, Новая Зеландия, Ирландия, Великобритания свидетельствует о том, что университетские библиотеки принимают на себя роль центрального аналитического подразделения университета, а наукометрия становится ведущим направлением деятельности и одной из наиболее востребованных услуг научной библиотеки [28, 29, 30, 31, 32].

Оценка эффективности научных организаций

Каждая организация заинтересована в повышении конкурентоспособности, инновационной и публикационной активности, достижении результатов и развитии в соответствии с глобальными тенденциями и запросами общества.

Ранее оценка результативности (эффективности) деятельности научных организаций осуществлялась преимущественно на основании цифровых наукометрических показателей, таких как количество публикаций автора и количество ссылок на его работы. На примере Методики расчёта качественного показателя государственного задания наблюдается постепенный переход от количественных показателей к качественным показателям публикационной результативности, таким как критерии влиятельности и рейтинг источника, начиная от изданий ВАК до журналов Web of Science первого квартили [12]. Однако, содержательная оценка науч-

ной деятельности организации возможна только в совокупности с дополнительными сведениями, отражающими всю полноту работы учреждения по различным критериям.

В качестве инструмента для содержательной оценки предлагается разработанная в ЦНБ УрО РАН «Методика оценки развития междисциплинарных исследований и деятельности научных организаций». Опыт работы над Методикой определил потребность, заключенную в одной из задач данного исследования – оценить качество результатов анализа по разработанной Методике и предложить дальнейшие варианты применения Методики в качестве основы единой системы наукометрического анализа деятельности научных организаций УрО РАН.

Опыт разработки Методики

Методика оценки развития междисциплинарных исследований и деятельности научных организаций разработана в период 2018–2019 гг. Впервые опробована на примере организаций Уральского федерального округа (институты и научные центры УрО РАН, УрФУ им. Б. Н. Ельцина). Более 50 организаций проанализированы по выбранным критериям с целью получения информации об их научной деятельности, отражённой в наукометрических базах данных (Web of Science и Scopus) [19].

В результате оценки проведенного анализа выявлен ряд недостатков при использовании комплексного подхода применительно к большому числу организаций:

— сложность проверки достоверности данных из-за опосредованного способа получения информации (из баз данных);

— масштабность проводимых исследований в условиях быстрых изменений данных;

— большие временные затраты и трудоёмкость обработки массива данных;

— общий поверхностный характер полученных результатов анализа.

Выявленные недостатки связаны с наличием следующих факторов, неблагоприятно влияющих на результаты анализа:

- нехватка специалистов одной библиотеки и недостаток профессиональных навыков [10, с. 48];
- отсутствие специалиста, осуществляющего контроль и редактирование профиля организации;
- отсутствие профилей некоторых организаций в наукометрических БД;
- наличие ошибок в отображении публикаций в профилях организаций;
- отсутствие в большинстве организаций библиотек или специализированных отделов, занимающихся наукометрическими исследованиями;
- отсутствие единой методики и системы сбора наукометрической информации;
- отсутствие налаженного взаимодействия между библиотеками и сотрудниками научных организаций.

Несмотря на недостатки и отрицательно воздействующие факторы, у Методики есть достоинства. Преимуществом данной Методики является то, что её результаты основаны на актуальных и объективных наукометрических данных. Критерии разработанной Методики дают возможность: сравнивать деятельность конкретной организации с другими научными учреждениями и странами; взглянуть на реальное положение научных исследований; прогнозировать развитие направлений науки; устанавливать научные коммуникации; оценивать уровень публикационной активности и востребованности результатов научной деятельности учёных. Результаты такого анализа являются мощной поддержкой руководству научных организаций при определении приоритетных направлений работы, разработке, реализации и корректировке стратегических документов.

По результатам оценки проведённого анализа на основании Методики, можно сделать вывод, методика применима:

- с учётом устранения недостатков и доработки слабых сторон;

– в условиях, способных компенсировать недостатки проведенного анализа;

– для анализа одной организации или для сравнения небольшого количества организаций.

Все выявленные недостатки, факторы воздействия и условия реализации указывают на необходимость в постоянном взаимодействии организации и сотрудников библиотеки. Поскольку сотрудники библиотеки обладают уникальной квалификацией в отношении библиометрии и анализа цитирования, а специалисты организаций знаниями своей научной области, результат такого анализа можно считать более достоверным и точным.

Конечным результатом использования разработанной Методики станет информационный продукт, содержащий аналитическую информацию о деятельности организации.

Научная библиотека в структуре организаций УрО РАН

Использование наукометрии для оценки деятельности институтов становится общепринятой практикой, и именно здесь открываются возможности участия научных библиотек.

Создание единой системы наукометрических исследований предполагает в организациях наличие библиотек или информационных центров, способных заниматься наукометрией.

Координация деятельности этих информационных подразделений станет задачей сотрудников ЦНБ УрО РАН. ЦНБ УрО РАН выступает в качестве единой точки сбора и обработки результатов анализа. Являясь методическим и одним из крупнейших информационно-библиотечных центров в регионе, ЦНБ УрО РАН в числе первых на Урале начали проводить наукометрические исследования, регулярное обучение работе с наукометрическими ресурсами, мониторинг состояния науки и информационную поддержку научно-

исследовательской деятельности учёных и организаций в целом.

Важно отметить, что не каждое предприятие или институт имеют в своей структуре подобное подразделение, хотя большинство осознают роль информации и информационного сопровождения для проведения научных исследований. В Уральском отделении РАН 23 научных учреждения (институты и научные центры) имеют в своей структуре библиотеки. Научно-исследовательская работа в них практически не ведется. Общая численность штатных сотрудников на 23 библиотеки составляет 47 человек, из которых 1 научный сотрудник (ТКНС УрО РАН). В 14 из 23 библиотек работает по 1 сотруднику [24]. Некоторые библиотеки УрО РАН проводят работу в наукометрических базах данных: помогают сотрудникам институтов овладеть основными навыками поиска в базах данных, осуществлять контроль корректного отображения публикаций в авторских профилях, определять и интерпретировать ключевые наукометрические показатели, а также осуществляют работу по корректировке профиля организации. Но основными направлениями работы по-прежнему остаются библиотечно-информационное и справочно-библиографическое обслуживание.

Подобные обстоятельства являются препятствием развития наукометрических исследований в научных организациях и наукометрии в целом. Для учреждений, где отсутствует научная библиотека или, если компетенций её сотрудников не хватает для проведения наукометрических исследований, решением является поиск такого специалиста или обучение сотрудников из коллектива.

Создание модели наукометрического анализа деятельности организаций УрО РАН на основании Методики

Исходя из наличия или отсутствия в организации научной библиотеки или лиц, ответственных

за наукометрические исследования, предложены две модели единой системы наукометрического анализа.

В ситуации отсутствия научных библиотек или информационных отделов в структуре организаций, предлагается следующая модель проведения наукометрического анализа на основании разработанной Методики (рис. 1).



Рис. 1. Модель наукометрического анализа на основании Методики.
Вариант 1

Первая модель наукометрического анализа деятельности организаций состоит из трёх этапов:

1. *Проведение анализа на основании Методики и взаимодействие библиотеки со специалистами организаций.*

Библиотека, в данном случае ЦНБ УрО РАН, проводит анализ деятельности каждого научного учреждения согласно Методике и отправляет результат на проверку и устранение несоответствий в организацию. Далее сотрудник организации (библиотека, информационная служба или отдел), проверяет полученные данные, вносит необходимые изменения и отправляет вариант с подтверждёнными сведениями обратно в ЦНБ УрО РАН.

2. *Создание аналитического отчёта о деятельности каждой организации УрО РАН.*

На основании полученных подтвержденных данных, ЦНБ УрО РАН подготавливает аналитический отчёт о научной деятельности каждой организации.

3. *Создание сводного отчёта о научной деятельности Уральского отделения РАН по результатам наукометрического анализа.*

На основании сформированных аналитических отчётов отдельных организаций УрО РАН, сотрудники ЦНБ УрО РАН готовят сводный отчёт о научной деятельности всего отделения.

Реализация первой модели проведения наукометрического анализа устраняет ряд недостатков, которые были описаны ранее. Например, проблему достоверности и проверки сведений путём прямого взаимодействия ЦНБ УрО РАН и специалистов из научных организаций.

Вторая модель наукометрического анализа предполагает проведение исследования в каждой организации специализированным отделом или специалистом со знанием специфики учреждения, направлений исследований, сотрудников, авторов и других особенностей. ЦНБ УрО РАН выступает в качестве конечной точки, принимающей от организации достоверную, проверенную информацию.

Такая организация работы реализуется второй моделью проведения наукометрического анализа на основании Методики (рис. 2).



Рис. 2. Модель наукометрического анализа на основании Методики. Вариант 2

Вторая модель включает два основных этапа:

1. Проведение анализа на основании Методики сотрудниками организаций.

Сотрудники каждой организации проводят наукометрический анализ деятельности своего института на основании разработанной Методики и отправляют отчёт с результатами в ЦНБ УрО РАН.

2. Создание сводного отчёта о научной деятельности Уральского отделения РАН по результатам наукометрического анализа.

Сотрудники ЦНБ УрО РАН на основании полученных отчётов организаций, создают Сводный отчёт по результатам анализа деятельности организаций УрО РАН.

Вторая модель проведения наукометрического анализа способна устранить ещё больше недостатков, чем первый вариант. Преимущества второй модели заключаются в следующем:

- сокращаются временные затраты. Анализ может выполняться одновременно всеми организациями, а результат представлен в ЦНБ УрО РАН к определенному сроку;

- меньшая нагрузка на сотрудников библиотеки. Основной задачей ЦНБ УрО РАН в данном варианте становится не анализ деятельности отдельной организации, а контроль, мониторинг, проверка, сбор и сведение информации;

- сохранение актуальности данных из-за уменьшения сроков проведения анализа;
- возможность постоянного отслеживания и корректировки сведений об организации и её сотрудниках в наукометрических базах данных;
- перспектива создания на основании Методики единой системы наукометрических исследований в каждой организации;
- повышение лояльности руководства научных организаций за счёт участия специалистов организаций в процессе проведения исследования.

На данный момент в связи с отсутствием в структурах большинства организаций научных библиотек, наиболее подходящим для реализации является первый вариант модели. Он предполагает проведение поэтапного наукометрического анализа каждой отдельной организации сотрудниками ЦНБ УрО РАН на основании Методики. Реализация второй модели возможна в случае создания соответствующих условий и проведения большой подготовительной работы, а именно: составление плана работы и взаимодействия с руководством организаций, согласование и утверждение Методики, определение в каждой организации ответственных лиц, организация обучения анализу в соответствии с Методикой.

Заключение

Сбор, обработка и интерпретация наукометрических данных, находящихся практически в открытом доступе, дают уникальные возможности по формированию представлений об активности и результативности научных организаций и исследователей [4].

Наряду с установленными количественными показателями, у руководства появляется возможность получения объективной содержательной информации, источниками которой выступают продукты информационно-аналитической деятельности, созданные в результате наукометрического анализа. Важным условием при создании информационно-

аналитических продуктов является прямое сотрудничество библиотечных работников и специалистов организаций, обладающих всей полнотой информации о научной деятельности учреждения.

Анализ зарубежных и российских исследований позволяет отметить растущий интерес к библиометрии для оценки исследований, однако, говорить об утвердившейся и зарекомендовавшей себя единой системе наукометрических исследований внутри организаций пока невозможно.

Представленные в исследовании модели наукометрического анализа могут стать основой единой системы наукометрического анализа организации с иерархической структурой, применительно к УрО РАН.

Внедрение единой системы позволит руководству институтов, научных центров и сотрудникам библиотек, обладающим опытом в области библиометрии, анализировать исследовательскую и публикационную деятельность организации и отдельных учёных. Это имеет значение, как на общем стратегическом уровне, так и на уровне взаимодействия между руководством и исследователями. Роль библиотеки в данной системе является ключевой, поскольку библиометрия за последние годы стала неотъемлемой частью библиотечно-информационной деятельности и именно научные библиотеки в большинстве случаев выступают инициаторами проведения наукометрических исследований. [28, с. 4]

Литература

1. Альперин Б. Л., Ведягин А. А., Зибарева И. В. SciAct – информационно-аналитическая система Института катализа СО РАН для мониторинга и стимулирования научной деятельности // Труды ГПНТБ СО РАН. 2015. № 9. С. 95–102.
2. Антопольский А. Б. Будущее научных коммуникаций и научной информации // Информация и инновации. 2019. Т. 14. № 1. С. 7–17.
3. Дудникова О. В., Богомолов А. А., Смирнова О. А. Роль библиотеки в информационно-аналитическом сопровождении научно-публикационной деятельности вуза // Университетская книга. <http://>

www.unkniga.ru/biblioteki/vuzbiblio/8200-rol-biblioteki-v-informatsionno-analiticheskom-soprovozhdaetsya.html

4. Интегральная оценка эффективности научно-исследовательских работ в научных учреждениях Минздрава России / Г. С. Лебедев, О. Б. Крылов, А. И. Леляков, Ю. Г. Миронов, В. В. Ткаченко // Журнал телемедицины и электронного здравоохранения. <http://jtelemed.ru/article/integralnaja-ocenka-jeffektivnosti-nauchno-issledovatelских-rabot-v-nauchnyh-uchrezhdenijah>

5. Использование РИНЦ и Science Index для анализа и оценки научной деятельности / В. Глухов, А. Блинкова, А. Ковалева [и др.] // Университетская книга. 2018. Май. С. 70. <http://www.unkniga.ru/biblioteki/vuzbiblio/8381-ispolzovanie-rinc-i-science-index-dlya-analiza-i-otsenki-nauchnoy.html>

6. Квелидзе-Кузнецова Н. Н., Морозова С. А., Матюшенко А. Д. Программный модуль сбора и анализа данных о наукометрических показателях сотрудников: интеграционные возможности и перспективы развития // Наука и научная информация. 2019. 2(4). С. 216–227. <https://doi.org/10.24108/2658-3143-2019-2-4-216-227>

7. Кычакова А. В., Трескова П. П. Отражение междисциплинарного развития науки через систему библиометрических показателей // Информационная школа молодого ученого. Екатеринбург: ЦНБ УрО РАН, 2011. С. 48–69.

8. Лаврик О. В., Плешакова М. А., Каложная Т. А. Информационно-аналитические продукты в научных библиотеках для информационного обеспечения научно-исследовательской работы // Вестник Томского государственного университета. Культурология и искусствоведение. 2018. № 32. С. 186–201.

9. Лаврик О. В., Плешакова М. А., Каложная Т. А. Систематический обзор как вид обзорно-аналитических продуктов // Библиотекосведение. 2019. № 2. С. 33–51.

10. Макеева О. В. Актуальные проблемы работы с персоналом библиотек на современном этапе развития библиотечного дела // Библиосфера. 2018. № 4. С. 48–55.

11. Методика наукометрической оценки развития научных организаций Уральского Федерального округа / Ю. Д. Прокофьева, А. С. Павлова, Т. А. Осипенко, Л. А. Оболенская // Информация и инновации. 2018. № 4. С. 76–88.

12. Методика расчёта качественного показателя государственного задания «Комплексный балл публикационной результативности: утверждена заместителем Министра науки и высшего образования Российской Федерации 30 декабря 2019 г. <http://нтр.рф/upload/iblock/f35/> Методи-

ка%20расчета%20комплексного%20балла%20публикационной%20результативности.pdf

13. Москалёва О. В., Акоев М. А. Наукометрия: немного истории и современные российские реалии // Управление наукой: теория и практика. 2019. Том 1. № 1. С. 135–148.

14. Научно-исследовательский портал Уральского федерального университета. <https://science.urfu.ru/ru/>

15. Отдел анализа наукометрических показателей ВГТУ. <https://cchgeu.ru/university/struktura/tsentr-publikatsionnoy-aktivnosti/otdel-analiza-naukometricheskikh-pokazateley/>

16. Отдел научного менеджмента и наукометрии МГЛУ. <https://linguanet.ru/ob-universitete/administrativnye-podrazdeleniya/otdel-nauchnogo-menedzhmenta-i-naukometrii/>

17. Пирожок Т. В., Трескова П. П. Библиометрический анализ публикаций сотрудников институтов Уральского отделения РАН, входящих в ОУС по гуманитарным и экономическим наукам за период 2008–2012 гг. По базам данных SCOPUS и РИНЦ // III Информационная школа молодого ученого. Екатеринбург : ЦНБ УрО РАН, 2013. С. 97–102.

18. Пирожок Т. В., Трескова П. П. Мировые и отечественные ресурсы как база наукометрических исследований // Информационная школа молодого ученого. Екатеринбург : ЦНБ УрО РАН, 2011. С. 37–47.

19. Прокофьева Ю. Д. Критерии оценок развития междисциплинарных исследований: к постановке проблемы // Библиосфера. 2018. № 4. С. 95–100.

20. Траулько М. В., Пашков П. М. Подходы к построению информационной системы текущих исследований в вузе: анализ, оценка и разработка методики выбора // Инновации в образовании. 2017. № 4(23). С. 139–161.

21. Трескова П. П. Наука в информационном измерении: анализ публикационной активности ученых с использованием баз данных «Web of Science» и «Scopus» // Информационное обеспечение науки: новые технологии. Москва, 2009. С. 253–262.

22. Трескова П. П. Применение библиометрических показателей для оценки результативности деятельности научных организаций Уральского отделения РАН // Новые технологии в информационно-библиотечном обеспечении научных исследований. Екатеринбург, 2010. С. 226–235.

23. Трубников В. С., Туральчук К. А. Проектирование системы сбора, анализа и визуализации наукометрических данных // Проблемы современной науки и образования. 2015. № 36. <https://cyberleninka.ru/article/n/proektirovanie-sistemy-sbora-analiza-i-vizualizatsii-naukometricheskikh-dannyh>

24. Федеральное государственное бюджетное учреждение науки «Центральная научная библиотека Уральского отделения Российской академии наук»: отчет о научно-исследовательской работе (организации) за 2019 год / Министерство науки и высшего образования Российской Федерации. Екатеринбург, 2019. 246 с.

25. Центр наукометрии и публикационной активности МГУПП. <http://www.mgupp.ru/science/center/>

26. Центр наукометрии и развития научных журналов РУДН (ЦН и РНЖ). <http://old.rudn.ru/?pagec=6084>

27. Центр наукометрического анализа и международных систем индексирования ВолГАУ. <http://www.volgau.com/деятельность/наука-и-инновации/центр-наукометрического-анализа-и-международных-систем-индексирования>

28. Åström F., Hansson J. How implementation of bibliometric practice affects the role of academic libraries // *Journal of Librarianship and Information Science*. 2013. V. 45. № 4. P. 1-7.

29. Čadovska I., Mitrović G. Uloga knjižnica u vrednovanju znanstvene produktivnosti: primjer Nacionalne i sveučilišne knjižnice u Zagrebu [The role of libraries in evaluating scientific productivity: the example of the National and University Library in Zagreb] // *Vjesnik bibliotekara Hrvatske*. 2018. V. 61. № 2. P. 179-198. <https://scholar.google.ru/citations?user=IUvMe1AAAAAJ&hl=ru>

30. Competencies for bibliometrics / A. Cox, E.Gadd, S. Petersohn, L. Sbaffi // *Journal of Librarianship and Information Science*. 2019. V. 51. № 3. P. 746-762. <https://journals.sagepub.com/doi/full/10.1177/0961000617728111>

31. Gumpenberger C., Wieland M., Gorraiz J. Bibliometric practices and activities at the University of Vienna // *Library Management*. 2012. V. 33. № 3. P. 174-183.

32. Kaarsted T. Catch-22: How do academic libraries position themselves between top level management expectations and scholars hopes and dreams? // *Library Management*. 2017. V. 38. № 2/3. P. 78-87.

References

1. Al'perin B. L., Vedyagin A. A., Zibareva I. V. SciAct – informacionno-analiticheskaya Sistema Instituta kataliza SO RAN dlya monitoringa i stimulirovaniya nauchnoj deyatel'nosti // *Trudy GPNTB SO RAN*. 2015. № 9. P. 95-102.

2. Antopol'skij A. B. Budushchee nauchnyh kommunikacij i nauchnoj informacii // *Informaciya i innovacii*. 2019. V. 14. № 1. P. 7-17.

3. Dudnikova O. V., Bogomolov A. A., Smirnova O. A. Rol' biblioteki v informacionno-analiticheskom soprovozhdenii nauchno-

publikacionnoj
deyatelnosti vuza // Universitetskaya kniga.
<http://www.unkniga.ru/biblioteki/vuzbiblio/8200-rol-biblioteki-v-informatsionno-analiticheskomp-soprovozhdaetsya.html>

4. Integral'naya ocenka effektivnosti nauchno-issledovatel'skih rabot v nauchnyh uchrezhdeniyah Minzdrava Rossii / G. S. Lebedev, O. B. Krylov, A. I. Lelyakov, YU. G. Mironov, V. V. Tkachenko // Zhurnal telemeditsiny i elektronnoho zdravoohraneniya.
<http://jtelemed.ru/article/integralnaja-ocenka-jeffektivnosti-nauchno-issledovatel'skih-rabot-v-nauchnyh-uchrezhdenijah->

5. Ispol'zovanie RINC i Science Index dlya analiza i ocenki nauchnoj deyatelnosti / V. Gluhov, A. Blinkova, A. Kovaleva [et al] // Universitetskaya kniga. 2018. May. P. 70.
<http://www.unkniga.ru/biblioteki/vuzbiblio/8381-ispolzovanie-rinc-i-science-index-dlya-analiza-i-otsenki-nauchnoy.html>

6. Kvelidze-Kuznecova N. N., Morozova S. A., Matyushenko A. D. Programmnyj modul' sbora i analiza dannyh o naukometricheskikh pokazatelyah sotrudnikov: integracionnye vozmozhnosti i perspektivy razvitiya // Nauka i nauchnaya informatsiya. 2019. V. 2. № 4. P. 216–227.
<https://doi.org/10.24108/2658-3143-2019-2-4-216-227>

7. Kychakova A. V., Treskova P. P. Otrazhenie mezhdisciplinarnogo razvitiya nauki cherez sistemu bibliometricheskikh pokazatelej // Informacionnaya shkola molodogo uchenogo. Ekaterinburg: Central Scientific Library of the Ural Branch of the RAS, 2011. P. 48–69.

8. Lavrik O. V., Pleshakova M. A., Kalyuzhnaya T. A. Informacionno-analiticheskie produkty v nauchnyh bibliotekah dlya informacionnogo obespecheniya nauchno-issledovatel'skoj raboty // Vestnik Tomskogo gosudarstvennogo universiteta. Kul'turologiya i iskusstvovedenie. 2018. № 32. P. 186–201.

9. Lavrik O. V., Pleshakova M. A., Kalyuzhnaya T. A. Sistematicheskij obzor kak vid obzorno-analiticheskikh produktov // Bibliotekovedenie. 2019. № 2. P. 33–51.

10. Makeeva O. V. Aktual'nye problemy raboty s personalom bibliotek na sovremennom etape razvitiya bibliotechnogo dela // Bibliosfera. 2018. № 4. P. 48–55.

11. Metodika naukometricheskoy ocenki razvitiya nauchnyh organizacij Ural'skogo Federal'nogo okruga / Y. D. Prokof'eva, A. S. Pavlova, T. A. Osipenko, L. A. Obolenskaya // Informatsiya i innovatsii. 2018. № 4. P. 76–88.

12. Metodika raschyota kachestvennogo pokazatelya gosudarstvennogo zadaniya «Kompleksnyj ball publikacionnoj rezul'tativnosti» / Ministry of Science and Higher Education of the Russian Federation.

<http://ntr.rf/upload/iblock/f35/Metodika%20rasceta%20kompleksnogo%20balla%20publikacionnoj%20rezul'tativnosti.pdf>

13. Moskal'yova O. V., Akoev M. A. Naukometriya: nemnogo istorii i sovremennye Rossijskie realii // Upravlenie naukoy: teoriya i praktika. 2019. № 1. P. 135–148.

14. Nauchno-issledovatel'skij portal Ural'skogo federal'nogo universiteta. <https://science.urfu.ru/ru/>

15. Otdel analiza naukometriceskih pokazatelej VGTU. <https://cchgeu.ru/university/struktura/tsentr-publikacionnoj-aktivnosti/otdel-analiza-naukometriceskikh-pokazatelej/>

16. Otdel nauchnogo menedzhmenta i naukometrii MGLU. <https://linguanet.ru/ob-universitete/administrativnye-podrazdeleniya/otdel-nauchnogo-menedzhmenta-i-naukometrii/>

17. Pirozhok T. V., Treskova P. P. Bibliometriceskij analiz publikacij sotrudnikov institutov Ural'skogo otdeleniya RAN, vhdoyashchih v OUS po gumanitarnym i ekonomicheskim naukam za period 2008–2012. Po bazam dannyh SCOPUS i RINC // III Informacionnaya shkola molodogo uchenogo. Ekaterinburg : Central Scientific Library of the Ural Branch of the RAS, 2013. P. 97–102.

18. Pirozhok T. V., Treskova P. P. Mirovye i otechestvennye resursy kak baza naukometriceskih issledovanij // Informacionnaya shkola molodogo uchenogo. Ekaterinburg : Central Scientific Library of the Ural Branch of the RAS, 2011. P. 37–47.

19. Prokof'eva Y. D. Kriterii ocenok razvitiya mezhdisciplinarnyh issledovanij: k postanovke problem // Bibliosfera. 2018. № 4. P. 95–100.

20. Traul'ko M. V., Pashkov P. M. Podhody k postroeniyu informacionnoj sistemy tekushchih issledovanij v vuze: analiz, ocenka i razrabotka metodiki vybora // Innovacii v obrazovanii. 2017. № 4(23). P. 139–161.

21. Treskova P. P. Nauka v informacionnom izmerenii: analiz publikacionnoj aktivnosti uchenyh s ispol'zovaniem baz dannyh «Web of Science» i «Scopus» // Informacionnoe obespechenie nauki: novye tekhnologii. Moscow, 2009. P. 253–262.

22. Treskova P. P. Primenenie bibliometriceskih pokazatelej dlya ocenki rezul'tativnosti deyatel'nosti nauchnyh organizacij Ural'skogo otdeleniya RAN. Novye tekhnologii v informacionno-bibliotechnom obespechenii nauchnyh issledovanij. Ekaterinburg, 2010. P. 226–235.

23. Trubnikov V. S., Tural'chuk K. A. Proektirovanie sistemy sbora, analiza i vizualizacii naukometriceskih dannyh // Problemy sovremennoj nauki i obrazovaniya. 2015. № 36. <https://cyberleninka.ru/article/n/proektirovanie-sistemy-sbora-analiza-i-vizualizatsii-naukometriceskih-dannyh>

24. Federal'noe gosudarstvennoe byudzhethoe uchrezhdenie nauki «Central'naya nauchnaya biblioteka Ural'skogo otdeleniya Rossijskoj

akademii nauk»: otchet o nauchno-issledovatel'skoj rabote (organizaciji) za 2019 god / Ministerstvo nauki i vysshego obrazovaniya Rossijskoj Federacii. Ekaterinburg, 2019. 246 p.

25. Centr naukometrii i publikacionnoj aktivnosti MGUPP. <http://www.mgupp.ru/science/center/>

26. Centr naukometrii i razvitiya nauchnyh zhurnalov RUDN. <http://old.rudn.ru/?pagec=6084>

27. Centr naukometriceskogo analiza i mezhdunarodnyh sistem indeksirovaniya VolGAU. <http://www.volgau.com/deyatel'nost'/nauka-i-innovacii/centr-naukometriceskogo-analiza-i-mezhdunarodnyh-sistem-indeksirovaniya>

28. Åström F., Hansson J. How implementation of bibliometric practice affects the role of academic libraries // Journal of Librarianship and Information Science. 2013. V. 45. № 4. P. 1-7.

29. Čadovska I., Mitrović G. Uloga knjižnica u vrednovanju znanstvene produktivnosti: primjer Nacionalne i sveučilišne knjižnice u Zagrebu [The role of libraries in evaluating scientific productivity: the example of the National and University Library in Zagreb] // Vjesnik bibliotekara Hrvatske. 2018. V. 61. № 2. P. 179-198. <https://scholar.google.ru/citations?user=IUvMe1AAAAAJ&hl=ru>

30. Competencies for bibliometrics / A. Cox, E.Gadd, S. Petersohn, L. Sbaffi // Journal of Librarianship and Information Science. 2019. V. 51. № 3. P. 746-762. <https://journals.sagepub.com/doi/full/10.1177/0961000617728111>

31. Gumpenberger C., Wieland M., Gorraiz J. Bibliometric practices and activities at the University of Vienna // Library Management. 2012. V. 33. № 3. P. 174-183.

32. Kaarsted T. Catch-22: How do academic libraries position themselves between top level management expectations and scholars hopes and dreams? // Library Management. 2017. V. 38. № 2/3. P. 78-87.

Профессионал научной инфосферы как субъект единого цифрового пространства научных знаний

М. А. Пекшева

ЦНБ УрО РАН

Аннотация. Зарубежные библиотеки, функционирующие в исследовательской среде, имеют большой опыт в части профессиональной адаптации к цифровой трансформации и изменениям инфраструктуры научных исследований и коммуникаций. В статье представлен обзор зарубежных материалов, посвященных вопросам актуальных компетенций информационно-библиотечных специалистов в области научных коммуникаций. На основе изученного материала представлен перечень знаний и навыков, информационных специалистов, востребованных в современной научной инфосфере. Сделан вывод о необходимости адаптации представленного перечня знаний и навыков к специфике отечественных библиотек и потребностям отрасли.

Ключевые слова: научные библиотеки, научные коммуникации, компетенции информационно-библиотечных специалистов, профессиональная подготовка, повышение квалификации, зарубежные библиотеки.

Professional of the scientific InfoSphere as a subject of the unified digital space of scientific knowledge

М. А. Peksheva

C SL UrB RAS

Abstract. Foreign libraries operating in the research environment have extensive experience in professional adaptation to digital transformation and changes in the research and communication infrastructure. The article presents a review of foreign materials on the issues of current competencies of information and library specialists in the field of scientific communications. Based on the studied material, a list of knowledge and skills of information specialists that are in demand in the modern scientific InfoSphere is presented. The conclusion is made about the need to adapt the

presented list of knowledge and skills to the specifics of domestic libraries and the needs of the industry.

Keywords: scientific libraries, scientific communications, competence of information and library specialists, professional training, professional development, foreign libraries.

1. Введение

Современная информационная парадигма определяет высокую значимость информационной инфраструктуры и её акторов для развития экономики и общества. В полной мере это касается инфраструктуры научно-технической информации, понятие которой может быть приравнено к понятиям информационного пространства науки и научной инфосферы [1]. Одним из ключевых субъектов научной инфосферы выступают научно-информационные службы, в том числе, научные библиотеки, функционирующие в исследовательской среде. В таких библиотеках накоплен богатый опыт информационной деятельности по сбору, обработке, передаче, хранению и предоставлению научной информации. Это определяет их высокую значимость для развития научной инфосферы. Однако, в условиях повсеместной цифровой трансформации и коренных изменений в сфере научных коммуникаций, накопленный опыт не является достаточным условием высокой эффективности и востребованности. Можно утверждать, что для полноценного участия в процессах роста и развития цифрового информационного пространства науки, научным библиотекам требуется прилагать значительные усилия для самосовершенствования. При этом в центре внимания должны находиться не только технологические и организационные аспекты, но и вопросы профессиональной подготовки и непрерывного развития сотрудников научных библиотек.

Отечественные авторы исследуют основные катализаторы изменений в сфере научных коммуникаций и возникающие на их фоне новые роли и компетенции [2-7]. В качестве катализаторов выступают: цифровизация; непрерывный рост объемов научных материалов и способов их распространения; популяризация идей открытой науки,

возрастание роли исследовательских данных и ряд других факторов. Анализ литературы показывает, что отечественные авторы, рассуждая о будущем научных библиотек и информационно-библиотечной профессии, активно ссылаются на зарубежных коллег. О передовом опыте зарубежных научных библиотек в области профессионального развития сотрудников также свидетельствуют проекты, направленные на повышение квалификации сотрудников и организацию непрерывного развития [8, 13, 28, 33].

Цель данного исследования – выявление компетенций, знаний и навыков информационных специалистов, востребованных в современной научной инфосфере, на основе анализа зарубежной профессиональной литературы. Исследование ограничивается знаниями и навыками, наиболее актуальными в области научных коммуникаций. Отдельным предметом, не рассматриваемым в данной статье, выступают общие компетенции и навыки, которые могут быть положены в основу работы информационно-библиотечного специалиста любой направленности. Базой исследования выступили статьи зарубежных авторов и материалы профессиональных ассоциаций, опубликованные в период с 2015 по 2020 г. Поиск публикаций по теме осуществлялся в базе Scopus, в открытых репозиториях и на сайтах профессиональных ассоциаций (IFLA Library, NASIG, American Library Association, Online Computer Library Center – OCLC и других).

2. Профессиональные компетенции в области научных коммуникаций

Современная система научных коммуникаций имеет сложный и динамично меняющийся характер. Зарубежные авторы активно обсуждают новые роли и компетенции информационно-библиотечных работников, возникающие на фоне трансформации системы научной коммуникации и отмечают их высокую значимость в современном исследовательском пространстве. Характеризуя деятельность информационно-библиотечных специалистов в области науч-

ных коммуникаций, авторы говорят о ней, как о «ключевой компетенции для академических библиотечных специалистов» [9]; «будущем академической библиотечной деятельности» [15]; «центральной услуге, которую могут предоставить библиотеки и в которой библиотекари должны быть квалифицированными специалистами» [10]. Наблюдается тенденция смещения акцента с обеспечения доступа и информационного обслуживания к активной поддержке исследовательского процесса [9, 10, 15, 24, 26, 30].

Зарубежные специалисты и организации ведут работу по разработке специализированных профилей компетенций в области научных коммуникаций. Одним из таких документов является «Профиль компетенций библиотечных специалистов для научной коммуникации и открытого доступа» (Librarians' Competencies Profile for Scholarly Communication and Open Access) [17], опубликованный целевой группой по компетенциям библиотечных специалистов в поддержку электронных исследований и научной коммуникации в июне 2016 года. В том же году отдельным документом опубликован «Профиль компетенций библиотечных специалистов для управления исследовательскими данными» (Librarians' Competencies Profile for Research Data Management, 2016) [25]. В 2017 году утвержден документ «Основные компетенции NASIG для библиотечных специалистов научной коммуникации» [21], разработанный независимой профессиональной ассоциацией библиотечных специалистов и специалистов в области академических изданий NASIG. Эти документы призваны структурировать и конкретизировать компетенции информационно-библиотечных специалистов в сфере научных коммуникаций. Несмотря на разницу в форматах представления и степени детализации, в документах можно выделить следующие общие группы ключевых компетенций в сфере научных коммуникаций: фоновые знания; знания и навыки в сфере ИКТ; компетенции в области управления институциональным репозиторием; компетенции в области издательских услуг; услуг по защите авторских прав; услуг по управлению данными; оценки научных исследований.

Важно отметить, что профили компетенций и иные аналогичные наборы компетенций в области научных коммуникаций представляют собой общие рамки, которые должны быть адаптированы к конкретным ролям и задачам в конкретной организации. Это отмечают многие зарубежные авторы [10, 14, 24, 30]. Об этом же свидетельствует вступление к профилю NASIG, в котором подчеркивается, что один специалист не может обладать полным набором компетенций в области научных коммуникаций и разнообразие является «единственной константой в должностных обязанностях библиотекаря научной коммуникации» [21].

Авторские концепции и компиляции во многом опираются на вышеперечисленные и иные профили компетенций, а также дополняются обзорами литературы, данными анализа объявлений о вакансиях, опросов и дискуссий. Heather Moulaison Sandy [24] и соавторы провели обширное исследование и сопоставили компетенции и навыки в области научных коммуникаций, закрепленные в трех документах:

«Профиль компетенций библиотекарей для управления исследовательскими данными» (Librarians' Competencies Profile for Research Data Management, 2016), «Матрица знаний и компетенций цифрового кураторства» (Matrix of Digital Curation Knowledge and Competencies, 2009) и «Подготовка кадров для цифрового курирования» (Preparing the Workforce for Digital Curation, 2015). Авторы отмечают высокий уровень согласованности документов и приводят 23 категории навыков в области цифрового кураторства. Сопоставив компетенции и навыки из изученных документов с жизненным циклом исследования, авторы приходят к выводу: все аспекты исследовательского процесса в той или иной степени могут быть поддержаны информационно-библиотечным специалистом в области научной коммуникации.

В основе исследования Maria Bonn [9] положен анализ объявлений о вакансиях, а также дискуссий с учеными и библиотекарями. Автор приводит перечень навыков и областей знаний, необходимых информационно-

библиотечным специалистам в сфере научных коммуникаций, включающий: компетенции в области авторского права, способов научной коммуникации, издательской деятельности, экономики научной коммуникации, цифровых гуманитарных наук, управления исследовательскими данными, цифровых хранилищ.

Во многих зарубежных публикациях не приведена структурированная информация о компетенциях специалиста в области научной коммуникации, но они подтверждают актуальность ранее перечисленных областей поддержки исследователей и являются источником интересного практического опыта [10, 15, 16, 26, 29, 32].

В качестве отдельного аспекта научных коммуникаций активно рассматриваются вопросы, связанные с управлением исследовательскими данными и компетенциями информационно-библиотечных специалистов, необходимыми для этой сферы. Единого терминологического подхода к определению понятий «управление исследовательскими данными», «кураторство данных», «цифровое кураторство» не выработано. По примеру исследования [14], посвященного терминологическому анализу сферы управления исследовательскими данными, далее эти термины используются как взаимозаменяемые. В зарубежной литературе часто цитируется определение цифрового кураторства, разработанное Центром цифрового кураторства (Digital Curation Centre, DCC). Данное определение фокусируется на «поддержании, сохранении и повышении ценности цифровых исследовательских данных на протяжении всего их жизненного цикла» [цит. по 29].

Anna Maria Tammaro [14] и соавторы провели масштабное исследование, посвящённое определению ролей и обязанностей кураторов данных во всем мире. Инициатором исследования выступает Секция библиотечной теории и исследований ИФЛА. Одним из предметов изучения выступают профессиональные компетенции, выводы о которых основаны на анализе объявлений о вакансиях и опросе специалистов, работающих в области курирования данных. Авторы приводят две ключевые группы компетенций: коммуникатив-

ные («мягкие») навыки и технические компетенции. Категория «мягких» навыков включила: способность эффективного взаимодействия с учёными, навыки межличностного общения и сотрудничества, презентационные навыки, навыки подготовки учебных и информационных материалов. Категория технических компетенций включила комплексы навыков по управлению данными, описанию данных и их документации, публикации, архивированию и сохранению. Одним из важных аспектов успешного выполнения роли куратора данных указано понимание процесса и методов исследования.

Andrew M. Cox [19] и соавторы в 2019 году опубликовали исследование, посвященное развитию в библиотеках служб, направленных на консультативную и техническую поддержку процессов управления исследовательскими данными. Результаты исследования основаны на опросе, проведенном в библиотеках Австралии, Канады, Германии, Ирландии, Нидерландов, Новой Зеландии и Великобритании. Авторы отмечают следующие профессиональные навыки сотрудников исследуемых служб: кураторство данных, технические навыки и навыки в области ИКТ, предметные и дисциплинарные знания, знание жизненного цикла и методов исследования, описание данных и документация, юридические, политические и консультационные навыки, понимание целостности исследований, принципов воспроизводимости и прозрачности.

Перечисленные навыки тем или иным образом представлены и в других статьях, посвященных деятельности библиотек в области управления исследовательскими данными [18, 31]. Многие зарубежные авторы придерживаются мнения, что кураторство данных — процесс, требующий консолидации усилий различных специалистов и отделов научной организации (непосредственно исследователей, отдела информационных технологий, юридического отдела и т. д.). Для оказания реальной поддержки исследовательскому процессу информационно-библиотечный специалист должен выступать частью большой квалифицированной команды.

В качестве отдельных направлений рассмотрения ролей и компетенций библиотечных специалистов выступают открытый доступ [23, 27] и библиометрия [20, 22, 28, 50]. Публикации в этих областях носят обобщенный характер и позволяют составить представление о наиболее актуальных путях развития библиотечных специалистов в данных областях.

На основе контент-анализа представленных в статьях материалов разработан перечень знаний и навыков информационно-библиотечного специалиста в сфере научных коммуникаций, востребованных в современной научной инфосфере (табл. 1). Для каждого элемента перечня знаком «+» отмечена его актуальность для одной или нескольких сфер профессиональной деятельности, характерных для современной научной коммуникации. Для возможности ознакомления с оригинальными исследованиями указаны ссылки на первоисточники. Представленный перечень является компиляцией материалов, опубликованных в профессиональной зарубежной печати за последние 5 лет и не претендует на полноту отражения всех возможных и необходимых знаний и навыков. Актуальность той или иной группы для конкретной сферы также носит дискуссионный характер, поскольку авторские концепции могут рассматривать компетенции в рамках конкретного вида деятельности, не учитывая их востребованность в смежных направлениях работы.

Необходимо ещё раз подчеркнуть, что один конкретный специалист не может обладать всеми представленными знаниями и навыками. Перечень отражает идеализированное представление о совокупных знаниях и навыках всех специалистов, связанных с научными коммуникациями. Глубина освоения того или иного навыка для различных сфер деятельности будет отличаться, их набор будет варьироваться в зависимости от занимаемой должности и конкретных функциональных обязанностей. Перечень может служить основой для определения компетенций и путей профессионального развития информационно-библиотечного специалиста в области научных коммуникаций.

Таблица 1

**Знания и навыки информационно-библиотечных специалистов, востребованные
в современной научной инфосфере**

Знания и навыки в области научных коммуникаций	Информационно-пропагандистская и научно-издательская деятельность, открытый доступ и авторское право	Управление репозиториями	Управление исследовательскими данными	Оценка научных исследований	Ссылка на источник
1	2	3	4	5	5
Знание национального и мирового законодательства в области авторского права, интеллектуальной собственности и защите данных	+	+	+	+	[20], [24], [25], [27]
Осведомленность о судебной практике в области авторского права	+				[24]
Знание политики и требований финансирующих организаций в области открытого доступа и авторского права	+		+		[18], [24]

Знания и навыки в области научных коммуникаций	Информационно-пропагандистская и научно-издательская деятельность, открытый доступ и авторское право	Управление репозиториями	Управление исследовательскими данными	Оценка научных исследований	Ссылка на источник
1	2	3	4	5	5
Способность разрабатывать политику, связанную с открытым доступом, развитием коллекций, цифровым сохранением, авторским правом и авторскими услугами, а также эмбарго издателей		+			[24]

Продолжение табл. 1

1	2	3	4	5	5
Знание лицензий открытого доступа и связанных с ними процессов лицензирования	+		+		[23], [24], [27]
Знание системы традиционной научной издательской системы, официальной публикации и редакционных процессов	+				[20]

Знание и опыт работы с коммерческими и открытыми издательскими платформами, программными средствами	+			+	[24], [30]
Знание и опыт работы с полным жизненным циклом публикации	+				[24]
Знания и опыт работы с созданием цифровых идентификаторов (Crossref, Datacite DOI, Handle, ORCiD, ISSN и т. д.)	+		+		[24], [27]
Предоставление технической поддержки в вопросах размещения публикаций, выполнение системного администрирования и программирования	+				[24]

Продолжение табл. 1

1	2	3	4	5	5
Знание принципов работы, проблем, возможностей цифровых хранилищ (дисциплинарных и институциональных репозиториях)	+	+	+	+	[12], [20], [24], [31]

Знание и навык работы с программным обеспечением для создания репозитория, его администрирования и обновления		+			[12], [24]
Навыки работы с исследователями по депонированию результатов исследований в хранилище, сбору и сохранению интеллектуальной продукции		+			[24]
Знание и умение применять политику издателя при архивировании		+			[24]
Навык сбора и распространения оценочных показателей, отчетной статистики	+	+	+		[24], [27]
Знание существующих центров обработки данных, хранилищ и коллекций данных и механизмы обнаружения данных			+		[23]
Знание политики и требований спонсоров в области управления данными и требований к публикации данных в журналах			+		[23], [24]

Продолжение табл. 1

1	2	3	4	5	5
---	---	---	---	---	---

Знание лучших практик для структур данных, типов, форматов, словарей, онтологий и метаданных или где найти информацию о них	+	+	+		[23], [24], [27]
Знание планов и инструментов управления данными			+		[38], [24], [27]
Навыки управления исследовательскими данными (навыки очистки и проверки файлов, обнаружения, организации, преобразования, лицензирования, цитирования, анализа, аудита и оценки данных)			+		[18], [20], [23], [24], [25], [27]
Навыки описания данных и создания документации (создание метаданных с использованием стандартизированных схем и словарей, интеграция и формирование связей данных)			+		[18], [20], [23], [24], [25], [27]
Навыки публикации и депозитного хранения данных (размещение в системы репозитория, назначение идентификаторов, анонимизация, архивирование, визуализация, безопасность данных)			+		[18], [20], [23], [24], [25], [27]

Продолжение табл. 1

1	2	3	4	5	5
Навыки информационно-пропагандистской деятельности в области управления исследовательскими данными			+		[18], [23], [27]
Способность четко сформулировать преимущества совместного использования и повторного использования данных, организовать доступ к наборам данных для третьих сторон			+		[23], [24]
Понимание дисциплинарных норм и стандартов для управления данными			+		[23]
Знание библиометрии как исследовательской специальности и практической области работы				+	[30]
Знания и навыки работы с базами данных цитирования, библиометрическими показателями, альтметриками				+	[30]
Навыки сбора данных, мониторинга и анализа публикационной активности и научной продуктивности				+	[31]
Навык оказания консультационных и справочных услуг в области библиометрии				+	[31]

2. Заключение

Анализ публикаций зарубежных авторов показывает, что при условии непрерывного повышения квалификации и развития актуальных компетенций информационно-библиотечные специалисты являются востребованными профессионалами современной информационной инфраструктуры науки. Сотрудники научных библиотек, функционирующих в исследовательской среде, находятся в значительной мере под влиянием развивающихся цифровых технологий и преобразований сферы научных коммуникаций. Это обуславливает трансформацию старых наборов компетенций и возникновение новых востребованных групп знаний и навыков, ориентированных на цифровую среду. Непрерывное развитие компетенций информационного специалиста должно опираться на следующие тенденции в области научных коммуникаций: необходимость квалифицированной поддержки исследовательских процессов, пропаганда открытого доступа и консультирования в области авторского права, лицензирование и управление исследовательскими данными, создание и поддержка репозиториев, анализ результатов научной деятельности. При этом специалистам в области научных коммуникаций важно развивать навыки и оказывать услуги таким образом, чтобы они облегчали работу исследователя, а не воспринимались как навязчивые или ограничивающие.

Представленный по итогам исследования перечень знаний и навыков информационного специалиста в области научных коммуникаций позволяет составить представление о компетенциях, которые, по мнению зарубежных авторов, востребованы сегодня и будут актуальны в ближайшем будущем. Как субъекты единого цифрового пространства научных знаний, информационные специалисты, обладающие перечисленными компетенциями, обеспечивают успешное функционирование и развитие всех остальных компонентов научной инфосферы в цифровой среде. Важно отметить, что полученная совокупность не учитывает специфику отечест-

венного информационного пространства. Это является основанием для дальнейшего изучения компетенций информационно-библиотечных специалистов в области научных коммуникаций с целью их уточнения и адаптации к потребностям отечественной научной инфосферы.

Литература

1. Антопольский А. Б. О разработке Единого российского электронного пространства знаний // Научное издание международного уровня – 2019: стратегия и тактика управления и развития: материалы 8-й Международ. науч.-практ. конф., Москва, 23–26 апреля 2019 г. Екатеринбург: Изд-во Урал. ун-та, 2019. С. 17–29.
2. Антопольский А. Б. Будущее научных коммуникаций и научной информации // Информация и инновации. 2019. Т. 14, № 1. С. 7–17.
3. Гуськов А. Е., Косяков Д. В., Макеева О. В. Матрица задач, ресурсов и компетенций для научных библиотек // Библиосфера. 2019. № 3. С. 35–46.
4. Парамонова И. Е. Модели компетенций для сотрудников специальных библиотек: зарубежный опыт // Научные и технические библиотеки. 2019. № 7. С. 3–13.
5. Редькина Н. С. «Надпрофессиональные» навыки и профессиональные знания библиотечного специалиста: требования времени // Библиотековедение. 2019. Т. 68, № 6. С. 647–658.
6. Редькина Н. С. Мировые тенденции развития библиотек: оптимизм vs пессимизм (по материалам зарубежной литературы). Часть 1 // Библиосфера. 2018. № 4. С. 87–94.
7. Редькина Н. С. Мировые тенденции развития библиотек: оптимизм vs пессимизм (по материалам зарубежной литературы). Часть 2 // Библиосфера. 2019. № 1. С. 49–58.
8. Barbuti N., Giorgio S. Di, Valentini A. The Project BIBLIO-Boosting Digital Skills and Competencies for Librarians in Europe: An Innovative Training Model for Creating Digital Librarian // International Information & Library Review. 2019. V. 51 (4). P. 300–304.
9. Bonn M. Tooling up: Scholarly communication education and training // College & Research Libraries News. 2014. V. 75 (3). P. 132–135.
10. Brantley S., Bruns T. A., Duffin K. I. Librarians in transition: Scholarly communication support as a developing core competency // Journal of Electronic Resources Librarianship. 2017. V. 29 (3). P. 137–150.
11. Competencies for bibliometrics / A. Cox, E. Gadd, S. Petersohn, etc. // Journal of Librarianship and Information Science. 2019. V. 51 (3). P. 746–762.

12. Competencies for Information Professionals / [Special Libraries Association]. <http://www.4info-management.com/pdf/SLACompetencies.pdf>
13. Cook E. I., Zanders T. An Interview with Tony Zanders, Founder of Skilltype // The Serials Librarian. 2020. <https://www.tandfonline.com/doi/pdf/10.1080/0361526X.2020.1722782?needAccess=true>
14. Data curator's roles and responsibilities: An international perspective / A. M. Tammaro, K. K. Matusiak, F. A. Sposito, etc. // Libri. 2019. V. 69 (2). P. 89-104.
15. Harrison R. Reshaping Academic Librarianship in the Context of Twenty-First Century Scholarly Communications and Information Fluency: A Case Study of the Scholarly Communications Management Team at Imperial College London Library Services // New Review of Academic Librarianship. 2018. V. 24 (3-4). P. 351-364.
16. Koltay T. Identifying new roles for academic libraries in supporting data-intensive research // Bibliosphere. 2020. V. 34 (4). P. 97-102.
17. Librarians' Competencies Profile for Scholarly Communication and Open Access / P. Calarco, K. Shearer, B. Schmidt, etc. 2016. https://www.coar-repositories.org/files/Competencies-for-ScholComm-and-OA_June-2016.pdf/
18. Martin C., Cadiou C., Jannès-Ober E. Data management: New tools, new organization, and new skills in a French research institute // Liber Quarterly. 2017. V. 27 (1). P. 73-88.
19. Maturing research data services and the transformation of academic libraries / A. M. Cox, M. A. Kennan, L. Lyon, etc. // Journal of Documentation. 2019. V. 75 (6). P. 1432-1462.
20. Mihaljevi J. The New Competency Profile of Academic Libraries in the Function of Evaluation of Scientific Productivity // Ekonomski Vjesnik. 2015. V. 28 (2). P. 535-549.
21. NASIG Core Competencies for Scholarly Communication Librarians / A. Wesolek, W. J. Thomas, A. Dresselhaus, etc. 2017. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1055&context=scholcom>.
22. Petersohn S. Professional competencies and jurisdictional claims in evaluative bibliometrics: The educational mandate of academic librarians // Education for Information. 2016. V. 32 (2). P. 165-193.
23. Rodriguez J. E. Scholarly Communications Competencies: Open Access Training for Librarians // New Library World. 2015. V. 116 (7/8). <http://dx.doi.org/10.1108/NLW-12-2014-0140/>
24. Sandy H. M., Million A. J., Hudson-Vitale C. Innovating support for research: The coalescence of scholarly communication? // College & Research Libraries. 2020. V. 81 (2). P. 193-214.

25. Schmidt B., Shearer K. Librarians' Competencies Profile for Research Data Management. 2016. https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf
26. Sewell C., Kingsley D. Developing the 21st Century Academic Librarian: The Research Support Ambassador Programme // *New Review of Academic Librarianship*. 2017. V. 23 (2-3). P. 148-158.
27. Sidney E. The Library Profession in the Time of Open Access // *The Serials Librarian*. 2017. V. 73 (3-4). P. 215-225.
28. Stevenson S., Hannaford J. Workplace-integrated-learning: Preparing tomorrow's academic library workforce // *Journal of Academic Librarianship*. 2019. V. 45 (3). P. 234-241.
29. Tammara A. M. Heritage Curation in the Digital Age: Professional Challenges and Opportunities // *International Information & Library Review*. 2016. V. 48 (2). P. 122-128.
30. Time to adopt: Librarians' new skills and competency profiles / B. Schmidt, P. Calarco, I. Kuchma, etc. // *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016*. 2016. P. 1-8.
31. Walek A. Is data management a new «digitisation»? A change of the role of librarians in the context of changing academic libraries' tasks // *IFLA Wlic* 2018. 2018. P. 1-11. <http://library.ifla.org/2247/1/139-walek-en.pdf>
32. Warren E. Is the doctor in? PhD to professional: Complementary perspectives in research libraries // *Liber Quarterly*. 2018. V. 28 (1). P. 1-21.
33. WebJunction: the learning place for librarian // [OCLC]. <https://www.webjunction.org/home.html>

References

1. Antopol'skij A. B. O razrabotke Edinogo rossijskogo jelektronnogo prostranstva znaniy // *Nauchnoe izdanie mezhdunarodnogo urovnja – 2019: strategija i taktika upravljenja i razvitija: materialy 8-j Mezhdunarod. nauch.-prakt. konf., Moskva, 23-26 aprelja 2019 g. Ekaterinburg: Izd-vo Ural. un-ta, 2019. P. 17-29.*
2. Antopol'skij A. B. Budushhee nauchnyh komunikacij i nauchnoj informacii // *Informacija i innovacii*. 2019. V. 14, № 1. P. 7-17.
3. Gus'kov A. E., Kosjakov D. V., Makeeva O. V. Matrica zadach, resurov i kompetencij dlja nauchnyh bibliotek // *Bibliosfera*. 2019. № 3. P. 35-46.
4. Paramonova I. E. Modeli kompetencij dlja sotrudnikov special'nyh bibliotek: zarubezhnyj opyt // *Nauchnye i tehničeskie biblioteki*. 2019. № 7. P. 3-13.

5. Red'kina N. S. «Nadprofessional'nye» navyki i professional'nye znaniya bibliotechnogo specialista: trebovaniya vremeni // Bibliotekovedenie. 2019. V. 68, № 6. P. 647–658.
6. Red'kina N. S. Mirovye tendencii razvitija bibliotek: optimizm vs pessimizm (po materialam zarubezhnoj literatury). Pt. 1 // Bibliosfera. 2018. № 4. P. 87–94.
7. Red'kina N. S. Mirovye tendencii razvitija bibliotek: optimizm vs pessimizm (po materialam zarubezhnoj literatury). Pt. 2 // Bibliosfera. 2019. № 1. P. 49–58.
8. Barbuti N., Giorgio S. Di, Valentini A. The Project BIBLIO-Boosting Digital Skills and Competencies for Librarians in Europe: An Innovative Training Model for Creating Digital Librarian // International Information & Library Review. 2019. V. 51 (4). P. 300–304.
9. Bonn M. Tooling up: Scholarly communication education and training // College & Research Libraries News. 2014. V. 75 (3). P. 132–135.
10. Brantley S., Bruns T. A., Duffin K. I. Librarians in transition: Scholarly communication support as a developing core competency // Journal of Electronic Resources Librarianship. 2017. V. 29 (3). P. 137–150.
11. Competencies for bibliometrics / A. Cox, E. Gadd, S. Petersohn, etc. // Journal of Librarianship and Information Science. 2019. V. 51 (3). P. 746–762.
12. Competencies for Information Professionals / [Special Libraries Association]. <http://www.4info-management.com/pdf/SLACompetencies.pdf>
13. Cook E. I., Zanders T. An Interview with Tony Zanders, Founder of Skilltype // The Serials Librarian. 2020. <https://www.tandfonline.com/doi/pdf/10.1080/0361526X.2020.1722782?needAccess=true>
14. Data curator's roles and responsibilities: An international perspective / A. M. Tammaro, K. K. Matusiak, F. A. Sposito, etc. // Libri. 2019. V. 69 (2). P. 89–104.
15. Harrison R. Reshaping Academic Librarianship in the Context of Twenty-First Century Scholarly Communications and Information Fluency: A Case Study of the Scholarly Communications Management Team at Imperial College London Library Services // New Review of Academic Librarianship. 2018. V. 24 (3–4). P. 351–364.
16. Koltay T. Identifying new roles for academic libraries in supporting data-intensive research // Bibliosphere. 2020. V. 34 (4). P. 97–102.
17. Librarians' Competencies Profile for Scholarly Communication and Open Access / P. Calarco, K. Shearer, B. Schmidt, etc. 2016. https://www.coar-repositories.org/files/Competencies-for-ScholComm-and-OA_June-2016.pdf

18. Martin C., Cadiou C., Jannès-Ober E. Data management: New tools, new organization, and new skills in a French research institute // *Liber Quarterly*. 2017. V. 27 (1). P. 73–88.
19. Maturing research data services and the transformation of academic libraries / A. M. Cox, M. A. Kennan, L. Lyon, etc. // *Journal of Documentation*. 2019. V. 75 (6). P. 1432–1462.
20. Mihaljevi J. The New Competency Profile of Academic Libraries in the Function of Evaluation of Scientific Productivity // *Ekonomski Vjesnik*. 2015. V. 28 (2). P. 535–549.
21. NASIG Core Competencies for Scholarly Communication Librarians / A. Wesolek, W. J. Thomas, A. Dresselhaus, etc. 2017. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1055&context=scholcom>
22. Petersohn S. Professional competencies and jurisdictional claims in evaluative bibliometrics: The educational mandate of academic librarians // *Education for Information*. 2016. V. 32 (2). P. 165–193.
23. Rodriguez J. E. Scholarly Communications Competencies: Open Access Training for Librarians // *New Library World*. 2015. V. 116 (7/8). <http://dx.doi.org/10.1108/NLW-12-2014-0140/>
24. Sandy H. M., Million A. J., Hudson-Vitale C. Innovating support for research: The coalescence of scholarly communication? // *College & Research Libraries*. 2020. V. 81 (2). P. 193–214.
25. Schmidt B., Shearer K. Librarians' Competencies Profile for Research Data Management. 2016. https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf
26. Sewell C., Kingsley D. Developing the 21st Century Academic Librarian: The Research Support Ambassador Programme // *New Review of Academic Librarianship*. 2017. V. 23 (2–3). P. 148–158.
27. Sidney E. The Library Profession in the Time of Open Access // *The Serials Librarian*. 2017. V. 73 (3–4). P. 215–225.
28. Stevenson S., Hannaford J. Workplace-integrated-learning: Preparing tomorrow's academic library workforce // *Journal of Academic Librarianship*. 2019. V. 45 (3). P. 234–241.
29. Tammara A. M. Heritage Curation in the Digital Age: Professional Challenges and Opportunities // *International Information & Library Review*. 2016. V. 48 (2). P. 122–128.
30. Time to adopt: Librarians' new skills and competency profiles / B. Schmidt, P. Calarco, I. Kuchma, etc. // *Positioning and Power in Academic Publishing: Players, Agents and Agendas : Proc. 20th Int. Conf. Electron. Publ. ELPUB 2016*. 2016. P. 1–8.
31. Walek A. Is data management a new «digitisation»? A change of the role of librarians in the context of changing academic libraries' tasks // *IFLA Wlic 2018*. 2018. P. 1–11. <http://library.ifla.org/2247/1/139-walek-en.pdf>

32. Warren E. Is the doctor in? PhD to professional: Complementary perspectives in research libraries // *Liber Quarterly*. 2018. V. 28 (1). P. 1-21.
33. WebJunction: the learning place for librarian // [OCLC].
<https://www.webjunction.org/home.html>

Методика оценки потенциала мобильных приложений для информационного обеспечения научных исследований

А. Ю. Герасименко

ЦНБ УрО РАН

Аннотация. Мобильные приложения все чаще упоминаются в библиотечном сообществе как перспективное направление развития информационного обслуживания. Однако разработка подобного программного обеспечения требует больших затрат, при этом не гарантируя достижения ожидаемых результатов.

Цель представленного исследования – разработка методики оценки потенциала мобильных приложений для совершенствования информационного обеспечения научных исследований. В ходе работы использовались сведения, представленные в Google Play Маркет. Проведен поиск, отбор и классификация мобильных приложений, содержащих научные материалы. Отмечена малая доля присутствия российских библиотек на рынке мобильных приложений.

Предложена двухэтапная методика оценки потенциала приложений. Первый этап – изучение спроса и предложения. Второй этап – изучение возможностей отобранных приложений. За основу анализа взяты мнения пользователей, выраженные через скачивания, оценку и отзывы. Выявлены характерные черты мобильного приложения, обладающего большим потенциалом для оптимизации информационного обеспечения науки.

Ключевые слова: мобильные приложения, информационное обеспечение, информационные технологии, научные библиотеки, научные электронные ресурсы.

Methodology for assessing the potential of mobile applications for information support of scientific research

A. Yu. Gerasimenko

C SL UrB RAS

Abstract. The library community increasingly is mentioning to mobile applications as a promising area for the development of the information community. But the development of this software is costly and does not guarantee the expected results.

The purpose of the study is to develop a methodology to assess the potential of mobile applications to improve the information support of scientific research. The work was based on the information provided in Google Play Market. Mobile applications containing scientific material have been searched, selected, and classified. There are few Russian libraries on the mobile app market.

A two-stage methodology for assessing the potential of applications has been proposed. The first stage is the examination of supply and demand. The second stage is exploring the possibilities of the selected applications. The analysis is based on the users' opinion, expressed through download, evaluation, and comment. The characteristic features of a mobile application with great potential for optimizing the information support of science are identified.

Keywords: mobile applications, information support, information technologies, scientific libraries, scientific electronic resources.

Научные библиотеки находятся в постоянном поиске новых подходов к формированию комфортной рабочей среды для ученых и специалистов. Одним из рассматриваемых направлений является внедрение в библиотеки современных технологий, основанных на применении мобильных устройств.

Интерес к мобильным устройствам как к инструментам для работы с информацией обусловлен их многофункциональностью, портативностью, прогрессивностью, ценовой доступностью, а также активным развитием и распространением беспроводных интернет-технологий.

На сегодняшний день данная категория гаджетов занимает лидирующую позицию среди средств доступа к информационным ресурсам и услугам. Согласно данным, предоставленным аналитическим агентством We Are Social и SMM-платформой Hootsuite, в 2019 году 56 % от общего объема мирового web-трафика приходилось на мобильные устройства (смартфоны и планшеты) [6, С. 54]. В сравнении с предыдущим годом наблюдается увеличение объема трафика, передаваемого через смартфоны (+8,6 %), и его умень-

шение для планшетов (-27 %), компьютеров и ноутбуков (-6,8 %).

Активное распространение мобильных устройств и беспроводных интернет-технологий подводит библиотеки к необходимости формирования собственных мобильных ресурсов [4], таких как:

- мобильный сайт – сайт, адаптированный под просмотр на мобильных устройствах (смартфонах, планшетах);
- мобильное приложение – программное обеспечение, специально созданное для работы на мобильных устройствах (смартфонах, планшетах).

Согласно статистике, пользователи мобильных устройств чаще обращаются к web-ресурсам через приложения, нежели через мобильные версии сайтов: в среднем из 3–4 часов [2], проводимых в мобильном интернете, около 2,5–3 часов тратится на приложения [7].

Мобильное приложение отличается от веб-сайта большей «отзывчивостью» (доступ в «одно касание»; возможность доступа к ресурсам в offline режиме), отсутствием лишних элементов в интерфейсе, наличием уникальных функций (например, управление через голосовые команды) [1, 5], встроенной маркетинговой поддержкой [8. С. 107]. При этом, существуют определенные препятствия, ставящие под сомнение целесообразность внедрения подобного программного обеспечения в процессы информационно-библиотечного обслуживания пользователей. Среди них выделяют большие временные и денежные затраты, наличие опытного специалиста-разработчика [5. С. 15], отсутствие гарантий достижения установленных целей. Таким образом, решение о создании собственного приложения должно быть эффективно оправданным.

Цель данного исследования – разработка методики оценки потенциала мобильных приложений для совершенствования информационного обеспечения научных исследований. Объектом изучения являются мобильные приложения – хранилища научной информации и сервисы для её предоставления.

Основой для проведения исследования послужили данные, представленные в магазине приложений Google Play

Маркет. В отличие от App Store, ассортимент и охват аудитории данной платформой значительно шире.

В процессе поиска подходящих мобильных приложений использовались следующие ключевые слова: *research, science, research journal, university library, ЭБС, научная библиотека, научная информация, научные журналы.*

В качестве критериев отбора установлены следующие характеристики: наличие материалов и функций, представляющих научную значимость; наличие оценки пользователей; количество скачиваний – не менее 1000.

Полученный в результате массив приложений разбит на четыре категории:

1. Мобильные приложения вузовских и академических библиотек – 11 наименований (2 – российских, 9 – зарубежных) (см. *табл. 1*). Предоставляют основные сведения о библиотеках, обеспечивают доступ к их информационным ресурсам и услугам.

2. Мобильные электронные библиотеки и библиотечные системы – 7 наименований (см. *табл. 2*). Содержат информационные материалы, предоставленные издательствами, институтами, библиотеками и прочими научными и образовательными организациями.

3. Мобильные приложения научных журналов – 54 наименования (см. *табл. 3*). Предоставляют доступ к отдельным наименованиям или коллекциям научных журналов.

4. Мобильные агрегаторы научной информации – 16 наименований (см. *табл. 4*). Наполнение приложений происходит за счет интеграции материалов из баз данных научной информации и журнальных платформ.

Таблица 1

**Мобильные приложения вузовских
и академических библиотек**

№	Наименование	Google Play Маркет
---	--------------	--------------------

		Кол-во ус- тановок	Кол-во оценок	Оценка пользо- вателей
1	2	3	4	5
	Sydney Uni Library	1 000+	5	4.8
	Личный кабинет читателя ГУНБК	1 000+	23	4.7
	Библиотека аль-Фараби	1 000+	29	4.7
	Journal Indexing Identifier (REVA University)	5 000+	32	4.5
	BARD Mobile	10 000+	386	4.3

Продолжение табл. 1

1	2	3	4	5
	UoL Library	5 000+	10	4
	Library Groningen University	5 000+	28	4
	University of Dundee Library	5 000+	23	3.7
	National Book Festival	5 000+	36	3.5
	Polimi Library	10 000+	192	3.3
	Фундаментальная библиотека ПМГМУ им. И. М. Сеченова	1 000+	15	2.7

Таблица 2

Мобильные электронные библиотеки и библиотечные системы

№	Наименование	Google Play Маркет		
		Кол-во установок	Кол-во оценок	Оценка поль- зователей
	IPR BOOKS Mobile Reader	10 000+	724	4.3
	Корпоративная библиотека	10 000+	140	3.8
	ЭБС Лань	10 000+	135	3.7
	Библиотека КУ РЖД	10 000+	103	3.5
	Юрайт.Библиотека	50 000+	240	2.5
	НЭБ.РФ Национальная биб- лиотека	10 000+	205	2.1
	ЭБС «Университетская биб- лиотека онлайн»	10 000+	112	1.9

Таблица 3

Мобильные приложения научных журналов

№	Наименование	Google Play Маркет		
		Кол-во установок	Кол-во оценок	Оценка пользователей
1	2	3	4	5
	Chemistry – A European Journal	1 000+	6	5
	Journal of Management Studies	1 000+	5	5
	TESOL Journal	1 000+	5	5

1	2	3	4	5
	AJMS (The American Journal of the Medical Sciences)	1 000+	5	5
	OBGYN Research Journal	1 000+	7	4.9
	Journal of Hospital Infection	1 000+	5	4.8
	Journal of Forensic Sciences	1 000+	10	4.8
	ESCMID Journals	1 000+	8	4.8
	The Journal of Pain	1 000+	12	4.8
	BJOG: An Intl. Journal of Obstetrics & Gynaecology	1 000+	13	4.8
	BJSM	5 000+	40	4.7
	JPD (The Journal of Prosthetic Dentistry)	1 000+	21	4.7
	The Journal of Hepatology	1 000+	10	4.6
	Journal of Hand Surgery	1 000+	7	4.6
	Pediatric Oncall Journal	5 000+	68	4.6
	JOMS (Journal of Oral and Maxillofacial Surgery)	1 000+	15	4.6
	AWHONN Journals	1 000+	5	4.6
	The Psychologist	10 000+	39	4.6
	Journal of Food Science	5 000+	13	4.5
	JACC Journals	10 000+	100	4.5
	JOE: Journal of Endodontics	5 000+	32	4.5
	CHEST® Journal App	1 000+	7	4.4
	ACS Mobile	50 000+	397	4.4
	Journal of Emergency Nursing	5 000+	8	4.4
	European Journal of Cancer	1 000+	9	4.4
	Urology, the Gold Journal	1 000+	8	4.4
	Journal of Renal Nutrition	1 000+	9	4.4
	Academic Journals & Conferences	10 000+	29	4.3
	The Journal of Pediatrics	10 000+	21	4.3
	Research Digest	10 000+	48	4.3
	Radiotherapy & Oncology	10 000+	35	4.3
	J Sport Sci & Med	1 000+	42	4.3
	Gastroenterology Journal	5 000+	13	4.3
	Medicine Journal	10 000+	89	4.2
	Journal of Vascular Surgery	1 000+	11	4.2
	American Journal	10 000+	66	4.2

	of Medicine			
--	-------------	--	--	--

Продолжение табл. 3

1	2	3	4	5
	Journal of Physiotherapy	5 000+	9	4.1
	Brain Journal	1 000+	8	4.1
	JAAD Journals	1 000+	11	4
	British Journal of Dermatology	1 000+	5	4
	Cell Press Journal Reader	10 000+	133	4
	The Lancet	50 000+	272	4
	The Journal of Finance	1 000+	6	3.8
	Elsevier Conferences	10 000+	29	3.7
	Pharmaceutical Journal	10 000+	101	3.6
	ESPEN Journals	1 000+	6	3.5
	Scientific American	50 000+	230	3.5
	Лечащий Врач. Научный журнал	10 000+	178	3.4
	American Journal of Cardiology	1 000+	5	3.4
	Наука.Журналы	1 000+	18	3.3
	The BMJ	10 000+	50	3.3
	Journal of VECC	1 000+	8	3.3
	The Economist	100 000+	2200	3.3
	ATS Journals	1 000+	12	3

Таблица 4

Мобильные агрегаторы научной информации

№	Наименование	Google Play Маркет		
		Кол-во установок	Кол-во оценок	Оценка пользователей
1	2	3	4	5
	Academia.edu	100 000+	7824	4.6
	WikiMed – Offline Medical Encyclopedia	100 000+	20065	4.6
	Science Magazines	10 000+	175	4.6
	Medical News & Journals	10 000+	68	4.6
	Prime: PubMed Journals & Tools	100 000+	2684	4.5

№	Наименование	Google Play Маркет		
		Кол-во установок	Кол-во оценок	Оценка пользователей
	Новости Науки – Наука Канала	100 000+	12729	4.5

Продолжение табл. 4

1	2	3	4	5
	Researcher: 16,000 Academic Publications	1 000 000+	5 784	4.5
	Medical Reads	1 000+	14	4.5
	Мир науки и техники	10 000+	310	4.3
	Research Core – Free Research Papers	5 000+	17	4.2
	FAG. Наука. Технологии. Образование	10 000+	84	4.2
	Science News Daily: Science Articles and News App	100 000+	1 487	4
	Paperity: Open Academic Papers Reader App	10 000+	63	4
	Science & Vie	100 000+	910	3.3
	Sciences et Avenir	500 000+	6870	3.2
	Scopus Discovery	5 000+	22	2.8

Из представленной информации видно, что российские библиотеки на сегодняшний день менее вовлечены в процесс развития мобильных технологий. Однако прежде чем говорить о сложившейся ситуации, как о проблеме и пытаться её устранить, необходимо удостовериться в целесообразности внедрения приложений в библиотечное пространство, то есть оценить их потенциал.

Потенциал приложения определяется как совокупность имеющихся в наличие ресурсов и средств к их развитию. Из этого складывается методика оценки потенциала, включающая анализ трех взаимодействующих факторов: предложение – имеющиеся ресурсы и функции; спрос – запросы активных и потенциальных пользователей; возможности –

пути дальнейшего развития для достижения поставленной цели.

Первый этап — изучение спроса и предложения. На данном этапе проведен сравнительный анализ отобранных приложений по следующим критериям (см. табл. 1–4):

1. Количество установок — отображает первоначальный интерес пользователей к мобильному приложению, позволяет проанализировать спрос на предлагаемые ресурсы и услуги.

2. Общее количество оценок — указывает на уровень интереса пользователей к приложению.

3. Средняя оценка пользователей (по 5-балльной системе) — рассчитывается на основе текущих отзывов. Так как приложения постоянно обновляются, такой подход позволяет получить наиболее точное и актуальное представление об отношении пользователей к предлагаемым функциям и ресурсам [3].

Отмечено, что наиболее высоким спросом пользуются мобильные агрегаторы научной информации, а также мобильные электронные библиотеки и библиотечные системы. Внимание пользователей к данным группам приложений объясняется тематической универсальностью предоставляемого материала.

Менее востребованными являются мобильные приложения вузовских и академических библиотек, что обусловлено их ориентированностью на определенную узкую аудиторию пользователей.

Второй этап — изучение возможностей дальнейшего развития мобильных приложений как инструментов информационно-библиотечного обслуживания. Для реализации данного этапа рассмотрены отзывы об исследуемых приложениях, оставленные пользователями на платформе Google Play Маркет.

На основании полученной информации выявлено, что для мобильных приложений, обладающих большим потенциалом в совершенствовании информационного обеспечения научных исследований, характерны следующие признаки:

- открытый доступ;
- большие массивы полных текстов документов;
- мультидисциплинарность материалов;
- простой и удобный интерфейс;
- гибкий навигационный аппарат;
- индивидуальный подход к подбору материалов (например, рассылка статей по выбранной тематике);
 - offline доступ к материалам;
 - наличие уникальных функций (например, поиск материалов по QR-коду).

Вывод

Разработанная методика оценки потенциала мобильных приложений универсальна и может применяться в любой научной, образовательной и культурной области. В результате применения данной методики заинтересованные организации получают рекомендации для создания и развития мобильных приложений, основанные на потребностях и запросах пользователей.

Такой подход позволяет расширить аудиторию пользователей и определить пути для развития и создания новых форм доступа к научно-информационным ресурсам и услугам организации.

Литература

1. Колосов К. А. Создание приложений для мобильных устройств с целью обеспечения потребностей удаленных пользователей библиотек // Библиотека будущего. Москва, 2014. С. 50–56.
2. WEB-Index. Общая аудитория интернета. <https://webindex.mediascope.net/general-audience>
3. Google Play. Оценки и отзывы в Google Play. <https://play.google.com/about/comment-posting-policy/>
4. Шрайберг Я. Л. Электронная информация, библиотеки и общество: что нам ждать от нового десятилетия информационного века? Ежегодный доклад конференции «Крым». Год 2011 // Научные и технические библиотеки. 2012. № 1. С. 11–62.
5. Bohyun K. The Present and Future of the Library Mobile Experience // Library Technology Reports. 2013. V. 49, Is. 6. pp. 15–28.

6. We are social. Digital in 2020. <https://wearesocial.com/digital-2020>
7. App Annie. Report: The State of Mobile 2020. <https://www.appannie.com/en/go/state-of-mobile-2020/>
8. Wong S. H. R. Which platform do our users prefer: website or mobile app? // Reference Services Review. 2012. Vol. 40, № 1. P. 103–115. <https://doi.org/10.1108/00907321211203667>

References

1. Kolosov K. A. Sozдание prilozhenij dlya mobilnyh ustrojstv s celyu obespecheniya potrebnostej udalennyh polzovatelej bibliotek // Biblioteka budushchego. M., 2014. P. 50–56.
2. WEB-Index. Obshchaya auditoriya interneta. <https://webindex.mediascope.net/general-audience>
3. Google Play. Ocenki i otzyvy v Google Play. <https://play.google.com/about/comment-posting-policy/>
4. Shrajberg YA. L. Elektronnaya informaciya, biblioteki i obshchestvo: chto nam zhdet ot novogo desyatiletija informacionnogo veka? Ezhegodnyj doklad konferencii «Krym». God 2011 // Nauchnye i tekhnicheskie biblioteki. 2012. № 1. P. 11–62.
5. Bohyun K. The Present and Future of the Library Mobile Experience // Library Technology Reports. 2013. V. 49, Is. 6, pp. 15–28.
6. We are social. Digital in 2020. <https://wearesocial.com/digital-2020>
7. App Annie. Report: The State of Mobile 2020. <https://www.appannie.com/en/go/state-of-mobile-2020/>
8. Wong S. H. R. Which platform do our users prefer: website or mobile app? // Reference Services Review. 2012. Vol. 40, № 1. P. 103–115. <https://doi.org/10.1108/00907321211203667>

Сохранение знаний и развитие научных школ в процессе цифровизации

О. М. Атаева, В. А. Серебряков, Н. П. Тучкова

ВЦ ФИЦ ИУ РАН

Аннотация. Проблема сохранения знаний в цифровом пространстве возникла в середине прошлого столетия и постепенно усугубляется с увеличением потока цифровой информации. На фоне возрастающего информационного шума поиск достоверных данных становится самостоятельным видом деятельности. На смену периоду накопления оцифрованной информации приходит время аккумуляции знаний в интегрированной форме, в виде структурированных данных. Тем не менее, увеличение количества информационных ресурсов не способствует улучшению качества предоставляемой информации. Одна из задач научного сообщества – дать экспертную оценку накопленным цифровым знаниям. Эта оценка должна базироваться на авторитетах научных школ. В работе обсуждается проблема поддержки информационных образов традиционных научных школ и их развития в цифровом пространстве. На примере интеграции математических знаний сформулированы задачи и пути их решения в процессе оценки и сохранения оцифрованных знаний. Предлагается вариант специализированных авторских указателей и тезаурусов математических школ.

Ключевые слова: авторский указатель, инструменты интеграции данных, тезаурус математической предметной области.

Knowledge preservation and development of scientific schools in the digitalization process

O. M. Ataeva, V. A. Serebryakov, N. P. Tuchkova

C C FRC CSC RAS

Abstract. The problem of preserving knowledge in digital space arose in the middle of the last century and is gradually aggravated with the increase in the flow of digital information. Against the background of increasing information noise, the search for reliable data becomes an independent activity. The period of digitized information accumulation is replaced by the time of knowledge accumulation in an integrated form,

in the form of structured data. However, the increase in the number of information resources does not improve the quality of the information provided. One of the tasks of the scientific community is to give an expert assessment of the accumulated digital knowledge. This assessment should be based on the authority of scientific schools. The work is focused the problem of supporting information images of traditional scientific schools and their development in the digital space. Using the example of integration of mathematical knowledge, tasks and ways to overcome them in the process of evaluating and preserving digitized knowledge are formulated. A version of specialized index and thesauruses of mathematical schools is offered.

Keywords: authors index, data integration tools, mathematical subject area thesaurus.

Введение

Период цифровизации, который определил развитие экономики, науки и культуры в начале 21 столетия, естественным образом отразился на получении и восприятии знаний. Классические научные школы, которые были основой университетского образования, также претерпели изменения, особенно в связи с дистанционными формами обучения. Дистанционное образование достаточно медленно входило в стандарты, пока в нем не появилась срочная необходимость, обусловленная пандемией COVID-19. В то же время стало понятно, что нельзя исключать живое общение в процессе передачи знаний от человека к человеку, и лекции переместились в пространство удаленного общения через известные приложения (*Zoom, Microsoft Teams* и др.). Доступ к бумажным учебникам сократился до минимума, и в этих условиях особенно важно *сохранить знания университетских и академических научных школ в доступном для восприятия современном цифровом виде, то есть, как структурированные данные* [1]. Это относится в полной мере и к сохранению понятия авторитетов в науке, поскольку современное восприятие знаний основано в большей степени на клиповом мышлении, восприятии фактов, а не на последовательном выстраивании логики познания [2]. Но без определения места научного открытия в общей картине научного познания и стоящего за ним уче-

ного или коллектива ученых невозможно в полной мере отразить смысл и значение того или иного научного факта. Для обеспечения такой возможности необходимо использовать инструменты интеграции данных и средства для их экспертной оценки со стороны научного сообщества. В этом процессе необходимо опираться на классические и вновь сформированные научные школы как на авторитеты, известные в своей области. Именно экспертные оценки могут служить ориентирами при поиске информации в период обучения и исследований. Тем не менее, существует значительный пробел в использовании экспертных оценок в информационных ресурсах. В работе предлагается вариант представления этой информации в виде авторских указателей, основанных на библиографии коллективов отечественных научных тематических школ.

1. Значение опыта научных школ в цифровом пространстве знаний

При поиске достоверных данных на первый план выходит использование опыта персональных исследований и соответствующих накопленных библиографических коллекций и результатов исследований, которые активно распространяются в сети интернет. Одним из первых, кто поставил вопрос о проблеме информационного обмена в научной среде, был профессор математики университета Торонто Кеннет Мэй (Kenneth O. May). Ему принадлежит авторство уникального цифрового проекта по созданию словарей и тезаурусов по математике совместно с издательством McGraw-Hill. Основной и очевидный для многих тезис К. О. Мэя, прозвучавший на одной из конференций, гласит: «Вы не можете сделать научной работы и систематически работать ни в какой области без словарей в этой области» [3]. То же можно сказать и про авторские коллекции и авторские библиографические указатели, которые отражают принадлежность предметной области (ПО), а также приверженность определенной научной школе,

в рамках которой **структурируются** знания, накопленные автором. Таким образом, можно отметить, что в цифровом пространстве знаний основной особенностью научных школ является общий «язык», на основе которого можно сформировать тезаурусы, характерные для научных школ. Общеизвестный тренд на экспоненциальный рост научной информации формирует проблему информационного шума при поиске. Это особенно негативно отражается на эффективности использования информационных ресурсов в научной работе. Составление обзоров при подготовке публикаций стало практически такой же частью научных исследований, как и сами научные исследования. Хотя обзоры всегда были необходимы для подтверждения актуальности предлагаемых разработок, сегодня, во многом, они опираются на анализ интернет-изданий, а не печатных (бумажных) изданий. Процесс поиска научной литературы занимает значительное количество времени, в отличие от ситуации, когда было достаточно ознакомиться с реферативными журналами, чтобы составить представительный список цитирований по изучаемой проблеме. Поиск и анализ данных в интернет-ресурсах становится самостоятельным видом деятельности [4]. В редких случаях современные научные библиотеки берут на себя такой труд, как подборка публикаций по определённой тематике. Исследования в науковедении стали отмечать поиск в интернете как необходимый элемент разделения научного труда, повышающий/замедляющий его производительность. Создание и поддержка цифровых образов научных школ, как ориентиров достоверных знаний, составляет важную задачу современных информационных технологий.

2. Интеграция математических знаний в цифровой среде

Информационные технологии не прекращают двигаться по пути предоставления поисковых ресурсов для научных и образовательных целей. Надо отметить, что некоторые онлайн библиотеки (например, <https://cyberleninka.ru/>)

и ресурсы (например, <https://www.researchgate.net/>, <https://www.semanticscholar.org/>) уже стали предлагать услугу подбора актуальных работ на аналогичные темы. Это оправданно, поскольку только сам специалист, исследователь может адекватно сформулировать свой информационный запрос, и необходимо предоставить ему такую возможность. Для математических работ это особенно важно, поскольку в них при поиске используется язык формул, как наиболее простой способ выражения специфики ПО.

Отличительные особенности математических публикаций – это наличие формул и специфика изложения (наличие утверждений, теорем, доказательств и др.). То есть, элементы предварительной аналитико-синтетической обработки математических текстов, кроме создания словарей, тезаурусов и наборов ключевых слов для индексации, должны дополняться списками формул и связей между ними, контекстов, где они встречаются. Связи утверждений и формул диктуются логикой самой математической ПО.

Исследования в этом направлении не теряют своей актуальности и складываются из нескольких разделов, характерных и для других естественнонаучных ПО, таких, как:

- разработка методов и технологий представления математических знаний в цифровых библиотеках;
- разработка технологий для использования символьной информации в поисковых запросах и для передачи данных;
- разработка технологий представления формул в системах компьютерной алгебры;
- разработка и использование систем доказательных данных;
- создание платформ для поддержки пользователей в изучении и использовании знаний математических ПО;
- разработка технологий поиска связей математических публикаций для выявления взаимовлияния исследований различных научных коллективов.

Все перечисленное можно обобщить, как создание систем поддержки актуальных математических знаний в научном цифровом пространстве и управления этими данными на основе семантических связей ПО.

Для формирования информационных образов разделов математики необходимо учитывать перечисленные аспекты. А именно, использовать представительные словари с толкованием терминов, формул и связей между ними как основы информационно-поискового тезауруса ПО и включение в них авторских указателей для отображения роли отдельных авторов и научных школ в онтологиях ПО [5].

2.1. Примеры реализации математических знаний в цифровой среде

В работе [6] был представлен тезаурус ПО «обыкновенные дифференциальные уравнения» и теперь разрабатывается его расширение на область «уравнений с частными производными», как части общего математического ресурса по «уравнениям математической физики и смежных областей» [7, 8], разработка которого реализуется на основании предлагаемого подхода. Тестовая реализация разрабатывается на базе цифровой библиотеки LibMeta [9, 12]. Для накопления персональных научных знаний в рамках математической предметной области используется понятие тезауруса адресата [10, 11]. Надо отметить, что теоретические разработки отечественных ученых в области обработки и представления цифровой информации ПО, традиционно опережают их практическую реализацию и некоторые зарубежные разработки. Тем не менее, отсутствие интегрированных данных отечественных публикаций зачастую не позволяет реализовать эти разработки в полной мере. Можно отметить некоторые, относящиеся к исследуемой теме, зарубежные и отечественные ресурсы:

Digital Library of Mathematical Functions (dlmf.nist.gov) – Электронная библиотека математических функций.

EMIS (www.emis.de) – Европейская математическая служба информации поддерживается FIZ Karlsruhe / Zentralblatt MATH. Сайт содержит электронную библиотеку

(журналы, труды научных обществ и т. д.), информацию о конференциях и конгрессах.

MacTutor History of Mathematics archive (turnbull.mcs.st-and.ac.uk) – ресурс Математической школы Университета святого Андрея, Шотландия, предлагающий массив материалов, связанных с математикой (история математики, биографии знаменитых математиков и т. п.).

MathDiss database (mathdiss.mathguide.de) – система, которая предоставляет доступ к диссертациям по математике Геттингенского университета. Эта база данных – часть Международного проекта Немецкого исследовательского совета.

SciGuide (<http://www.prometeus.nsc.ru/sciguide/>) – веб-навигатор зарубежных и отечественных научных электронных ресурсов открытого доступа, элемент поддержки научной коммуникации в Сибирском отделении РАН. Навигатор помогает вести поиск научных ресурсов мирового уровня. Структура навигатора и его наполнение поддерживаются сотрудниками Отдела комплектования Отделения ГПНТБ СО РАН

NIST Digital Library of Mathematical Functions <https://dlmf.nist.gov/> – каталог математических функций, представляет собой электронную версию справочника «NIST Handbook of Mathematical Functions»

The European Mathematical Information Service <https://www.emis.de/> – включает: Mathematics Literature Databases, ELibM – Electronic Library of Mathematics – портал предоставляющий доступ к европейским математическим ресурсам.

Zentralblatt MATH (<https://zbmath.org>, Mathematical abstracts and reviews) содержит данные, начиная с 1826 года по настоящее время, и представляет собой ресурс, который осуществляет сбор, систематизацию и распространение библиографических данных и рефератов книг и статей, посвящённых всем разделам математики и её приложениям. В частности, этот проект свидетельствует, что интерес к проблеме использования символьной записи в виде формул в поисковом запросе сформировался достаточно давно. Zentralblatt MATH демонстрирует на своей библиографиче-

ской базе данных преимущества использования символьной информации, формул.

ResearchGate (<https://www.researchgate.net/>) – интегратор научной библиографии, основанный на авторских профайлах.

<https://zbmath.org>, – наиболее близок к заявляемому подходу, и содержит, в частности, поиск по формулам в аннотациях к публикациям. Методика наполнения, предлагаемая в настоящем подходе, так же как в ResearchGate, предполагает использование открытых авторских источников.

Среди всех прочих выделяется насыщенный вычислительными методами и снабженный «математическим языком» портал Wolfram Research (<http://www.wolfram.com>).

Среди отечественных ресурсов, посвященных математике, можно отметить также электронную библиотеку math.ru – собрание оцифрованных версий книг, журналов, и брошюр по математике для школьников и всех, интересующихся математикой.

«Общероссийский математический портал» (<http://www.mathnet.ru/>) математического института им. В. А. Стеклова РАН представляет собой ресурс интегрирующий различную информацию о математической жизни в России. Для классификации контента используются классификаторы MSC, УДК и др. На странице Math-Net.Ru содержатся ссылки на актуальные программные продукты, такие как MikTeX (пакет для работы с TeX'ом в операционной системе Windows), Magma (компьютерно-алгебраическая система для вычислений в алгебре, теории чисел и геометрии), PARI/GP (компьютерно-алгебраическая система, разработанная для быстрых вычислений в теории чисел), GAP (система для вычислений в дискретной алгебре, в частности, в вычислительной теории групп) и др. Имеются также ссылки на известные классификаторы: УДК и MSC 2000 и ряд полезных ссылок на известные издательства, математические ресурсы в Интернете, реферативные базы и математические научные сообщества.

Интернет-проект (<http://eqworld.ipmnet.ru/indexr.htm>) «Мир математических уравнений» (EqWorld) поддерживает шесть языков (русский, английский, немецкий, французский, итальянский, испанский). Научный редакторский коллектив состоит из представителей различных математических коллективов России, США, Канады и др. EqWorld содержит многочисленные ссылки на вспомогательный материал, библиографию, вычислительные средства (Maple, Mathematica, MATLAB и др.). Указатель (предметный и именной) математических уравнений включает разделы: обыкновенные дифференциальные уравнения, дифференциальные уравнения с частными производными (уравнения математической физики), интегральные уравнения, функциональные уравнения. EqWorld наиболее приближен к российской классической математической школе как учебник и справочник с формулами. Это реализуется за счёт ссылок на справочник авторов, т. е. по ссылке из указателя есть переход к страничке справочника.

Ресурс «OntoMathpro Ontology classes» (<http://ontomathpro.org/ontology>) Казанского университета содержит англо-русский алфавитный указатель терминов с горизонтальными связями.

Есть еще энциклопедия математики Encyclopedia of Mathematics https://www.encyclopediaofmath.org/index.php/Main_Page – оцифрованный вариант энциклопедии под редакцией И. М. Виноградова, который, к сожалению, существует отдельно от других информационных ресурсов, как в русскоязычном, так и в англоязычном варианте.

Надо заметить, что теперь эта энциклопедия доступна из цифровой библиотеки LibMeta [12]. Это стало возможным благодаря структуре данных этой библиотеки как интегрированной системы, где реализована связь математических понятий с источниками. Одним из таких источников выступает Encyclopedia of Mathematics.

Многие известные группы в научном сообществе заняты созданием информационных систем, так или иначе связанных с представлением математических знаний, исследований и публикаций в цифровом пространстве (например,

[13–14]). Это определяется востребованностью математических знаний, компетенций и результатов исследований во многих областях науки, образования и бизнеса. В работе [15] представлена история создания информационных ресурсов отделения ГПНТБ СО РАН. В ней рассматривается путь развития баз данных, посвященных выдающимся ученым и возглавляемым ими научным школам и их представлению в цифровом виде.

В настоящей работе рассматривается задача выявления научных школ на основе контента цифровой библиотеки. Анализ контента, связанного как с отдельным автором, так и всего контента в целом, предполагает выделение следующих этапов:

- предварительная обработка текстов автора, для приведения их к виду пригодному для семантического анализа;
- выделение области интересов автора на основе имеющихся данных о публикациях и связях;
- формирование тезауруса адресата автора на основе связей тезауруса предметной области;
- анализ сети соавторов и библиографических списков.

На основе результатов прохождения этих этапов можно сформировать списки ученых из ПО и оценить их вклад в развитие ПО в рамках контента цифровой библиотеки.

Заметим, что понятие научных школ в обычном понимании в реальном научном мире более широко, и мы не претендуем на замещение признанных авторитетов в предметных областях, на основе накопленного контента библиотеки. В работе рассматривается только имеющийся ограниченный цифровой контент, на основе которого и делаются соответствующие выводы. Очевидно, что присутствие признанных авторитетов при достаточно большом объеме данных должно проявиться при анализе цифрового контента любой ПО и будет коррелировать с ситуацией в реальном мире.

2.2. Обозначения и формулы как отражение подходов к исследованиям

Язык формул отражает подход к постановкам задач, характерный для различных математических школ, поэтому он имеет важное значение, которое не требует обоснований. Известно, что различные ученые в разных концах света проводили аналогичные математические исследования и записывали результаты по-разному. Самая громкая история о споре Ньютона и Лейбница по вопросу вклада каждого в развитие математического анализа [16].

В работе [17] даётся обзор источников, посвящённых основным тенденциям исследований в области использования индексирования публикаций на основе таксономии и включению формул в поисковый запрос. Так как в качестве основного языка математики все-таки используется представление формул, то анализа текста без учета представления формул недостаточно. А представление формул в контексте научной школы является основной характеристикой ее языка. Поэтому, специалисты отделяют проблемы представления формул для хранения, визуализации от проблем поиска публикаций с формулами и представление формул играет определяющую роль. Для сохранения данных используются методы кодирования и анализа символьных выражений, индексация математических статей ключевыми словами и формулами. Разрабатываются системы контекстного поиска с учётом таксономии. Эти разработки составляют направление, которое сочетает использование семантических связей и рекурсивный поиск индексированных документов с формулами. Такой подход предполагает создание предварительной структуры данных и разметку документов для быстрого поиска и дальнейшего рекурсивного уточнения запроса после предварительной выборке.

Цифровая семантическая библиотека LibMeta, развивается на этих принципах. В ней реализован тезаурус по обыкновенным дифференциальным уравнениям, который

встроен в онтологию ПО и связан с публикациями и авторами (рис. 1).

Просмотр связей тезауруса

Тезаурус ОДУ

- [Невозмущенное движение](#)
Ассоциация > DE0136 DE0142 DE0137 DE0174
Синоним > RDE0139 RDE0140 RDE0143
[Невозмущенное движение устойчивое](#)
Ассоциация > RDE0118 RDE0122 RDE0124 RDE0126 RDE0129 RDE0130 RDE0121
[Невозмущенное движение неустойчивое](#)
Ассоциация > RDE0119 RDE0120 RDE0123 RDE0125 RDE0127 RDE0131
- [Невозмущенное решение](#)
Ассоциация > DE0136 DE0142 RDE0110 DE0137 DE0174
Синоним > RDE0140 RDE0143
- [Фазовый портрет](#)
Ассоциация > DE0136 RDE0104
- [Метод решения Якоби ОДУ первого порядка](#)
Ассоциация > DE0037
- [Автономная система ОДУ](#)
Ассоциация > DE0136 DE0103 DE0128
Синоним > DE0142 RDE0110 RDE0135 RDE0139 RDE0140 RDE0143 RDE0104 DE0174

Рис. 1. Представление тезауруса в цифровой библиотеке LibMeta

В библиотеке присутствует модуль, который позволяет производить поиск по формулам с учётом их семантики. Для поддержки поиска был смоделирован класс *Формула*, который позволяет хранить оригинальную строку формулы из источника, где она получена. Строка может быть в формате Content MathML, Presentation MathML, LaTeX. При необходимости, количество типов для представления формул в различных нотациях легко расширяется. Класс *Формула* связан отношениями с *информационными объектами*, составляющими контент библиотеки, и с *понятиями* тезауруса. Таким образом, можно построить сеть связей формулы с различными объектами, составляющими контент библиотеки. Каждая формула может быть дополнена ключевыми словами.

Непосредственно для реализации интерфейса поиска был использован open-source редактор математических формул VisualMathEditor, автор — David Grima [18]. Данный редактор позволяет пользователю набирать формулу, не зная тонкостей работы с LaTeX или MathML.

Поиск по формулам предполагает поиск по структуре формулы и использование ее семантических связей.

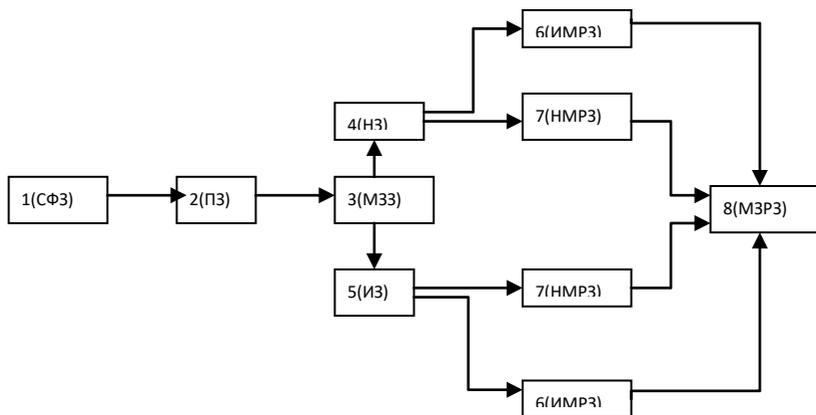


Рис. 2. Схема связей для поиска по формулам в общем виде

Пояснения для схемы связей (рис. 2):

1. Словесная формулировка задачи (СФЗ).
2. Представление задачи в виде совокупности ключевых слов (ПЗ) (дескриптор, синонимы, ассоциативные термины).
3. Математическая запись задачи (МЗЗ).
4. Новая задача (НЗ).
5. Известная задача (ИЗ).
6. Известный метод решения (ИМРЗ).
7. Новый метод решения (НМРЗ).
8. Математическая запись решения (МЗРЗ).

Информация об авторах реализована в цифровой библиотеке LibMeta через понятие «Ресурс». С этим понятием связаны публикации, ключевые слова и понятия тезауруса.

2.3. Состав авторского указателя для математических ПО

Механизм формирования авторских указателей в той или иной степени реализован во всех современных библиографических базах данных. Тем не менее, этого недостаточно для того, чтобы отразить в должной мере значение персона-

лии в науке и использовать в образовательном процессе как значимый элемент в ПО.

Проведенные исследования показывают, что персона в науке должна быть отражена в контексте связей с конкретными достижениями в ПО [20]. Конечно, нельзя всех авторов объявлять родоначальниками научных школ, но можно указывать связи с понятиями тезауруса ПО. Это позволяет учитывать логику развития исследований и принадлежность некоторой научной школе. Такая возможность реализуется только при наличии интегрированной системы [21], в которой кроме библиографических данных присутствуют онтологии ПО, что отличает цифровую библиотеку LibMeta.

В состав авторского указателя, таким образом, необходимо включить, кроме списка публикаций, *список ключевых слов со связями тезауруса* предметной области. Эти минимальные связи с ПО дадут представление о характере исследований автора и, в том числе, позволят отразить развитие ПО и научной школы во времени. Для математических ПО такой указатель необходимо снабдить *указателем формул*, что облегчит *узнавание работ автора* среди аналогичных публикаций.

3. Анализ данных на основе публикаций цифровой библиотеки

3.1. Цели и назначение анализа данных цифровой библиотеки

Количественные и качественные оценки данных и знаний цифровой библиотеки связаны, в первую очередь, с целеполаганием и выбором критериев. Для сохранения и развития научных школ вопрос достоверности цифровых данных является одним из ключевых. Наряду с *количественными метриками*, задаваемыми данными по цитированию, современные исследования направлены на *качественный анализ накопленных библиографических масси-*

вов. Особенности цифровой библиотеки LibMeta как интегратора данных, позволяют представить топологию связей, учесть их количество и провести качественный анализ накопленных данных. Данные представляются как многомерная матрица и анализируются в направлении векторов связей [22]. Такой подход отличается от учета в оценках только «линейных» связей, так как позволяет анализировать данные по совокупности признаков.

Необходимо также определить, что понимается под *качеством данных и знаний в цифровой библиотеке*. Одно из определений *качества данных* заключается в том, что речь идет о «плохих» данных, если данные отсутствуют, неверны или недопустимы в каком-либо контексте [23]. Поскольку цифровые знания, это структурированные данные [1], то сказанное распространяется на качество знаний, с той только разницей, что для оценки *качества знаний* можно привлечь дополнительные связи (ассоциативные, семантические) и экспертные оценки.

Цели качественного анализа библиографических массивов в цифровой библиотеке составляет получение целого ряда оценок:

- «вклад» авторов в публикации и исследования в данной ПО и ПО авторов;
- установление авторства в спорных случаях;
- выявление рейтинга авторов и публикаций;
- выявление ведущих ученых и базовых исследований;
- выявление новых направлений в исследованиях в ПО и науки в целом.

Назначение полученных оценок:

- верификация баз данных и знаний;
- выявление новых и ключевых исследований;
- выявление исследователей, авторитетов в ПО, на знания которых можно опираться в процессе обучения.

Для подсчета и оценок используются такие *исходные данные*, как тезаурусы, ключевые слова, классификаторы, списки публикаций и авторов. Все показатели учитываются с распределением по годам и содержатся в связанном виде в онтологии предметных областей цифровой библиотеки.

3.2. Качество данных

Для оценки данных приведем некоторые таблицы их сравнения в цифровой библиотеке, в целом, и для математических ПО, в частности, для определения соответствия публикации, авторов и ключевых слов описаниям онтологии ПО.

Таблица 1

Соответствие Публикации предметной области

Публикация	Авторы (au)	сравнение {au} и {auПО}	ПО цифровой библиотеки	содержит au%
	Ключевые слова (kw)	сравнение {kw} и {kwПО}		содержит kw%
	UDC udc (udc)	сравнение {udc} и {udcПО}		содержит udc%
	MSC (msc)	сравнение {kw} и {kwПО}		содержит msc%
	Аннотация (abs)	сравнение {abs} и {absПО}		содержит abs%
	Цитируемая библиография (ref)	сравнение {ref} и {refПО}		содержит ref%
	Идентификаторы (id)	сравнение {id} и {idПО}		Содержит id%
	Год (yr)	сравнение {yr} и {yrПО}		содержит yr%
	Издание (pub)	сравнение {pub} и {pubПО}		содержит pub%
Формулы (form)	сравнение {form} и {formПО}	содержит form%		

Табл. 1 многомерная, так как сравниваются множества полей вторичной информации публикации с аналогичным множеством данных в ПО. В результате получаем количество (или проценты, как в таблице) вхождений данных публикации в некоторую ПО цифровой библиотеки. Далее,

в зависимости от строгости критериев, можно относить публикацию к данной ПО или к другой.

Второе назначение таких оценок заключается в *верификации данных о публикации*. При хорошем совпадении атрибутов с ПО, можно дополнить публикацию любыми недостающими признаками (*обратная связь* из ПО), если они отсутствуют, например UDC или MSC, или связать с ключевыми словами тезауруса ПО. То есть, благодаря связям в ПО, обеспечить *полноту данных о публикации*. После этого, если авторы не вызывают сомнений и соответствуют предметной области, то *публикация* со всеми атрибутами вносится в соответствующие списки автора и становится частью тезауруса адресата.

Таблица 2

Соответствие *Автора* предметной области

	Соавторы (coa)	сравнение {coa} и {coaПО}	ПО цифровой библиотеки	содержит coa %
<i>Автор</i>	Ключевые слова (kw)	сравнение {kw} и {kwПО}		содержит kw %
	Публикации (art)	сравнение {art} и {artПО}		содержит art %

Табл. 2 также многомерная, поскольку там также анализируются пересечения множеств выбранных свойств автора и ПО.

На основе сравнения аналогично выясняется соответствие автора предметной области и выполняется верификация данных об авторе. То есть, при наличии «спорного» автора (полный тезка, ошибочное написание, «двойникование» в цифровой библиотеке и т. д.) сведения из ПО позволяют «оценить» автора (*обратная связь* из ПО) и установить недостающие данные, идентифицировать персону и т. д. Наличие соавторов из предметной области составляет одно из важных свойств *автора* как участника совместных исследований

и претендента на «спорное» авторство в рамках цифровой библиотеки.

Авторы, у которых наибольший процент «пересечений» с онтологией ПО, могут считаться «ключевыми» исследователями, ведущими учеными научных школ.

Замечание 1: в нашем исследовании не дается никакой оценки обоснования исследований авторов и качества научных работ.

Замечание 2: все оценки делаются только на основе публикаций, вторичной информации или полных текстов (если они доступны) и авторских методов, отслеживания связей в цифровой библиотеке.

Замечание 3: реальный вклад автора в публикацию и исследования может оценить только научное сообщество. В цифровой библиотеке можно только установить количество связей по выбранным признакам и на основе того массива данных, который уже есть в библиотеке. Это дает картину вклада публикации и рейтинг автора в масштабах имеющихся данных, а не качества публикации и знаний автора в целом.

Замечание 4: в библиотеке LibMeta есть технология создания предметного тезауруса и на его основе можно получить представление о тезаурусе адресата [10] как участника обмена информацией в информационной среде. Эта технология позволяет рассматривать *значение и вклад* публикаций автора применительно к различным предметным областям, которые составляют пересечение множеств в рамках предметного авторского тезауруса.

Табл. 3 представляет наибольшее количество связей, поскольку в ней присутствует связь *ключевых слов* с авторским предметным тезаурусом (если он есть) и тезаурусом предметной области, который заложен в основу онтологии предметной области в цифровой библиотеке. Все эти связи имеют наибольшее значение для верификации данных в самой предметной области.

**Соответствия *Ключевых слов* предметной области
публикациям и авторам**

<i>Ключевые слова</i>	публикации {art}	сравнение {art} и {artПО}	ПО цифровой библиотеки	содержит art%
	авторы (au)	сравнение {au} и {auПО}		содержит au%
	тезаурусы автора (auths)	сравнение {auths} и {thsПО}		содержит auths%
	UDC (udc)	сравнение {udc} и {udcПО}		содержит udc%
	MSC (msc)	сравнение {msc} и {mscПО}		содержит msc%
	Формулы {form}	сравнение {form} и {formПО}		содержит form%

**3.3. Значение оценок для сохранения и развития
научных школ в цифровом пространстве**

Запрос на качество данных и знаний в цифровых библиотеках актуален со всех точек зрения, но особенно в контексте их сохранения и дальнейшего использования. Выделим следующее использование качественных оценок:

- по учету соавторов определяется принадлежность коллективу, научной школе, исследовательской группе;
- распределение по годам публикаций автора используется для оценки вклада в предметную область;
- распределение авторов по публикациям соавторов – для оценки вклада в исследование (публикацию);
- распределение по годам ключевых слов используется для характеристики развития предметной области.

Вычисление максимумов, минимумов и средних (медианных) характеристик полученных значений оценок и распределение их по годам дает характеристики различных периодов в развитии предметных областей, научных исследований и научных школ.

Эти исследования и выводы можно получить исходя из особенностей цифровой библиотеки LibMeta, заложенных в архитектуре связей интеграции данных и отраженных в *таблицах (1–3)* сравнения множеств полей вторичной информации о публикациях и персонах.

Заключение

Идея объединения классических знаний и новых достижений в науке основана на использовании тезауруса адресата и автоматизированного авторского указателя как части онтологии предметной области. Отличие от представленных в обзоре современных ресурсов, состоит в том, что наряду с автоматизированным указателем, разрабатывается функционал для анализа публикаций и других научных материалов, для дальнейших исследований и оценки их качества. Это с одной стороны способствует сокращению числа дублированных цифровых данных, работ и результатов, и с другой стороны не требует многократных повторных поисков в интернете. Предложенные методы оценки качества данных демонстрируют, как используются эти связи для определения принадлежности исследований и авторов публикаций предметным областям и научным школам.

Собственная (авторская, персональная или коллективная) коллекция включает словари терминов, которые семантически связаны со словарем предметной области, что позволяет поддерживать актуальность персональных данных и предметной области. Такой подход приводит к сокращению лагун и дублирования в онтологиях предметных областей за счет публикационной активности адресата и учета новых терминов и связей в процессе накопления результатов авторских исследований, что важно для развития знаний в цифровом представительстве научных школ.

Создание персональной среды для научного исследования на базе индивидуальных библиографических коллекций и результатов, собранных пользователем и автором в процессе исследований, позволяет повысить эффективность использования цифровой информации. Это сократит

также время и количество обращений к информационным ресурсам, что важно, в то числе, в образовательном процессе.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты 18–29–10085мк, 20–07–00324.

Литература

1. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.
2. Семеновских Т. В. Феномен «клипового мышления» в образовательной вузовской среде // Интернет-журнал «НАУКОВЕДЕНИЕ», Выпуск 5 (24), [Электронный ресурс] М.: Наукоедение, 2014, <https://naukovedenie.ru/PDF/105PVN514.pdf> (доступно 17.11.2020).
3. May K. O. Historiography: A Perspective for Computer Scientists. Invited address to International Research Conference on the History of Computing, June 10, 1976, Los Alamos.
4. Ландэ Д. В. Поиск знаний в Internet. Профессиональная работа. М: Издательский дом «Вильямс». 2005. 272 с.
5. Grimm S., Abecker A., Völker J., Studer R. Ontologies and the Semantic Web. In: Domingue J., Fensel D., Hendler J. A. (eds) Handbook of Semantic Web Technologies. 2011. Springer, Berlin, Heidelberg
6. Моисеев Е. И., Муромский А. А., Тучкова Н. П. Тезаурус информационно-поисковый по предметной области: обыкновенные дифференциальные уравнения. М.:МАКС Пресс. 2005. 116 с.
7. Моисеев Е. И., Муромский А. А., Тучкова Н. П. О тезаурусе предметной области смешанные уравнения математической физики // CEUR Workshop Proceedings. 2018. Vol. 2260, pp. 395–405. <https://doi.org/10.20948/abrau-2018-43>.
8. Моисеев Е.И., Муромский А. А., Тучкова Н. П. Интернет и математические знания: представление уравнений математической физики в информационно-поисковой среде. М.: МАКС Пресс. 2008. 80 с.
9. Ataeva O., Serebryakov V., Tuchkova N. Query Expansion Method Application for Searching in Mathematical Subject Domains // CEUR Workshop Proceedings, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen (Aachen, Germany), 2020. Vol. 2543, pp. 38–48. <http://ceur-ws.org/Vol-2543/paper04.pdf> urn:nbn:de:0074-2543-4
10. Шрейдер Ю. А. Тезаурусы в информатике и теоретической семантике // Научно-техническая информация. Сер. 2. 1971. № 3. С. 21–24.

11. Жданова Г. С., Колобродова Е. С., Полушкина В. А., Черный А. Н. Словарь терминов по информатике на русском и английском языках. М.: Наука. 1971, 359 с.
12. Атаева О. М., Серебряков В. А., Онтология цифровой семантической библиотеки LibMeta // Информатика и её применения. 2018. Т. 12. № 1. С. 2–10.
13. Kohlhase M. Mathematical knowledge management: Transcending the one-brain barrier with theory graphs // European Mathematical Society (EMS) Newsletter, June 2014, pp. 22–27.
14. Miller, B. R. Technical aspects of the digital library of mathematical functions / B. R. Miller, A. Youssef // Ann. Math. Artif. Intell. 2003. Vol. 38. № 1, pp. 121–136.
15. Юдина И. Г., Базылева Е. А. Комплексные информационные ресурсы академической библиотеки: модернизация и развитие // Библиосфера. 2018. № 4. С. 56–63.
16. Meli D. B. Equivalence and Priority: Newton versus Leibniz: Including Leibniz's Unpublished Manuscripts on the Principia. Clarendon Press, 1993. P. 318.
17. Ion P. D. F. The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. / P.D. F. Ion, S. M. Watt // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (eds). Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science. 2017. Vol. 10383. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-62075-6_5 (доступно 17.11.2020).
18. <http://visualmatheditor.equatheque.net/> (доступно 17.11.2020)
19. Тучкова Н. П., Атаева О. М. Подходы к извлечению знаний в научных предметных областях // Информационные и математические технологии в науке и управлении. 2020. № 2 (18). С. 5–18.
20. Тучкова Н. П. Научные школы в цифровом пространстве // Информационные ресурсы России. 2020. № 5. С. 49–51.
21. Каленов Н. Е., Серебряков В. А. Об онтологии Единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 10–12.
22. Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И. Методы и модели анализа данных: OLAP и Data Mining. БХВ-Петербург, 2004.
23. Singh R., Singh K. A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing // IJCSI International Journal of Computer Science Issues, 2010. Vol. 7. Issue 3, No 2.

References

1. Gavrilova T. A., Horoshevskij V. F. Bazy znaniy intellektual'nyh sistem. SPb.: Piter, 2000. 384 p.
2. Semenovskikh T. Fenomen «klipovogo myshleniya» v obrazovatel'noj vuzovskoj srede (The phenomenon of «clip-thinking» in the educational high school environment) // Internet-zhurnal «Naukovedenie», 2014. № 5 (24) [Jelektronnyj resurs]. M.: Naukovedenie, 2014. <https://naukovedenie.ru/PDF/105PVN514.pdf> (support 17.11.2020).
3. May K. O. Historiography: A Perspective for Computer Scientists. Invited address to International Research Conference on the History of Computing, June 10, 1976, Los Alamos.
4. Lande, D. V. Poisk znaniy v Internet. Professional'naya rabota. M: Izdatel'skij dom «Vil'yams». 2005. 272 s.
5. Grimm S., Abecker A., Völker J., Studer R. Ontologies and the Semantic Web. In: Domingue J., Fensel D., Hendler J. A. (eds) Handbook of Semantic Web Technologies. 2011.Springer, Berlin, Heidelberg.
6. Moiseev E.I, Muromskij A. A., Tuchkova N. P. Tezaurus informacionno-poiskovyj po predmetnoj oblasti: obyknovennye differencial'nye uravneniya. M.:MAKS Press. 2005. 116 s.
7. Moiseev E. I., Muromskij A. A., Tuchkova N. P. O tezauruse predmetnoj oblasti smeshannye uravneniya matematicheskoy fiziki // CEUR Workshop Proceedings. 2018. Vol. 2260. P. 395–405. <https://doi.org/10.20948/abrau-2018-43>
8. Moiseev E. I., Muromskij A. A., Tuchkova N. P. Internet i matematicheskie znaniya: predstavlenie uravnenij matematicheskoy fiziki v informacionno-poiskovoj srede. M.:MAKS Press. 2008. 80 s.
9. Ataeva O., Serebryakov V., Tuchkova N. Query Expansion Method Application for Searching in Mathematical Subject Domains // CEUR Workshop Proceedings, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen (Aachen, Germany), 2020. Vol. 2543, pp. 38–48. <http://ceur-ws.org/Vol-2543/rpaper04.pdf> urn:nbn:de:0074-2543-4. Vserossiiskaia nauchnaia konferentsiia «Nauchnyi servis v seti Internet». <http://agora.guru.ru/abrau2020>
10. Shrejder Yu. A. Tezaurusy v informatike i teoreticheskoy semantike // Nauchno-tekhnicheskaya informaciya. Ser. 2. 1971. № 3. S. 21–24.
11. Zhdanova G. S., Kolobrodova E. S., Polushkina V. A., CHernyj A. N. Slovar' terminov po informatike na russskom i anglijskomazykah. M.: Nauka. 1971, 359 s.
12. Ataeva O. M., Serebryakov V. A., Ontologiya cifrovoy semanticheskoy biblioteki LibMeta //Informatika i eyo primeneniya. 2018. T. 12. № 1. S. 120.

13. Kohlhase M. Mathematical knowledge management: Transcending the one-brain barrier with theory graphs // European Mathematical Society (EMS) Newsletter, June 2014. P. 22–27.
14. Miller, B. R. Technical aspects of the digital library of mathematical functions / B. R. Miller, A. Youssef // Ann. Math. Artif. Intell. 2003. Vol. 38. № 1. P. 121–136.
15. Yudina I. G., Bazyleva E. A. Kompleksnyye informacionnye resursy akademicheskoy biblioteki: modernizaciya i razvitie // Bibliosfera. 2018. № 4. S. 56–63.
16. Meli D. B. Equivalence and Priority: Newton versus Leibniz: Including Leibniz's Unpublished Manuscripts on the Principia. Clarendon Press, 1993. P. 318.
17. Ion P. D. F. The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. / P.D. F. Ion, S. M. Watt // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (eds). Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science. 2017. Vol. 10383. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-62075-6_5 (support 17.11.2020).
18. <http://visualmatheditor.equatheque.net/> (support 17.11.2020).
19. Tuchkova N. P., Ataeva O. M Podhody k izvlecheniyu znanij v nauchnyh predmetnyh oblastyah // Informacionnye i matematicheskie tekhnologii v nauke i upravlenii. 2020. № 2 (18). S. 5–18.
20. Tuchkova N. P. Nauchnye shkoly v cifrovom prostranstve // Informacionnye resursy Rossii. 2020. № 5. S. 49–51.
21. Kalenov N. E., Serebryakov V. A. Ob ontologii Edinogo cifrovogo prostranstva nauchnyh znanij // Informacionnye resursy Rossii. 2020. № 5. S. 10–12.
22. Barsegyan A. A., Kupriyanov M. S., Stepanenko V. V., Holod I. I. Metody i modeli analiza dannyh: OLAP i Data Mining. BHV-Peterburg, 2004.
23. Singh R., Singh K. A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing // IJCSI International Journal of Computer Science Issues, 2010, vol. 7, issue 3, No 2.

**Особенности динамики формирования
цифрового кластера знаний
о результативности и востребованности итогов
исследований российских учёных в области
медиации**

В. В. Арутюнов

РГГУ

Аннотация. Рассматриваются особенности динамики изменения в 2013–2019 гг. показателей публикационной активности, цитируемости и востребованности результатов исследований российских учёных в области медиации. Отмечается взрывной рост числа публикаций за рассматриваемый период в этой области знаний, что свидетельствует о возрастающем интересе специалистов к результатам исследований в этой сфере. Выявлены научные направления, итоги исследований по которым отличались повышенной востребованностью. В их числе истоки и становление института медиации как примирительной процедуры в странах западной Европы и США, правовые технологии медиации в России, медиация в уголовном процессе, проблемы применения медиации при разрешении трудовых споров, медиативная деятельность в образовательном процессе и ряд др. Результаты исследования получены на основе баз данных Российского индекса научного цитирования (РИНЦ).

Ключевые слова: медиация, публикационная активность, востребованность результатов исследований, цитируемость, РИНЦ.

Features of dynamics on development of digital knowledge cluster on the impact and relevance of research findings Russian scientists in the field of mediation

V. V. Arutyunov

Russian State University for the Humanities

Abstract. The features of the dynamics of changes in 2013–2019 are considered for indicators of publication activity, citation and demand for research results of Russian scientists in the field of mediation. There is an explosive growth in the number of publications in this area of knowledge during the period under review, which indicates the growing interest of specialists in the results of research in this area. Scientific directions were identified, the results of research in which were in high demand. These include the origins and development of the institution of mediation as a conciliation procedure in Western Europe and the USA, legal technologies of mediation in Russia, mediation in criminal proceedings, and problems of using mediation in resolving labor disputes, mediation in the educational process, and a number of others. The research results are based on the Russian science citation index (RSCI) databases.

Keywords: mediation, publication activity, relevance of research results, citation, RSCI.

Уровень развития науки в стране в определённой мере определяет экономические показатели развития государства. В XXI веке он всё чаще выявляется на основе наукометрических показателей (публикационной активности, индексам цитирования и Хирша). Невысокие значения этих индексов для российских учёных, отражаемые в известных зарубежных журналах и системах (например, Web of Science или Scopus), в значительной степени связаны с некоторой изоляцией по ряду причин отдельных направлений российской науки от международного научного сообщества.

Российские журналы довольно часто не переводятся на английский язык, который является в наше время международным языком общения учёных. Российские ученые, по ряду причин, недостаточно активно участвуют

в международных конференциях; при этом наблюдается ощутимый недостаток международных конференций высокого уровня, проводимых в России с приглашением ведущих ученых мира. В результате, успехи и достижения российских ученых остаются вовсе неизвестными или становятся известны мировому научному сообществу со значительным опозданием. Такая изоляция приводит к нарастающему отставанию российской науки в некоторых отраслях знаний и потере способности в среде учёных создавать, воспринимать и развивать новые направления исследований в ряде отраслей науки, техники и экономики.

В последнее десятилетие в России результативность научных исследований оценивается не только с помощью индексов Хирша, публикационной активности I_p , цитируемости I_c , итогов исследований [1–3], но и на основе востребованности результатов исследований I_v , определяемой соотношением $I_v = I_c / I_p$ [4, 5].

Публикации в различных отраслях знаний, определяющие публикационную активность учёного или организации, и их цитирование образуют так называемые сети цитирования. Ежегодный их анализ для конкретной предметной области позволяет выявить механизмы эволюции направлений исследований, определять учёных и организации — лидеров в конкретной отрасли науки и взаимосвязи между ними [6].

Под термином «медиация» обычно понимается процедура достижения консенсуса конфликтующих сторон на основе их вступления в добровольные переговоры с использованием третьей стороны — посредника (медиатора), который оказывает содействие обеим сторонам для урегулирования конфликта (спора). В наши дни медиацию можно рассматривать в качестве альтернативного, всё более распространяемого в современных условиях способа урегулирования споров и конфликтов в различных сферах экономики и коммуникаций [7]. Эффективность использования данного института подтверждается международной практикой [8, 9].

Процедура медиации проводится при взаимном волеизъявлении сторон на основе принципов добровольности, конфиденциальности, сотрудничества и равноправия сторон, а также беспристрастности и независимости медиатора.

На основе баз данных Российского индекса научного цитирования (РИНЦ) была проанализирована динамика публикаций в 2013–2019 гг. российских учёных, участвующих в решении задач медиации, изменение цитируемости результатов этих работ в 2013–2019 гг., а также востребованность итогов исследований в этой отрасли знаний.

На *рис. 1* и *2* представлена динамика изменения в 2013–2019 гг. вышеуказанных индексов I_p , I_c и I_v .

Как следует из *рис. 1*, публикационная активность ученых в рассматриваемой области исследований непрерывно росла практически до 2019 г., увеличившись в этом году практически в 2,5 раза по сравнению с 2013 г. и достигнув значения более 1000 единиц, что свидетельствует о «взрывном» росте интереса исследователей к результатам работ в этой области знаний (как показала оценка числа публикаций, в начале текущего десятилетия оно составляло одну-две сотни единиц в год).

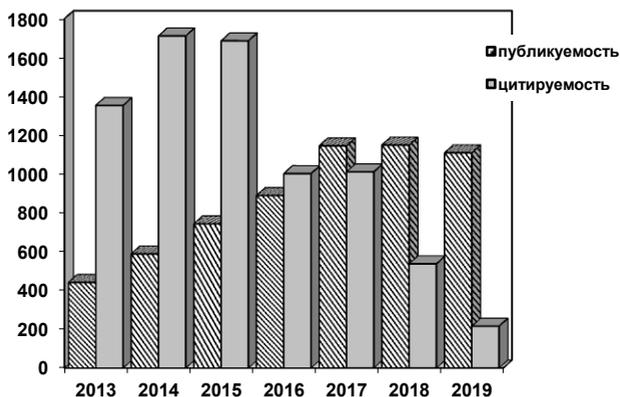


Рис. 1. Динамика публикационной активности и цитируемости российских учёных в области медиации

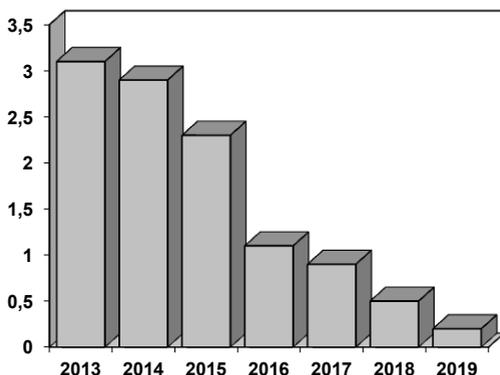


Рис. 2. Динамика востребованности итогов исследований российских учёных в области медиации

Что касается цитируемости работ, то в этой области отмечался максимум в 2014 г. с последующим уменьшением показателей цитирования в 2018 г. более чем в три раза по сравнению с 2014 г. (этот спад, возможно, объясняется кризисом 2014 г., когда во многих научных отраслях началось уменьшение финансирования исследовательских работ).

Малые значения в 2019 г. показателей цитирования для данной отрасли науки объясняются, как и для других естественнонаучных отраслей науки, известной закономерностью: слабым откликом научного сообщества на публикации текущего года.

Что касается востребованности результатов исследовательских работ, то, как следует из рис. 2, наибольшее и наименьшее значение этого показателя отмечалось на границах рассматриваемого периода (соответственно, в 2013 г. и 2019 г.).

В 2018 г. имело место нелинейное уменьшение этого показателя почти в шесть раз по сравнению с 2013 г., что, возможно, является «эхом» кризиса 2014 г.

Небольшие значения показателей востребованности в 2019 г. объясняются, очевидно, той же причиной, что и для показателей цитирования в этом году — замедленным откликом российского и мирового научного сообщества на публикации этого года.

В заключение необходимо отметить, что выявленный по базе данных РИНЦ индекс Хирша для множества публикаций за указанный период в области медиации оказался весьма высоким и равным 28. Этот факт свидетельствует, во-первых, о том, что и в дальнейшем в этой отрасли знаний следует ожидать стабильно высокую публикационную активность российских учёных и специалистов. Во-вторых, — уровень научной активности отечественных исследователей в данной научной отрасли более чем в полтора раза превышает минимальный порог мирового уровня научной активности учёных — докторов наук, равный 16 [10]. При этом, наибольшая востребованность результатов исследований отмечалась по следующим направлениям: истоки и становление института медиации как примирительной процедуры в странах западной Европы и США, правовые технологии медиации в Российской Федерации, медиация в уголовном процессе, безопасность подростков в интернете: риски и родительская медиация, проблемы применения медиации при разрешении трудовых споров, медиативная деятельность в образовательном процессе и ряду др.

Литература

1. О мерах по реализации государственной политики в области образования и науки. Указ Президента РФ от 07.05.2012, № 599. Режим доступа: <http://base.garant.ru/70170946/>
2. Правила оценки и мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения (утв. постановлением Правительства РФ от 08.04.2009, № 312 (в редакции постановления Правительства РФ от 01.11.2013 г., № 979). Режим доступа: <http://base.garant.ru/195302/>
3. Приказ Минобрнауки РФ от 05.03.2014, № 161 «Об утверждении типового положения о комиссии по оценке результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения и типовой методики оценки результативности деятельности научных организаций, выполняющих эти работы». Режим доступа: <http://legalacts.ru/doc/prikaz-minobrnauki-rossii-ot-05032014-n-161/>

4. Арутюнов В. В. О результативности научной деятельности в области приоритетных направлений науки, технологий и техники // Научно-техническая информация, сер. 1. 2013. № 10. С. 12–19.
5. Арутюнов В. В., Цветкова В. А. Сравнительный анализ показателей публикационной активности и цитируемости российских учёных в отдельных естественнонаучных областях знаний по данным РИНЦ и WoS СС // Информация и инновации. 2018. Т. 13, № 1. С. 76–91.
6. Маршакова-Шайкевич И. В. Россия в мировой науке. РАН, Институт философии. М.: ИФРАН, 2008. 227 с.
7. Арутюнов В. В. Типология и особенности современных коммуникаций в работе библиотек. М.: Литера, 2009. 216 с.
8. Минеева Т. Г., Назарова Н. Л., Сальников В. П., Соловьева С. В. Медиация в России. (Памяти Нильсы Кристи) // Правовое поле современной экономики. 2015, № 8. С. 195–202.
9. Бегаева А. А. Корпоративные слияния и поглощения: проблемы и перспективы правового регулирования / отв. ред. Н. И. Михайлов. М.: Инфотропик Медиа, 2010. 256 с.
10. Ершова С. К. Инструкция по использованию РИНЦ. Режим доступа: <https://eip.ru/about-organization/rints.html>

References

1. O merah po realizacii gosudarstvennoy politiki v oblasti obrazovaniya i nauki. Ukaz Prezidenta RF ot 07.05.2012, № 599. <http://base.garant.ru/70170946/>
2. Prikaz Ministerstva obrazovaniya i nauki RF ot 5 marta 2014 g., № 161 «Ob utverzhdenii tipovogo polozheniya o komissii po otsenke rezultativnosti deyatel'nosti nauchnykh organizatsiy, vypolnyayushchikh nauchno-issledovatel'skie, opytно-konstruktorskie i tehnologicheskie raboty grazhdanskogo naznacheniya i tipovoy metodiki otsenki rezultativnosti deyatel'nosti nauchnykh organizatsiy, vypolnyayushchikh nauchno-issledovatel'skie, opytно-konstruktorskie i tehnologicheskie raboty grazhdanskogo naznacheniya»/ – <http://legalacts.ru/doc/prikaz-minobrnauki-rossii-ot-05032014-n-161/>
3. Pravila otsenki i monitoringa rezultativnosti deyatel'nosti nauchnykh organizatsiy, vypolnyayushchikh nauchno-issledovatel'skie, opytно-konstruktorskie i tehnologicheskie raboty grazhdanskogo naznacheniya (utv. Postanovleniem Pravitel'stva RF ot 8 aprelya 2009 g., № 312; v redaktsii Postanovleniya Pravitel'stva RF ot 1 noyabrya 2013 g. № 979). <http://base.garant.ru/195302/>
4. Arutyunov V. V. O rezultativnosti nauchnoy deyatel'nosti v oblasti prioritetnykh napravleniy nauki, tehnologiy i tehniki // Nauchno-tehnicheskaya informatsiya, ser. 1. 2013. № 10. С. 12–19.

5. Arutyunov V. V., Tsvetkova V. A. Sravnitelnyy analiz pokazately publikatsionnoy aktivnosti i citiruemosti rossiyskikh uchenykh v otdelnykh estestvennonauchnykh oblastiakh znaniy po dannym RINTS i WoS CC // Informatsiya i innovatsii. 2018. T. 13, № 1. P. 76–91.
6. Marshakova-Shaykevich I. V. Rossiya v mirovoy nauke. RAN, Institut filosofii / – M.: IFRAN, 2008. 227 p.
7. Arutyunov V. V. Tipologiya i ossobennosti sovremennykh kommunikatsiy v rabote bibliotek. M.: Litera, 2009. 216 p.
8. Mineeva T. G., Nazarova N. L., Salnikov V. P., Soloveva S. V. Meditsiya v Rossii (Pamyati Nilsa Kristi) // Pravovoe pole sovremennoy ekonomiki. 2015, № 8. P. 195–202.
9. Begaeva A. A. Korporativnyye sliyaniya i poglosheniya: problemy i perspektivy pravovogo regulirovaniya / otv. red. Mihaylov N. N. M. Infotropik Media, 256 p.
10. Ershova S. K. Instruktsiya po ispolzovaniyu RINTS. <https://eeip.ru/about-organization/rints.html>

Экспериментальное обоснование критериев количественной оценки инновационности объекта

В. К. Иванов

ТвГТУ

Аннотация. Представлены результаты экспериментов, подтверждающие основные идеи предлагаемого подхода к определению инновационности объектов. Этот подход основан на предположении об адекватности отображения жизненного цикла продуктов, описания которых размещены в различных хранилищах данных. Предложенная формальная модель позволяет вычислить количественное значение аддитивного оценочного критерия инновационности объектов. Критерий включает в себя показатели новизны, востребованности и имплементируемости продуктов и/или технологий. Их значения вычисляются на основе данных об объектах и частотных характеристик доступа к ним, взятых в ретроспективе. Представленные данные экспериментов дают возможность оценить корректность принятого подхода. Так, получены сходные тренды изменения показателей для различных объектов, нормирующих функций и источников данных. Отмечена цикличность изменения показателей в течение значимого периода. Это является отражением типичной ситуации падения спроса после достижения его максимума, следствием чего может быть улучшение конструкции, технологии использования, эксплуатационных характеристик объекта. Эксперименты показали совпадение оценки объектов с помощью вычисленных показателей с экспертными оценками тех же объектов. Использование многих источников информации об объектах для сбора исходных данных позволяет получить более адекватные значения. Использование таких методов, как теория свидетельств, дает возможность обоснованно выполнить комбинирование отличающихся результатов.

Ключевые слова: хранилище данных, инновационность, аддитивный критерий, функция полезности, поисковый запрос.

Experimental Justification of the Object Innovativeness Quantitative Evaluation Criteria

V. K. Ivanov

Tver State Technical University

Abstract. The paper presents the results of the experiments that were conducted to confirm the main ideas of the proposed approach to determining the objects innovativeness. This approach assumed that the product life cycle of whose descriptions are placed in different data warehouses is adequate. The proposed formal model allows us to calculate the quantitative value of the additive evaluation criterion of objects innovativeness. The criterion includes indicators of novelty, demand, and implementability of products or technologies. Their values are calculated based on data about objects and frequency characteristics of access to them taken retrospectively. The obtained experimental data make it possible to evaluate the adopted approach correctness. Thus, similar trends of changes in indicators for various objects, normalizing functions, and data sources were obtained. The cyclical nature of indicators values over a significant period is noted. This reflects a typical situation when falling demand after reaching its maximum, and then may result in an improvement in the design, technology of use, and operational characteristics of the object. Experiments have shown that evaluating objects using calculated indicators coincide with expert estimates of the same objects. Using many sources of information about objects to collect source data allows you to get more adequate values. The proposed use of evidence theory makes it possible to combine different results more reasonably.

Keywords: Data Warehouse, Innovation, Additive Criterion, Utility Function, Search Query.

Введение

Условия современного развития производственных процессов предполагают очевидный интерес к объектам (продуктам и технологиям), обладающим значимым инновационным потенциалом. Отсюда возникает необходимость в количественной оценке такого потенциала. Предлагается подход, который позволяет оценить инновационный потенциал объектов с помощью индекса инновационности, модель вычис-

ления которого основана на таких показателях инновационности, как новизна, востребованность и имплементируемость объекта.

Исходными данными для вычисления показателей инновационности объекта служат его описания, в том или ином виде хранящиеся в различных гетерогенных базах данных. Отметим, что указанные показатели вычисляются не для самого объекта, а для его лингвистической модели. Лингвистическая модель объекта используется для генерации множества запросов к доступным базам данных. Характеристики множеств релевантных описаний объекта, найденных при выполнении запросов, используются при вычислении показателей инновационности. Вычисленные показатели сводятся к глобальному аддитивному критерию — индексу инновационности объекта.

При вычислении показателей инновационности используются формальные выражения, основанные на ряде предположений, которые описаны далее. Например, предположение о том, что имплементируемость или реализуемость продукта зависит от величины периода восстановления спроса на его улучшенную модификацию. Или принятие в качестве одного из параметров востребованности продукта количества запросов пользователей к базам данных с упоминанием характеристик этого продукта. Естественно, эти предположения требуют экспериментального подтверждения с помощью соответствующих методик и инструментов для измерения, обработки и сравнения значений обсуждаемых показателей.

Цель статьи — экспериментально подтвердить некоторые предположения, использованные при формализации модели для вычисления аддитивного оценочного критерия инновационности объектов. Некоторые результаты экспериментальной проверки этой модели, впервые представленные в настоящей статье, дают возможность в определенной степени оценить корректность принятого подхода. Настоящая статья является продолжением работ автора и его коллег по обсуждаемой тематике [1, 2 и др.], которые направлены на развитие методов и средств повышения эффективности

информационного поиска в различных прикладных областях (см., например, [3]).

1. Лингвистическая модель объекта

$$LM = \{A_S, A_C, A_T, M, C\}, \quad (1)$$

где A_S , A_C и A_T – множества или классы архетипов объекта, определяющих соответственно структуру, условия применения и результаты функционирования этого объекта, при этом $|A_S| = |A_C| = |A_T|$. Архетип объекта $a \in A$ – это концепции предметной области для рассматриваемого объекта, которые реализуются терминами, определяющими ключевые свойства объекта, и группируются в классы A_S , A_C и A_T ;

M – множество термов-маркеров $t \in M$, задающее область определения архетипов объекта.

C – дополнительные локальные ограничения: умолчания, синонимы термов, веса термов, предельное количество запросов, количество термов в запросе и т. п. Эти ограничения используются в алгоритмах генерации поисковых запросов.

Примеры архетипов и маркера приведены в *табл. 1*.

Таблица 1

Примеры архетипов и маркера для объекта «смартфон»

Архетипы A_S	Архетипы A_C	Архетипы A_T	Маркер t
<i>камера экран</i>	<i>NFC 4G</i>	<i>производительность аккумулятор</i>	<i>смартфон Galaxy S</i>

Примечание к *табл. 1*: маркеры $t \in M$ ограничивают множество анализируемых смартфонов только устройствами одного семейства *Samsung Galaxy S*.

Лингвистическая модель (1) используется как поисковый паттерн для генерации набора запросов для поиска информации о потенциально инновационном объекте. Поисковые запросы – логические выражения, где множества операндов есть различные комбинации термов $a \in A$ и $m \in M$. Релевантные документы, найденные после исполнения всех сгенерированных запросов, используются для вычисления отдельных показателей инновационности объекта и индекса инновационности объекта в целом.

2. Модель вычисления индекса инновационности объекта

Количественная оценка показателей инновационности объекта основывается на предположении об адекватности отображения жизненного цикла продуктов в виде их описаний, размещенных в различных хранилищах данных [1].

Показатель технологической новизны Nov вычисляется следующим образом:

$$Nov = 1 - 1/N \sum_{k=1}^N f_k^{01}(R_k, \dots), \quad (2)$$

где N – общее количество выполненных запросов; R_k – число документов, найденных в результате выполнения k -го запроса; $f_k^{01}(R_k, \dots)$ – вариативная функция, нормирующая значение R_k на диапазон $[0;1]$.

Показатель востребованности Dem вычисляется следующим образом:

$$Dem = 1/S \sum_{k=1}^S f_k^{01}(F_k, \dots), \quad (3)$$

где S — общее количество выполненных запросов; F_k — частота выполнения k -го запроса; $f_k^{01}(F_k, \dots)$ — вариативная функция, нормирующая значение F_k на диапазон $[0;1]$.

Показатель имплементируемости Imp вычисляется следующим образом:

$$Imp = 1 - 1/2 \cdot f_n^{01}(\Delta t_N(Nov(t)) + \Delta t_D(Dem(t))), \quad (4)$$

где $Nov(t)$ — функция, показывающая зависимость Nov от времени на временном интервале $[t_0; t_m]$; $Dem(t)$ — функция, показывающая зависимость Dem от времени на том же $[t_0; t_m]$; Δt_N и Δt_D — средние расстояния между двумя последовательными точками временного ряда $t_i, t_{i+1} \in [t_0; t_m]$ локальных максимумов функций $Nov(t)$ и $Dem(t)$; f_k^{01} — вариативная функция, нормирующая значение Δt на диапазон $[0;1]$. При этом Nov , Dem и Imp рассчитываются для точки t_{m+1} .

Индекс инновационности Ix имеет вид аддитивного критерия:

$$Ix = w_{Nov}Nov + w_{Dem}Dem + w_{Imp}Imp, \quad (5)$$

где w_{Nov} , w_{Dem} , w_{Imp} — весовые коэффициенты для Nov , Dem и Imp соответственно и $w_{Nov} + w_{Dem} + w_{Imp} = 1$.

В случае неполной и неточной информации об объектах вводятся нечеткие показатели вероятности того, что объект обладает новизной, востребованностью и имплементируемостью. Для вычисления указанных вероятностей применяется теория свидетельств Демпстера — Шафера [4]. Определяются базовые вероятности m попадания результатов измерения Nov , Dem и Imp в k -й интервал значений A_k .

Результаты из различных источников рекурсивно комбинируются по парам источников: из двух источников свидетельств образуется один условный источник, свидетельства которого комбинируются с очередным фактическим источником.

Рассчитываются функция доверия $Bel(A) = \sum_{A_k: A_k \subseteq A} m(A_k)$ и функция правдоподобия

$Pl(A) = \sum_{A_k: A_k \supseteq A} m(A_k)$, которые определяют верхнюю и

нижнюю границу вероятности обладания объектом свойства, заданного соответствующим показателем. Тогда выражение (5) приобретает вид мультипликативного оценочного критерия:

$$I_x = [Bel_{Nov}(A), Pl_{Nov}(A)]^{w_{Nov}} * [Bel_{Dem}(A), Pl_{Dem}(A)]^{w_{Dem}} * [Bel_{Imp}(A), Pl_{Imp}(A)]^{w_{Imp}}, \quad (6)$$

которое сводится логарифмированием I_x к аддитивному критерию:

$$\ln I_x = w_{Nov} * \ln([Bel_{Nov}(A), Pl_{Nov}(A)]) + w_{Dem} * \ln([Bel_{Dem}(A), Pl_{Dem}(A)]) + w_{Imp} * \ln([Bel_{Imp}(A), Pl_{Imp}(A)]) \quad (7)$$

Так как $\ln()$ возрастающая функция, рассуждения, касающиеся I_x , справедливы для $\ln(I_x)$.

3. Методика проведения экспериментов

В ходе проведения экспериментов планировалось получить данные, подтверждающие следующие предположения, использованные при формулировке выражений (2), (3) и (4):

– Необходимость использования многих источников информации об объектах для сбора исходных данных при вычислении показателей инновационности.

– Применимость предложенной модели вычисления индекса инновационности объекта для различных типов объектов.

– Зависимость показателя новизны объекта от времени есть очевидное уменьшение его значения. Следовательно, количество информации об объекте, найденной в базах данных, должно со временем увеличиваться.

– Цикличность изменения показателей инновационности вследствие улучшений конструкции, технологии использования, эксплуатационных характеристик объектов.

– Использование в качестве параметров востребованности объекта различных характеристик пользовательского доступа к информации о нем.

– Зависимость показателя имплементируемости объекта от величины периода создания его улучшенной модификации и соответствующего восстановления спроса.

В качестве объектов с инновационным потенциалом были отобраны 37 продуктов и технологий. В их число входили: смартфоны популярных моделей известных производителей, способы получения новых материалов, космические аппараты, медицинские технологии, технологии приготовления продуктов питания, технические устройства.

После этого экспертами были сформированы лингвистических модели для каждого объекта. Эти модели были использованы для генерации поисковых запросов.

В качестве источников данных об объектах использовались следующие хранилища данных: ACM Digital Library (<https://dlnext.acm.org>), AliExpress (<https://aliexpress.com>), Google (<https://google.com>), Google Patents (<https://patents.google.com>), Google Scholar (<https://scholar.google.ru>), IEEE Explore Digital Library (<https://ieeexplore.ieee.org>), ЕГИСУ НИОКТР (<https://rosid.ru>), Научная электронная библиотека (<https://elibrary.ru>), поисковая система ФИПС (<https://www1.fips.ru/iiss>), Яндекс (<https://yandex.ru>), а также базы данных системы электронного обучения и учебно-

методических комплексов ТвГТУ
(<http://elearning.tstu.tver.ru>).

Показатели *Nov* и *Dem*, измеренные по результатам выполнения совокупности сгенерированных поисковых запросов, усреднялись среднеарифметическим и медианным значениями, а также нормализовались на диапазон [0;1]

функциями f_n^{01} для линейного $f_{nk}^{01} = \frac{F - \min F}{\max F - \min F}$

и экспоненциального $f_{ne}^{01} = 1 - \exp(1 - \frac{F}{\min F})$ нормирования.

Здесь F использовано для обозначения *Nov* или *Dem* в соответствующих случаях.

4. Результаты экспериментов

На *рис. 1* представлены средние значения показателя *Nov* для всех исследованных объектов, вычисленные на основании информации, накопленной за 23 года во всех использованных источниках данных. По совокупности результатов выполнения запросов вычислялись:

– среднеарифметические значения *Nov_1c* и *Nov_4c*, нормированные функциями f_{nk}^{01} и f_{ne}^{01} соответственно;

– медианные значения *Nov_1m* и *Nov_4m*, нормированные функциями f_{nk}^{01} и f_{ne}^{01} соответственно.

Для каждого ряда значений показаны линии трендов (линейная аппроксимация).

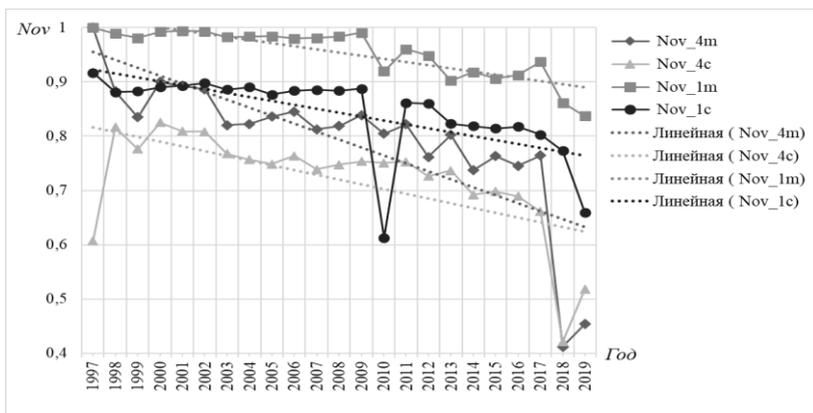


Рис. 1. Средние значения показателя *Nov* по годам (все исследованные объекты, все использованные источники данных)

На рис. 2 показаны значения показателя *Nov* для двух исследованных объектов из области современной медицины: ген-активированного материала для регенерации тканей и технологии лечения пародонта. Вычисления выполнены на основании информации, накопленной за 21 год в двух источниках данных: elibrary.ru и Google Scholar. По совокупности результатов выполнения запросов вычислялись среднеарифметические значения Nov_{4c} , нормированные функцией f_{ne}^{01} . Для каждого ряда значений показаны линии трендов (полиномиальная аппроксимация).

Рис. 3 иллюстрирует динамику изменения средних значений показателя *Dem* для ген-активированного материала для регенерации тканей и технологии лечения пародонта. Вычисления выполнены на основании информации, накопленной за 20 лет в двух источниках данных: elibrary.ru и Google Scholar. По совокупности результатов выполнения запросов вычислялись среднеарифметические значения Dem_{4c} , нормированные функцией f_{ne}^{01} . Для каждого ряда значений показаны линии трендов (полиномиальная аппроксимация).

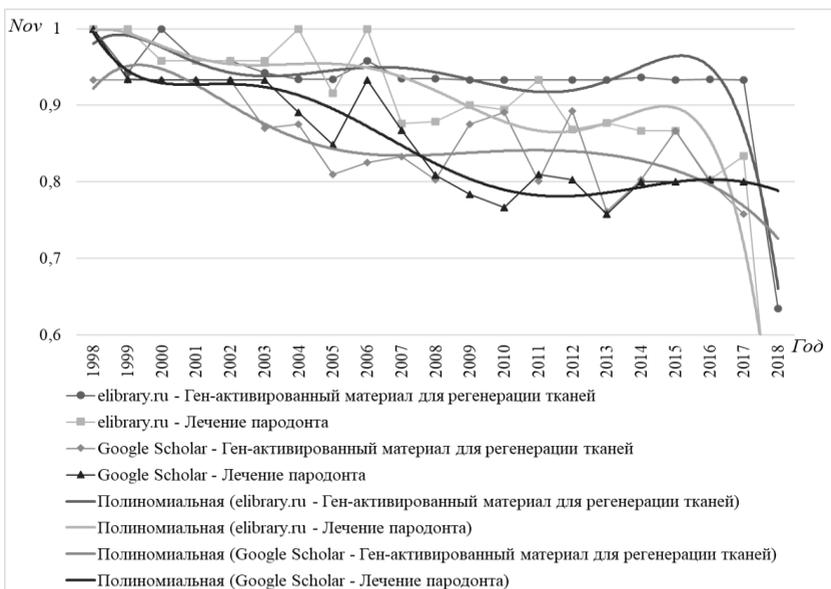


Рис. 2. Значения показателя *Nov* по годам для двух объектов и двух источников данных

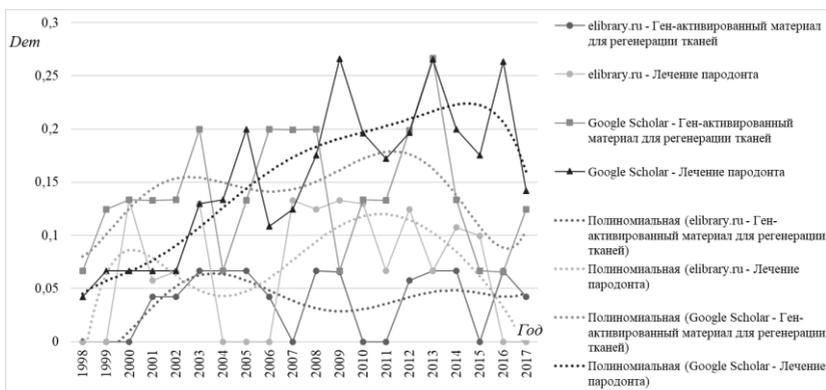


Рис. 3. Средние значения показателя *Dem* по годам для двух объектов и двух источников данных

На рис. 4 представлены средние значения показателя *Dem* для двух исследованных объектов, вычисленные на основании информации, накопленной за 20 лет в двух источ-

никах данных. По совокупности результатов выполнения запросов вычислялись:

- среднеарифметические значения Dem_{1c} и Dem_{4c} , нормированные функциями f_{nk}^{01} и f_{ne}^{01} соответственно;
- медианные значения Dem_{1m} и Dem_{4m} , нормированные функциями f_{nk}^{01} и f_{ne}^{01} соответственно.

Для каждого ряда значений показаны линии трендов (линейная аппроксимация).

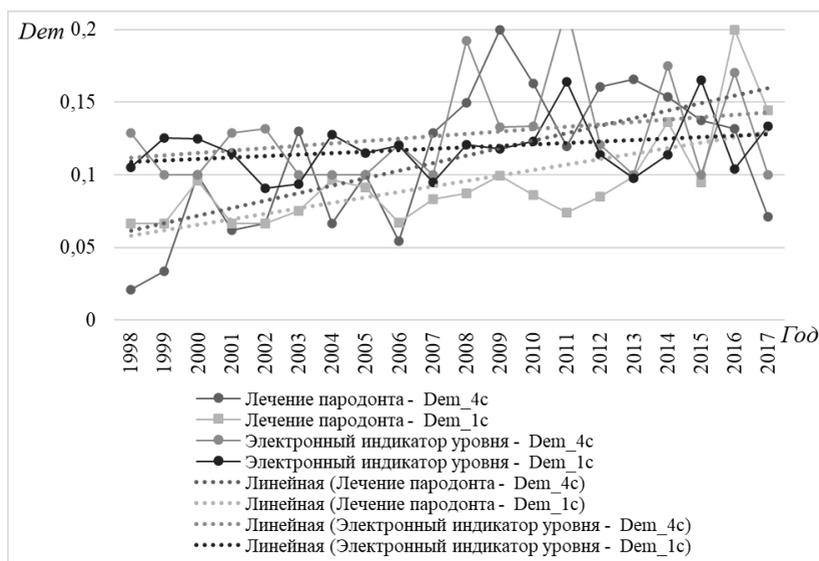


Рис. 4. Значения показателя Dem по годам (два исследованных объекта, два источника данных)

На рис. 5 показано сравнение вычисленных показателей Nov и Dem для объектов, отобранных экспертами как инновационные, и случайно отобранных объектов. Для проведения экспериментов использовались описания изобретений из базы данных ФИПС (<http://www1.fips.ru/iiss>): 10 самых значимых изобретений 2017 года, выбранных экспертами «Роспатента» [5]. Вторая группа – 10 случайно выбранных в базе

данных Роспатента. По совокупности результатов выполнения запросов вычислялись:

– средние значения Nov_4c и Dem_4c , нормированные функцией f_{ne}^{01} ;

– медианные значения Nov_4m и Dem_4m , нормированные функцией f_{ne}^{01} .

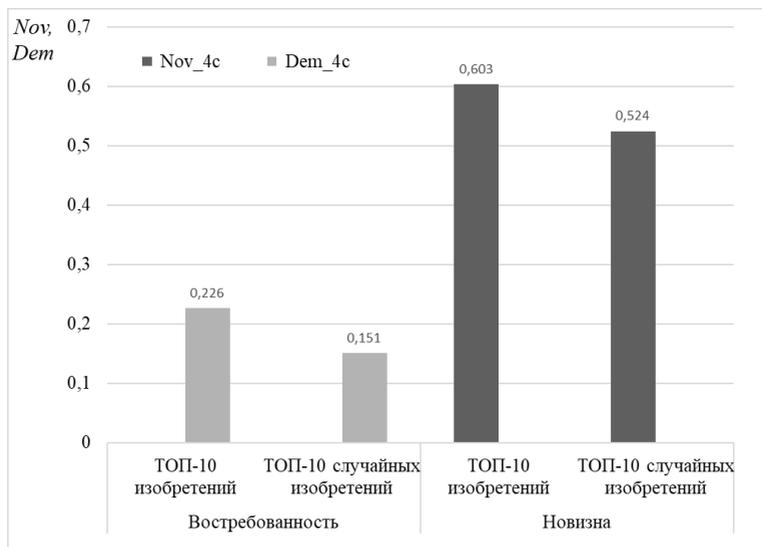


Рис. 5. Сравнение вычисленных показателей Nov и Dem для объектов, отобранных экспертами как инновационные, и случайно отобранных объектов

5. Обсуждение результатов

Очевидно, что использование многих источников информации об объектах для сбора исходных данных при вычислении показателей инновационности позволяет получить более адекватные значения. Например, на рис. 2 для технологии лечения пародонта данные из баз данных elibrary.ru и Google Scholar дают похожие результаты по виду аппроксимирующей кривой, но абсолютные значения показателя Nov отличаются. Аналогичная ситуация с другим объектом, дан-

ные о котором представлены на этом рисунке. *Рис. 3* дает сходную картину для показателя *Dem*: похожие аппроксимирующие кривые, но отличающиеся (иногда существенно) значения показателей. Простое усреднение измеренных значений может улучшить окончательный расчет. Предлагаемое использование таких методов, как теория свидетельств Демпстера – Шафера [4], позволяет еще более обосновано выполнить комбинирование отличающихся результатов.

Как следует из графика на *рис. 1* зависимость показателя новизны объекта от времени есть очевидное уменьшение его значения – линейная аппроксимация рядов экспериментальных значений есть монотонно убывающая функция. Аналогичная ситуация наблюдается практически для всех исследованных объектов (например, *рис. 2*). При этом характер тренда не зависит ни от вида используемой нормирующей функции, ни от используемого источника данных. Достаточно большой временной интервал, на котором измерялись показатели времени, позволяет говорить о приемлемой точности результатов.

На *рис. 2, 3* и *4* показана цикличность изменения показателей *Nov* и *Dem* для различных объектов в течение более, чем 20-летнего периода. Отметим на *рис. 2* приблизительные локальные экстремумы *Nov* для технологии лечения пародонта: 2005 и 2015 гг. (elibrary.ru) и 2003 и 2016 гг. (Google Scholar). На *рис. 4* отметим приблизительные локальные экстремумы *Dem* для того же объекта: 2000 и 2012. Вероятная интерпретация этих данных такова: падение спроса после достижения его максимума вызывает улучшения конструкции, технологии использования, эксплуатационных характеристик объекта. Следствием этого является улучшение показателя новизны объекта, что мы и видим на графиках.

Цикличность изменения показателей *Nov* и *Dem* используется при вычислении показателя имплементируемости объекта *Imp* в соответствии с (4). Рассмотрим расстояния между двумя последовательными точками временного ряда локальных максимумов функций $Dem(t)$ для объекта «ген-активированного материала для регенерации тканей». Как следует из *рис. 3*, эти точки соответствуют 2011 и 2003 гг. Ана-

логичные расстояния для объекта «электронный индикатор уровня» как следует из *рис. 4* соответствуют точкам 2011 и 2002 гг. Как видно, период восстановления спроса вследствие различных улучшений обоих объектов в данном примере достаточно большой (8–9 лет). Нормализованные значения $Imp=0,42$ в первом случае и $Imp=0,47$ — во втором.

Проведенные эксперименты показали совпадение результатов оценки инновационности объектов с помощью вычисленных показателей с экспертными оценками тех же объектов. *Рис. 5* хорошо иллюстрирует это. Среднее значения вычисленных показателей Dem для всех объектов, отобранных экспертами как инновационные, превышает аналогичное значение для случайно отобранных объектов на 49 %. Аналогичное превышение среднего значения вычисленного показателя Nov — 15 %.

Рассмотрим применимость предложенной модели вычисления индекса инновационности объекта для различных типов объектов. Данные экспериментов, некоторые из которых представлены в настоящей статье, показывают сходную динамику поведения измеряемых показателей для различных объектов и источников информации о них. Это касается снижения значений новизны Nov (*рис. 1* и *2*) и повышения значений востребованности Dem (*рис. 4*) со временем, цикличности изменений значений Nov и Dem (*рис. 2* и *3*).

Отметим использование в выполненных экспериментах различных показателей в качестве параметров востребованности объекта. Так, базы данных ACM Digital Library, Google Patents, Google Scholar, IEEE Explore Digital Library и НЭБ использовались для получения количества цитирований материалов об объекте, Яндекс и AliExpress — для получения количество запросов с характеристиками объекта и количество продаж объекта. Во всех случаях полученные данные позволили выполнить требуемые вычисления.

В целом анализ представленных результатов показывает, что основные предположения, использованные при формулировке выражений (2), (3) и (4), подтверждены.

Заключение

Приведенные в статье результаты исследований являются частью проекта РФФИ «Организация и поддержка хранилища данных на основе интеллектуализации поискового агента и эволюционной модели отбора целевой информации» [6]. Они будут использованы в разработке программного обеспечения для селекции информации об инновационных объектах. В целом автоматизация процессов количественной оценки ключевых характеристик современных продуктов и технологий (включая их динамику) должна найти применение в таких областях, как: выбор направлений бизнеса и объектов инвестиций, экспертиза проектов, проведение сравнительного анализа различных продуктов.

Работа выполнена при финансовой поддержке РФФИ (проект № 18-07-00358-а).

Литература

1. Ivanov V. K. Computational model to quantify object innovativeness. In: CEUR Workshop Proceedings, vol. 2258, pp. 249-258 (2019). <http://ceur-ws.org/Vol-2258/paper31.pdf> (last accessed 2020/07/12)
2. Ivanov V. K., Palyukh B. V., Sotnikov A. N. Features of Data Warehouse Support Based on a Search Agent and an Evolutionary Model for Innovation Information Selection, Proceedings of the Fourth International Scientific Conference "Intelligent Information Technologies for Industry" (ITI'19), Springer Nature Switzerland AG, pp. 120-130, (2020), DOI: <https://doi.org/10.1007/978-3-030-50097-9>
3. Hawking D., Moffat A. & Trotman A. Efficiency in information retrieval: introduction to special issue. *Inf Retrieval J* 20, 169-171 (2017). <https://doi.org/10.1007/s10791-017-9309-7>
4. Yager R., Liping Liu. *Classic Works of the Dempster-Shafer Theory of Belief Functions*, London: Springer, 2010.
5. «Электрический глаз», блокчейн-технологии и биохакинг: Роспатент составил топ-10 российских разработок 2017 года [Электронный ресурс]. АНО «ТВ-Новости». 2017 – URL: <https://ru.rt.com/9zqx> (Дата обращения 12.07.2020).
6. Организация и поддержка хранилища данных на основе интеллектуализации поискового агента и эволюционной модели отбора

целевой информации. URL: https://www.rfbr.ru/rffi/ru/project_search/o_2071601 (Дата обращения 12.07.2020).

References

1. Ivanov V. K. Computational model to quantify object innovativeness. In: CEUR Workshop Proceedings, vol. 2258, pp. 249–258 (2019). <http://ceur-ws.org/Vol-2258/paper31.pdf> (last accessed 2020/07/12).
2. Ivanov V. K., Palyukh B. V., Sotnikov A. N. Features of Data Warehouse Support Based on a Search Agent and an Evolutionary Model for Innovation Information Selection, Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry” (ИИТ’19), Springer Nature Switzerland AG, pp. 120–130, (2020), DOI: <https://doi.org/10.1007/978-3-030-50097-9>
3. Hawking D., Moffat A. & Trotman A. Efficiency in information retrieval: introduction to special issue. *Inf Retrieval J* 20, 169–171 (2017). <https://doi.org/10.1007/s10791-017-9309-7>
4. Yager R., Liping Liu. *Classic Works of the Dempster-Shafer Theory of Belief Functions*, London: Springer, 2010.
5. “Elektricheskij glaz”, blokchejn tekhnologii i biohaking: Rospatent sostavil top 10 rossijskih razrabotok 2017 goda. ANO TV Novosti — 2017 — URL: <https://ru.rt.com/9zqx> (last accessed 12.07.2020).
6. Organizaciya i podderzhka hranilishcha dannyh na osnove intellektualizacii poiskovogo agenta i ehvolyucionnoj modeli otbora celevoj informacii. RFFI — 2018 — URL: https://www.rfbr.ru/rffi/ru/project_search/o_2071601 (last accessed 12.07.2020).

Мягкие измерения и вычисления при эволюционном управлении непрерывным многостадийным производством

Б. В. Палюх, А. Н. Ветров

ТвГТУ

Аннотация. Рассматривается механизм эффективного управления эволюцией многостадийного производства, основанный на применении мягких измерений и вычислений. Для снижения чувствительности к шумовым помехам и запаздываниям по динамическим каналам между входом и выходом непрерывного многостадийного производства предлагается система диагностики, основанная на анализе интервалов изменения значений диагностических переменных за заданный интервал времени. Система принятия решений базируется на методах теории Демпстера – Шафера. Модель архитектуры управления процессом описывается как многоагентная система, в которой агенты взаимодействуют между собой и с центром. Обосновываются методы и средства управления эволюцией многостадийных непрерывных процессов для повышения эффективности и долгосрочной устойчивости в течение всего жизненного цикла путем внедрения инновационных решений.

Ключевые слова: эволюционное управление, непрерывное многостадийное производство, мягкие измерения и вычисления, интервальный анализ, многоагентная система, теория Демпстера – Шафера, инновационные решения.

Soft measurement and calculation in evolutionary control of continuous multistage production

B. Polykh, A. Vetrov

Tver State Technical University

Abstract. The mechanism of effective management of the evolution of multi-stage production, based on the use of soft measurements and calcu-

lations, is considered. To reduce the sensitivity to noise interference and delays in dynamic channels between the input and output of continuous multi-stage production, a diagnostic system based on the analysis of the intervals of changes in the values of diagnostic variables for a given time interval is proposed. The decision-making system is based on the methods of the Dempster-Schafer theory. The process management architecture model is described as a multi-agent system in which agents interact with each other and with the center. Methods and tools for managing the evolution of multi-stage continuous processes to improve efficiency and long-term sustainability throughout the life cycle through the introduction of innovative solutions are substantiated.

Keywords: evolutionary control, continuous multi-stage production, soft measurements and calculations, interval analysis, multi-agent system, Dempster-Schafer theory, innovative solutions.

Эволюция непрерывного многостадийного производственного процесса рассматривается как элемент её самоорганизации, определяемый структурными изменениями в процессе функционирования отдельных стадий технологического процесса и снижением показателей инновационности и конкурентоспособности выпускаемых продуктов. Критическое состояние производства, при котором необходимы его структурные изменения, определяет точку бифуркации, т. е. момент принятия решений по поиску и внедрению инноваций для повышения эффективности его функционирования. Предлагается определять критическое состояние производственной системы на основе непрерывного контроля диагностических переменных. Это дает возможность адекватно определять локацию структурных изменений системы при отсутствии полной информации о корреляционных связях между отдельными процессами, протекающими в производственной системе и управлять эволюцией непрерывных многостадийных процессов в условиях неопределенности.

Состояние отдельных стадий непрерывного производства определяется методами параметрической диагностики, которые предполагают непосредственное измерение большого количества диагностических переменных и вычисление не измеряемых по косвенным. Из-за шумовых помех

и запаздываниям сигнала по динамическим каналам следует применять интервальный анализ [1].

Вероятность ошибочных решений можно уменьшить, если применить интервальный анализ. Пусть имеется интервал значений диагностического параметра

$$X = [x; \bar{x}] = \{x | x \leq \bar{x}; x, \bar{x} \in R\},$$

где R – множество вещественных чисел.

Если анализировать не мгновенные значения диагностического параметра, а интервалы его изменения за заданный интервал времени

$$T = [t; \bar{t}] = \{t | t \leq \bar{t}; t, \bar{t} \in R\},$$

то вероятности ложной тревоги и пропуска дефекта уменьшаются.

Теорема 1. Вероятность ложной тревоги системы диагностики непрерывного производства, построенной по принципу анализа мгновенного значения диагностического параметра, больше вероятности ложной тревоги системы диагностики, построенной по принципу анализа интервала значений диагностического параметра.

Теорема 2. Вероятность пропуска дефекта системы диагностики непрерывного производства, построенной по принципу анализа мгновенного значения диагностического параметра, больше вероятности пропуска дефекта системы диагностики, построенной по принципу анализа интервала значений диагностического параметра.

Обоснование теорем приведено в [1].

Вторую трудность, связанную с запаздыванием динамического канала между входом и выходом, можно преодолеть, если выбрать интервал времени T таким образом; чтобы $t = to$ и $\bar{t} = to + \tau$, где to – время начала фиксирования величины изменения диагностических параметров, τ – время запаздывания динамического канала между входом и выходом объекта. За выбранный интервал времени измеряются ин-

тервалы наблюдаемых технологических переменных $Z = [z; \bar{z}]$. Необходимо найти гарантированную оценку вычисляемых диагностических показателей в виде интервалов $X = [x; \bar{x}]$. Математическая модель объекта диагностики обычно задается в виде системы алгебраических или дифференциальных уравнений. В нашем случае уравнения будут с интервальными переменными и коэффициентами.

В качестве примера построения параметрических моделей в виде интервальных обыкновенных дифференциальных уравнений возьмем модель каталитического окисления аммиака в контактном аппарате производства слабой азотной кислоты:

$$\frac{dX_{28}}{dt} = 13500 \left(\frac{Z_3 + Z_4}{172800} \right)^{0.56} (2.266X_5)^{0.59} \times (1 - X_{28}) \exp \left[-793 \left(\frac{1}{273 + Z_5} - \frac{1}{1123} \right) \right], X_{28}(0) = 0,$$

$X_5 = [0.655; 0.678]$ — давление в контактном аппарате, МПа.

$Z_3 = [5200; 5300]$ — расход газообразного аммиака для аммиачно-воздушной смеси (АВС), м³/ч;

$Z_4 = [48000; 49000]$ — расход воздуха для АВС, м³/ч;

$Z_5 = [898; 902]$ — температура нитрозного газа под сетками контактного аппарата, К.

X_{28} — общая степень превращения аммиака.

Решая это уравнение, получим зависимость общей степени превращения аммиака X_{28} от времени t . Уравнение решается двусторонними численными методами. Решение уравнения по исходным данным, взятым из приведенного выше примера, иллюстрируется на *рис. 1*.

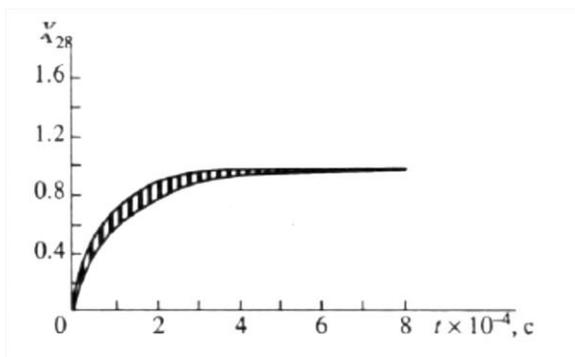


Рис. 1. Гарантированная оценка общей степени превращения аммиака от времени t

Проведенные исследования позволяют сделать вывод о необходимости построения систем управления эволюцией многостадийных непрерывных производств не на традиционных принципах анализа мгновенных значений диагностических параметров, а на анализе интервалов изменения этих параметров за определенный интервал времени [2]. Поэтому задачи управления эволюцией непрерывных производств наиболее целесообразно решать с помощью мягких измерений и вычислений [3].

Задача эволюционного управления описывается как многоагентная система. Пусть химико-технологическая система состоит из центра Z и n групп агентов $A = \{A_1, A_2, \dots, A_n\}$, каждый агент $a = \{a_1, a_2, \dots, a_N\}$ обладает свойствами активности и автономности. Каждый агент a_k группы агентов A_n (каждая группа агентов состоит из k агентов) управляет каким-либо технологическим узлом. Возможности k -го агента по выпуску продукции описываются технологическим множеством — областью X_k в пространстве размерности n_k измерений. При этом любой допустимый план X_k k -го агента должен принадлежать области X_k .

План всей активной системы будет описываться вектором, имеющим размерность $N = \sum_{k=1}^n n_k$. Допустимый план

x должен удовлетворять не только локальным ограничениям, но так же ряду ограничений вида $G(x) \geq b$, где

$$G(x) = \{g_i(x), i = \overline{1, L}\}, b = \{b_i, i = \overline{1, L}\}$$

Суть ограничений состоит в том, что за их выполнение отвечает и следит только центр, управляющий всей системой, а не отдельный агент. Общие ограничения отражают задания по выпуску продукции, наличие ресурсов, потребляемых производством из внешней среды, связи по материальным потокам между технологическими стадиями и т. п.

Будем считать, что функции $g_i(x), i = \overline{1, L}$, — вогнутые, дифференцируемые функции, а множество X — выпуклое множество. Тогда задача, решаемая центром Z , может рассматриваться как двойственная задача выпуклого программирования

$$\begin{aligned} \Phi(x) &\rightarrow \max, \\ G(x) &\geq b, \end{aligned}$$

Практическая невозможность для центра Z получить точное представление о технологических возможностях агентов предполагает необходимость разработки специальных процедур обмена информацией между центром и агентами. Это позволяет сформировать согласованный план без необходимости решения задачи математического программирования в полном объеме с учетом всех глобальных и локальных ограничений, отражающих интересы агентов.

При принятии решений используется как формализованная, так и неформализованная информация, поэтому важнейшим вопросом является объединение различного вида информации в пользу той или иной гипотезы. В соответствии с представленными выше принципами эволюционного управления неопределенность в задачах принятия решений учитывается путем использования теории сви-

детельств (теория Демпстера – Шафера) [4] и аппарата интервальной математики. В этом случае вероятность утверждения A определяется как интервал $[S(A), P(A)]$, где $S(A)$ изменяется в пределах $[0;1]$ – вероятность очевидности утверждения A (достаточность). Величина $P(A)$ представляет собой вероятность правдоподобия утверждения A (необходимость) и вычисляется как $P(A) = 1 - S(\bar{A})$. Разница $u(A) = P(A) - S(A)$ представляет собой неопределенность диагноза. Здесь \bar{A} – есть отрицание A . Таким образом, доказательство подтверждения гипотезы не отрицает ее отрицания, как это происходит при байесовом подходе, что больше соответствует структуре человеческого мышления. Байесов подход возникает как частный случай при равенстве $S(A) = P(A)$.

Оказывается, что, несмотря на неизвестность конкретных мер вероятности, закрепленных за отдельными объектами с заданными свойствами, можно сделать некоторые выводы, исходя только из известного распределения мер вероятности. Поэтому принятие решений для поставленной задачи наиболее целесообразно осуществлять на основе теории Демпстера – Шафера [4].

При выявленном неэффективном функционировании конкретной стадии многостадийного производства необходимо осуществить тематический поиск инноваций для устранения причины кризисного состояния [5]. Проведенные исследования показали целесообразность применения подхода к генерации поисковых запросов, основанного на генетическом алгоритме. Этот подход использован для уточнения семантического ядра искомого множества документов и генерации множества эффективных запросов. Постановка задачи предусматривает организацию эволюционного процесса, формирующего устойчивую и эффективную популяцию поисковых запросов, образующую соответствующий поисковый образ документа. Целевое множество результатов поиска должно формироваться такими адресами документов, которые (а) находятся в первых позициях ранжированного списка, построенного поисковой системой; (б) присутствуют в списках результатов, полученных при вы-

полнении большинства запросов; (в) семантически близки к эталонным текстам, формируемым в ходе эволюции запросов; (г) соответствуют условиям среды, задаваемыми поисковому агенту параметрами пользователя.

Предлагаемые принципы эволюционного управления непрерывным многостадийным производством позволяют оперативно определять кризисные состояния (точки бифуркации) конкретной стадии производства, прогнозировать их развитие на заданном интервале времени, своевременно принимать решение по внедрению инновационных технологий, современного оборудования и организации гибкой системы управления многостадийным технологическим процессом.

Литература

1. Кафаров В. В., Палюх Б. В., Перов В. Л. Решение задачи технической диагностики непрерывного производства с помощью интервального анализа // Докл. АН СССР. 1990. Т. 311. № 3. С. 677–680.
2. Палюх Б. В., Виноградов Г. П., Егерев И. А. Управление эволюцией химико-технологической системы // Теоретические основы химической технологии / Наука. 2014. Т. 48. № 3. С. 349–355.
3. Палюх Б. В. Принципы управления эксплуатационной надежностью непрерывного производства на основе мягких измерений и вычислений // Мягкие измерения и вычисления. 2018. № 8(9). С. 6–16.
4. Ronald R. Yager, Liping Liu, «Classic Works of the Dempster – Shafer Theory of Belief Functions». Springer, 2008.
5. Ivanov V. K., Palyukh B. V., Sotnikov A. N. Approaches to the Intelligent Subject Search // Federated Conference on Computer Science and Information Systems (FedCSIS'2014) (September 7–10, 2014. Warsaw, Poland). Annals of Computer Science and Information Systems. Volume 3. Position Papers, DOI 10.15439/978–83–60810–57–6. Warsawa, 2014. P. 13–20.

References

1. Kafarov V. V., Palyukh B. V., Perov V. L. Reshenie zadachi tekhnicheskoy diagnostiki nepreryvnogo proizvodstva s pomoshchyu intervalnogo analiza // Dokl. AN SSSR. 1990. T. 311. № 3. С. 677–680.
2. Palyukh B. V., Vinogradov G. P., Egereva I. A. Upravlenie evolyuciej himiko- tekhnologicheskoy sistemy // Teoreticheskie osnovy himicheskoy tekhnologii / Nauka. 2014. T. 48. № 3. С. 349–355.

3. Palyukh B. V. Principy upravleniya ekspluatacionnoj nadezhnostyu nepreryvnogo proizvodstva na osnove myagkih izmerenij i vychislenij // Myagkie izmereniya i vychisleniya. 2018. № 8(9). С. 6–16.
4. Ronald R. Yager, Liping Liu, "Classic Works of the Dempster-Shafer Theory of Belief Functions". Springer, 2008.
5. Ivanov V. K. Palyukh B. V. Sotnikov A. N. Approaches to the Intelligent Subject Search // Federated Conference on Computer Science and Information Systems (FedCSIS'2014) (September 7–10, 2014. Warsaw, Poland). Annals of Computer Science and Information Systems. Volume 3. Position Papers, DOI 10.15439/978-83-60810-57-6. Warsawa, 2014. P. 13–20.

Информационное пространство современности: реалии, проблемы и перспективы

Н. А. Коровникова

ФИЦ ИУ РАН

Аннотация. В статье рассматриваются основные подходы к определению понятия информационное пространство. Показана его взаимосвязь с ментальным пространством. Проанализированы проблемы и перспективы их развития в контексте современности.

Ключевые слова: информация, информационное пространство, ментальное пространство, современность.

Modern information space: realities, problems and prospects

N. A. Korovnikova

FRC "Computer Science and Control" RAS

Abstract. The article considers main approaches to the definition of the concept of information space. Shows its interconnection with mental space. Analyzes the problems and prospects of their development in the context of modernity.

Key words: information, information space, mental space, modernity.

В современном глобальном контексте особенности информационного пространства, его элементов и процессов стали определяющими факторами условий жизни населения, цифровизации всех сфер жизнедеятельности, мирового развития в целом. Особую значимость состояние, качество и транспарентность информационного пространства на всех уровнях (от микроиндивидуального до макрореждународного) приобрели в условиях общечело-

веческой угрозы (пандемии коронавирусной инфекции COVID-19), противостояние которой представляется возможным только при условии эффективного сотрудничества всех субъектов и объектов информационного пространства.

В современном общественно-научном дискурсе циркулируют различные коннотации и дефиниции понятия «информационное пространство». Отталкиваясь от одного из довольно широких определений «информации» как «сведений (данных) независимо от формы их представления» [1, с. 141], наиболее универсальной представляется трактовка информационного пространства, как эволюционно сформировавшейся структуры, детерминированной правовыми инструментами и предоставляющей высокую степень открытости и достоверности информационных ресурсов, полученных в результате коммуникационной деятельности большинства информационных субъектов [1, с. 141–142].

В академической среде выделяют несколько подходов к пониманию информационного пространства современности, например:

– *материально-технический (функциональный) подход*, в рамках которого информационное пространство представляет собой информационную инфраструктуру и/или «техническую» систему передачи, обработки и хранения информации с использованием технического инструментария и материальных носителей [1, с. 142];

– *территориальный подход*, в русле которого информационное пространство представляет собой «информатизированную территорию», где «физически» располагаются субъекты, объекты и источники информации [2, с. 270];

– *социальный подход*, определяющий информационное пространство как «сферу деятельности» информационных субъектов и их сообществ, которая заключается в реализации информационно-виртуального обмена [2, с. 271];

– и, наконец, *гуманитарный подход*, согласно которому информационное пространство аккумулирует знания и опыт ментальной эволюции современного человеческого сообщества [1, с. 142], учитывая, что в сегодняшнем контексте существенную роль в создании и трансформации менталь-

ных (когнитивных и аксиологических) структур общественного сознания реализуют информационные структуры, например, такие компании, как Apple, Google, Facebook, Youtube, Instagram и т. п., детерминирующие современную ментальность человечества [3].

Очевидная взаимосвязь информационного и ментального пространств современности также объясняется их схожей *структурой*, основными элементами которой являются: «аудитория», способная к восприятию и трансляции информационно-ментальных данных; динамично развивающаяся система ИКТ; и, что немало важно в глобальных (пост)пандемических условиях, международная система информационного и ментального взаимодействия [1, с. 143].

Исходя из вышеизложенного, можно констатировать, что информационное и ментальное пространства современного социума столкнулись с тождественными *трендами и проблемами* их дальнейшего развития, в том числе: а) ментальные вирусы и эпидемии (психические и психосоматические расстройства общественного сознания) [4, с. 85], например, «синдром рассеянного внимания» [5, с. 134]; «деструктивные социогенез, психогенез, анимогенез» (социально-ментальная деформация, распространение депривационно-изоляционных тенденций, информационная деградация общественной среды) [4, с. 89–90]; б) информационная гипертекстуальность, тотальная визуализация культурно-ментального пространства [5, с. 134–136]; в) формирование субкультур деструктивной информационно-ментальной направленности (например, группа NEET) [6, с. 583]; г) формирование «клипового мышления», для которого характерна информационная гетерогенность, фрагментарность и алогичность [5, с. 137–138]; д) превалирование «метаинформации», которое чревато нарушением словесно-логической памяти [5, с. 140], а также «утечкой» ментальных ресурсов в социальные сети [3] и т. д.

И, наконец, конструктивное развитие ментально-информационного пространства должно способствовать преодолению поляризации и крайних проявлений индиви-

дуализации, экономического неравенства и дезинтеграции, девальвации основополагающих аксиологических ориентиров, что, в свою очередь, является обязательным условием преодоления и предотвращения (пост)пандемических угроз и последствий в мировом масштабе.

Литература

1. Добровольская И. А. Понятие «информационное пространство»: различные подходы к его изучению и особенности // Вестник РУДН, серия Литературоведение. Журналистика. 2014. № 4. С. 140–147.
2. Чайковский Д. В. Информационное пространство: анализ определений // Вестник Бурятского государственного университета. Педагогика. Филология. Философия. 2010. С. 269–274. Режим доступа: <https://cyberleninka.ru/article/n/informatsionnoe-prostranstvo-analiz-opredeleniy>
3. Почепцов Г. Как информационные технологии атакуют ментальное пространство человечества // ResearchGate GmbH. 2018. 05.11. Режим доступа: https://www.researchgate.net/publication/328738753_Kak_informacionnye_tehnologii_atakuut_mentalnoe_prostranstvo_celovecestva
4. Сидоров П. И. Синдром приобретенного ментального иммунодефицита // Медицинский академический журнал. 2015. Т. 15. № 4. С. 82–95.
5. Современная когнитология и когнитивная аналитика в контексте философской инноватики. Монография / Научн. ред. проф. А. М. Старостин. Ростов н/Д.: Изд-во ЮРИУ РАНХиГС, 2014. 228 с. Режим доступа: https://www.google.ru/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwjYjMyFtrDoAhWE-yoKHW7yBcsQFjADegQIBVAB&url=https%3A%2F%2Fhub.lib.sfedu.ru%2Fstorage%2F1%2F389108%2F2194b49c-bddb-4714-80ab-9c53ec7b1e15%2F&usg=AOvVaw2v_bEITwrFoSgkV1TE5JI
6. Ломоносовские чтения-2019. Секция экономических наук. Экономические отношения в условиях цифровой трансформации: сборник тезисов выступлений. М.: Экономический факультет МГУ имени М. В. Ломоносова, 2019. 1046 с. Режим доступа: <https://www.econ.msu.ru/sys/raw.php?o=56275&p=attachment>

Международный научно-практический журнал «Программные продукты и системы» — более 30-ти лет в цифровом пространстве научных знаний

Н. А. Семенов

ТвГТУ

Аннотация. Международный научно-практический журнал «Программные продукты и системы» был организован в 1988 году по инициативе Генерального директора НПО «Центрпрограммсистем» проф. Тихомирова В. П. (г. Тверь) и руководителя Главной редакции международного журнала «Проблемы теории и практики управления» к.э.н. Силина В. М. (г. Москва). В период с 1988 по 2018 гг. главным редактором журнала был академик РАН С. В. Емельянов. С 2020 года международную редакционную коллегию, в которую входят известные ученые в области информационных технологий РФ, Белоруссии, Украины, Азербайджана, Германии, Финляндии, Мексики, Вьетнама и других стран, возглавляет научный руководитель Межведомственного суперкомпьютерного центра НИИСИ РАН, академик РАН Савин Г. И. Ассоциированными членами редакции являются Национальный исследовательский университет «МЭИ» (г. Москва), Технологический институт ЮФУ (г. Таганрог), ТвГТУ и НИИ «Центрпрограммсистем». Журнал реферируется в библиографических БД РИНЦ, CrossRef, внесен в Перечень ведущих рецензируемых журналов и изданий, рекомендуемых ВАК РФ для публикаций результатов исследований при защите кандидатских и докторских диссертаций по специальностям «Информатика, вычислительная техника и управление» (05.13.XX).

В журнале публикуются результаты оригинальных исследований в области информационных технологий, прикладных систем искусственного интеллекта, в том числе, разработки баз данных и баз знаний, экспертных и обучающих систем, систем поддержки принятия решений и распознавания образов, многоагентных систем, интеллектуального анализа данных, нейронных сетей, роботов, телекоммуникационных систем. При этом предметная область исследований не ограничивается.

Журнал «Программные продукты и системы» регулярно выступает в качестве информационного спонсора при проведении международных и национальных научно-технических мероприятий в области информационных технологий и искусственного интеллекта.

Ключевые слова: прикладные системы искусственного интеллекта, экспертные системы, системы поддержки принятия решений, нейронные сети, многоагентные системы, интеллектуальный анализ данных.

**International scientific and practical magazine
«Software products and systems» — more than 30
years
in the digital space of scientific knowledge**

N. A. Semenov

Centerprogramsystem Research Institute

Abstract. The international scientific and practical journal «Software products and systems» was organized in 1988 on the initiative of the General Director of the NGO «Centerprogramsystem» prof. V. P. Tikhomirov (Tver) and V. M. Silina (Moscow), head of the Main editorial office of the international journal «Problems of management theory and practice». In the period from 1988 to 2018, the editor-in-chief of the journal was academician S. V. Yemelyanov. Since 2020, the international editorial Board, which includes well-known scientists in the field of information technologies of the Russian Federation, Belarus, Ukraine, Azerbaijan, Germany, Finland, Mexico, Vietnam and other countries, is headed by academician G. I. Savin. Associate members of the editorial Board are the national research University «MEI» (Moscow), the technological Institute of the southern Federal University (Taganrog), TvSTU and the research Institute «Centerprogramsystem». The journal is reviewed in the bibliographic databases of RSCI, CrossRef, and is included in the List of leading peer-reviewed journals and publications recommended by the higher attestation Commission of the Russian Federation for publication of research results in the defense of candidate and doctoral theses in the fields of computer Science, computer engineering and management (05.13.XX).

The journal publishes the results of original research in the field of information technologies, applied artificial intelligence systems, including the development of databases and knowledge bases, expert and training systems, decision support systems and image recognition, multi-agent sys-

tems, data mining systems, neural networks, robots, and telecommunications systems. However, the subject area of research is not limited.

The magazine «Software products and systems» regularly acts as an information sponsor for international and national scientific and technical events in the field of information technology and artificial intelligence.

Keywords: applied artificial intelligence systems, expert systems, decision support systems, neural networks, multi-agent systems, data mining.

Международный научно-практический журнал «Программные продукты и системы» был организован в 1988 году по инициативе Генерального директора НПО «Центрпрограммсистем» проф. Тихомирова В. П. (г. Тверь) и руководителя Главной редакции международного журнала «Проблемы теории и практики управления» к.э.н. Силина В. М. (г. Москва). В период с 1988 по 2018 гг. главным редактором журнала был академик РАН С. В. Емельянов. С 2020 года международную редакционную коллегию, в которую входят известные ученые в области информационных технологий РФ, Белоруссии, Украины, Азербайджана, Германии, Финляндии, Мексики, Вьетнама и других стран, возглавляет научный руководитель Межведомственного суперкомпьютерного центра РАН – филиала ФГУ ФНЦ НИИСИ РАН, академик РАН Савин Г. И. Ассоциированными членами редакции являются Национальный исследовательский университет МЭИ (г. Москва), Технологический институт ЮФУ (г. Таганрог), ТвГТУ и НИИ «Центрпрограммсистем» (г. Тверь). Журнал реферируется в библиографических БД РИНЦ, CrossRef, внесен в Перечень ведущих рецензируемых журналов и изданий, рекомендуемых ВАК РФ для публикаций результатов исследований при защите кандидатских и докторских диссертаций по специальностям 05.13.XX – информатика вычислительная техника и управление.

По состоянию на 01.01.2020 г. в рейтинге Science Index по тематике «Автоматика и вычислительная техника» журнал занимает десятое место, пятилетний индекс Херфиндаля по цитирующим журналам составляет 310, десятилетний индекс Хирша – 19, число цитирований – 8001. Только

за период с 2015 по 2019 гг. опубликовано 602 статьи общим объемом 250 условных печатных листов.

В журнале, выходящем один раз в квартал (объем до 250 страниц, формат А4), публикуются результаты оригинальных исследований в области информационных технологий, прикладных систем искусственного интеллекта, в том числе, разработки баз данных и баз знаний, экспертных и обучающих систем, систем поддержки принятия решений и распознавания образов, многоагентных систем, интеллектуального анализа данных, нейронных сетей, роботов и телекоммуникационных систем. При этом предметная область исследований может варьироваться от медицины и сельского хозяйства до космических исследований и нанотехнологий. Многие исследования авторов поддержаны грантами Российского научного фонда и Российского фонда фундаментальных исследований.

Все статьи, поступающие в редакцию, подлежат обязательному рецензированию в течение месяца со дня поступления. Рецензирование проводится конфиденциально одним из членов редакционной коллегии. При необходимости статья отправляется на доработку. Требования к публикации и архив электронных версий журнала с 1988 по 2020 гг. приведены на сайте журнала: www.swsys.ru. Статьи, одобренные редакционным советом, публикуются бесплатно в течение года.

Проведем дайджест-анализ отдельных статей российских и зарубежных авторов, опубликованных в журнале за последние несколько лет. В качестве зарубежных авторов, как правило, выступают аспиранты из стран Азии, Ближнего Востока, Африки и Латинской Америки, обучающиеся в российских вузах.

Для реализации прецедентного подхода в статье **«Реализация прецедентного модуля для интеллектуальных систем»** (Зо Лин Кхаинг, Ар Кор Мью, Варшавский П. Р., Алехин Р. В., НИУ МЭИ, Москва. № 2/2015) авторы предлагают использовать сетевую (онтологическую) модель представления прецедентов и гибридный алгоритм извлечения прецедентов, базирующийся на теории структурного отображения и методе ближайшего соседа. Описаны особенно-

сти программной реализации основных модулей прототипа CBR-системы в среде программирования MS Visual Studio 2010.

«Архитектура пользовательского интерфейса для взаимодействия с мультиагентной средой» (Сидоров И. А., Институт динамики систем и теории управления СО РАН, Иркутск. № 4/2016). В статье предлагается оригинальная архитектура пользовательского веб-интерфейса, основанная на концепции Agent-as-a-Service, для организации взаимодействия между пользователями и агентами сети. Архитектурный каркас веб-приложения спроектирован с применением схемы MVC. Предложенный подход снижает нагрузку на узел агента, улучшает кроссплатформенность и встраиваемость компонентов пользовательского интерфейса в различные подсистемы мультиагентной среды, облегчает тестирование и поддержку веб-приложений.

В статье **«Обогащение модели Bag-of-words семантическими связями для повышения качества классификации текстов предметной области»** (Нугуманова А. Б., Бессмертный И. А., Пецина П., Байбурин Е. М., Восточно-казахстанский ГТУ им. Д. Сергибаева, НИЦ ИТМО, Карлов университет. Усть-Каменогорск (Казахстан), Санкт-Петербург, Прага (Чехия). № 2/2016) авторы предлагают обогатить модель Bag-of-words семантическими связями, которые извлекаются из текстов на основе статистики совместной встречаемости слов, то есть новый способ построения и применения матрицы семантических связей, которые затем используются для отображения представлений текстов в пространство связанных слов. Серия экспериментов на стандартной коллекции Reuters 21578, проведенных авторами, показала повышение качества классификации текстов по сравнению с известными методами.

Статья **«Гибридные когнитивные нечеткие системы управления автономным роботом на основе нейроинтерфейса и технологии мягких вычислений»** (Ульянов С. В., Решетников А. Г., Мамаева А. А., Международный университет природы, общества и человека. Дубна. № 3/2017) посвящена рассмотрению возможности управления типовым дви-

жением объекта посредством когнитивного шлема с помощью стандартного блока распознавания команд и различных типов систем управления, в том числе, на основе оптимизатора баз знаний на мягких вычислениях.

«Программный комплекс решения задач кластеризации» (Григораш А. С., Курейчик В. М., Курейчик В. В., Технологический институт ЮФУ. Таганрог. № 2/2017). Особенностью предлагаемого авторами подхода к решению задач кластеризации является использование модифицированных методов эволюционного моделирования и роевого интеллекта, которые адаптируются к изменениям внешней среды. Выделение модификации для нахождения квазиоптимальных решений, позволило сократить время формирования кластеров.

Особенность постановки задачи в статье **«Системы автоматического управления с запаздыванием: робастность, быстродействие, синтез»** (Гхан В. З., Берчук Д. Ю., Томский политехнический институт. № 1/2017) связана с формированием и решением уравнения синтеза численным методом без аппроксимации передаточной функции звена запаздывания с целью повышения точности синтеза регуляторов. Модель желаемой системы, совместно с не аппроксимированным описанием объекта управления и принятой структурой регулятора, позволили сформировать более точное уравнение по сравнению с традиционным подходом.

В статье **«Грамматика запросов для хранилища разнородных данных в проактивных системах»** (Чан Ван Фу, Сай Ванг Квонг, Щербаков М. В., Волгоградский ГТУ. № 4/2018) предложена грамматика, основанная на расширениях DML языка SQL. Для обработки сформированных запросов сгенерирован парсер с использованием библиотеки ANTLR 3.0. В результате генерации созданы классы на языке Java, объекты которых используются для разбора запросов.

Особенностью предложенной методики в статье **«Методика оценки качества гетерогенных робототехнических комплексов моделью Banker-Charnes-Cooper»** (Сильвестров Д. Е., Русаков К. Д., Хиль С. Ш., Савилкин С. Б., Военная

академия РВСН им. Петра Великого, ИПУ им. В. А. Трапезникова РАН, НИУ МАИ. № 4/2018) является то, что исследуемой объект обладает множеством входных и выходных разнородных параметров, а решение задачи осуществляется в условиях этой разнородности. Разработанная модель основана на выходной модели Banker-Charnes-Cooper и модели синтеза, основанной на аддитивной модели методологии анализа среды функционирования.

В статье **«Нейросетевой метод обнаружения вредоносных программ на платформе Android»** (Татарников Т. М., Журавлев А. М., ГУАП, Санкт-Петербург. № 3/2018) с целью повышения вероятности обнаружения вредоносных программ операционной системы Android, авторами предложено использование классификационных признаков, полученных как на основе статистического анализа кода, так и с помощью анализа поведения программы в виртуальной среде.

«О реализации средств машинного обучения в интеллектуальных системах реального времени» (Еремеев А. П., Кожухов А. А., Голенков В. В., Гулякина Н. А., НИУ МЭИ, Белорусский ГУИР, Москва, Минск. № 2/2018). В статье описана реализация алгоритмов обучения с подкреплением (RL) на основе темпоральных различий и гибких алгоритмов, которые способны находить приемлемые решения в условиях жестких временных ограничений. Предложенный авторами алгоритм включает статистический модуль прогнозирования и мультиагентный модуль RL-обучения. В работе приводится обоснование технологии OSTIS, используемой для разработки интеллектуальных систем реального времени.

«Исследование оптимального количества процессорных ядер для алгоритма многократной маркировки перколяционных кластеров на суперкомпьютерных вычислительных системах» (Лапшина С. Ю., Сотников А. Н., Логинова В. Е., Юдинцев К. Ю., МСЦ РАН филиал ФГУ ФНЦ НИИСИ РАН, Москва. № 4/2019). Статья посвящена выбору оптимального количества запрашиваемых процессорных ядер для запуска алгоритма многократной маркировки пер-

коляционных кластеров в процессе проведения имитационных экспериментов задачи мультиагентного моделирования процессов распространения массовых эпидемий на современных компьютерных системах.

В статье **«Реализация экспертной системы для оценки инновационности технических решений»** (Иванов В. К., Образцов И. В., Палюх Б. В., ТвГТУ, Тверь. № 4/2019) авторами разработано специальное программное обеспечение на основе интервальных оценок в соответствии с теорией свидетельств, используемое для анализа сложных многокомпонентных систем, агрегации больших объемов нечетких и неполных данных различной структуры.

«Метод и программные средства интеллектуальной поддержки принятия логистических решений» (Борисов В. В., Рязанов А. В., Смоленский филиал НИУ МЭИ. № 4/2019). Авторами предложен и программно-реализован метод интеллектуальной поддержки принятия логистических решений на основе генетической кластеризации, нечеткого оценивания и назначения логистических средств для выполнения заказов на основе модифицированного метода Г. Куна.

В статье **«Алгоритмическое и программное обеспечение когнитивного агента на основе методологии Д. Пойма»** (Курбатов С. С., Фоминых И. Б., Воробьев А. Б., НИЦ ЭВТ. НИУ МЭИ, Москва. № 1/2019) описывается оригинальный подход к созданию интеллектуальной системы решения задач обучения. Когнитивный агент (система) предполагает тесную интеграцию этапов лингвистической обработки, онтологического представления задачи, эвристически-ориентированного решения и концептуальной визуализации. При этом онтология реализована в среде СУБД Progress, а программы визуализации на Java Script с использованием JSXGraph и MathJax.

«Интегрированная среда разработки с поддержкой структурного редактирования на языке программирования Go» (Ванясин Н. Н., Сидоркина И. Г., Поволжский ГТУ, Йошкар-Ола. № 1/2020). В статье представлена программная реализация интегрированной среды разработки

со структурным редактором, позволяющей создавать и редактировать проекты с использованием языка программирования Go. Программная реализация отличается от известных способов представления хранимого состояния исходного кода, а также пользовательским интерфейсом структурного редактора, что позволило повысить производительность среды разработки и эффективность работы программиста.

В статье **«Реализация логического вывода в продукционной экспертной системе с использованием Rete-сети и реляционной базы данных»** (Массель Л. В., Ан Г. В., Пестерев Д. В., Институт систем энергетики им. Л. А. Мелентьева СО РАН, Иркутск. № 2/2020) для ускорения времени реализации логического вывода на продукционной модели Поста, предлагается использовать Rete-сеть. Авторы иллюстрируют применение Rete-сети для логического вывода на продукциях применительно к когнитивной модели одной из угроз энергетической безопасности. В результате сокращается время логического вывода при большом объеме правил в базе знаний по сравнению с алгоритмами наивного вывода.

Редакция международного научно-практического журнала «Программные продукты и системы» в своей работе руководствуется сводом правил Кодекса этики научных публикаций, разработанным и утвержденным Комитетом по этике научных публикаций (Committee on Publication Ethics – COPE), которые изложены на сайте журнала.

Журнал «Программные продукты и системы» регулярно выступает в качестве информационного спонсора при проведении международных и национальных научно-технических мероприятий (конгрессов, конференций, семинаров) в области искусственного интеллекта и информационных технологий, организуемых учреждениями РАН, Российской ассоциацией искусственного интеллекта и вузами РФ.

Авторы сборника

Абрамов Алексей Геннадьевич, к. ф.-м. н., в. н. с., Санкт-Петербургское отделение Межведомственного суперкомпьютерного центра РАН – филиала ФГУ ФНЦ НИИСИ РАН, abramov@runnet.ru

Антопольский Александр Борисович, д. т. н., профессор, гл. н. с., Институт научной информации по общественным наукам РАН, ale5695@yandex.ru

Арутюнов Валерий Вагаршакевич, д. т. н., профессор, Российский государственный гуманитарный университет, warut698@yandex.ru

Атаева Ольга Муратовна, к. т. н., н. с., Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, oli@ultimeta.ru

Белоозеров Виктор Николаевич, к. ф. н., доцент, в. н. с., Всероссийский институт научной и технической информации РАН

Бескаравайная Елена Вячеславовна, с. н. с., Библиотека по естественным наукам РАН, elenabesk@gmail.com

Бодрова Алина Сергеевна, к. филол. н., доцент, н. с., Национальный исследовательский университет «Высшая школа экономики», abodrova@hse.ru

Ветров Александр Николаевич, к.т.н., доцент, профессор кафедры «Информационные системы», Тверской государственный технический университет, vetrov_48@mail.ru

Власова Светлана Александровна, к. т. н., в. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, vlas.svetlana2013@yandex.ru

Второв Иван Петрович, к. г. н., рук. гр., Геологический институт РАН, vip@gmail.com

Галеев Равиль Марветович, м. н. с., Институт этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН, ravil.galeev@gmail.com

Герасименко Алена Юрьевна, м. н. с., Центральная научная библиотека УрО РАН, elbook@cbibl.uran.ru

Глушановский Алексей Валерианович, с. н. с., Библиотека по естественным наукам РАН, avglush@mail.ru

Гуськов Сергей Николаевич, к. ф. н., зам. дир., Институт русской литературы РАН (Пушкинский Дом), sgouskov@gmail.com

Гончар Андрей Андреевич, зам. дир., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, andrey.gonchar@jscs.ru

Дмитриева Елена Юрьевна, к. т. н., зав. отд., Всероссийский институт научной и технической информации РАН, tyro@viniti.ru

Евсеев Антон Владимирович, дир., Санкт-Петербургское отделение Межведомственного суперкомпьютерного центра РАН – филиала ФГУ ФНЦ НИИСИ РАН, evseev@runnet.ru

Ефременко Дмитрий Валерьевич, д. полит. н., зам. дир., Институт научной информации по общественным наукам РАН, efdv2015@mail.ru

Залаев Геннадий Захарович, д. т. н., зам. дир., науч. рук., Российский государственный архив научно-технической информации, gzalaev@mail.rgantd.ru

Иванов Владимир Константинович, к. т. н., доцент, нач. упр., Тверской государственный технический университет, mtivk@tstu.tver.ru

Ильина Ирина Николаевна, к. и. н., с. н. с., рук. лаб., Архив РАН, ilina_irina_ran@mail.ru

Каленов Николай Евгеньевич, д. т. н., профессор, гл. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, nekalenov@mail.ru

Кириллов Сергей Александрович, с. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, skirillov@jscs.ru

Коровникова Наталья Александровна, к. полит. н., с. н. с., Институт научной информации по общественным наукам РАН, natalia.kor@list.ru

Лапшина Светлана Юрьевна, с. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, lapshina@jscs.ru

Леонов Валерий Павлович, д. пед. н., профессор, науч. рук., Библиотека Российской академии наук, valleo@mail.ru

Лейбов Михаил Борисович, к. ф-м. н., с. н. с., Институт этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН, m_leybov@mail.ru

Лопатина Наталья Викторовна, д. пед. н., профессор, зав. каф., Московский государственный институт культуры, dreitser@yandex.ru

Малахова Ирина Геннадьевна, к. г.-м. н., в. н. с., Геологический институт РАН, mig@ginras.ru

Митрошин Иван Андреевич, с. н. с., Библиотека по естественным наукам РАН, imitros@gmail.com

Мохначева Юлия Валерьевна, к. пед. н., в. н. с., зав. отд., Библиотека по естественным наукам РАН, j-v-m@yandex.ru

Никсон Муганда Охара (Nixon Muganda Ochara), Professor, University of Venda, South Africa, nixon.muganda@gmail.com

Палюх Борис Васильевич, д. т. н., профессор, зав. каф., Тверской государственной технической университет, pboris@tstu.tver.ru

Пекшева Мария Андреевна, м. н. с., Центральная научная библиотека УрО РАН, maria.ulasovets@yandex.ru

Погорелко Константин Павлович, к. т. н., в. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, konstpog@yandex.ru

Прокофьева Юлия Дмитриевна, н. с., Центральная научная библиотека УрО РАН, Yulia9011@mail.ru

Савин Геннадий Иванович, д. ф.-м. н., профессор, академик РАН, науч. рук., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, savin@jscs.ru

Семенов Николай Александрович, д. т. н., профессор кафедры «Информационные системы», Тверской государственной технической университет, slt1155@mail.ru

Серебряков Владимир Алексеевич, д. ф.-м. н., профессор, гл. н. с., Вычислительный центр им. А. А. Дородницына ФИЦ «Информатика и управление» РАН, serebrvas@gmail.com

Скворцов Николай Алексеевич, н. с., ФИЦ «Информатика и управление» РАН, nskv@mail.ru

Соболевская Ирина Николаевна, к. ф.-м. н., с. н. с., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, ins@jscs.ru

Соловьева Татьяна Николаевна, с. н. с., Библиотека по естественным наукам РАН, ya.solo-tata@yandex.ru

Сотников Александр Николаевич, д. ф.-м. н., профессор, зам. дир., Межведомственный суперкомпьютерный центр РАН – филиал ФГУ ФНЦ НИИСИ РАН, asotnikov@jscs.ru

Сютнюренко Олег Васильевич, д. т. н., профессор, в. н. с., Всероссийский институт научной и технической информации РАН, olegasu@mail.ru

Сысоев Александр Николаевич, гл. спец., Всероссийский институт научной и технической информации РАН, vims-sysoev@mail.ru

Тучкова Наталия Павловна, к. ф.-м. н., с. н. с., Вычислительный центр им. А. А. Дородницына ФИЦ «Информатика и управление» РАН, natalia_tuchkova@mail.ru

Угринович Евгений Витальевич, ген. дир., Международный центр научной и технической информации, icsti@icsti.int

Федоров Игорь Григорьевич, д.э.н., профессор, Российский экономический университет им. Г. В. Плеханова, Fedorov.IG@rea.ru

Функ Дмитрий Анатольевич, д. и. н., профессор, дир., Институт этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН, director@iea.ras.ru

Харыбина Татьяна Николаевна, с. н. с., Библиотека по естественным наукам РАН, Natsl@vega.protres.ru

Хохлов Никита Викторович, к. б. н., н. с., Институт этнологии и антропологии им. Н. Н. Миклухо-Маклая РАН, ethno@yandex.ru

Цветкова Валентина Алексеевна, д. т. н., профессор, гл. н. с., Библиотека по естественным наукам РАН, vats08@mail.ru

Шабурова Наталья Николаевна, к. пед. н., зав. науч. библ., Институт физики полупроводников СО РАН, shaburova@isp.nsc.ru

Шапкин Александр Владимирович, к. т. н., нач. упр., Всероссийский институт научной и технической информации РАН, ss@viniti.ru

**Единое цифровое пространство
научных знаний:
проблемы и решения**

Сборник научных трудов

16+

Ответственный редактор *А. Иванова*
Верстальщик *С. Лобанова*

Издательство «Директмедиа Паблишинг»
117342, Москва, ул. Обручева, 34/63, стр. 1
Тел/факс + 7 (495) 334-72-11
E-mail: manager@directmedia.ru
www.biblioclub.ru
www.directmedia.ru



Издайте свою книгу у нас!

Издательство «Директ-Медиа» публикует учебники, монографии, литературу NON-FICTION, аудиокниги, новые издания и те, что с годами не утратили своей актуальности, коллективные научные сборники.

Наше издательство берет свои корни в книгоиздательских традициях и технологиях Германии. Мы – лидеры современного книгоиздательского процесса, охватывающего цифровые образовательные платформы для школ и вузов, издание электронных и печатных книг. Нашу продукцию отличает высокое полиграфическое качество и высокотехнологичный процесс продвижения книги.

Наши авторы – ведущие ученые и преподаватели страны. За 20 лет работы в России нами издано более 10 000 изданий учебной, академической и научно-популярной литературы.

Приобрести наши книги можно в интернет-магазине DIRECTMEDIA.RU и в ЭБС «Университетская библиотека онлайн» (BIBLIOTHECA.RU), в книжных и в интернет-магазинах страны.

***Хотите приобрести книгу издательства «Директ-Медиа» или
издать свое произведение?***

Мы ждем Вас!

www.directmedia.ru

Email: manager@directmedia.ru

Tel.: 8-800-333-6845 (звонок бесплатный)

