



О разработке Единого российского электронного пространства знаний

А. Б. Антопольский

ИНИОН РАН, г. Москва, Россия

Резюме: В докладе предлагается понятие Единого российского электронного пространства знаний приравнять к понятиям инфраструктуры научно-технической информации, информационного пространства науки или научной инфосферы. Констатируется, что современное информационное пространство науки переходит в цифровую форму, становится открытым, информационные процессы дезинтегрируются, меняется роль большинства участников научной коммуникации, в том числе научных библиотек. Перечисляются некоторые ведущие ресурсы и основные институты сферы научно-технической информации. Делается вывод, что существующая нормативная база государственной системы научно-технической информации устарела и не соответствует реальности, а новые проекты, появляющиеся в программных правительственных документах, не предполагают решения накопившихся проблем. Формулируются основные принципы и задачи формирования Единого российского электронного пространства знаний, предлагается план формирования, а также описывается проект, предусматривающий моделирование некоторых процессов его создания.

Ключевые слова: пространство знаний; информационные ресурсы; представление знаний; социальные и гуманитарные ресурсы; интеграция; навигация; проекты; ИНИОН РАН, Россия

Благодарности: Данное исследование поддержано комплексным грантом РФФИ № 18-00-00376 (К) «Исследование и разработка принципов, методов и средств конвергенции естественнонаучных и социогуманитарных ресурсов как составляющих единого электронного пространства знаний» и входящим в его состав грантом РФФИ № 18-00-00218 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН»

Для цитирования: 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. DOI: 10.24069/konf-23-26-04-2019.02.

On the development of a Single Russian electronic space of knowledge

A. B. Antopolsky

INION RAS, Moscow, Russia

Abstract: The report proposes to equate the concept of a Single Russian electronic space of knowledge with the concepts of infrastructure of scientific and technical information, information space of science or scientific InfoSphere. It is stated that the modern information space of science goes into digital form, becomes open, information processes are disintegrated, the role of the majority of participants of scientific communication, including scientific libraries, is changed. Some of the leading resources and major institutions in the field of scientific and technical information are listed. It is concluded that the existing regulatory framework of the state system of scientific and technical information is outdated and does not correspond to reality, and new projects appearing in government policy documents do not involve solving the accumulated problems. The basic principles and tasks of formation of the Unified

Russian electronic space of knowledge are formulated, the plan of formation is offered, and also the project providing modeling of some processes of its creation is described.

Keywords: knowledge space; information resources; knowledge representation; social and human sciences; integration; navigation; projects; INION RAS; Russia

Acknowledgment: This study was supported by RFBR grant № 18-00-002\18 “Integration of scientific and information resources of RAS institutions (on the example of linguistics) as part of a single digital space of RAS» This study was supported by RFBR grant № 18-00-002\18 “Integration of scientific and information resources of RAS institutions (on the example of linguistics) as part of a single digital space of RAS”.

For citation: Antopolsky A. B. On the development of a Single Russian electronic space of knowledge. In: *World-Class Scientific Publication – 2019: Strategy and Tactics of Management and Development: Proc. 8th Int. Sci. & Pract. Conf., Moscow, April 23–26, 2019*. Ekaterinburg: Ural University Press, 2019, pp. 17–29. DOI: 10.24069/konf-23-26-04-2019.02.

Вводные замечания

В ряде нормативных документов последних лет появилось понятие «Единое российское электронное пространство знаний» (ЕРЭПЗ)¹, которое рассматривается как некоторая цель или перспектива создания, развития и интеграции информационных ресурсов и сервисов в области науки, культуры и образования в современной цифровой среде. Недавно появилось Положение о Национальной электронной библиотеке², в котором предлагается следующее определение:

«Пространство знаний – единое российское электронное пространство знаний, представляющее собой совокупность взаимно интегрированных на основе Национальной электронной библиотеки информационных систем и иных информационных ресурсов, сформированных на базе научного, исторического и культурного достояния народов Российской Федерации, образцов зарубежных научных, культурных и исторических ценностей и функционирующих на основе единых информационных технологий и принципов, которые обеспечивают семантическую взаимосвязь их содержимого, а также инструменты поиска и извлечения информации по запросу пользователей»³.

Ранее в некоторых публикациях уже обсуждались отдельные аспекты создания ЕРЭПЗ, в частности, правовые вопросы [1], а также вопросы формирования лингвистического обеспечения [2].

В настоящем докладе делается попытка предложить некоторый план действий по созданию ЕРЭПЗ. Частично моделирование этих действий производится в рамках комплексного гранта РФФИ «Исследование и разработка принципов, методов и средств конвергенции естественнонаучных и социогуманитарных ресурсов как составляющих единого электронного пространства знаний».

¹ Впервые это понятие введено в указе Президента РФ «Об утверждении Основ государственной культурной политики» от 24 декабря 2014 г. № 808 и затем использовано в Федеральном законе «О библиотечном деле» от 29 декабря 1994 г. № 78-ФЗ ст. 18.1 (ред. от 03.07.2016).

² Положение о федеральной государственной информационной системе «Национальная электронная библиотека»: утв. Постановлением Правительства РФ от 20 февраля 2019 г. № 169. Режим доступа: <http://government.ru/docs/35825/>

³ Положение о федеральной государственной информационной системе «Национальная электронная библиотека»: утв. Постановлением Правительства РФ от 20 февраля 2019 г. № 169. Режим доступа: <http://government.ru/docs/35825/>

Существующее положение

Прежде всего, попытаемся определить некоторые понятия. Под ЕРЭПЗ¹ будем понимать совокупность ресурсов, институций и сервисов научно-технической информации (НТИ), или инфосферы научной информации, как она может или должна выглядеть в обозримом будущем. Очевидно, что к такому состоянию инфосфера НТИ, может прийти в результате тех процессов и тенденций, которые протекают в настоящее время в области научных коммуникаций. Многие из этих тенденций рассмотрены в докладе группы экспертов Европейской Комиссии «Будущее научных коммуникации и научных изданий» [3]. Важно в данном контексте подчеркнуть, что автор рассматривает понятие ЕРЭПЗ, с одной стороны шире, а с другой – конкретней, чем это предлагается в цитируемом выше Положении о НЭБ.

Последние годы автор посвятил исследованию инфосферы академического сектора общественных наук России, включая проведение мониторинга и инвентаризации этих ресурсов. Результаты этих исследований изложены в ряде статей [4–6], а также в монографии, подготовленной в соавторстве с Д. В. Ефременко [7].

Главным выводом из этих исследований является определение основных тенденций развития инфосферы. Кратко их можно сформулировать следующим образом:

1. Научные и образовательные коммуникации неуклонно переходят в цифровую форму, хотя этот процесс оказался медленнее ожидаемого и, очевидно, в обозримом будущем сохранятся обе формы коммуникации.

2. Наука и образование переходят к открытым формам коммуникации (открытому доступу), что существенно меняет экономические модели функционирования научных журналов и других научно-информационных продуктов.

3. Ресурсы НТИ, отражающие результаты научной деятельности демонстрируют увеличение разнообразия форм или «библиоразнообразия» [8]: роль традиционных публикаций сокращается, а роль первичных и промежуточных результатов, условно называемых «данными», возрастает.

4. Библиотеки демонстрируют тенденцию к разделению на два типа (культурно-досуговый и научно-информационный), причем второй тип функционально сближается с органами НТИ (ГПНТБ России, БЕН РАН, ВПТБ, ЦНМБ, ЦНСХБ и др.)

5. Генераторами общедоступных ресурсов НТИ стали множество научных и образовательных структур разного уровня, вплоть до отдельных ученых и преподавателей, в результате число таких ресурсов разного типа возросло многократно (десятки тысяч сайтов: электронные библиотеки (ЭБ), базы данных (БД) и другие автоматизированные информационные системы (АИС)).

6. Ресурсы НТИ играют важную роль в оценке эффективности научных и образовательных организаций (инфометрика, вебометрика, альтметрика), однако принципы такой оценки должны изменяться.

Заметим, что появление новых форм информационных коммуникаций и ресурсов практически не привело к исчезновению прежних. Хотя роль некоторых инфор-

¹ Некоторые участники исследования считают более корректным название «Единое цифровое пространство научных знаний».

мационных продуктов, например, печатных библиографических и реферативных изданий, значительно сократилась за последние 20 лет. Таким образом, в настоящее время мы наблюдаем сосуществование традиционных и новых форм информационных ресурсов, которое, очевидно, сохранится в обозримом будущем.

В ходе упомянутого мониторинга этого сектора научной инфосферы был создан Навигатор информационных ресурсов по общественным наукам (НИРОН) [9], содержащий описание свыше 3 тыс. информационных ресурсов и ссылки еще на 13 тыс. ресурсов, принадлежащих академическим учреждениям и аффилированным с ними структурам.

Напомним типологию информационных ресурсов, выявленных в рамках мониторинга:

- сайты учреждений – владельцев;
 - библиотеки;
 - архивы;
 - музеи;
 - электронные коллекции и ЭБ;
 - информационные системы (кроме ЭБ);
 - справочные, энциклопедические ресурсы;
 - персональные (просопографические) ресурсы;
 - периодические, продолжающиеся, сериальные издания;
 - библиографические ресурсы;
 - мероприятия (конференции, конгрессы, симпозиумы, выставки и др.);
 - неопубликованные документы;
 - лингвистические ресурсы;
 - медиаресурсы;
- прочие интернет-ресурсы, не отнесенные к вышеперечисленным типам.

Представляется очевидным, что формирование ЕРЭПЗ должно основываться на существующих информационных ресурсах, как традиционных, так и новых, поскольку именно в этих ресурсах зафиксирован опыт многих поколений информационной деятельности по сбору, обработке научной информации и представлению знаний, заключенных в этой информации.

Соответственно, способ и форма использования существующих ресурсов должны различаться в зависимости от типа ресурсов. Автор считает, что предлагаемая типология поможет решить эту проблему.

Конечно, при этом необходимо учитывать и зарубежный опыт, по возможности включаясь в международное разделение труда при создании информационного пространства знаний. Например, применительно к социальным и гуманитарным наукам следует учитывать идеи, методы и технологии, возникающие в рамках цифровой гуманитаристики как определяющего направления при формировании и развитии инфосферы этих наук [10].

В последние годы ведется целый ряд проектов, которые призваны в той или иной степени агрегировать (интегрировать) научную и образовательную информацию в электронной форме. Перечислим некоторые из них:

1. Национальная электронная библиотека <https://нэб.рф>
2. Национальный энциклопедический портал на основе Большой российской энциклопедии <http://www.encyclopedia.ru/cat/books/book/73437/3>
3. Национальный портал открытого образования <http://npoed.ru/about>
4. Электронная библиотека диссертаций <http://diss.rsl.ru>
5. Интегратор научной периодики: Научная электронная библиотека и Российский индекс научного цитирования <https://elibrary.ru>
6. Интеграторы библиографической информации: Либнет <http://www.nilc.ru>, ЭКБСОН <http://www.vlibrary.ru>, АРБИКОН <https://arbicon.ru>
7. Интеграторы материалов научных конференций <http://konferencii.ru>
8. Интегратор научных ресурсов открытого доступа «Киберленинка» <https://cyberleninka.ru>
9. Репозитории научно-образовательной информации: НОРА <https://openrepository.ru>, СОЦИОНЕТ <https://socionet.ru>, Федеральная резервная система банков знания <https://noosphere.ru>
10. Государственный каталог Музейного фонда РФ <https://goskatalog.ru/portal>
11. Центральный фондовый каталог Росархива <http://cfc.rusarchives.ru/CFC-search>
12. Электронная библиотека «Научное наследие России» <http://e-heritage.ru/index.html>

Одновременно создается множество других ресурсов НТИ, в том числе полнотекстовых ЭБ, энциклопедических, словарных, аналитических, фактографических БД, тематических, проблемных, видовых АИС. Каталог «Наука в Рунете» (<https://elementy.ru/catalog>) содержит около 10 тыс. ссылок, Каталог Рамблера «Наука» (<https://top100.rambler.ru/navi/?theme=1113>) – 2,3 тыс. ссылок, Навигатор информационных ресурсов по общественным наукам (<http://niron.inion.ru>) – свыше 3 тыс. ссылок. Число российских периодических изданий, индексируемых в РИНЦ (<https://elibrary.ru>) – 5,3 тыс., а общее количество российских журналов свыше 14 тыс.

По оценке автора, общее число электронных информационных ресурсов по науке и образованию, функционирующих в России, – не менее 50 тыс. Среди них много информационных систем федерального уровня и значения, которые поддерживаются такими авторитетными организациями как МГУ, ВИНТИ, ИНИОН, РГБ, ГПНТБ России, БЕН РАН, ФИПС, ЦНМБ, ЦНСХБ, Президентская библиотека им. Б. Н. Ельцина, или таких популярных ресурсов как «Википедия», «Интегрум», «Киберленинка».

Если оценивать нормативную базу национальной системы НТИ, то следует констатировать, что Постановление Правительства РФ «Об утверждении Положения о государственной системе научно-технической информации» от 24 июля 1997 г. № 950 в настоящее время, безусловно, устарело, хотя формально действует. Более подробная позиция автора по этому вопросу изложена в монографии [7]. В целом можно сказать, что состав участников и актуальных ресурсов, цели, задачи, форма управления Государственной системой научно-технической информации (ГСНТИ) не отвечают современным требованиям.

Нужно отметить, что последнее время в рамках вновь принятых нормативных и концептуальных правительственных документов появилось несколько новых проектов, которые можно рассматривать как мероприятия, направленные на

организацию национальной системы НТИ. Так, в Национальном проекте «Наука» и в Федеральном проекте «Развитие передовой инфраструктуры для проведения исследований и разработок в Российской Федерации» п. 1.4 сформулирован так «Цифровая система управления сервисами научной инфраструктуры коллективного пользования, включая цифровые коллекции и банки данных». К этой сфере относятся также нормативные акты, регламентирующие создание Национального энциклопедического портала, Государственного каталога Музейного фонда и других проектов

Исходные принципы построения ЕРЭПЗ

Сформулируем некоторые принципы создания ЕРЭПЗ, которые считаем принципиальными. Конечно, они отражают личное мнение автора и не претендуют на завершенность и истинность:

- ЕРЭПЗ должно формироваться из существующих в России ресурсов на основе их конвергенции, не нарушая их функционирования;
- для формирования ядра ЕРЭПЗ должны быть отобраны ресурсы по критериям качества;
- ЕРЭПЗ должно строиться на принципах открытой науки;
- онтология научного знания должна наследовать существующие информационно-поисковые языки;
- информационные объекты, включаемые в Базу знаний ЕРЭПЗ, должны отвечать критериям достоверности, проверяемости, авторитетности и научности;
- технологии формирования ядра ЕРЭПЗ должны обеспечивать возможность коллаборации участников;
- ЕРЭПЗ должно выполнять функции ГСНТИ в части координации деятельности по сбору, обработке, предоставлению в общественный доступ и обеспечению сохранности НТИ;
- ЕРЭПЗ должно использоваться для оценки результатов научно-технической деятельности, но с учетом модернизации систем оценки.

Цели создания ЕРЭПЗ

Целью работ по формированию ЕРЭПЗ является создание современной системы НТИ, ориентированной на решение приоритетных задач социально-экономического и научно-технологического развития России, обеспечение свободного доступа российских пользователей к актуальной, полной, достоверной научной и образовательной информации, создаваемой в России, преимущественно в электронной (цифровой) форме.

Основные задачи ЕРЭПЗ:

- научно-информационное обеспечение достижения целей национального развития России;
- информационное обеспечение стратегического планирования и управления научно-технологическим и инновационным развитием России, разработки и реализа-

ции научно-технологической и инновационной политики, приоритетных программ и проектов социально-экономического и научно-технологического развития России;

- координация и оптимизация научно-информационной деятельности в России, в том числе деятельности по созданию научных и образовательных электронных ресурсов;

- международное сотрудничество в области информационного обеспечения науки и образования;

- популяризация и пропаганда науки, научных знаний и научно-технических достижений;

- наукометрические исследования научно-образовательных институций России, а также выявление новых и перспективных научных направлений.

Перечисленные выше основные задачи ЕРЭПЗ можно назвать *внешними*.

Кроме того, при создании ЕРЭПЗ необходимо обеспечить решение задач, направленных на оптимизацию самой научной инфосферы, которые можно назвать *внутренними*. К последним можно отнести следующие задачи:

- мониторинг научных и образовательных электронных ресурсов (НОЭР), создаваемых в России;

- экспертизу и отбор наиболее качественных НОЭР как составной части ЕРЭПЗ. К этому направлению можно отнести и активную деятельность по оптимизации системы научных журналов;

- согласованное развитие качественных и востребованных НОЭР;

- организацию доступа к НОЭР, навигацию и поиск по ним из единого окна;

- создание новых информационных продуктов и сервисов, обеспечивающих предоставление актуальной, достоверной информации в современной и удобной форме;

- разработка логико-лингвистических и программных средств, обеспечивающих обработку и преобразование НОЭР, извлечение знаний, проверку научной информации на новизну и других задач, диктуемых развитием науки;

- разработка согласованной программы оцифровки традиционных НОЭР;

- архивирование и сохранность НОЭР;

- обмен с международными системами представления знаний;

- стандартизация информационных процессов и продуктов;

- разработка и реализация экономической модели функционирования научно-информационных сервисов;

- совершенствование правового обеспечения ЕРЭПЗ.

Принципы и последовательность формирования ЕРЭПЗ

Изложенное выше современное состояние научной инфосферы диктует стратегию ее развития на основе конвергенции существующего хаотического множества информационных систем и ресурсов в упорядоченную совокупность знаний и данных. Реализацию этой стратегии можно представить себе в несколько этапов.

Первый этап – мониторинг инфосферы – включает проведение инвентаризации научно-образовательной инфосферы, организацию постоянного мониторинга и учета научно-образовательных ресурсов и сервисов, а также научных институций – генера-

торов и операторов этих ресурсов и сервисов. Результатом первого этапа должен стать общегосударственный каталог НОЭР. Этот каталог может быть единым, но вероятнее, будут реализованы несколько каталогов по основным научным направлениям:

- естественные и точные науки;
- социальные и гуманитарные науки,
- технические науки;
- медицинские науки;
- сельскохозяйственные науки.

Прототипом такого ресурса может послужить каталог НИРОН, созданный в ИНИОН РАН.

Конечно, мониторинг НОЭР и формирование каталогов может осуществляться и на более детальном уровне – по отдельным научным дисциплинам или по отдельным категориям НОЭР. Например, в настоящее время разрабатывается проект создания специализированного Центра лингвистических ресурсов на базе Института русского языка им. В. В. Виноградова РАН. Очевидно, что первичной задачей этого центра должен быть мониторинг лингвистических ресурсов России, их учет и каталогизация.

Также возможно организовать учет и мониторинг по типам НОЭР, или по региональным отделениям научных учреждений. Например, учет и мониторинг университетских репозиторий осуществляет НЭИКОН в рамках проекта НОРА. Важно, чтобы мониторинг научно-информационных продуктов осуществлялся согласованно.

Мониторинг инфосферы и формирование каталога должны производиться регулярно, например, ежегодно.

Очевидно, что ресурсов, которые их создатели относят к категории НОЭР, слишком много, их качество далеко не всегда соответствует современным стандартам и пользовательским требованиям. Поэтому необходим **второй этап – оценка и отбор наиболее качественных НОЭР.**

Критериями отбора НОЭР должны стать различные показатели – в том числе:

- пользовательские (открытость, видимость, цитируемость, скачиваемость);
- содержательные (уникальность данных, полнота, актуальность, достоверность);
- правовые (легитимность, правовой статус, режим доступа);
- информационно-технологические (например, качество поиска, надежность ПО).

Важным критерием для отбора должны стать целесообразность включения НОЭР в процессы интеграции и агрегации, а также их использования для формирования ядра системы НТИ, иначе называемого ядром ЕРЭПЗ (или ЕЦПНЗ – Единого цифрового пространства научных знаний). Назовем отобранные НОЭР основными ресурсами ЕРЭПЗ.

В состав НОЭР цифровой инфраструктуры научно-технической информации могут быть включены ресурсы различных типов, включающих как первичную, так и обработанную научную и образовательную информацию, в том числе:

- электронные библиотеки;
- архивы периодических изданий;
- репозитории, другие коллекции неопубликованных документов;

- оцифрованные архивные и музейные коллекции, фонды, выставки и пр.
- фактографические БД, реестры, регистры, каталоги, перечни;
- энциклопедии, справочники, указатели;
- учебная и методическая литература, онлайн-курсы;
- первичные данные научных исследований (журналы наблюдений и др.);
- географические информационные системы;
- экспертные системы, базы знаний;
- банки изображений и аудиоинформации;
- интерактивные продукты, 3d-модели, продукты виртуальной и дополненной реальности.

Возможно, конечно, создание и других научных информационных продуктов в электронной форме, поскольку развитие так называемых E-science происходит весьма динамично.

В настоящее время количество российских НОЭР, претендующих на научную, образовательную и просветительскую функцию, весьма велико, по экспертной оценке составляет 50–100 тыс., поэтому процедура их отбора для включения в ЦИНТИ представляется необходимой.

Отбор НОЭР может предполагать также их агрегацию. Например, многочисленные электронные библиографические указатели, каталоги и БД, создаваемые в научных и образовательных учреждениях, могут входить в состав основных ресурсов НТИ через библиографические агрегаторы, такие как ЛИБНЕТ, ЭКБСОН, АРБИКОН.

Вероятно, в число основных НОЭР без отбора должны быть включены ресурсы, имеющие официальный статус государственных информационных систем, такие как Национальная электронная библиотека, Автоматизированная информационная система Федерального института промышленной собственности (АИС ФИПС) или Государственный каталог Музейного фонда.

Вероятно, целесообразным является создание специализированных научно-информационных центров.

Дискуссионным является включение в основные НОЭР российских ресурсов на иностранных языках (например, переводные журналы), НОЭР, создаваемые совместно с иностранными участниками, русскоязычные ресурсы, владельцами которых являются нерезиденты, и другие спорные категории ресурсов.

В любом случае методика отбора НОЭР и организация экспертизы НОЭР должны стать результатом широкого обсуждения научно-информационным сообществом и быть зафиксированы в нормативно-техническом документе.

Основные НОЭР, прошедшие экспертизу и отбор, должны получить официальный статус и соответствующую поддержку, включая государственное архивирование. Как уже было отмечено, основные НОЭР должны стать источником ядра ЕРЭПЗ (ЕЦПНЗ). Вероятно, в рамках экономической модели ЕРЭПЗ основные НОЭР должны получить определенные преференции.

Третий этап – формирование ядра ЕРЭПЗ (ЕЦПНЗ) должно стать наиболее сложным этапом создания национальной системы научной информации. Это ядро включает две части – Онтологию научных знаний и Базу знаний.

Онтология научного знания должна унаследовать терминологию, номенклатуру, парадигматику и концептуальные схемы информационных языков и систем метаданных всех основных НОЭР ЕРЭПЗ с тем, чтобы обеспечить поиск и навигацию в этих ресурсах из единого окна. Основное условие – формирование онтологии не должно приводить к переработке основных НОЭР. Поиск по основным НОЭР следует обеспечивать с помощью поисковых средств, которые были заложены в этих НОЭР.

Естественно, кроме специфических видов поиска, должен быть предложен и обычный сквозной лексический поиск для тех НОЭР, которые формируют поисковый лексический индекс.

Онтология должна обеспечивать поиск и навигацию в базе знаний ЕРЭПЗ.

База знаний ЕРЭПЗ представляет собой совокупность энциклопедических и справочных сведений, представленных как в виде формализма, пригодного для компьютерной обработки, так и в человекочитаемом виде. В качестве прототипа базы знаний ЕРЭПЗ можно указать на DBpedia.

Процедуру формирования исходной версии Базы знаний можно представить себе следующим образом:

- формирование перечня наиболее авторитетных и популярных энциклопедий и справочников по различным наукам;
- загрузку этих энциклопедий в единую среду и их интеграция;
- выявление противоречий, ошибок;
- содержательное редактирование выявленных ошибок;
- индексирование информационных объектов Базы знаний при помощи онтологии;
- разработка технологии пополнения Базы знаний.

База знаний может быть как централизованной, так и разделенной на области наук. В последнем случае должна быть сформирована междисциплинарная (общенаучная) база знаний.

Формирование Базы знаний должно осуществляться на принципах открытости, актуальности, достоверности, авторизованности, тематического универсума НТИ.

При формировании Базы знаний могут различаться просветительский, базовый (образовательный) и профессиональный уровни. Навигация в Базе знаний на основе онтологии также должна различаться по этим уровням.

В Базе знаний должна быть возможность отражения различных (в том числе противоречивых) научных теорий и концепций. Однако должен быть предложен механизм предотвращения включения в Базу знаний лженаучной и паранаучной информации. Очевидно, решающая роль в этом должна принадлежать РАН, как основной экспертной организации страны.

Информация, содержащая государственную или коммерческую тайну, в Базу знаний не включается.

На основе Базы знаний и онтологии могут разрабатываться и реализовываться дополнительные (монетизированные) сервисы, сокращающие затраты на функционирование ЕРЭПЗ.

Четвертый этап – разработка системы выявления нового знания на основе потока научной информации, генерируемого российскими научными организациями.

Эта система должна быть основана на существующих информационно-аналитических и реферативных службах. Специалисты-аналитики должны иметь в своем распоряжении сервис, позволяющий определить наличие новизны в документе или ином научно-информационном продукте по сравнению с Базой знаний. Если такая новизна обнаружена, новые информационные объекты (энциклопедические статьи и справочные данные) вводятся в Базу знаний. При появлении новых понятий, терминов, имен или связей между ними они также вводятся в онтологию. Этот сервис может опираться на алгоритмы существующих систем выявления плагиата, хотя должен иметь более высокий уровень семантических технологий.

Процедуры выявления нового знания и включение их в ядро ЕРЭПЗ должны непосредственно отражаться в системах оценки эффективности научной деятельности. Эта система должна постепенно заменять существующие способы оценки на основе публикационной эффективности, которые демонстрируют устойчивое снижение адекватности и полезности.

Очевидна сложность и противоречивость предлагаемого подхода. Однако необходимость формирования новой системы НТИ, отражающей реальность современного этапа научно-технического развития, не вызывает сомнений.

Четвертый этап реально может быть реализован только после появления приемлемого прототипа Базы знаний. До тех пор существующая система аналитических и реферативных служб должна сосуществовать с процедурами формированием Базы знаний.

Моделирование создания ЕРЭПЗ

Как уже упоминалось, моделирование некоторых аспектов создания ЕРЭПЗ предусмотрено в рамках Комплексного гранта РФФИ «Исследование и разработка принципов, методов и средств конвергенции естественнонаучных и социогуманитарных ресурсов как составляющих единого электронного пространства знаний». В этом гранте участвуют четыре организации РАН (ФИЦ ИУ, НИИСИ, БЕН и ИНИОН), каждая из которых реализует свой собственный проект.

В данном докладе описывается исследование, которое выполняется ИНИОН РАН совместно с Институтом русского языка (ИРЯ) РАН в рамках гранта № 18-00-002\18 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН». В рамках этого проекта предусмотрено выполнение следующих работ:

- разработка стратегии для действий создателей информационных ресурсов РАН в области языкознания по интеграции информационных ресурсов в единое цифровое информационное пространство РАН;
- разработка методики оценки научных ресурсов РАН на примере ресурсов по языкознанию;
- проведение инвентаризации, классификации и каталогизации информационных ресурсов РАН по языкознанию;
- участие в комплексном проекте по разработке онтологии научного знания на основе имеющихся в ИНИОН РАН и ИРЯ РАН информационных ресурсов по языкознанию, тезаурусов, рубрикаторов и систем метаданных.

Стратегия интеграции научно-информационных ресурсов должна строиться на основе анализа современного состояния инфосферы по гуманитарным и социальным наукам и прогноза ее развития на обозримый период. Стратегия должна включать комплекс правовых, экономических, технологических и организационных мер, обеспечивающих в перспективе интеграцию научно-информационных ресурсов учреждений РАН в цифровое информационное пространство знаний. В стратегии должен быть учтен опыт предыдущих разработок – Единого научного информационного пространства РАН и Интегрированной системы информационных ресурсов РАН (ЕНИП РАН и ИСИР РАН). Однако в условиях современного состояния инфосферы РАН дорожная карта разрабатывается впервые.

Методика оценки качества научно-информационных ресурсов должна строиться на основе сочетания экспертных и инфометрических (альтметрических, вебометрических и др.) показателей, обеспечивающих комплексную оценку научных информационных ресурсов. Методика оценки качества научно-информационных ресурсов должна учитывать наиболее перспективные их типы (ЭБ, БД, интерактивные и коллаборативные сервисы и др.). Методика должна быть ориентирована на использование в различных процессах интеграции, обеспечения сохранности информационных ресурсов и управления инфосферой. Методика оценки качества научно-информационных ресурсов разрабатывается впервые.

Мониторинг научно-информационных ресурсов учреждений РАН должен предусматривать различные методы сбора данных, обеспечивающих полноту и достоверность результатов инвентаризации, и завершаться формированием электронного каталога информационных ресурсов с фиксацией их организационной принадлежности, тематики и структурного типа. Мониторинг научно-информационных ресурсов учреждений РАН проводится впервые.

Участие в комплексном проекте по интеграции ресурсов должно предусматривать формирование фрагментов различных категорий информационных ресурсов по лингвистической тематике, имеющих в ИНИОН и ИРЯ РАН в соответствии с требованиями единой онтологии научного знания. Форматы и состав метаданных фрагментов различных категорий информационных ресурсов разрабатываются совместно с другими участниками комплексного проекта. Также совместно осуществляются формирование онтологии на основе представленных лингвистических средств и систем метаданных и загрузка представленных образцов в единую программно-технологическую среду. Комплексный проект интеграции учитывает опыт предыдущих разработок и обладает существенной новизной с учетом состава информационных ресурсов и новых технологий формирования онтологии и единого информационного пространства РАН.

Список литературы

1. Вершинин А. П. Единое российское электронное пространство знаний: вопросы права. *Университетская книга*. 2016;декабрь. Режим доступа: <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossijskoe-elektronnoe-prostranstvo-znaniy-voprosy-prava.html>
2. Антопольский А. Б., Белоозеров В. Н., Маркарова Т. С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей. *Информационные ресурсы России*. 2017;(5)2–7.

3. *Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission*. Available at: <https://publications.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1>

4. Антопольский А. Б. Инфосфера общественных наук: структура, границы, функции. *Научно-техническая информация*. Серия 1: Организация и методика информационной работы. 2017;(4):14-20.

5. Антопольский А. Б. Информационные ресурсы общественных наук. Опыт организации мониторинга. *Библиосфера*. 2017;(3):78–84.

6. Антопольский А. Б. Определение границ при проведении мониторинга информационных ресурсов социально-гуманитарных наук. *Информационные ресурсы России*. 2017;(3):6–10.

7. Антопольский А. Б., Ефременко Д. В. *Инфосфера общественных наук России*. М.; Берлин: Директ-Медиа; 2017. DOI: [10.23681/468227](https://doi.org/10.23681/468227)

8. *Призыв Жюсье к открытой науке и библиоразнообразию*. Available at: <https://jussieucall.org/jussieu-call/>

9. *Навигатор информационных ресурсов по общественным наукам (НИРОН)*. Available at: <http://niron.inion.ru>

10. Schreibman S., Siemens R., Unsworth J. (eds) *A New Companion to Digital Humanities*. Wiley-Blackwell; 2016.

Информация об авторе

Антопольский Александр Борисович – главный научный сотрудник, ИНИОН РАН, г. Москва, Россия; e-mail: ale5695@yandex.ru.

Information about the author

Aleksander B. Antopolsky – Chief Researcher, INION RAS, Moscow, Russia; e-mail: ale5695@yandex.ru.