

РОССИЙСКАЯ АКАДЕМИЯ НАУК

**ИНСТИТУТ НАУЧНОЙ ИНФОРМАЦИИ
ПО ОБЩЕСТВЕННЫМ НАУКАМ**

**ТЕОРИЯ
И ПРАКТИКА
ОБЩЕСТВЕННО-
НАУЧНОЙ
ИНФОРМАЦИИ**

ВЫПУСК 22

**МОСКВА
2014**

Журнал «Актуальные проблемы Европы»

Коллектив авторов

**Теория и практика общественно-
научной информации. Выпуск 22**

«Агентство научных изданий»

2014

УДК 002.63 (470+571)
ББК 78.647 (2Рос)

Коллектив авторов

Теория и практика общественно-научной информации. Выпуск
22 / Коллектив авторов — «Агентство научных изданий»,
2014 — (Журнал «Актуальные проблемы Европы»)

ISBN 978-5-248-00763-9

Основное содержание выпуска составляют материалы секции
«Информационно-библиотечное обеспечение науки и образования»
XIII Международной научной конференции «Модернизация России:
Ключевые проблемы и решения» (Москва, ИНИОН РАН, 21 декабря
2012 г.) и заседаний 7-9 совместного семинара ИПИ РАН и ИНИОН РАН
«Методологические проблемы наук об информации» (май- декабрь 2012 г.).
Представлены материалы, освещающие зарубежный опыт информационной
деятельности, и статьи памяти коллег. Для специалистов в области
информатики, библиотековедения и библиографоведения, науковедения и
философии науки.

УДК 002.63 (470+571)
ББК 78.647 (2Рос)

ISBN 978-5-248-00763-9

© Коллектив авторов, 2014
© Агентство научных изданий, 2014

Содержание

От редакции	5
Информационно-библиотечное обеспечение науки и образования	7
Принципы формирования единого информационного пространства науки и образования	7
Интеграция информационных ресурсов Сибирского отделения РАН как шаг к формированию единого научно-образовательного информационного пространства 11	15
Интегральная система ВИНТИ РАН по информационному обеспечению научных исследований в области точных, естественных и технических наук: Концепция создания и перспективы развития	24
Конец ознакомительного фрагмента.	32

Теория и практика общественно-научной информации № 22 / 2014

От редакции

Очередной 22-й выпуск сборника научных трудов ИНИОН РАН «Теория и практика общественно-научной информации» отражает результаты научно-исследовательской деятельности специалистов в области информатики, библиотековедения, библиографоведения, книговедения и ряда других дисциплин.

21 декабря 2012 г. в рамках XIII Международной научной конференции «Модернизация России: Ключевые проблемы и решения» состоялась секция «Информационно-библиотечное обеспечение науки и образования»¹. Ее организаторами стали Информационно-библиотечный совет РАН и ИНИОН РАН. Активную поддержку в подготовке работы секции оказала Библиотека по естественным наукам РАН, с которой ИНИОН связывают партнерские отношения.

Всего в работе секции принял участие 41 человек, в том числе два действительных члена РАН (академики Ю.С. Пивоваров и А.Б. Жижченко), два члена-корреспондента РАН (А.А. Саранин и А.М. Федотов), семь докторов наук (д-р техн. наук Н.Г. Алексеев, д-р техн. наук А.Б. Антопольский, д-р филол. наук Л.И. Госина, д-р техн. наук Б.С. Елепов, д-р техн. наук Н.Е. Каленов, д-р техн. наук И.Ю. Никольская, д-р культурол. О.В. Шлыкова) и 17 кандидатов наук. Присутствовали 10 членов Информационно-библиотечного совета РАН, в том числе шесть членов Бюро Совета.

На секции было предложено обсудить следующий круг вопросов:

- формирование единого информационного пространства для науки и образования;
- роль библиотечных и информационных центров в организации и структурировании научных ресурсов, а также навигации по ним;
- интернет-технологии и электронные ресурсы в информационном обеспечении образования и научных исследований;
- новые сервисы в информационном обеспечении научных исследований;
- научно-исследовательская деятельность библиотек.

Представленные для обсуждения материалы соответствовали широте заявленной тематики и вызвали большой интерес. На их основе сформирован первый раздел настоящего издания – **«Информационно-библиотечное обеспечение науки и образования»**.

Эти публикации дают синхронический срез современной академической информационно-библиотечной теории и практики. Лейтмотивом обсуждения стали проблемы разработки и внедрения электронных информационных технологий и предоставления доступа пользователей к электронным ресурсам. Выявились особенности подходов к решению этих проблем в разных организациях Академии наук – ВИНТИ РАН, ИНИОН РАН, Библиотеке РАН (БАН), БЕН РАН и ГПНТБ СО РАН.

¹ См.: Информационно-библиотечное обеспечение науки и образования (2012) [Электронный ресурс]. – Режим доступа: <http://www.inion.ru/conf.Russia-2012.iopo>; обзор работы секции см. в статье директора Научной педагогической библиотеки им. К.Д. Ушинского РАО канд. филол. наук, доц. Т.С. Маркаровой (Маркарова Т.С. Заседание секции «Информационно-библиотечное обеспечение науки и образования» XIII Международной конференции «Модернизация России: ключевые проблемы и решения» (21 дек. 2012 г., ИНИОН РАН) // Науч. и техн. б-ки. – М., 2013. – № 4. – С. 97–103. См. также: http://www.gpntb.ru/ntb/ntb/2013/4/ntb_4_16_2013.pdf).

Во втором разделе, который называется **«Зарубежный опыт информационной деятельности»**, представлены работы канд. филос. наук М.Е. Соколовой и канд. полит. наук П.А. Шарикова. В статье М.Е. Соколовой рассказывается о деятельности Канадского института научной и технической информации (КИНТИ) при Национальном научном совете (ННС) Канады. П.А. Шариков привлекает внимание к уникальной модели взаимодействия государства, бизнеса и общества, способствующей развитию национального информационного ресурса США.

В третьем разделе сборника публикуются **материалы семинара «Методологические проблемы наук об информации»**. Речь идет о седьмом, восьмом и девятом заседаниях совместного семинара Института проблем информатики (ИПИ) РАН и ИНИОН РАН, состоявшихся 21 мая, 11 октября и 13 декабря 2012 г.

Материалы семинара размещаются в электронном виде на сайте ИНИОН РАН². Наряду с этим редколлегия приняла решение представлять их в традиционной печатной форме. Первый опыт такой публикации (вступительная статья руководителей семинара, список участников заседаний 1–6, тексты докладов, а также «Послесловия» к ним) был реализован в 21-м выпуске настоящего издания³.

Завершает выпуск раздел **«In memoriam»**, посвященный памяти коллег, – сотрудникам ИНИОН РАН Б.П. Гинзбургу, О.А. Жирнову, Г.С. Померанцу, В.В. Спивак и Л.А. Рожновой, а также сотруднику ВИНТИ РАН А.И. Черному, одному из создателей и классика отечественной информатики.

² Семинар «Методологические проблемы наук об информации» [Электронный ресурс]. – Режим доступа: <http://www.inion.ru/seminars.mprn>

³ Материалы семинара «Методологические проблемы наук об информации» (г. Москва, 2011–2012 гг.) // Теория и практика обществ.-науч. информации: Сб. науч. тр. / РАН. ИНИОН; Редкол.: Черный Ю.Ю. (гл. ред.) и др. – М., 2013. – Вып. 21. – С. 7–169.

Информационно-библиотечное обеспечение науки и образования

Принципы формирования единого информационного пространства науки и образования

*А.Б. Антопольский*⁴

Обсуждается проблема совершенствования и развития информационного пространства науки и образования. Констатируется его неудовлетворительное состояние. Предлагается метод решения проблемы, основанный на разработке идеальной модели такого пространства. Обсуждается структура информационного пространства, лежащие в его основе социально-экономические модели. Приводятся данные по состоянию пространства, основанные на исследованиях автора информационного пространства по психологии и педагогики. Предлагаются различные меры организационного, экономического и правового характера, способные улучшить ситуацию.

Ключевые слова: информационное пространство науки; социально-экономические модели информационной деятельности; информационное законодательство; государственная информационная политика; эндаументы.

Последние годы характеризуются значительными изменениями характеристик информационного пространства науки и образования. Современное состояние и тенденции изменения этого пространства требуют осмысления. Специалисты по научной информации обязаны предложить властям, а также широкому научному и образовательному сообществу внятную концепцию развития информационного пространства в сложившихся социально-экономических и технологических условиях.

Причинами изменений в указанной сфере стали следующие факторы:

- фактический развал государственной системы научно-технической информации и отсутствие ясно сформулированной государственной информационной политики;
- беспорядочная конкуренция в информационном пространстве государственных, коммерческих и общественных структур;
- переход в цифровую форму большинства информационно-коммуникационных каналов в сфере науки и образования, их интеграция средствами Интернета;
- появление большого количества новых типов электронных информационных ресурсов, сервисов и медиа (электронные библиотеки, социальные сети, прочие средства, которые принято относить к e-learning и e-science);
- отсутствие системности в бюджетных проектах в сфере информационного пространства науки и образования, крайне неудовлетворительный менеджмент этих проектов;
- включение отдельных информационных сервисов в систему электронного государства;
- тенденции изменения системы управления наукой и образованием, в том числе интеграция научных и образовательных учреждений в различных формах;
- усилия правительства по созданию системы оценки результатов научной деятельности на основе специализированных электронных ресурсов, прежде всего индексов цитирования.

⁴ Антопольский Александр Борисович – доктор технических наук, заведующий лабораторией Института научной и педагогической информации Российской академии образования (г. Москва). E-mail: ale5695@yandex.ru

Для дальнейшего рассмотрения следует уточнить, что понимается под информационным пространством. Мы считаем, что информационное пространство образуется совокупностью каналов коммуникации, информационных ресурсов и систем. К числу важнейших из них относятся:

- электронные журналы, книги, учебно-методические пособия, авторефераты и другие документы, циркулирующие по каналам коммуникации;
- научно-образовательные веб-сайты;
- электронные библиотеки и электронно-библиотечные системы;
- библиографические и реферативные БД, электронные каталоги библиотек, архивов и музеев;
- социальные сети и другие ресурсы Web 2.0;
- фактографические БД и другие структурированные научные и образовательные ресурсы;
- ресурсы для e-learning, прежде всего интерактивные обучающие сервисы;
- ресурсы управления наукой, в том числе используемые для оценки эффективности науки и образования;
- традиционные ресурсы печатной продукции и традиционные каналы их доставки (библиотеки, книжная торговля).

В настоящее время традиционные ресурсы научно-образовательной информации преобладают. Однако динамика информационных потоков и информационных потребностей свидетельствует о неуклонном и достаточно быстром переходе различных ресурсов и сервисов в электронную форму. Поэтому любые прогнозы и предложения по развитию информационного пространства должны ориентироваться главным образом на электронные коммуникации.

В то же время нельзя спешить с ликвидацией и сокращением традиционных форм распространения научной информации, прежде всего библиотек. Если сокращение традиционных потоков происходит по 3–5% в год (данные книжной торговли), то сокращение функций традиционных библиотек тоже не должно превышать этой цифры. Причем сокращение должно замещаться новыми функциями, связанными с электронными ресурсами и сервисами. Следует также продумать меры по обеспечению сохранности библиотечных фондов, хотя бы на базе крупнейших научных библиотек.

Важной особенностью современного информационного пространства является его сложная социально-экономическая структура. Здесь достаточно беспорядочно конкурируют акторы (учреждения, люди, группы) существенно различающиеся по моделям функционирования и мотивации (3). Ограничимся перечнем типов информационных ресурсов и сервисов в сфере науки и образования, которые существенно различаются по экономическим моделям:

- коммерческие ресурсы с платным доступом;
- коммерческие ресурсы с бесплатным доступом;
- государственные ресурсы с платным доступом;
- открытые архивы;
- публичные ЭБ на основе фондов библиотек и архивов;
- электронные библиотечные системы вузов;
- общественные ресурсы (сайты и коллекции, создаваемые отдельными лицами и коллективами из альтруистических соображений).

Важной особенностью модели информационного пространства, принципы которого обсуждаются в настоящей статье, должно стать ясное представление о правовом статусе и экономической модели функционирования каждого ресурса и сервиса. К сожалению, такое представление сегодня отсутствует. Рассмотрим несколько примеров.

Одним из важных новых ресурсов последних лет стала Научная электронная библиотека eLibrary.ru, на базе которой функционирует Российский индекс научного цитирования

(РИНЦ). Этот индекс рассматривается в ряде правительственных документов как обязательный инструмент оценки эффективности научного исследования. Автор убежден в важности и полезности национального индекса цитируемости. В то же время статус самого ресурса непонятен. Юридически он принадлежит коммерческой организации и может быть в любой момент закрыт. Его разработка и сдача в эксплуатацию происходили в условиях, приведших к совершенно неудовлетворительному качеству, на что неоднократно указывали специалисты (7).

Другой пример. На основе грантов, выданных российскими научными фондами (РФФИ и РГНФ) за последние годы, было разработано множество (несколько сотен) информационных систем, сервисов и ресурсов, ориентированных на информационное обеспечение науки. Многие из них содержали интересные технологические решения, ценную информационную базу и могли стать очень полезными для научного сообщества. Однако подавляющее большинство из них прекратило существование после завершения гранта. Это произошло потому, что грантовая система финансирования, весьма полезная сама по себе, не приспособлена к поддержке информационных ресурсов. Что же касается других источников финансирования, прежде всего Минобрнауки, то оно тратит большие средства на разработку новых информационных систем. Уже действующие, похоже, министерство не интересуют. Во всяком случае, систем, выживших в этих условиях, насчитывается мало.

Безусловным признанием и заслуженной известностью пользуется созданная в РГБ Электронная библиотека диссертаций. Но каков ее правовой статус? Насколько правомочно фактически коммерческое использование документов, поступивших в РГБ по закону об обязательном экземпляре документов? И как соотносится с этим использованием заявление Д.А. Медведева, что диссертации должны стать общественным достоянием?

Еще раз подчеркнем, насколько важно разработать адекватную модель функционирования общественно значимых ресурсов науки и образования и обеспечить финансовые и правовые условия реализации такой модели.

Рассмотрим некоторые особенности информационного пространства в гуманитарной сфере. Здесь мы будем опираться на исследования информационных ресурсов по педагогике и психологии, проведенные автором по заданию Российской академии образования (1; 2).

Прежде всего, удивляет очень большое количество российских журналов по педагогике и психологии. По подсчетам автора, их более 1180, в том числе в Научной электронной библиотеке и соответственно в РИНЦ представлено 570. При этом в ведущем международном индексе цитирования Web of Science, точнее, в индексе Social Science Citation Index обрабатывается только один (!) российский журнал.

С другой стороны, большая часть публикаций по этим наукам приходится не на журналы, а на монографии, методические пособия и другую серую литературу. Так, публикационная активность академиков и членов-корреспондентов РАО, отражаемая в библиотечном каталоге НПБ им. К.Д. Ушинского, почти в 2 раза больше, чем по журналам, представленным в РИНЦ (4).

Конечно, это говорит о том, что методы оценки эффективности научных исследований, предлагаемые Минобрнауки и РАН и основанные на использовании международных индексов цитируемости, малопригодны для оценок в гуманитарной сфере, в частности в педагогике и психологии.

В связи с этим можно процитировать академика А.Р. Хохлова, который на замечание корреспондента, что численные показатели применимы к естественным наукам, но не к гуманитарным, ответил:

«Разумеется, гуманитарные науки оценивать сложнее. Например, в Arts & Humanities Citation Index не высчитывают импакт-фактор журнала. Поэтому в ИСТИНЕ (6) мы отмечаем любую публикацию наших сотрудников в журналах из этой реферативной базы.

Для обсуждения затронутого вами важного вопроса я бы считал полезным пригласить на “круглый стол” наиболее авторитетных российских гуманитариев с мировым именем, чтобы они разработали критерии для своих наук. Важно, чтобы эти критерии были количественными, проверяемыми и чтобы научная деятельность была видна вне России – даже когда речь идет об изучении русской истории или русской литературы» (8).

К сказанному добавим еще некоторые цифры. В рассматриваемой сфере (педагогика и психология) ежегодно издается до 30 тыс. книг и брошюр, выходит до 20 тыс. неопубликованных документов (отчеты и диссертации), которые также имеются в электронном виде. В наиболее полном каталоге образовательных ресурсов Интернета зарегистрировано около 60 тыс. объектов.

Эти цифры подтверждают известный факт, что объем электронной научной информации сопоставим или даже превосходит объем традиционной печатной информации.

Далее попытаемся перейти от констатации к некоторым конструктивным утверждениям. В качестве базового приема следует представить себе некое идеальное состояние информационного пространства науки и образования, а уже затем обсуждать необходимые правовые, экономические, организационные и технологические условия, позволяющие достичь этого идеального состояния или, по крайней мере, приблизиться к нему.

Для построения этой идеальной модели попытаемся ответить на некоторые принципиальные вопросы:

- какова должна быть роль государства и стоит ли вспоминать ГСНТИ;
- нужна ли единая система информационных ресурсов науки и образования;
- нужна ли система внешнего управления информационного пространства или возможно самоуправление;
- какая должна быть система финансирования, точнее, как в системе финансирования должны сочетаться разные источники и способы финансирования;
- как следует строить систему оценок качества создаваемых и проектируемых ресурсов;
- что нужно изменить в правовой базе?

Ответить на эти вопросы однозначно, конечно, затруднительно. Попытаемся рассмотреть факторы, которые могут повлиять на поиск таких ответов.

Рассмотрим вопрос о государственной роли. С одной стороны, в настоящее время как в науке, так и в образовании государство доминирует, что видно, например, из анализа доли государства в общих затратах на науку и образование. У большинства специалистов не вызывает сомнения, что государственные структуры, например библиотеки и органы научной информации, должны играть важную роль в информационном пространстве.

С другой стороны, в создании научно-образовательных ресурсов Интернета весьма значительна роль негосударственных акторов. Для некоторых типов ресурсов, например электронных библиотек, виртуальных музеев, доля негосударственного сектора достигает 75%. Большинство ресурсов создается не коммерческими организациями, а добровольцами-энтузиастами, иногда организованными в различные сообщества. Качество таких ресурсов с научной точки зрения не всегда бывает удовлетворительным.

Следует также учитывать роль межведомственной конкуренции и слабую способность государственных институций находить общий язык. Среди научно-образовательных информационных ресурсов почти нет таких, которые были бы созданы на основе добровольной кооперации нескольких государственных организаций. Поскольку управление наукой в стране фактически разделено между Минобрнаукой и РАН, соответственно разделяются и научно-информационные институции. Кроме того, собственные ресурсы часто имеют отдельные научные учреждения и вузы, федеральные ведомства и регионы.

Очевидно, что реанимация ГСНТИ в сколько-нибудь масштабном варианте невозможна. Однако некоторые идеи ГСНТИ, например придание системообразующей роли видовым феде-

ральным системам, сохраняют определенную жизнеспособность. Действительно, большинство центральных функционирующих научно-информационных ресурсов носят видовой характер: библиотека диссертаций, ИНИОН, ВИНТИ, НЭБ⁵, ЦИТИС⁶, ФИПС⁷. А вот предпринятая несколько лет назад попытка Минобрнауки организовать информационно-аналитические центры на основе перспективных направлений (по аналогии с ранее существовавшими отраслевыми институтами информации) успеха не имели. Однако есть пример прекрасного тематического ресурса (УИС РОССИЯ⁸). Хотя он создан и функционирует скорее вопреки помощи государства, чем благодаря ей.

В этих условиях говорить о действительно единой системе научно-образовательных ресурсов, действующей по единым правилам и снабженной некоей системой управления, не приходится. Вероятно, можно надеяться на некоторые способы самоорганизации, в том числе самоуправления, но только если для этого возникнут серьезные стимулы.

Лучшим из них, конечно, является финансирование. Интегральные ресурсы реально появляются, если есть соответствующая финансовая база. Хорошим примером здесь могут служить система образовательных порталов, разработанная в 2003–2007 гг. при головной роли Федерального государственного автономного учреждения «Государственный научно-исследовательский институт информационных технологий и коммуникаций» (до 1999 г. – Центр «Информика»). После завершения действия программы «Развитие ЕИОС»⁹, в рамках которой была создана система порталов, большая часть порталов прекратила свое существование, да и головной портал «Российское образование» поддерживается гораздо хуже, чем раньше. Это еще одно доказательство того факта, что научно-образовательные информационные ресурсы могут жить только при постоянной поддержке.

Другой пример подобной интегрированной системы – это «Электронная Земля». Задуманная как мощная интегрированная система для научных геоданных, она погибла, как только закончилась соответствующая целевая программа. Правда, здесь можно указать на еще одну причину неудачи: возглавивший проект ВИНТИ РАН уже давно не обладает кадровыми и ресурсными возможностями для выполнения роли головной организации крупных проектов.

По-видимому, крайне ограниченную перспективу имеет и модель аутсорсинга для научно-информационного пространства. Примером может служить нынешняя ситуация с электронными библиотечными системами (ЭБС). Целый ряд коммерческих структур, которые попытались выйти на этот рынок, не имеют должных информационных ресурсов и поэтому не могут удовлетворить вузов-заказчиков. Большинство из них не переживет нынешнего бума с ЭБС. А крупнейшие владельцы информационных ресурсов – библиотеки – не обладают финансовыми возможностями для разработок ЭБС, а также ресурсами для маркетинга в данной области.

Примерами удачной бизнес-модели в информационном пространстве могут служить правовые информационные системы (Гарант, Консультант+ и др.). Они победили в конкурентной борьбе 1990-х годов и демонстрируют неплохую финансовую устойчивость. Следует, конечно, учитывать, что российский рынок правовой информации развит гораздо больше, чем другие сектора информационного рынка.

Вероятно, реальным инструментом для организации финансирования научно-информационного пространства могли бы стать один или несколько эндаументов, законодательство о

⁵ Научная электронная библиотека eLibrary.ru

⁶ Федеральное государственное автономное научное учреждение «Центр информационных технологий и систем органов исполнительной власти».

⁷ Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности».

⁸ Университетская информационная система РОССИЯ.

⁹ ЕИОС – Единая информационная образовательная среда.

которых нацелено на решение именно этих задач (9; 10; 11). Однако кажется, такая возможность даже не рассматривалась руководителями нашей науки и образования.

Принципы организации финансирования по данной модели очевидны: во главе эндаумента стоит наблюдательный совет из специалистов, который решает, какие ресурсы стоит финансировать на временной или постоянной основе, а какие не следует. В качестве основания для принятия решений следует использовать некие количественные характеристики ресурсов, например посещаемость и цитируемость сайтов или страниц.

Правда, есть один существенный нюанс. Финансировать деятельность информационной системы из средств эндаумента на постоянной основе разумно, только если можно установить принятую всеми зависимость между объемами финансирования и полученными реальными результатами. А это означает необходимость принятия некоторых показателей качества для научно-информационных систем. Дело это вовсе не простое. Это показала, с одной стороны, неудача Российской ассоциации электронных библиотек (ЭЛБИ) с разработкой таких показателей, а с другой – показатели, принятые Минобрнауки для электронно-библиотечных систем (ЭБС) вузов. Эти показатели вызвали активную дискуссию, поскольку в их число не вошли ни функциональность ЭБС, ни тематическая полнота, ни качество оцифровки и визуализации. Но при этом в число показателей ЭБС включены совершенно бессмысленные требования, вроде регистрации ЭБС как СМИ.

Возможно, приемлемыми окажутся показатели, основанные на вебметрических методах. Эти методы получили распространение и в России. В связи с этим следует упомянуть вебметрический индекс научных и образовательных организаций России, разработанный при участии автора (5).

Важным направлением выработки принципов конструирования будущего идеального информационного пространства науки и образования должна стать политика в отношении научных журналов. Цели этой политики в общем виде сформулировать несложно: журналов должно быть меньше, они должны быть лучше, удовлетворять требованиям международных систем, распространяться в электронном формате и быть общедоступными. Однако конкретные механизмы достижения этих целей совершенно неочевидны. В частности, вызывает пессимизм нынешнее состояние научного рецензирования.

Основные баталии, связанные с научно-информационным пространством, разворачиваются в последние годы вокруг проблем авторского права в связи с принятием 4-й части Гражданского кодекса (ГК). Действительно драконовские нормы ГК являются основным препятствием для нормального развития научного информационного пространства. Хотя по существу проблемы почти нет. Согласно исследованиям, до 95% ученых согласны на некоммерческое свободное распространение своих произведений, конечно, при соблюдении личных неимущественных прав, в частности защиты от плагиата. Поэтому включение в ГК института свободных лицензий, аналога широко распространенного в мире «Common creative license», позволило бы снять большую часть существующих барьеров. Можно надеяться, что этот институт все-таки будет утвержден, поскольку соответствующие изменения включены в тексты поправок к 4-й части ГК, которые администрация президента (еще Д.А. Медведева) направила весной 2012 г. в Государственную думу.

Главным препятствием здесь является позиция издательского сообщества, противостоящего любым изменениям 4-й части ГК, направленным на более свободное распространение научной и образовательной информации. Эту позицию можно понять, поскольку доходы от книгоиздательства и книжной торговли неуклонно сокращаются. Но сокращать доступ к научной информации гораздо опаснее, поскольку это угрожает будущему страны.

Еще одной возможной альтернативой существующей ныне системе копирайта является идея открытых архивов, или систем открытого доступа, активно развивающихся в стране и в мире¹⁰.

В любом случае, очевидно, что для научной информации открытые общедоступные каналы коммуникации обладают заметным преимуществом перед коммерческими механизмами. В то же время российскому научно-образовательному обществу предстоит разработать механизм компенсации затрат на поддержку таких каналов. Одним из возможных организационных решений могло бы стать включение соответствующих сервисов в число государственных информационных услуг в рамках электронного правительства. Чем более открытым является канал коммуникации, тем больше возможностей его использовать для оценки результативности научной деятельности, и именно эта задача стоит сегодня перед нами.

В число сервисов, которые необходимо поддерживать на государственном уровне и которые должны использоваться для контроля качества научных результатов, а также научных коммуникаций и научных институций в целом, входит система антиплагиата. Ее необходимость стала очевидной и научному сообществу, и руководству страны. Дело осталось за внедрением. Но наличие таких мощных ресурсов, как Электронная библиотека диссертаций и Научная электронная библиотека, позволяет надеяться на положительный результат.

Самостоятельная задача, которую нужно решить при моделировании информационного пространства науки и образования, – это создание системы долговременного хранения информационных ресурсов. Существующая государственная система архивов как по идеологическим, так и по организационно-технологическим причинам непригодна для решения этой задачи. В то же время очевидна необходимость решения этой задачи, причем принципы сохранности должны быть приемлемы для научного сообщества и реализуемы практически.

Таким образом, научно-информационному сообществу следует разработать и согласовать следующие компоненты идеального информационного пространства:

- минимальную номенклатуру общедоступных информационных ресурсов, необходимых для научно-образовательного сообщества и претендующих на государственную постоянную поддержку;
- полную номенклатуру полезных ресурсов, рассчитанных на различные бизнес-модели, в том числе грантовые или коммерческие;
- критерии качества информационных ресурсов, на которых должна быть основана система финансирования;
- механизмы формирования и обновления корпуса экспертов, обеспечивающих контроль качества информационно-образовательных ресурсов и сервисов;
- механизмы сохранности научно-образовательных информационных ресурсов;
- изменения в правовой базе, кроме авторско-правовых, в частности правовой статус разных компонентов информационного пространства.

Возможно, для построения модели информационного пространства потребуются и другие инструменты, например, стандартизация, мониторинг, учет и статистика, типовые решения (особенно по онтологиям, классификации, метаданным). Для нынешнего состояния информационного пространства эти механизмы непригодны, но нельзя исключать их полезности в будущем.

¹⁰ См., например, цикл статей А.И. Земскова и Я.Л. Шрайберга, посвященных системам открытого доступа, в журнале «Научно-технические библиотеки» за 2008–2009 гг.

Литература

1. Антопольский А.Б. О концепции системы научной информации в сфере педагогики и психологии // Информационное обеспечение науки: новые технологии: сб. науч. тр. / Под ред. Н.Е. Каленова. – М.: Научный мир, 2009. – С. 42–54.
2. Антопольский А.Б. О моделях доступа к российской педагогической периодике // Библиотека и чтение в структуре современного образования: Материалы Межрегиональной научной конференции (Москва, 29 октября 2009 г.). – М.: Наука, 2009. – С. 10–23.
3. Антопольский А.Б. Правовые, социальные и экономические модели информационного общества в сфере культуры // Информационное право и становление основ гражданского общества в России: Материалы теоретического семинара по информационному праву 2007 г. – М.: ИГП РАН, 2008. – С. 62–90.
4. Антопольский А.Б., Поляк Ю.Е. Об исследованиях публикационной активности ученых (на примере членов Российской академии образования) // Информационные ресурсы России. – М., 2011. – № 1. – С. 4–16.
5. Антопольский А.Б., Поляк Ю.Е., Усанов В.Е. О российском индексе веб-сайтов научно-образовательных учреждений // Информационные ресурсы России. – М., 2012. – № 4. – С. 2–7.
6. Истина – Интеллектуальная Система Тематического Исследования Научно-технической информации, созданная в МГУ им М.В. Ломоносова [Электронный ресурс]. – Режим доступа: <http://istina.imec.msu.ru/>
7. Каленов Н.Е., Селюцкая О.В. Некоторые оценки качества Российского индекса научного цитирования на примере журнала «Информационные ресурсы России» // Информационные ресурсы России. – М., 2010. – № 6. – С. 2–13.
8. Огнев А. Академию наук взбудрит прозрачный рейтинг [Электронный ресурс]. – Режим доступа: http://www.strf.ru/material.aspx?CatalogId=221&d_no=51747
9. Федеральный закон № 275-ФЗ «О порядке формирования и использования целевого капитала некоммерческих организаций» от 30 декабря 2006 года.
10. Федеральный закон № 276-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации в связи с принятием Федерального закона № 275-ФЗ «О порядке формирования и использования целевого капитала некоммерческих организаций» от 30 декабря 2006 года.
11. Федеральный закон № 328-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации в части формирования и использования целевого капитала некоммерческих организаций» от 21 ноября 2011 года.

Интеграция информационных ресурсов Сибирского отделения РАН как шаг к формированию единого научно-образовательного информационного пространства ¹¹

Б.С. Елепов, О.Л. Жижимов, А.М. Федотов, Ю.И. Шокин ¹²

*Послушайте, ребята,
Что вам расскажет дед.
Земля наша богата,
Порядка в ней лишь нет.*

А.К. Толстой

Представлен опыт интеграции информационных ресурсов Сибирского отделения РАН для формирования единого научно-образовательного информационного пространства. Описывается платформа массовой интеграции данных ZooSPACE, разрабатываемая в Институте вычислительных технологий СО РАН.

Ключевые слова: информационное пространство науки и образования; информационные ресурсы; информационная система; распределенная информационная система; платформа массовой интеграции данных ZooSPACE.

Одним из основных результатов социальной и интеллектуальной человеческой деятельности является создание и накопление информационных ресурсов с целью их дальнейшего использования и недопущения утраты опыта предыдущих поколений. На протяжении всей истории человечества уровень развития технологий накопления информации и эффективности ее использования значительно влиял на уровень развития производительных сил. Утеря информации приводила к отбрасыванию цивилизации на века назад. Чтобы эффективно пользоваться накопленной ранее информацией, необходимы специальные инструменты и технологии.

Наверное, не будет большим преувеличением утверждение о том, что проблема поиска информации является для человеческого сообщества одной из главных. Великий аргентинский писатель Хорхе Луис Борхес в эссе «Четыре цикла» писал, что в мировой литературе вечны четыре темы.

1. Падение города.
2. Возвращение героя.
3. Поиск.
4. Самопожертвование бога.

Нетрудно заметить, что наиболее часто встречается как в литературе, так и в реальности третья тема – поиск, ибо четвертая тема выходит за рамки обычного человеческого опыта, а две первые проявляются лишь в «минуты мира роковые».

¹¹ Работа выполняется при частичной финансовой поддержке Министерства образования и науки Российской Федерации (грант № «2012–07.514.11.4130», а также при поддержке РФФИ (проекты 12-07-00 472, 13-07-00 859).

¹² *Елепов Борис Степанович* – доктор технических наук, профессор, директор Государственной публичной научно-технической библиотеки СО РАН (г. Новосибирск). E-mail: office@spsl.nsc.ru; *Жижимов Олег Львович* – доктор технических наук, заведующий Лабораторией информационных ресурсов Института вычислительных технологий СО РАН (г. Новосибирск). E-mail: zhizhim@mail.ru; *Федотов Анатолий Михайлович* – член-корреспондент РАН, профессор, заместитель директора Института вычислительных технологий СО РАН (г. Новосибирск). E-mail: fedotov@sbras.ru; *Шокин Юрий Иванович* – академик РАН, профессор, директор Института вычислительных технологий СО РАН (г. Новосибирск). E-mail: dir@ict.nsc.ru

Любой производственный или научный процесс порождает огромные объемы данных, и работать с ними по мере увеличения объемов становится все сложнее. Количество данных когда-нибудь превысит способность человека их обрабатывать, поэтому необходимы новые инструментальные средства и алгоритмы для анализа. Вместе с тем предъявляются серьезные требования к обеспечению прозрачного доступа и долговременной сохранности информации. В результате вопросы «что хранить?», «как хранить?» и «как найти?» остаются самыми существенными: без ответа на них все остальные теряют актуальность (10).

Взаимодействие человека с информацией, существующей в машиночитаемом виде (данными), осуществляется при помощи специализированных программных комплексов – информационных систем (ИС), предназначенных для:

- организации хранения информации (организация хранилищ, поддержка систем хранения данных);
- управления информацией (добавление, модернизация, изменение данных);
- управления доступом к информации (контроль исполнения правил регламентации доступа к данным), идентификация данных;
- поиска информации;
- извлечения информации и предоставления ее пользователю (приложению) в необходимом ему виде;
- визуализации (представления) информации в соответствии с требованиями пользователя (4; 7).

Исходя из тенденций развития систем коммуникаций и практики организации локальных хранилищ данных современная ИС должна удовлетворять некоторым общим требованиям, а именно:

- она должна быть распределенной – РИС¹³;
- РИС должна обеспечивать контролируемый доступ к различным информационным ресурсам для пользователей и администраторов различных уровней;
- РИС должна обеспечивать сквозной поиск, в том числе полнотекстовый, по различным критериям;
- РИС должна предоставлять информацию в требуемом пользователем виде;
- внутренние технологии РИС должны быть скрыты от пользователей;
- внешние интерфейсы должны быть стандартизованы (6).

При попытках разработки РИС, удовлетворяющих перечисленным требованиям, возникают проблемы, связанные с различными аспектами их функционирования. Эти проблемы можно сгруппировать в следующие классы (2; 6):

- *модели и стандарты представления информации и метainформации* .

К сожалению, следование рекомендациям международного сообщества о стандартизации схем данных не является характерной чертой подавляющего большинства разрабатываемых и функционирующих информационных систем¹⁴;

- *автоматическая классификация информации* .

Разработчики ИС не используют модели классификации хранимой информации, затрудняя тем самым включение механизмов, в том числе внешних, для ее автоматической классификации;

- *доступ к распределенным и разнородным коллекциям (интероперабельность, масштабируемость, обнаружение релевантной информации, интеграция метainформации)* .

¹³ РИС – распределенная информационная система.

¹⁴ Сами рекомендации тоже являются весьма противоречивыми. К примеру, можно сравнить концептуальные модели DELOS и FRBR.

Интероперабельность создаваемых информационных систем не выдерживает даже минимальной критики (2);

- *интерфейсы пользователей, визуализация и анализ данных.*

К сожалению, подавляющее количество существующих ИС оперирует только с графическими веб-интерфейсами доступа к данным. При этом возможность доступа к данным вне графических интерфейсов, как правило, не реализуется. Это исключает возможность интеграции разрабатываемых ИС в крупные информационные системы (4);

- *вопросы интеллектуальной собственности.*

Это вечная проблема информационных ресурсов, тиражирование которых не связано с материальными затратами;

- *анализ и обработка естественного языка, изображений, видео- и аудиоданных.*

Обработка контента требует специальных подходов к каждому типу данных. В частности, необходима реализация поиска по фрагментам (фрагмент текста, фрагмент изображения и т.п.) с использованием шаблонов (шаблон текста, шаблон изображения и т.д.). Реализация специфических шаблонов для такого поиска представляет собой отдельную проблему;

- *многоязыковый доступ к данным и обслуживание данных на нескольких языках.*

Доступ такого рода требует многовариантности индексирования данных и реализации динамического преобразования данных;

- *мобильные технологии и агенты.*

Речь идет о реализации интерфейсов доступа к информации, ориентированных на возможности клиентских устройств;

- *методы и средства поиска, каталогизация, индексирование, поддержка целостности и непротиворечивости коллекций, безопасность и защита информации.*

Это вечные проблемы, от способа решения которых зависит общая функциональность ИС.

Решение сформулированных проблем предполагает создание развитой инфраструктуры для представления и обмена метаданными (данными о ресурсах), без которой невозможно формирование единого информационного пространства (9). Это можно рассматривать как первый шаг к интеграции и интероперабельности информационных систем. Под интероперабельностью любой ИС, в том числе и электронной библиотеки, понимается степень ее способности взаимодействовать с другими ИС, в том числе и с человеком. Но если в последнем случае основная нагрузка на обеспечение взаимопонимания ложится на человека, который в состоянии обработать даже очень плохо организованную информацию, то для обеспечения эффективного взаимодействия между автоматическими ИС требуются специальные технологические методы и общие соглашения (5).

В основе интеграции РИС лежит интеграция метаданных, основанная на стандартах формата для представления метаданных, одновременно с унификацией нормативно-справочной информации (профиля информационных систем) (7).

Под интеграцией данных с точки зрения пользователя следует понимать:

- возможность свободно группировать любые имеющиеся разнородные данные по любому признаку в произвольные реальные и / или виртуальные коллекции;
- возможность организовывать по всем массивам данных прозрачный для конечного потребителя сквозной поиск информации.

Реализация механизмов интеграции данных немыслима без их стандартизации: данные одного типа должны описываться и предоставляться единым образом в соответствии с нормативными документами. В частности, в стандартизованном виде должны предоставляться следующие типы информационных ресурсов:

- географические информационные ресурсы (картографические материалы, спутниковые снимки, данные полевых наблюдений и т.п.), а также соответствующие базы метаданных;

- фактографические базы данных и метаданных;
- библиографические базы данных и электронные каталоги;
- полнотекстовые базы данных и электронные библиотеки;
- авторитетные базы данных (описывающие субъекты информационного взаимодействия: персоны, организации и т.п.);
- другие ресурсы (аудио- и видеозаписи, электронные презентации и др.), снабженные стандартизованными метаданными.

Исходя из вышеперечисленных особенностей, на наш взгляд, РИС должна содержать следующие отдельные функциональные компоненты (5):

- систему идентификации информационных ресурсов;
- систему идентификации, аутентификации и авторизации пользователей;
- систему управления метаданными;
- систему управления информационными ресурсами, в том числе полнотекстовыми;
- систему сбора статистики;
- систему мониторинга доступности сервисов и ресурсов.

Реализация этих подсистем должна основываться на открытых спецификациях, связанных с международными стандартами. В распределенной среде должны быть задействованы механизмы синхронизации данных, например на основе репликаций. При этом в качестве протоколов сетевого взаимодействия должны выступать стандартные протоколы, например OAI, Z39.50, SRW/SRU, LDAP и др. (1; 3; 4; 8) (см. рис. 1).

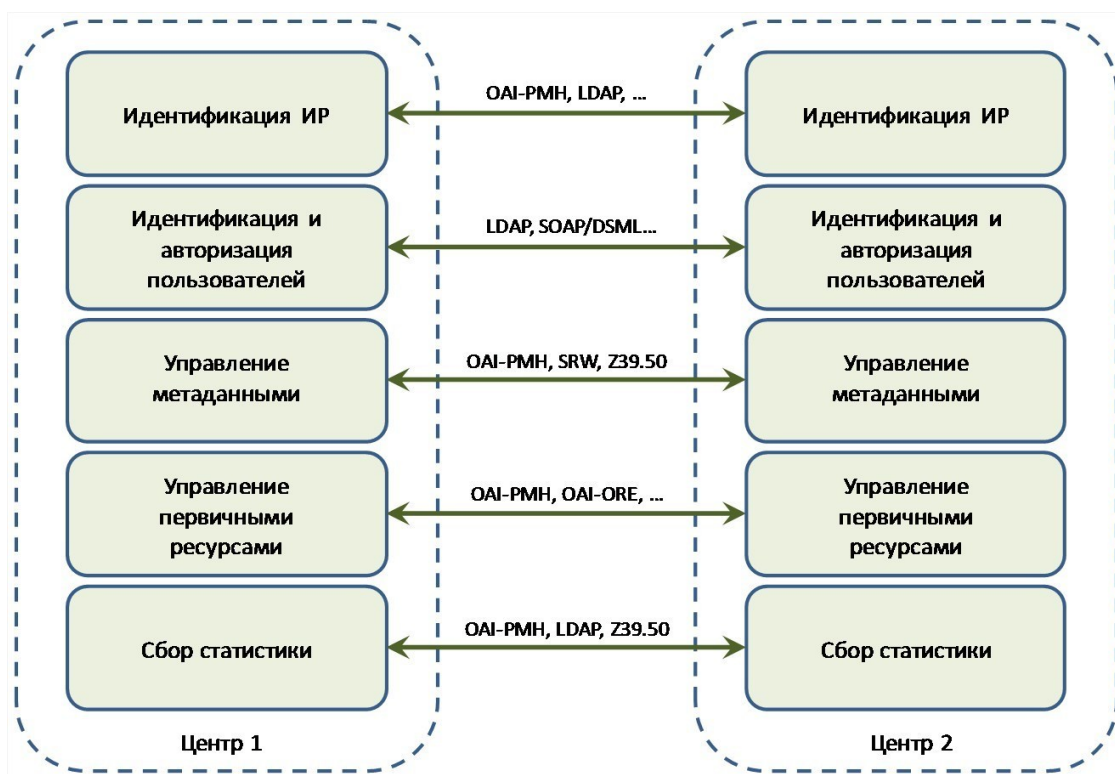


Рис. 1. Сетевое взаимодействие подсистем РИС

Практическая реализация сервисов SRW/SRU даст существенно новое качество информационной системы – возможность включения ее ресурсов в глобальные поисковые системы на более высоком уровне, нежели уровень внешней индексации статических веб-страниц другими системами. Иные возможные типы поиска связаны с поиском по заданным шаблонам и поиском с привлечением онтологии. Последний является более интеллектуальным типом поиска.

Для его реализации требуется дополнительная информация о предметной области, включающая определения терминов, сущностей и связей. Следует отметить, что представление этой дополнительной информации должно соответствовать глобальным договоренностям и международным стандартам. Иначе поиск с привлечением словарей, тезаурусов и онтологии всегда будет ограничен текущей системой, а интероперабельность не будет реализована.

В настоящее время существуют достаточно мощные информационные системы, в той или иной степени удовлетворяющие потребности научных работников в информации. Однако основной недостаток большинства систем – ограниченность возможностей обеспечения интеграции ресурсов как внутри каждой из систем, так и вовне. Отметим, что основу разработки ИС составляют, прежде всего, стандарты и международные рекомендации, формирующие профиль ИС. Под ним понимается набор из одного или нескольких базовых нормативно-технических документов (стандартов и спецификаций), ориентированных на решение определенной задачи (реализацию заданной функции либо группы функций приложения или среды) с указанием при необходимости выбранных классов, подмножеств, опций базовых стандартов, требуемых для выполнения конкретной функции. Наиболее важным являются профили метаданных циркулирующей в системе информации. Выбор профиля должен основываться на выполнении следующих требований:

- включать в себя основные типы информации, требующейся для поддержки научной работы;
- быть открытыми, т.е. обеспечивать доступ к соответствующей информации по этим описаниям;
- быть расширяемыми, т.е. обеспечивать возможность детализации описаний;
- обеспечивать возможности интеграции информации;
- обеспечивать возможности уникальной идентификации информации;
- обеспечивать возможности размещения и поиска информации в распределенной среде;
- быть ориентированными на современные и перспективные технологии описания и использования информации;
- обеспечивать возможности интероперабельности с внешней средой.

Собственно реализация каждой подсистемы при стандартизованных внешних интерфейсах не очень существенна. Однако базовые технологии их реализации вытекают естественным образом из их общей функциональности (см. пример на рис. 2).



Рис. 2. Базовые технологии компонент РИС

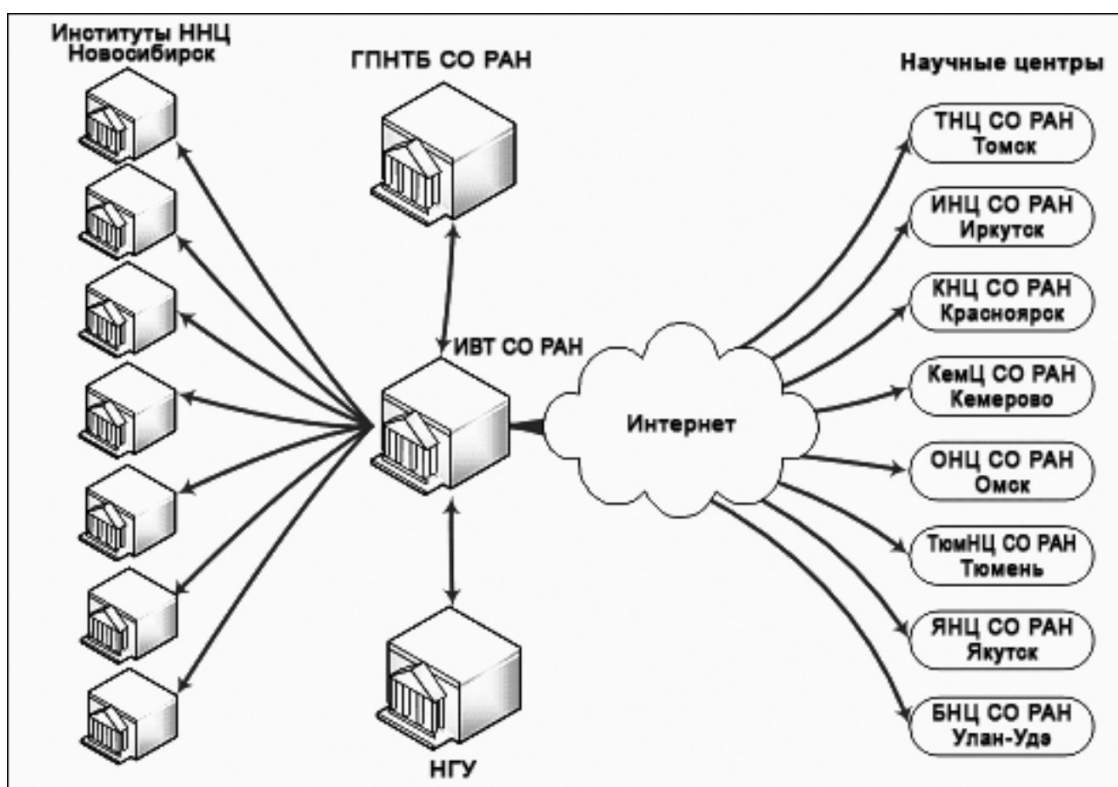


Рис. 3. Инфраструктура РИС библиотек СО РАН

В качестве иллюстрации РИС на основе перечисленных правил может выступать создаваемая РИС библиотек СО РАН, инфраструктура которой изображена на рисунке 3.

Другой иллюстрацией описываемого подхода может служить платформа массовой интеграции данных ZooSPACE, разрабатываемой в рамках Проекта ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса Рос-

сии на 2007–2013 годы» по теме «Разработка принципов и программных средств виртуальной интеграции распределенных источников данных на основе международных стандартов для создания масштабных информационных инфраструктур» (контракт № 07.514.11.4130), заказчиком которого является Министерство образования и науки Российской Федерации.

Платформа массовой интеграции предназначена для создания и поддержки функционирования масштабных, динамически формирующихся информационных инфраструктур из большого числа автономных баз данных. ZooSPACE должна обеспечивать функциональные характеристики:

1) поддержку унифицированного по информационной инфраструктуре представления данных, которое позволяет выполнять поисковые запросы, не зависящие от физического расположения данных;

2) предоставление прикладных программных интерфейсов для выполнения массовых поисковых запросов и управления информационной инфраструктурой;

3) обработку массовых запросов к совокупности баз данных реляционного и иерархического типов;

4) выбор поискового пространства запроса по метаданным, описывающим характеристики баз данных информационной инфраструктуры;

5) синтаксический контроль запроса с соответствующей диагностикой до начала его выполнения;

6) подключение / отключение баз данных и вычислительных ресурсов по инициативе их администраторов в процессе функционирования инфраструктуры;

7) защиту хранимых в информационной инфраструктуре данных от несанкционированного доступа.

Предполагается, что система ZooSPACE будет состоять из произвольного количества слабосвязанных самодостаточных узлов, функционирующих в соответствии с единой политикой. Взаимодействие узлов между собой осуществляется посредством сетевых протоколов прикладного уровня на основе транспортного протокола TCP / IP, как показано на рисунке 1.

Количество узлов в ZooSPACE не нормируется и может быть любым. Система ZooSPACE может состоять из одного-единственного узла.

Такой выбор инфраструктуры узлов позволяет обеспечить достаточно гибкую распределенную информационную систему и реализовать всю необходимую функциональность, которая обеспечивается подсистемами ZooSPACE. В качестве подсистем ZooSPACE должны выступать следующие (см. рис. 4):

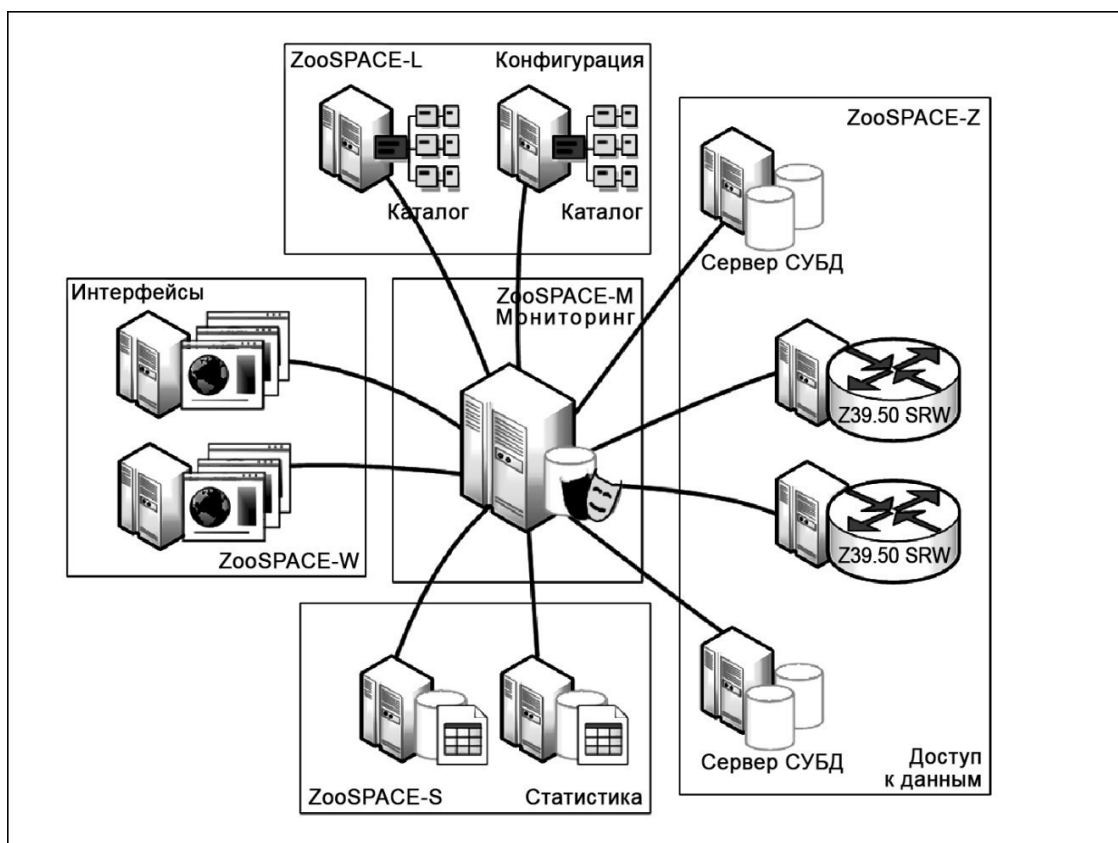


Рис. 4. Основные подсистемы ZooSPACE

- **ZooSPACE-L** – обеспечение функционирования справочной и административной подсистемы ZooSPACE;
- **ZooSPACE-Z** – обеспечение функционирования подсистемы доступа к базам данных системы ZooSPACE;
- **ZooSPACE-M** – обеспечение функционирования системы мониторинга всех компонент ZooSPACE;
- **ZooSPACE-S** – обеспечение функционирования подсистемы сбора статистики работы всех компонент ZooSPACE;
- **ZooSPACE-W** – обеспечение реализации пользовательских и административных веб-интерфейсов доступа к другим подсистемам ZooSPACE.

Система ZooSPACE в настоящее время находится на уровне предварительных испытаний.

В заключение отметим, что разрабатываемый в ИВТ СО РАН подход к построению распределенных информационных систем позволяет обеспечить возможность интеграции разнородных и разнотипных информационных ресурсов в единую информационную среду и унифицированного поиска благодаря использованию унифицированной модели работы с данными (в идеологии протокола Z39.50). Созданная система сервисов предоставляет широкому кругу потенциальных пользователей стандартизированный доступ к данным и алгоритмам их обработки. Такой подход позволяет обеспечить высокую степень информационной поддержки междисциплинарных научных исследований.

Литература

1. Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50 / ОИГГМ СО РАН. – Новосибирск: ИВТ СО РАН. – 2004. – 361 с.
2. Жижимов О.Л., Мазов Н.А., Федотов А.М. Некоторые заметки об эволюции цифровых репозитариев традиционных библиотек к полнофункциональным электронным библиотекам // Вестник Владивостокского гос. ун-та экономики и сервиса. Территория новых возможностей. – Владивосток, 2010. – Т. 3. – № 7. – С. 55–63.
3. Жижимов О.Л., Пестунов И.А., Федотов А.М. Структура сервисов управления метаданными для разнородных информационных систем [Электронный ресурс] // Электронные библиотеки: рос. науч. электр. журнал. – 2012. – Т. 15. – № 5. – <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2012/part6/ZPF>
4. Жижимов О.Л., Федотов А.М., Федотова О.А. Построение типовой модели информационной системы для работы с документами по научному наследию // Вестник НГУ. Сер.: Информац. технологии. – 2012. – Т. 10. – № 3. – С. 5–14.
5. Жижимов О.Л., Федотов А.М., Шокин Ю.И. Разработка принципов и программных средств виртуальной интеграции распределенных источников данных на основе международных стандартов для создания масштабных информационных инфраструктур // XIV Российская конференция с участием иностранных ученых «Распределенные информационные и вычислительные ресурсы» – DICR-2012 (Новосибирск, Россия, 26.11 – 30.11.2012): Программа конференции и тезисы докладов. – Новосибирск: ИВТ СО РАН, 2012. – С. 20.
6. Жижимов О.Л., Федотов А.М., Юданов Ф.Н. Модель управления информационными ресурсами организации // Вестник НГУ. Сер.: Информац. технологии. – 2010. – Т. 8. – № 4. – С. 81–95.
7. Федотов А.М., Барахнин В.Б., Жижимов О.Л., Федотова О.А. Технология создания корпоративных информационных систем учета трудов научных работников // Вестник НГУ. Сер.: Информац. технологии. – 2011. – Т. 9, вып. 2. – С. 31–41.
8. Федотов А.М., Шокин Ю.И., Жижимов О.Л., Молородов Ю.И. Служба директорий LDAP как единая информационная среда // Открытое и дистанционное образование. – Томск, 2007. – № 4 (28). – С. 31–41.
9. Шокин Ю.И., Федотов А.М. К вопросу о развитии информационной инфраструктуры СО РАН // Вычислительные технологии. – Новосибирск, 2009. – Т. 6. – № 6. – С. 127–137.
10. Шокин Ю.И., Федотов А.М., Барахнин В.Б. Проблемы поиска информации. – Новосибирск: Наука, 2010. – 198 с.

Интегральная система ВИНТИ РАН по информационному обеспечению научных исследований в области точных, естественных и технических наук: Концепция создания и перспективы развития

*И.Ю. Никольская*¹⁵

Рассмотрен опыт проектирования и основные направления развития интегральной информационной системы (ИИС) ВИНТИ РАН по точным, естественным и техническим наукам. Характеризуются информационные ресурсы, представленные в ИИС, поисковые возможности и средства их реализации. Показаны новые тематические фрагменты в составе ресурсного комплекса ИИС, обозначены пути формирования механизмов информационной навигации. Обосновывается необходимость дальнейшего развития классификационных систем по точным, естественным и техническим наукам и работы по совершенствованию рубрикаторов ВИНТИ РАН, относящихся к этим областям научного знания.

Ключевые слова: база данных; банк данных; информационная система; информационный ресурс; информационная навигация; классификационная система; реферативный журнал; рубрикатор; предметный указатель.

1. Основные характеристики интегральной информационной системы ВИНТИ РАН

Интегральная информационная система (ИИС) ВИНТИ РАН – это комплекс информационных ресурсов и научно-информационные средства для их эффективного использования. Ядром проектируемой ИИС станет Банк данных (БнД) ВИНТИ (1). Он представляет собой совокупность подсистем, функционирующих на различных программно-аппаратных платформах в технологическом сегменте локально-вычислительной сети (ЛВС) ВИНТИ. Характерной особенностью БнД является постоянное совершенствование механизма его функционирования и способов представления в нем информации. К другим составляющим элементам ИИС относятся архив опубликованных первичных и вторичных документов, а также разработанные процедуры информационного обмена с отечественными и зарубежными информационными организациями. Эти процедуры позволяют говорить о создании в ВИНТИ уникальной информационной навигационной системы, которая будет функционировать в рамках создаваемого в ВИНТИ специального портала.

Общее представление об информационных потоках и ресурсах ВИНТИ дает рисунок 1.

¹⁵ *Никольская Инга Юрьевна* – доктор технических наук, заместитель директора по научной работе Всероссийского института научной и технической информации РАН (г. Москва). E-mail: noz@viniti.ru

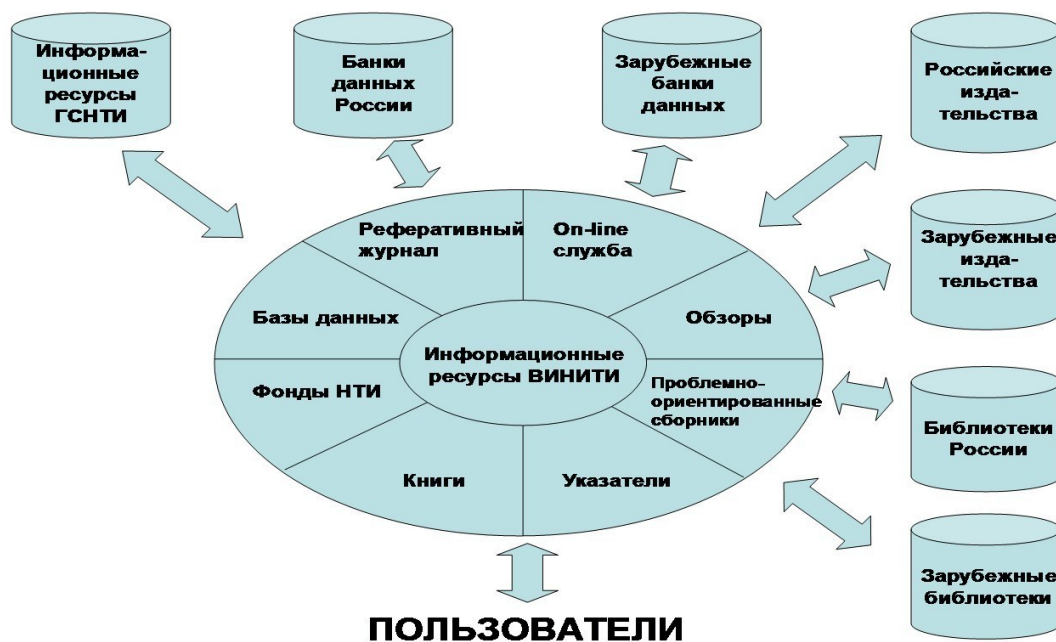


Рис. 1. Информационные потоки и ресурсы ВИНТИ

Специализированный научно-информационный портал ВИНТИ¹⁶ призван обеспечивать многоцелевую навигацию по электронным ресурсам в различных областях науки и техники. Благодаря использованию сети Интернет обмен информацией может происходить между регионами России, странами СНГ, а также ближнего и дальнего зарубежья. Система информационной навигации ВИНТИ представлена на рисунке 2.

¹⁶ <http://www2.viniti.ru>



Рис. 2. Средства информационной навигации ВИНИТИ

Составной частью в ИИС ВИНИТИ войдет архив ВИНИТИ. В своей основе – это сотни тысяч поступивших в ВИНИТИ за многие годы книг, статей, материалов конференций, специальных выпусков и изданий, депонированных рукописей, авторефератов кандидатских и докторских диссертаций. Архив будет постоянно пополняться огромным количеством ежегодных поступлений, которые обрабатываются и систематизируются с использованием УДК (13–14), ГРНТИ (4–6) и Рубрикатора ВИНИТИ (11–12).

Доступ к архиву будет обеспечиваться через фонды ВИНИТИ с помощью систематизированных информационных кодов документов (СИДов) и поисковых образов документов (ПОДов). Сегодня очевидна необходимость постепенного перевода наиболее используемой части архива в электронную форму. Поэтому требуется дальнейшее развитие информационной системы в указанном направлении. Система должна базироваться на современных информационных технологиях и располагать богатым банком знаний (9, с. 21–23).

Для решения этих задач была предложена концепция создания ИИС ВИНИТИ. На базе ВИНИТИ как головного института Государственной системы научной и технической информации (ГСНТИ) и базовой организации по обмену научной и технической информацией между государствами – участниками СНГ была начата разработка специальной информационно-поисковой системы по точным, естественным и техническим наукам. В соответствии с концепцией вся информационная система ВИНИТИ будет состоять из нескольких информационных комплексов, каждый из которых, в свою очередь, будет содержать несколько информационных подсистем:

1. Точные науки
 - 1.1. Математика
 - 1.2. Физика
 - 1.3. Астрономия
 - 1.4. Механика

2. Естественные науки
 - 2.1. Науки о жизни
 - 2.2. Науки о Земле
 - 2.3. Химия
3. Технические науки
 - 3.1. Автоматика и радиоэлектроника
 - 3.2. Машиностроение
 - 3.3. Металлургия
 - 3.4. Транспорт
 - 3.5. Энергетика и электротехника

Тематическое деление комплексов соответствует внутренней структуре ВИНТИ (15, с. 72). Пути создания интегрированной системы информационных ресурсов рассмотрены в (8). Работы по созданию ИИС ВИНТИ будут проходить в несколько этапов. На первом этапе планируется охватить комплекс точных наук (в первую очередь, математических и физических). Совершенно очевидно, что формирование ИИС ВИНТИ должно определяться новыми перспективами развития информационной системы ВИНТИ (обновленными представлениями об информационных ресурсах соответствующих БД, выходами на различные классификационные системы, широким использованием данных из смежных областей знаний и др.).

2. Архитектура ИИС ВИНТИ РАН по математическим и физическим наукам

Качество информационной системы в первую очередь определяется спектром информационных услуг, предоставляемых пользователю. При этом имеется в виду как удовлетворение запросов самого пользователя, так и информационный сервис, предлагаемый системой.

Информационная система должна обеспечивать выполнение следующих основных требований (9, с. 21–23):

- гарантированная полнота вводимого массива документов по объявленной тематике и данным, содержащихся в документе;
- высокая информативность рефератов и формализация изложения;
- многоаспектность отражения содержания документа как на естественном языке в реферате, так и на языке классификационных кодов и ключевых слов;
- наличие ряда дополнительных сведений: место выполнения работы, указание характера документа (оригинальная работа, обзор, персоналия и т.д.), вида документа (статья, диссертация, монография, препринт).

Внедрение современных технологий, использующих высокоорганизованную электронную технику, предоставляет новые возможности по части расширения спектра информационных услуг (быстрота поиска нужной информации, использование для поиска больших массивов документов, возможность вхождения в информационные банки данных и передачи информации по различным каналам связи и др.).

В то же время автоматизацию процессов подготовки, хранения, поиска и выдачи информации можно рассматривать как средство совершенствования обслуживания пользователей. Основой совершенствования обслуживания и расширения сервиса являются:

- обогащение информационного фонда данными, в первую очередь, получаемыми путем углубленной интеллектуальной обработки документа;
- обеспечение средств доступа к хранящейся в банке информации;
- возможность пополнения информации, содержащейся в базах данных, новыми фактами, полученными путем применения экспертных систем, и в конечном итоге – создание базы знаний.

Автоматизация информационных процессов должна обеспечить пользователю возможность не только получить ответ на поставленный вопрос, но и провести диалог с базой знаний. Поэтому наряду с документальными базами данных должен быть создан ряд сервисных (обслуживающих информационный поиск) и справочных баз данных, имеющих корреляцию с документальными базами. Все эти базы будут загружены в систему и сформируют ее единый, политематический аспектный информационный фонд. При этом предполагается, что информация, предназначенная для вывода из баз данных системы, может предоставляться как на бумажных, так и на электронных носителях.

Таким образом, можно сформулировать цели создания интегральной информационной системы ВИНТИ (ИИС ВИНТИ).

А. Создание политематического информационного фонда, содержащего:

- документальные базы первичных и вторичных информационных источников;
- сервисные (обслуживающие информационный поиск) базы, необходимые для формирования документальных баз данных (классификационные системы, различного вида указатели, многоязычные терминологические словари, тезаурусы, дескрипторные списки и др.);
- справочные базы (научных центров, тематических исследовательских планов и их реализации, научных конференций, регистрационной системы ученых-математиков, планов изданий математических издательств и др.).

Б. Организация оперативного и качественного информационного обслуживания индивидуальных и коллективных пользователей, которым необходима многоаспектная информация по спектру проблем и задач, исследование и решение которых возможно с использованием информационных ресурсов ИИС.

ИИС будет выполнять оперативное обслуживание своих постоянных пользователей, обращающихся с запросами. Система сможет обеспечить выдачу любой информации, содержащейся в ее информационном фонде. Поиск информации будет проходить в режиме диалога пользователя с системой.

Обращение к системе будет происходить на естественном языке предметной области. Системой будут обеспечиваться «дружелюбный» интерфейс и возможность доступа к информационному фонду пользователей с разным уровнем подготовки.

В качестве рабочего естественного языка первоначально будет принят русский язык, а затем – английский и другие языки. Английский язык будет использоваться как язык-посредник для дальнейшего расширения набора рабочих языков.

Сведения могут выдаваться в любой заказанной компоновке информационных данных, содержащихся в системе.

Пользователь сможет стать постоянным абонентом издания, скомпонованного по его личному заказу (избирательное распространение информации (ИРИ)). В зависимости от набора заказанных данных (информационных полей) такие издания будут иметь вид экспресс-оповещения.

Система будет обеспечивать получение информационных изданий в электронном виде, допускающем их тиражирование, как на электронных, так и на бумажных носителях. В частности, смогут компоноваться и выдаваться на электронных носителях:

- реферативный журнал (русская версия, в перспективе и английская версия);
- сигнальная информация (библиографический сборник тематически упорядоченных библиографических описаний новых документов, поступивших в систему);
- указатели к реферативному журналу (к каждому номеру реферативного журнала), а также кумулятивные годовые указатели (авторские, систематические, указатели журналов, конференций, монографий, персоналий и т.д.).

В режиме диалога, используя также экспертную систему, пользователь сможет последовательно установить интересующую его многоплановую информацию. Эта форма сервиса осно-

вана на том, что все базы данных, включенные в систему (и те, которые могут быть включены в дальнейшем), образуют Единый информационный фонд.

Программное обеспечение системы позволит выдавать комплексный ответ на запрос, который компонуется путем обращения к разным базам данных. Банк данных обеспечит взаимодействие всех баз, погруженных в Единый информационный фонд. Система предусматривает возможность расширения числа баз данных и организации взаимодействия между ними.

В результате анализа информационных ресурсов в области математических и физических наук и средств доступа к ним стала очевидной необходимость формирования в составе ИИС ВИНТИ принципиально новых фрагментов этой системы. Создание этих фрагментов должно обеспечить процесс формирования у пользователей интегральных проблемно-ориентированных информационных полей и, следовательно, переход на более высокий уровень информационного сопровождения научных исследований. Для решения поставленной задачи необходимо:

- разработать архитектуру и необходимый интерфейс для пользователей ИИС;
- расширить зону поиска релевантной информации и существенно снизить уровень «информационного шума»;
- разработать интегральные классификационные системы (рубрикаторы по математическим и физическим наукам), позволяющие вести сквозной поиск необходимой информации в отечественных и зарубежных источниках;
- разработать и обеспечить дальнейшее развитие системы лингвистического оснащения, обеспечивающей поиск релевантной информации;
- обеспечить взаимодействие электронных информационных ресурсов (ЭИР) ИИС с информационными ресурсами в традиционной (печатной форме), включая первоисточники.

Обобщенная архитектура ИИС ВИНТИ РАН по математическим и физическим наукам представлена на рисунке 3. В составе ИИС центральную часть составляют БД «Математика» и «Физика».

3. Электронные информационные ресурсы ИИС ВИНТИ РАН по математическим и физическим наукам

При разработке БД «Математика» и «Физика» необходимо было полностью переосмыслить архитектуру новых БД в составе политематического БД ВИНТИ, а также перспективы развития обеих БД и в полной мере использовать многолетний опыт известных математических и физических школ России.

Для эффективной работы системы информационного обеспечения научных исследований в области математических и физических наук БД «Математика» и «Физика» должны обладать следующими характеристиками:

- расширяемость, т.е. возможность переноса как системы, так и всей совокупности данных на другие программно-аппаратные платформы;
- возможность эффективной работы в гетерогенных средах;



Рис. 3. Обобщенная архитектура ИИС ВИНТИ по математическим и физическим наукам

- обеспечение работы с современными средствами универсального доступа к данным (www-браузерами, CASE-средствами разработки и др.);
- наращиваемость, т.е. возможность расширять объемы доступных данных без уменьшения пропускной способности системы.

Обобщенная структурная схема БД ВИНТИ по математическим и физическим наукам представлена на рис. 3. Она содержит в своем составе три функциональных комплекса: ресурсный, навигационный и программно-технологический.

3.1. Ресурсный комплекс БД ВИНТИ «Математика» и «Физика»

Приведем описание содержания баз данных ресурсного комплекса БД ВИНТИ «Математика» и «Физика» (рис. 4).



Рис. 4. Обобщенная структурная схема ресурсного комплекса БД «Математика» и «Физика»

3.1.1. Документальные базы данных

БД первоисточников содержит полнотекстовые электронные копии документов, отраженных в РЖ «Математика» и «Физика». Система обеспечивает получение копий документов в любом количестве экземпляров как на электронных, так и на бумажных носителях информации.

БД вторичных документов

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.