

Единая электронная система библиотечных сервисов

© С.К. Шикота

sveta@chg.ru

© С.А. Крашаков

Институт теоретической физики им. Л.Д. Ландау РАН
Научный центр РАН в Черногловке

Черногловка

sakr@chg.ru

© Л.Н. Щур

shchur@chg.ru

Аннотация

Излагается опыт по разработке единой электронной системы библиотечных сервисов Научного центра РАН в Черногловке. Особое внимание уделено разработке подсистемы научных публикаций сотрудников Научного центра и связанной с ней подсистемой подсчета индекса цитирования.

Работа выполнена при поддержке РФФИ, грант 11-07-00385, и в рамках программы РАН “Высокопроизводительные вычислительные системы и научные телекоммуникации”.

1 Введение

Для успешного выполнения научной работы необходимо иметь доступ к информации о результатах научных исследований. Информация по узкой области содержится в тематических периодических изданиях и доступна на рабочих совещаниях, конференциях и симпозиумах. Получаемая из этих источников информация дает удовлетворительное представление о проводимых в узкой области исследованиях. Следует отметить, что для проведения успешных исследований необходим также доступ к научным результатам исследований, проводимых в смежных областях науки. Это особо важно для исследований в области критических технологий, таких, например, как наука о новых материалах и новые информационные технологии. В этих областях практически трудно не упустить важный научный результат или разработку.

Настоящая работа посвящена одной из составляющих решения этой проблемы. Предлагается создание единой информационной системы библиотечных сервисов, предоставляющей, в частности, открытый доступ к полным текстам научных статей и разработок сотрудников учреждений Научного центра РАН в Черногловке (НЦЧ РАН). Назначение разрабатываемой с этой целью базы данных публикаций двунаправленное. Первое состоит в привлечении внимания широкой аудитории к научным достижениям сотрудников НЦЧ РАН. Второе состоит в предоставлении сотрудникам НЦЧ возможности для поиска научных контактов по смежным областям внутри структуры НЦЧ РАН.

Можно привести весьма показательный пример малого уровня информированности научных сотрудников о проводимых в соседних учреждениях и лабораториях исследованиях и разработках. Нобелевская премия по физике 2010 года была присуждена А. Гейму и К. Новоселову за проведенные ими в Университете Манчестера пионерские исследования физических свойств графена [1]. Исследования проводились на образцах, изготовленных в Институте проблем технологии микроэлектроники и особо чистых материалов РАН (ИПТМ РАН), одном из учреждений НЦЧ РАН (часть соавторов в цитируемом источнике – штатные научные сотрудники ИПТМ РАН). Этот факт практически неизвестен широкой научной общественности в стране и мире. Удивительно то, что он также неизвестен и в самой Черногловке. О чем он свидетельствует? Первое, о высоком уровне технологий, имеющихся в ИПТМ РАН для создания экспериментальных образцов наноразмера. Второе, о наличии распределенного международного научного коллектива. Третье, о том, что в ИПТМ проводятся работы на мировом уровне по одной из критических технологий. Четвертое, о том, что в настоящее время отсутствует эффективно работающий механизм обмена научной информацией.

Разрабатываемая электронная система библиотечных сервисов призвана внести вклад в решение проблемы распространения научной информации.

Интерфейс пользователя должен предоставлять: возможности поиска и навигации, возможность организации личного кабинета с набором настраиваемых функций и иметь механизмы реализации других встраиваемых функциональных возможностей.

План статьи таков. В разделе 2 мы приводим информацию об организации библиотек НЦЧ РАН. В разделе 3 описана телекоммуникационная инфраструктура сети НЦЧ РАН. В разделе 4 излагается общая постановка задачи. В разделе 5 перечисляются отдельные элементы системы и их основные функции. Подсистема электронного каталога публикаций сотрудников изложена в разделе 6. В разделе 7 обсуждаются технические аспекты организации подсистемы подсчета индекса цитирования статей.

2 О Научном центре РАН в Черногловке

Учреждения Научного центра РАН в Черногловке [2] входят в состав четырех отделений

РАН: отделения физических наук, отделения химии и наук о материалах, отделения нанотехнологий и информационных технологий и отделения наук о земле. В состав НЦЧ РАН входит 16 учреждений Научного центра. Из общей численности работников в 4500 человек более половины составляют научные сотрудники и аспиранты - 2700. Каждое учреждение Научного центра имеет свою собственную библиотеку. Организационная сложность заключается в том, что библиотеки имеют различный юридический статус: 1) филиал Библиотеки естественных наук РАН (БЕН РАН); 2) филиал Библиотеки Научного центра (БНЦ РАН), которая в свою очередь входит в структуру БЕН; 3) собственная библиотека организации. Информация об имеющихся ресурсах, в том числе публикациях сотрудников, также разрознена. Отсутствуют механизмы поиска. Для создания единой системы, в первую очередь, необходима инфраструктура,:

- обеспечивающая объединение существующих ресурсов в единое целое;
- предоставляющая свободный доступ пользователей к распределенной системе по протоколу http.

3 Инфокоммуникационная инфраструктура Научного центра РАН в Черноголовке

Академические организации Черноголовки объединены в Научный центр РАН в Черноголовке. Это дает уникальную возможность по построению единой информационной и телекоммуникационной инфраструктуры, в которой принципиально возможна координация действий специалистов различного профиля. В Научном центре создана академическая компьютерная сеть ChANT (Chernogolovka Academic Network) [3], которая обеспечивает пользователям сети, сотрудникам учреждений, входящих в структуру Научного центра, высокоскоростной доступ к опорной оптоволоконной сети. Опорная сеть обеспечивает доступ к инфокоммуникационным ресурсам учреждений НЦЧ РАН: вычислительным кластерам; системе проведения распределенных видеоконференций; базам данных; архиву свободно распространяемого программного обеспечения и др. Опорная сеть предоставляет высокоскоростной доступ к удаленным ресурсам научных сетей России, Европы и Северной Америки, а также к общедоступным ресурсам сети Интернет (рис. 1).

4 Единая электронная система библиотечных сервисов

Единая электронная система библиотечных сервисов (ЕБС) предназначена для обеспечения доступа сотрудников учреждений Научного центра РАН в Черноголовке к ресурсам Библиотеки естественных наук РАН (БЕН РАН) и библиотек Научного центра. Элементы ЕБС должны иметь функции:

- обеспечения доступа к единому электронному каталогу неперіодических изданий БЕН РАН и библиотек Центра;
- предоставления возможности удаленного заказа литературы;

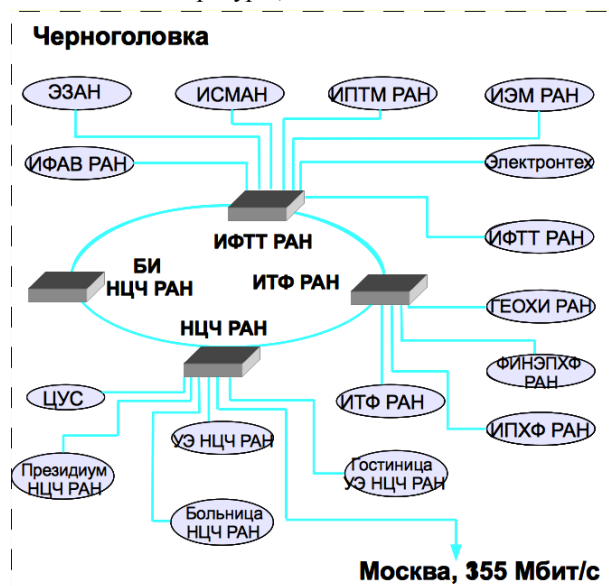


Рис. 1 - Схема опорной сети ChANT

- осуществления доступа к единой базе данных публикаций сотрудников учреждений Научного центра;
- предоставления доступа к единому каталогу подписки учреждений НЦЧ РАН к полнотекстовым ресурсам периодических изданий.

Основной целью создания системы является обеспечение возможности унифицированного доступа сотрудников учреждений Научного центра к научной информации разного типа непосредственно с рабочих мест научных сотрудников, аспирантов, студентов и библиотечных работников. Под различными типами научной информации считаются:

- каталоги неперіодических изданий библиотек БЕН, БНЦ и учреждений НЦЧ РАН;
- каталоги периодических изданий библиотек БЕН, БНЦ и учреждений НЦЧ РАН;
- электронные каталоги и полные тексты статей научных издательств по подписке БЕН, НЦЧ РАН и учреждений НЦЧ РАН;
- электронные каталоги и полные тексты научных работ научных сотрудников учреждений НЦЧ РАН.

На первом этапе мы планируем создание двух основных составных частей общей библиотечной системы. Одна часть, которая не обсуждается в настоящей работе, связана с созданием системы единого каталога неперіодических изданий. Такая система позволит объединить в единое целое распределенные ресурсы библиотек, что не только облегчит пользователю поиск необходимого издания, но и позволит координировать внутробиблиотечный учет. В качестве средства создания каталога

была выбрана автоматизированная система библиотечного учета «Библиобус», разработанная в Библиотеке естественных наук РАН [4]. Система «Библиобус» позволяет осуществлять все основные технологические процессы на пути книги от формирования заказа на литературу до выдачи изданий читателю. Также предусматривается возможность просмотра каталогов непериодических изданий, имеющихся в библиотеках РАН и заказ непериодических изданий по МБА.

Вторая составная часть библиотечной системы связана исключительно с публикациями сотрудников научного центра. В мире имеется большое число информационных систем для работы с научными и библиографическими данными. Многие научные учреждения имеют собственные информационные системы, представляют в электронной форме данные о публикациях сотрудников, о проводившихся или ведущихся научных исследованиях и проектах, о результатах исследований. Нами предусматривается возможность конвертации уже имеющихся данных в единую базу. По нашим данным, региональных информационных систем на уровне научного центра, подобных разрабатываемой в Научном центре РАН в Черноголовке, не имеется.

Таким образом, мы ставим задачу создания единой библиотечной системы научного центра, которая обеспечит доступ к электронным каталогам и полным текстам научных статей научных сотрудникам, студентов и аспирантов непосредственно с их рабочего места.

5 Разработка структуры единой библиотечной системы

Проект технического задания на разработку Единой библиотечной системы НЦЧ РАН (ЕБС) выделяет следующие основные компоненты системы:

- подсистему электронного каталога непериодических изданий. Подсистема предназначена как для персонала библиотек центра (внутрибиблиотечный учет), так и для сотрудников Научного центра (поиск литературы, заказ литературы по межбиблиотечному абонементу);

- подсистему доступа к единой библиотечной системе НЦЧ РАН по протоколу http. Подсистема обеспечивает различный уровень доступа к ресурсам для зарегистрированного пользователя (он же читатель библиотеки), гостя и сотрудника библиотеки, а также позволяет выполнять поиск по всему информационному пространству ЕБС;

- подсистему электронного каталога периодических изданий. Подсистема предназначена для поиска изданий в ЕБС и одновременно является частью подсистемы учета публикаций сотрудников и аспирантов учреждений Научного центра;

- подсистему электронного каталога публикаций сотрудников и аспирантов учреждений Научного центра. Подсистема предназначена для ввода, хране-

ния, поиска и отображения информации о публикациях сотрудников и институтов НЦЧ;

- подсистему подсчета индексов цитирования сотрудников институтов НЦЧ;

- подсистему резервного копирования, обеспечивающую сохранность данных путем ежедневного дублирования информации, хранящейся в ЕБС, в облачную систему хранения данных.

Далее мы рассматриваем более подробно две ключевые подсистемы: 1) подсистему электронного каталога публикаций сотрудников учреждений Научного центра и 2) подсистему подсчета индексов цитирования публикаций отдельных сотрудников и организаций в целом.

6 Подсистема электронного каталога публикаций сотрудников НЦЧ РАН

Электронный каталог публикаций сотрудников предназначен для ввода, хранения, поиска и отображения информации о публикациях сотрудников и учреждений НЦЧ. Кроме того, в связи с возникшей в последние годы необходимостью заполнения отчетности для АСУ РИД РАН и вычисления индексов ПРНД, система также предназначена для обеспечения выдачи актуальной информации об ipract факторе журнала и индексе цитирования статьи.

При создании подобной системы необходимо учитывать ряд особенностей современных научных публикаций и российских, в частности.

Во-первых, большинство российских журналов параллельно переводятся и издаются за рубежом. Однако, ни одна из существующих библиографических баз и баз данных о цитировании, включая Российский индекс научного цитирования – РИНЦ [5], не учитывает эту особенность и считает статью в российском журнале и статью в соответствующем переводном журнале за две различных публикации.

Во-вторых, многие статьи параллельно публикуются в виде препринтов (особенно это свойственно для публикаций в области физико-математических наук и computer science), наиболее известна существующая с августа 1991 года электронная библиотека препринтов arXiv.org [6]. В большинстве существующих библиографических баз публикации в этой библиотеке препринтов игнорируются. Исключениями являются разве что библиографические базы ADS по физике и астрономии [7] и INSPIRE (SPIRES) по физике высоких энергий [8], которые считают публикации в журнале и в библиотеке препринтов за одну публикацию и суммируют их цитирование.

И, наконец, многие публикации могут цитироваться различными способами: как статья (глава) в книге и как статья в периодическом издании, если данная книга является сериальным изданием; как публикация в трудах конференции и как статья в журнале или сериальном издании и т.д. Большин-

ство существующих библиографических баз не учитывают эту особенность.

Отдельным можно поставить вопрос о необходимости привязки публикации к конкретным авторам и организациям. Это далеко не просто по нескольким причинам. Во-первых, имеются однофамильцы и даже полные тезки (особенно по ФИО сотрудников), нередко даже в одной организации. Во-вторых, имеется множество вариантов написания фамилий (и имен) в переводных публикациях. В-третьих, имеются различные варианты написания названия организации в публикуемых статьях. Система должна различать такие публикации при вводе и выводе информации. Многие существующие библиографические базы пытаются решать эту проблему, однако ни одна из них не решает ее в полной степени правильно и в каждой из них встречаются ошибки. По нашему мнению, единственным способом точной привязки статей к конкретным авторам является децентрализованный экспертный подход.

Исходя из этого, основу предлагаемой нами подсистемы составляет общая база публикаций (без разделения на тип публикации) с привязкой каждой публикации к одному или нескольким авторам, работающих или работавших (на момент публикации) в одном из институтов НЦЧ и к одному или нескольким институтам НЦЧ, в которых выполнялась работа. Корректность данных по каждому из институтов должна контролироваться уполномоченными представителями соответствующих учреждений.

Предложенный подход уже частично реализован в базе публикаций сотрудников Института теоретической физики им. Л.Д. Ландау РАН [9].

7 Подсистема подсчета индекса цитирования статьи

Ключевым моментом подсистемы подсчета индекса цитирования является обеспечение наиболее полного охвата цитируемых источников и цитирующих их публикаций. При этом следует обеспечить также максимально корректный поиск и отбор публикаций конкретных сотрудников, работающих в НЦЧ, что, как было отмечено выше, является совсем не тривиальной задачей.

В мире имеется несколько источников информации о цитировании публикаций. Каждый из них имеет определенные достоинства и недостатки. Наиболее полным и широко признанным источником информации о цитировании, охватывающем все области знаний, по которым ведутся исследования в институтах НЦЧ РАН, является ISI Web of Knowledge (бывший Science Scitation Index) компании Thomson Reuters [10]. Другой источник информации – Scopus [11] охватывает меньший период времени, хотя и имеет несколько большее число цитируемых источников. Российский индекс научного цитирования (РИНЦ) находится в стадии развития и имеет серьезные

недостатки [12], обсуждение которых выходит за рамки настоящей статьи.

Системы WoK и Scopus позволяют проводить поиск по большинству полей и по тем данным в том виде, в котором они были опубликованы в статьях и далее введены в базу данных систем. Система РИНЦ позволяет проводить поиск только по названиям статей, фамилиям из имеющегося списка (при этом в случае однофамильцев отнесение публикации к конкретному автору из списка не всегда корректно), тематике, журналу по списку и году. Поиск по названию или адресу организации возможен только по организациям, имеющимся в списке (при этом отнесение статьи к организации также не всегда правильно).

Отметим, что только система WoK имеет программный интерфейс, что позволяет делать запросы из программы.

Эти соображения привели к выбору системы WoK в качестве источника информации о цитировании. Нами получено предварительное согласие о возможности использования доступа к WoK из ЕБС по программному интерфейсу.

В Web of Knowledge (далее - WoK) имеется два типа поиска:

- Search — поиск в публикациях изданий, реферируемых WoK, по авторам, ключевым словам, названию, источникам (например, название журнала), году, адресам авторов и некоторым другим параметрам поиска;
- Cited Reference Search — поиск в полной базе, включающей дополнительные базы (типа Chinese Citation Index, к которой мы не имеем подписки) и также ссылки на другие публикации, найденные в статьях из реферируемых источников; данный поиск возможен только по фамилии, источнику (например, названию журнала) и году публикации.

Соответственно, найденные ссылки (и соответственно, их индексы цитирования) могут быть следующих типов:

- полные ссылки, включающие полное библиографическое описание публикации, имеющие уникальный ISI-ID, по которому с использованием соответствующего API можно осуществлять запрос на полное описание публикации или на ее цитирование;
- частичные ссылки, найденные в других базах; для этих ссылок выдается фамилия первого автора, фамилия И.О. искомого автора (не всегда), название статьи, источник, ISI-ID, DOI и количество цитирований;
- неполные ссылки, найденные в ссылках из статей из реферируемых источников; для этих ссылок выдается фамилия первого автора, источник и количество цитирований;

сюда же относятся ссылки на статьи в реферируемых журналах, которые были неполными (например, отсутствовал номер журнала, полный список авторов и т.п.), в результате чего WoK не смог отнестись к конкретному цитированию к конкретной статье;

- ошибочные ссылки — аналогичны предыдущим, но с ошибками в написании фамилии автора, названии источника, года, тома, страницах и т.п.

Вычисление индексов цитирования по полным (а также частичным) ссылкам в журналах, реферируемых WoK и имеющим уникальный ISI-ID, возможно только в тех случаях, когда практически все статьи опубликованы в зарубежных изданиях, реферируемых WoK.

Необходимо подчеркнуть, что полные ссылки выдаются только на публикации в изданиях, реферируемых WoK, вне зависимости от того, какой режим поиска используется — Search или Cited Reference Search. Следует также отметить, что из русских журналов в WoK реферируется около 100 (по всем областям знаний). При наличии двух параллельных изданий, русского и переводного, в базу реферируемых попадает только одно из них.

Доступная нам база публикаций включает полные ссылки только на работы после 1980 года, хотя в режиме Cited Reference Search выдаются ссылки (неполные и ошибочные) и на более ранние работы.

8 Заключение

В работе обсуждена необходимость создания единой электронной системы библиотечных сервисов Научного центра РАН в Черноголовке.

Подходы, использованные нами в данной работе, могут оказаться полезными для улучшения библиотечных сервисов.

Литература

- [1] K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, and A.A. Firsov, *Science* 306, 666 (2004)
- [2] <http://scc.chant.ru>
- [3] С.К. Шикота, Л.Н. Щур, С.А. Крашаков, А.Ю. Меньшутин, М.В. Григорьева. Региональная сеть для науки и образования ChANT как инфраструктура для Грид-приложений. Труды 4-ой международной конференции ГРИД'2010 "Распределенные вычисления и Грид-технологии в науке и образовании", Дубна, 28 июня - 03 июля 2010 г., с 345-351.
- [4] Автоматизированная библиотечная система «Библиобус» <http://www.eks3.ru/>
- [5] Российский индекс научного цитирования (РИНЦ), http://elibrary.ru/project_risc.asp
- [6] Электронный архив <http://arxiv.org/>
- [7] The SAO/NASA Astrophysics Data System (ADS), <http://www.adsabs.harvard.edu/>
- [8] High-Energy Physics Literature Database INSPIRE, <http://inspirehep.net/>
- [9] <http://itp.ac.ru/>
- [10] Web of Knowledge, <http://wokinfo.com/>
- [11] Scopus, <http://www.scopus.com/>
- [12] Н.Е.Каленов, О.В.Селюцкая. Некоторые оценки качества Российского индекса научного цитирования на примере журнала «Информационные ресурсы России» // Информационные ресурсы России, 2010, № 6. — С. 2-13

Information System for Library Services

Svetlana Shikota, Sergey Krashakov, Lev Shchur

We present experience with the development of system for integration of library services of Scientific Center RAS in Chernogolovka. We pay special attention to the development of subsystem of publications of researchers of Scientific Center as well as to the subsystem of citation indexing.