

ОПИСАНИЕ НАУЧНЫХ ИССЛЕДОВАНИЙ И ПРЕДСТАВЛЕНИЯ ПРЕДМЕТНЫХ ОБЛАСТЕЙ В КОММУНИКАЦИОННОЙ СРЕДЕ НА ОСНОВЕ КОНТРОЛИРУЕМОЙ ЛЕКСИКИ

Е. И. Моисеев, А. А. Муромский, Н. П. Тучкова, А. М. Меденников

Проблема распространения результатов научной деятельности в электронной информационно-поисковой среде является составной частью общей проблемы: создание научного информационного пространства РАН. Этой проблеме посвящено множество публикаций [1]. Фактически научное информационное пространство РАН имеется, если под этим понимать совокупность данных, сведений научного и иного характера, которые институты РАН помещают в интернете, соотносясь с особенностями научных направлений, традициями и возможностями. Проблема, следовательно, заключается в том, чтобы модернизировать уже имеющуюся систему информирования на основе новых современных информационных технологий, используя принцип интеграции данных и адресное распространение информации. Имеющиеся в распоряжении каждого пользователя системы Интернет возможности, позволяют индивидуально представлять свои труды, используя вторичную информацию, вторичные документы с предоставлением авторского и документального поиска для заинтересованных коллег, с последующим получением ими первичных документов в электронном/печатном виде. Такого рода «инициатива снизу» предоставляет определенные возможности и для коллективов авторов, например, сборников трудов научных организаций.

Вопросами адресной доставки информации, как библиотечной услуги, занимались и занимаются достаточно давно (режим ИРИ, рассылка по подписке и пр.). Создание такой коммуникационной среды представляется авторам насущной задачей информационной поддержки научных исследований. Частичная реализация такой задачи обеспечивается имеющимися в сети электронными библиотечными системами (например, БЕН РАН, e-library), системами цитирования (например, ISI WebKnowledge, Scopus), административными информационно-библиографическими системами (ЕНИП, math-net, ИСТИНА и др.). Тем не менее, вопросами адресной доставки информации занимаются в основном коммерческие ресурсы (amazon.com, например), в которых, как правило, мало или совсем отсутствуют научные публикации (исключение составляют учебники, монографии и компьютерные руководства) и издательские группы (Elsevier, Springer и др.). Известна и часть открытых библиографических ресурсов в Интернет (бесплатных электронных журналов), которые, как правило, не занимаются адресной доставкой информации, вероятно, в силу специфики организации (отсутствие коммерческой составляющей).

Основу информационного наполнения предлагаемого портала составляет тезаурус адресата [2]. Совокупность накопленной информации представляет описание предметной области исследований автора или авторского коллектива. Информация, помещенная в коммуникационную систему (КС), становится доступной научному сообществу, заинтересованному в разработках по определенной предметной области. Пересечение предметных областей авторов и авторских коллективов позволяет получить дополнительную информацию об актуальности исследований и приоритетах полученных результатов. Если у автора имеется некоторое количество трудов, разнообразных как по виду, так и по тематике в смысле какой-либо классификации, то возникает обычная проблема, называемая нередко проблемой «Найти». Обычная практика доступа к первоисточникам в режиме поиска, как правило, основана на процедуре, где в качестве поискового материала используются компоненты вторичных документов первоисточников. Вторичный документ первоисточника, представляемый как совокупность компонентов, каждый из которых воплощает определенную информацию о первоисточнике, позволяет принимать решение об обращении к первоисточнику в зависимости от условий режима доступа. Цель обращения к сервисам коммуникационной среды для автора состоит в использовании возможностей адресной доставки информации для определения новизны полученных результатов, актуальности научных исследований, поиска коллег и новых результатов (например, если работа реферируется в РЖ ВИНТИ, Annual Reviews или Mathematical Reviews без замечаний, то можно считать, что результат новый).

Состав и структура данных. Данные, накапливаемые в КС, представляют собой вторичные документы публикаций авторов [3]: названия, автор(ы), аннотации, ключевые слова и др. Вторичный документ рассматривается здесь как совокупность (множество) компонентов, имеющих в зависимости от назначения и применения определенное значение, связанное с задачей создания структурированного краткого и свернутого описания первоисточника – это, по определению, одно из назначений вторичного документа. Списки цитируемой литературы и сами тексты составляют предмет авторского, интеллектуального и имущественного права и предоставляются при соблюдении этих прав. Для наполнения системы предполагается использовать также информацию о метаданных из различных информационных ресурсов, в частности, Издательского отдела ЕНИП ВЦ РАН (<http://ccas.enip.ras.ru>), ПРНД ВЦ РАН [4], доступ ко вторичной информации научных публикаций зарубежных и отечественных библиотечных систем и издательских баз данных, что,

обеспечивается в системе РАН за счет средств РФФИ и других фондов. Для получения актуальной информации о публикациях предполагается, в частности, интеграция с данными библиотечных ресурсов РАН.

Классификационные схемы предметных областей формируются на основе:

- УДК (UDC)5/54, содержащая деление и соответствующую терминологию по естественным наукам, в том числе по математике, физике, химии и др.
- Рубрикатора (отечественного/иностранного)
- Mathematical Subject Classification (MSC), Zentralblatt (MATH и др.), Mathematical Reviews
- Тезаурусов (классификаций и терминологий)
- Некоторых классификаций в виде отдельных кодов, принятых в РЖ, издательствах (подчеркнем, что имеется в виду предметная область и смежные области конкретного адресата).

Основные функциональные компоненты КС. Функциональные сервисы КС направлены на организацию информационной поддержки исследований автора по некоторой заданной предметной области. Персональная страница автора формируется на основе авторизованного доступа к системе накопления публикаций. Для работы с публикациями автору предоставляются сервисы: ввода и редактирования списка публикаций, персонального тезауруса (тезауруса адресата), поиска публикаций с ключевыми словами из персонального тезауруса; расширения (объединения) списка ключевых слов за счет списков других авторов, авторизованных в данной коммуникационной среде; поиска коллег по ключевым словам; ввода и редактирования списка цитируемой литературы; обмена списками ключевых слов и списками публикаций с авторами, авторизованными в КС; услуга «подписка на публикации в системе» с ключевыми словами из заданного списка. Поисквые образы (запросы) содержат как термины, так и коды.

Поиск по ключевым словам позволяет учесть специфику и тематику публикаций [5], [6]. Основная функция по обмену информацией между авторами реализуется благодаря процедуре контроля ключевых слов. Для этого в системе хранится информация о всех ключевых словах авторизованных пользователей системы и ссылки на их публикации. При пополнении индивидуальных коллекций публикаций и ключевых словах осуществляется рассылка по подписке на публикации с данными ключевыми словами.

Реализация аналогичных проектов уже ведется, например, в европейском научном сообществе, например, немецкий проект ResearchGate (<http://www.researchgate.net/>) или американский проект профессиональных сообществ <http://www.linkedin.com/>. Однако в них в большей степени реализуются задачи социальной сети, а не адресной доставки информации о научных публикациях, что наиболее важно для информационной поддержки научных исследований.

Работа поддерживается Российским фондом фундаментальных исследований (проект № 13-07-00334-а).

ЛИТЕРАТУРА:

1. Cawkell T., Garfield E. Chapter 15. Institute for Scientific Information // A century of science publishing: a collection of essays / Einar N. Fredriksson (Ed.). IOS Press, 2001. С. 149-160. 312 с. ISBN 9781586031480
2. Шрейдер Ю. А. Тезаурусы в информатике и теоретической семантике // «Научно-техническая информация. Сер. 2», 1971, № 3.
3. Моисеев Е.И., Муромский А.А., Тучкова Н.П. О представлении и поиске научных результатов современными средствами в электронной информационно-поисковой среде. М.: МАКС Пресс, 2009. 98 с.
4. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Интернет и математические знания. Представление уравнений математической физики в информационно-поисковой среде. М.:МАКС Пресс, 2008. 80 с.
5. Гиляревский Р.С., Шапкин А.В., Белоозеров В.Н. Рубрикатор как инструмент информационной навигации. СПб.: Профессия, 2008. 352 с.
6. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Поиск научных публикаций для предметных областей с неустоявшейся терминологией // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (17-22 сентября 2012 г., г. Новороссийск). М.: Изд-во МГУ, 2012. 752 с. (С. 599-601).