

УДК 021(045)+655.4(045)

А. Н. Гребнев

КОММУНИКАЦИОННАЯ НАУЧНО-ОБРАЗОВАТЕЛЬНАЯ СРЕДА

В данной статье рассмотрены основные принципы и стандарты построения территориально распределенной информационной системы с открытым доступом, основанной на самоархивировании, автоматизирующей процессы обмена научными знаниями (преимущественно публикациями) и этапы жизненного цикла научной публикации в образовательных учреждениях.

Ключевые слова: электронные научные публикации, научные коммуникации, жизненный цикл научной публикации, самоархивирование, открытый доступ.

1 Введение

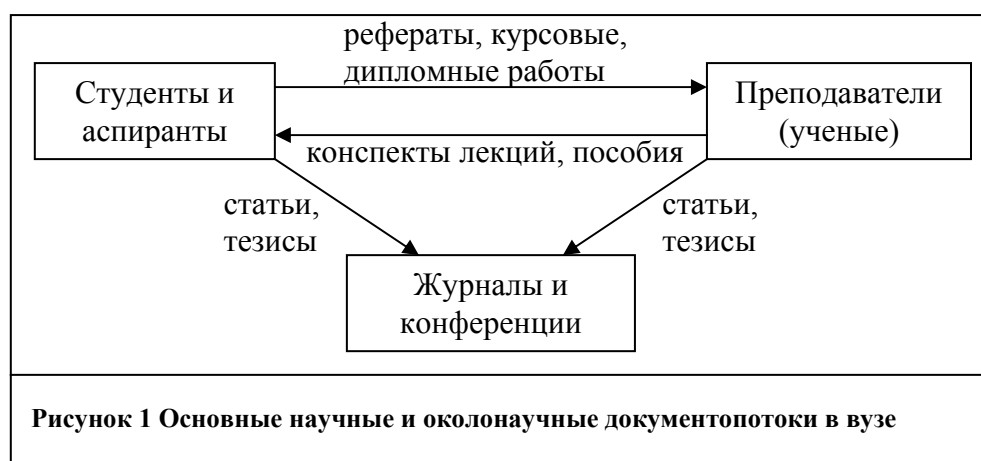
В эпоху научно-технического прогресса, по мере накопления громадных объемов информации, проблема медлительности обмена знаниями начинает все более и более остро стоять перед учеными, исследователями, преподавателями, студентами и прочими деятелями науки. Классические бумажные способы обмена научными и околонаучными знаниями уже не отвечают большинству предъявляемых требований. К настоящему моменту, в подавляющем большинстве случаев, документы (рефераты, тезисы, статьи и т.п.) подготавливаются в электронной форме, однако способы распространения и обмена информацией отнюдь не всегда столь прогрессивны. Развитие сети Интернет и ее технологий имеют огромные потенциальные возможности в области обмена научной информацией, но на данный момент они задействованы лишь на малую долю. Причем менее развитой оказалась научная коммуникация в сфере высшего профессионального образования. Если хорошо финансируемые

крупные научные учреждения имеют возможность использовать дорогие информационные системы и содержать достаточный штат сотрудников для обеспечения научных коммуникаций на требуемом уровне, то в среднестатистическом российском вузе данные процессы проходят при минимальной финансовой поддержке, а порой и исключительно на энтузиазме сотрудников. А ведь именно здесь, в истоках науки, в вузах растут и подготавливаются молодые ученые, приобщается к процессам формального научного обмена молодое перспективное поколение.

Необходима единая среда информационного обмена, основанная на прогрессивных технологиях, доступная каждому вузу, обеспечивающая более глубокую интеграцию науки в образование и тем самым достигающую не только их развития в отдельности и в совокупности, но и более плотного взаимодействия с мировыми научными и образовательными сообществами.

2 Научные и образовательные коммуникации сегодня

Рассмотрим (рис. 1) основные научные и околонаучные документопотоки в среднестатистическом вузе. Очевидной является необходимость обеспечения доступности максимально широкой аудитории всех обозначенных документов. В пример можно привести огромную популярность, среди студентов (да и не только), Интернет коллекций с рефератами, курсовыми и дипломными работами. В случае же статей и тезисов задача их прочтения наибольшим количеством заинтересованных



людей, привлечения внимания и получение откликов коллег на важные результаты исследований является первостепенной.

Научные бумажные публикации переживают не лучшие времена. Тираж большинства не крупных журналов и сборников, доступных начинающим ученым, невелик, он не позволяет обеспечить всех желающих требуемой информацией, территориально удаленные читатели порой вообще не имеют доступа. Частота и скорость выхода публикаций являются еще одним недостатком бумажных изданий, нередко случаи того, что еще до выхода бумажной версии в свет информация статьи потеряла свою былую актуальность, а идеи автора приобрели новые интерпретации. К сожалению, на данный момент научное общество, в силу своей инертности и консерватизма еще не готово перейти на полностью электронную форму научной публикации. Однако уже существует необходимость, возможность и попытки сочетания бумажной версии, обеспечивающей защиту прав автора публикации, и электронной версии, обеспечивающей доступ широкой аудитории читателей, быстроту выхода, удобство поиска информации.

Можно выделить два основных способа самостоятельного размещения научных документов в сети Интернет. Первый – это ручное размещение, предварительно подготовленных материалов, на домашней страничке автора. Однако для этого необходимо обладать соответствующими техническими знаниями и опытом подготовки документов, иметь под рукой специальный инструментарий. К примеру, скажем что, в большинстве случаев, представители гуманитарных наук не обладают ни тем, ни другим и как следствие не могут предоставить свои материалы широкой аудитории. Хочется заметить и еще один не маловажный отрицательный момент размещения материалов на домашних страничках: сложность поиска информации и как следствие более узкая аудитория читателей. Второй способ подразумевает использование специальных информационных систем, автоматизирующих процессы публикации научных документов. В качестве

примера можно привести портал ИОС ОО РФ¹ (Информационно-Образовательная Среда Открытое Образование Российской Федерации) [1]. Эта среда уже внедрилась почти во все крупные образовательные учреждения страны в виде Виртуальных Университетов или Виртуальных Представительств. Однако с точки зрения развития науки в образовании она имеет несколько весомых недостатков:

1. платная основа доступа к информации;
2. сложная процедура регистрации читателей;
3. невозможность самостоятельной публикации материалов (самоархивирование).

Основная ориентация системы на дистанционное платное образование не позволяет использовать ее в качестве среды для быстрого и широкого доступа к научным знаниям.

В соответствии с Будапештской инициативой «Открытый доступ» (Budapest Open Access Initiative) [2] только посредством самоархивирования (self-archiving) и предоставления открытого доступа (open access) (общего бесплатного и оперативного) мы можем обеспечить развитие науки. «Устранение барьеров на пути к научным публикациям ускорит прогресс в науке, обогатит образовательный процесс, уравнивает в возможностях богатых и бедных, которые смогут свободно обмениваться знаниями, и сделает научную литературу настолько полезной, насколько это вообще возможно, закладывая фундамент для объединения человечества в общем интеллектуальном общении и научном поиске» [2].

Итак, необходима единая коммуникационная научно-образовательная среда (КНОС), построенная в соответствии с принципами Будапештской инициативы «Открытый доступ».

¹ <http://www.openet.ru>

3 Коммуникационная среда

Перспективная КНОС – это единая федеральная территориально распределенная информационная система [3], состоящая из типовых независимых программно-информационных комплексов (ПИК), внедренных в центры научно-образовательных учреждений (преимущественно вуз), обеспечивающая, посредством Интернет-технологий, автоматизацию научных коммуникаций: внутри учреждения; между учреждениями, в рамках среды; и с внешними, по отношению к среде, аналогичными системами других стран.

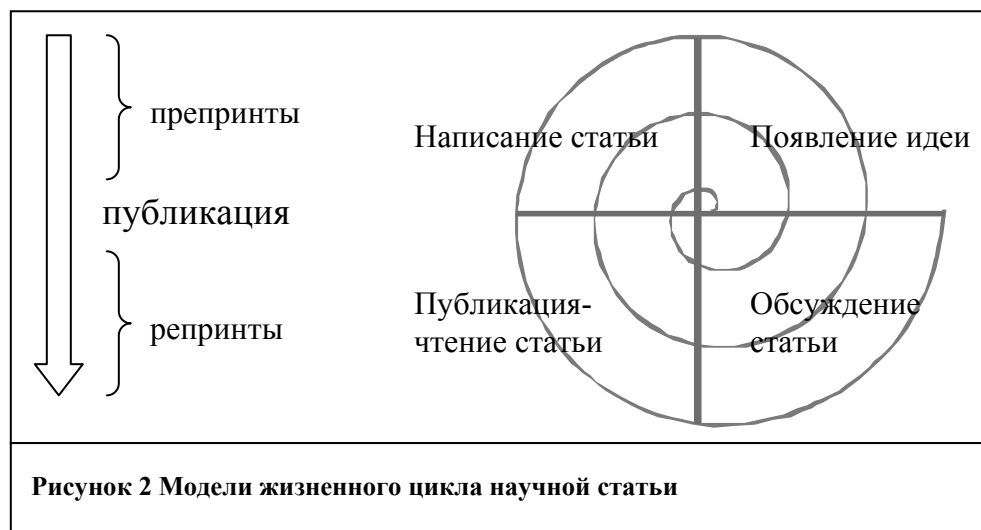
Каждый ПИК является технологически и экономически независимым и поддерживается за счет перераспределения и адаптации ресурсов, уже, как правило, существующей, инфраструктуры научных коммуникаций учреждения.

Здесь и далее под научными коммуникациями будут подразумеваться все виды обмена научными и околонучными знаниями между индивидами, независимо от формы представления (бумажная, электронная), например: выпуски сборников статей (тезисов), научных и научно-популярных журналов; проведение конференций, дискуссий (синхронных и асинхронных) и т.п.

4 Жизненный цикл научных публикаций

Основу научных коммуникаций составляют научные публикации. Рассмотрим подробнее их жизненный цикл и его основные этапы, в двух моделях (спиральной и линейной), на примере научной статьи (см. рис. 2).

На основе изучения опыта предыдущих исследований (как правило, изложенных в научных статьях), у автора будущей публикации *появляется идея*. Наступает длительный этап *написания статьи* автором, в процессе него публикация может переживать значительные изменения, корректировки, дополнения, в этот момент она уже может быть доступна некоторому кругу читателей в виде так называемых *препринтов*. По завершению



«оттачивания» содержимого, статья появляется в свет в виде законченной публикации, наступает этап публикации статьи и прочтения ее широким кругом читателей. Чтение статьи равномерно сочетается с процессом ее обсуждения между читателями и между читателями и автором. Процесс обсуждения неминуемо наводит автора на новые корректировки исходной идеи, начинают появляться репринты (или постпринты) статьи. В конечном итоге обсуждение статьи приводит к появлению конструктивно новой идеи (хотя и не обязательно у автора исходной статьи). В этот момент появляется новый виток спиральной модели жизненного цикла.

Вообще говоря, понятие препринт означает допечатную электронную версию публикации, а репринт – послепечатную. Они отличаются от оригинальной бумажной статьи только форматом представления, содержимое публикации остается неизменным. Однако в эти понятия вкладывается дополнительный смысл – содержимое репринтов и препринтов может отличаться как от самой публикации, так и между собой. Публикация непосредственно соотносится с ее бумажной версией, если же публикация чисто электронная, то понятия препринтов и репринтов можно объединить и назвать просто версиями публикации. Именно наличие различных версий публикации позволит сохранить стройность замысла первоначальной идеи, единообразную и элегантную структуру, на всей протяженности ее развития.

5 Задачи и требования

Выделим две основные *задачи* КНОС:

1. Автоматизация обмена научными документами между исследователями (учеными) работающих, как в одном учреждении, так и территориально удаленными (см. рис. 1). Только расширив круг общения ученого можно *качественно* улучшить его научные результаты.
2. Автоматизация всех этапов жизненного цикла научных документов (см. рис. 2). Только переложив рутинные работы с плеч автора на машину, дав автору публикации возможность сосредоточиться на самой сути идеи, можно сократить время между появлениями конструктивно новых научных идей, а значит *количественно* улучшить его научные результаты.

Выделим основные *требования* к проектируемой среде, с точки зрения отдельного учреждения обеспечивающего работоспособность ПИК.

Дешевизна системы – полное отсутствие необходимости покупать дополнительное программное обеспечение для установки и использования системы. Система должна быть полностью построена на бесплатных решениях.

Кроссплатформенность системы обеспечивает легкость ее установки на любую операционную систему, что позволяет без особых усилий интегрировать ее в уже существующую инфраструктуру организации.

Дешевизна использования – это минимизация расходов на обслуживание (администрирование, редакторская деятельность над публикациями и т.п.) системы, достигается за счет максимальной автоматизации всех процессов.

Легкость обслуживания системы, позволяет обеспечить быстрое внедрение системы и сэкономить средства на обучении обслуживающего персонала.

Обозначим требования к системе с точки зрения ее пользователей.

Дешевизна использования – это отсутствие необходимости покупать дополнительное программное обеспечение (ПО) для работы с системой, достигается за счет использования либо стандартного широкого используемого ПО или за счет использования бесплатного легкодоступного ПО.

Простота использования достигается за счет интеграции с общеизвестными распространенными приложениями, отсутствие необходимости сложной и длительной установки и настройки приложений.

Среди множества проблем (экономических, технических) возникающих при разработке подобной системы, наиболее сложной является организационная: проблема вовлечения пользователей и образовательных учреждений в процессы научного обмена. Только простота и дешевизна системы может склонить их к этому важному шагу.

6 Архитектура

Как уже было сказано КНОС – это распределенный набор взаимодействующих независимых типовых программно-информационных комплексов. В своей основе, каждый, из которых представляет собой, с точки зрения хранения – архив документов, с точки зрения систематизации – коллекцию документов, с точки зрения использования – портал. Однако наиболее общим определением будет электронная библиотека [4]. Итак, КНОС – это подкласс электронной библиотеки, это распределенная система с открытым доступом, объединяющая коллекции, сервисы и людей для поддержки полного жизненного цикла создания, распространения, использования и сохранения полнотекстовых научных публикаций, представленных в слабоструктурированной гетерогенной форме.

Распределенность и взаимодействие отдельных ПИК поддерживается на принципах Инициативы Открытые Архивы (Open Archives Initiative) [5]. Которая, по сути, является протоколом сбора метаданных. Когда протокол используется в качестве поисковой машины, то информация из отдельных архивов обрабатывается так, как будто бы она взята из одного архива. К настоящему моменту данный протокол используют уже более 70 архивов мира. В их число входит и один из наиболее крупных бесплатных архивов arXiv.org² содержащий более 200 тыс. документов по физике, математике и компьютерным наукам. К наиболее крупным, так же можно отнести CogPrints³ (когнитивные науки), Cryptology⁴, BioMed Central⁵. Из российских можно привести архив по экономике, социологии, политике Socionet.ru⁶.

Вливание российских вузов в среду научного обмена таких гигантских архивов позволит в значительной степени *качественно* улучшить научные результаты.

Совместная авторская разработка и контроль версий в Web (Web Distributed Authoring and Versioning) [6] – так называется стандарт, расширяющий функции HTTP 1.1, если говорить коротко, возможностями сохранения информации на сервере. В настоящее время данный стандарт бурно развивается, в качестве примера можно привести то, что его поддерживают такие продукты как: Microsoft Internet Explorer 5.0 и выше, Windows 2000 и выше, Office 2000 и выше; Adobe Photoshop 6.0 и выше; Atlovas XML Spy; и т.д. В WebDAV все файлы хранятся в абстрактной файловой системе, которая легко может быть распределена. Физически же файлы могут находиться в обычной файловой системе, СУБД или другом хранилище информации. Каждый ресурс может иметь неограниченный набор атрибутов, в нашем случае набор атрибутов Дублинское ядро (Dublin Core –

² <http://www.arxiv.org>

³ <http://cogprints.ecs.soton.ac.uk>

⁴ <http://wprint.iacr.org>

⁵ <http://www.biomedcentral.com>

⁶ <http://www.socionet.ru>

DC⁷) сочетающийся с Инициативой Открытые Архивы. С помощью расширения DASL (DAV Searching & Locating) стандартизован поиск по этим атрибутам. Для организации совместной авторской разработки реализуется механизм блокировок. Расширение DeltaV позволяет хранить несколько версий файла, разветвлять дерево ревизий и ставить метки. С помощью спецификации ACL (Access Control Extension) возможно с легкостью реализовать гибкую систему управления правами доступа.

Использование этого протокола в системе позволит в значительной степени упростить техническую сторону процесса создания научного документа, а значит ускорит процесс их написания и как следствие *количественно* улучшить научные результаты.

7 Модули

В системе необходимо наличие множества дополнительных модулей (сервисов) обеспечивающих решение двух основных задач системы. Это могут быть и различные электронные дискуссии (синхронные и асинхронные) между пользователями системы, модули, обеспечивающие персонализацию (ручную и автоматическую) системы, модуль классификации публикаций на основе УДК (Универсальная Десятичная Классификация) и т.д. Рассмотрим два наиболее важных сервиса.

Модуль преобразования форматов и отображения математических формул. С точки зрения использования системы, форматы подразделяются на *входные* (авторские) и *выходные* (читательские). Входные форматы являются либо *общеиспользуемыми* (Word, TeX, HTML), либо *открытыми* (DocBook, TEI), основанными на XML. Читатель будет иметь возможность получить публикации в зависимости от его требований HTML, DjVu, LuraDocument для ознакомления; PDF, PS для печати; WML для беспроводных устройств; eBook (OpenBook) для чтения в электронных книгах; подмножество XML для обмена; либо исходные Word, TeX для дальнейшей работы. При этом

⁷ <http://purl.org/dc/>

получаемый документ, при необходимости, может быть адаптирован под физические возможности читателя или программные и аппаратные возможности его системы. Способ же хранения публикаций в самой системе слабоструктурированный и гетерогенный, т.е. публикации хранятся в различных исходных авторских вариантах, подчас имеющих не до конца проработанную структуру.

При отображении публикаций в браузерах для быстрого просмотра система должна обеспечивать способ отображения мат. формул без потери качества при масштабировании и без потери информации о содержимом формулы для ее дальнейшего использования. Для этого уже существует готовое решение с несколькими способами отображения на основе MathML.

Модуль построения метрик научных публикаций. Количество публикаций в мире растет с каждым днем и в месте с ним растет и количество так называемого «научного шума». Необходимы возможности для более или менее корректного получения информации о важности статьи и ее рейтинге и других числовых параметрах. Рейтинг статьи (публикации) может оцениваться как по *внешним* параметрам: кол-во прочтений статьи, кол-во реплик в форуме ассоциированным со статьей, учитывая рейтинг их авторов, общее кол-во статей автора данной статьи и т.д.; так и по *внутренним*: кол-во ссылок данной статьи на другие, учитывая их рейтинг, кол-во ссылок на эту статью и т.д.

Построение метрик статьи в открытом архиве должно происходить в полностью автоматическом режиме, либо в автоматизированном при помощи только автора публикации. На данный момент уже существуют разработки по автоматическому извлечению библиографических данных статьи и построению взаимосвязей с ними [7].

8 Заключение

Назревшая потребность автоматизации научных коммуникаций не реализована в сфере высшего профессионального образования. Не

существует простой и дешевой информационной системы доступной для образовательных учреждений основанной на принципах самоархивирования и открытого доступа. В вузах для решения данной проблемы порой используются подручные не адекватные средства.

К настоящему моменту существуют отдельные стандарты, технологии и инструменты, с помощью которых возможно построить систему отвечающую предъявляемым требованиям. Однако их необходимо связать воедино, адаптировать к использованию в среде русскоязычных ресурсов, к использованию в российских образовательных учреждениях, необходимо обеспечить интеграцию с мировыми ресурсами. Необходимо выработать концепцию и разработать проект построения коммуникационной научно-образовательной среды.

Автоматизация процессов создания научного знания и обмена им в образовательных учреждениях не только качественно и количественно улучшит состояние российской науки, но и приведет к положительным результатам в образовательной сфере путем приобщения к процессам формального научного обмена молодого перспективного поколения.

9 Литература

1. С.Л. Лобачев, А.А. Поляков, В.И. Солдаткин. Концепция информационно-образовательной среды открытого образования Российской Федерации
(<http://www.informika.ru/text/exhibit/portal/consept.zip>)
2. Будапештская Инициатива «Открытый доступ»
(<http://www.soros.org/openaccess/ru/read.shtml>)
3. А.Н. Гребнев. Научные информационные системы // Вестник УдГУ: Математика.- Ижевск: Изд-во УдГУ, 2003. С.99-106.

4. М.Р. Когаловский. Научные коллекции информационных ресурсов в электронных библиотеках. Институт проблем рынка РАН. 1999. (<http://www.dl99.nw.ru/PDF/02.pdf>)
5. Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0 (<http://www.openarchives.org/OAI/openarchivesprotocol.html>)
6. HTTP Extensions for Distributed Authoring WEBDAV RFC 2518 (<http://asg.web.cmu.edu/rfc/rfc2518.html>)
7. Donna Bergmark, Paradee Phenmpoonpanich, Shumin Zhao. Scraping the ACM Digital Library. (<http://www.cs.cornell.edu/cdlrg/ReferenceLinking/ScrapingACM.pdf>)

A.N. Grebnev

This article considers foundations and standards of developing geographically distributed information system with open access, based on self-archiving, which automate processes of scholarly knowledge's exchange (mainly publications) and life circle phases of science publications in educational institutions.

Гребнев А.Н. Коммуникационная научно-образовательная среда// Вестник УдГУ: Математика.- Ижевск: Изд-во УдГУ, 2004.- С. 29-38.

Гребнев Андрей Николаевич

Удмуртский Государственный Университет,

Кафедра Математического Обеспечения ЭВМ,

Ижевск, ул. Университетская, 1 (корп. 4)

426034, Россия

E-mail: ag@blandware.com

<http://fox.blandware.com>