

УДК 021(045)+655.4(045)

А. Н. Гребнев

НАУЧНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

В данной статье предложено определение научной информационной системы (НИС) в аспекте научной коммуникации. Определены задача, субъекты, объекты и основные функции НИС. Дано ее объяснение в терминах электронной библиотеки, издательства и консалтинга. Описаны наиболее актуальные логические блоки (подсистемы).

Ключевые слова: научные информационные системы, электронные библиотеки, электронные издательства, электронные научные публикации.

1 Введение

Стремительно развивающаяся научно-техническая революция стала основой глобального процесса информатизации всех сфер жизни общества. От уровня информационно-технологического развития и его темпов зависят состояние экономики, качество жизни людей, национальная безопасность, роль в мировом сообществе.

Развитие современных технологий зависит, в первую очередь, от интеллектуального потенциала общества и, следовательно, от уровня науки в стране. В научном сообществе важнейшую роль играют качество и темпы информационного обмена. Основу формальной научной коммуникации составляет научная публикация. «Публикация выступает как первичный источник сведений о научном знании, отношениях между учеными, строении и динамике научных объединений и т.п. Для науковеда, философа, логика, методолога, специалиста по информатике, социолога науки той конечной реальностью, из которой исследователь черпает свои представления о науке, выступают публикации. Отличающиеся друг от друга изображения науки в различных исследовательских традициях ... становятся объектами изучения

лишь постольку, поскольку сведения о них имеются в научной публикации» [2].

2 Научные публикации сегодня

Обычные бумажные публикации (издания), будь то журналы, сборники статей, не успевают за темпами научно-технического прогресса [13], [14]. Они уже не могут обеспечить требуемого качества и скорости информационного обмена. В обществе наметились тенденции частичного перехода (а в будущем возможно и полного) с бумажных публикаций на электронные. Но почему же это все еще не произошло? Почему огромное количество возможностей электронных публикаций (быстрота выхода, широта охватываемой аудитории, дешевизна, расширенные возможности поиска и извлечения информации и др.) пока не смогли обеспечить главенство над бумажными публикациями?

В настоящее время в российском Интернет научные публикации (статьи) можно найти в трех основных местах:

1. На домашних страничках самих авторов, в виде «как есть» (как набрали, так и положили) без какого либо поиска и вообще чего бы там ни было.
2. На сайтах конференций. Это, как правило, доклады их участников. В данном случае, иногда присутствует классификация и поиск.
3. На сайтах электронных журналов или специализированных тематических сайтах. Здесь в редких случаях существует несколько форм для отображения (например, версия для печати), классификация и поиск.

Основную массу этих статей составляют препринты (тезисы, программы) или репринты – перепечатка, полная или частичная, уже изданного в бумажном виде. В случае если статья все же публикуется в чисто

электронном виде то это, как правило, не более чем хобби, иногда самореклама.

Причин недостаточной развитости электронных публикаций конечно несколько, но главная из них состоит в том, что сам процесс электронного издания не поставлен на формальную научную основу. Нет единой системы для осуществления публикаций, есть лишь отдельные технологии и инструменты, которые либо не доступны, либо для работы с ними не хватает знаний. Второй главной проблемой является отсутствие, какой либо защиты авторских прав.

Нужна простая в использовании, но в тоже время мощная, основанная на прогрессивных технологиях, система, защищающая права авторов, охватывающая широкий круг читателей. Нужна единая научная информационная система.

3 Основы научной информационной системы

Научная информационная система (НИС) – система, автоматизирующая процессы научного информационного обмена¹. Существуют попытки альтернативного определения данного термина связанные с англоязычным обозначением подобных систем CRIS (Current Research Information System)[1].

Главной *задачей* НИС является ускорение темпов и улучшение качества информационного обмена в научном сообществе.

Обозначим основные категории *субъектов* (пользователей) с которыми взаимодействует НИС:

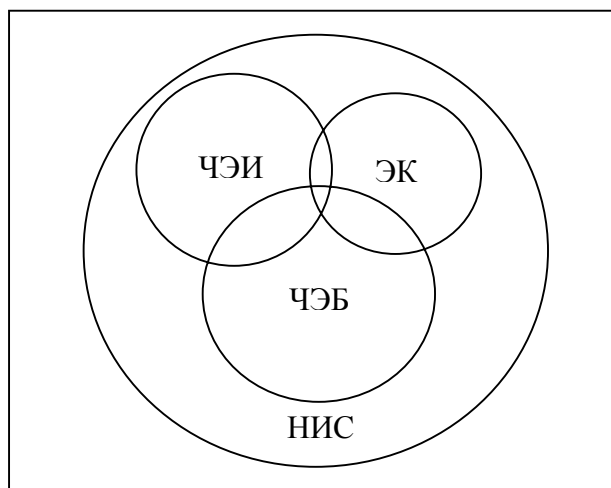
1. научные работники (ученые, исследователи, аналитики);
2. работники образования (преподаватели, учителя, студенты);
3. работники управления (руководители организаций, учреждений, отделов, проектов);

¹ Предложенное определение не претендует на полноту и рассматривает НИС только с точки зрения информационного обмена.

4. инвесторы (финансирующие организации и физические лица);
5. прочая публика (журналисты, работники статистики, обозреватели).

Определим основные *функции* НИС, используя уже знакомую устоявшуюся терминологию. Итак, НИС – это:

1. чисто электронная библиотека (ЧЭБ)²;
2. чисто электронное издательство (ЧЭИ);
3. электронный консалтинг (дискуссия) (ЭК).



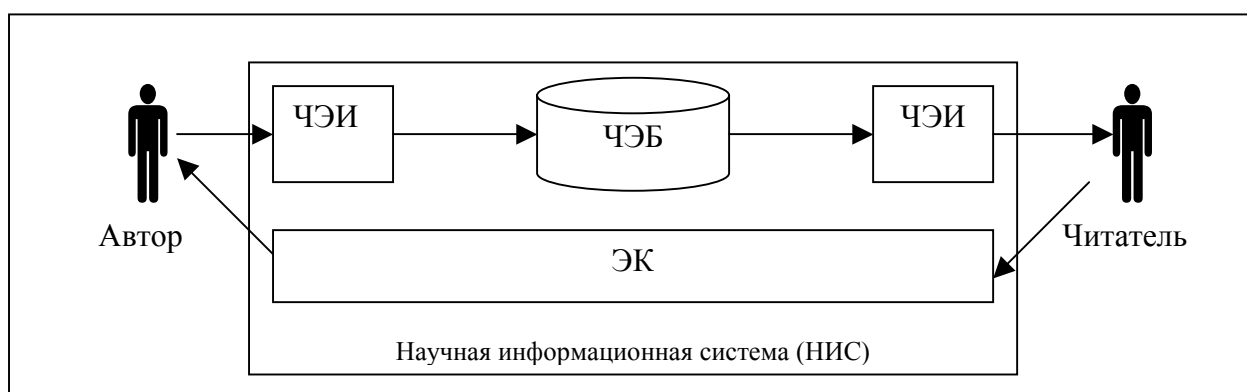
ЧЭБ является основой НИС. Термин «электронная библиотека» (ЭБ) возник в 90-е годы на основе англоязычного термина «цифровые библиотеки» (digital library [3]). «ЭБ – это система, обеспечивающая сообществу пользователей доступ понятным для них образом к большим репозиториям мультимедийной информации и знаний, организованным при отсутствии каких либо сведений о способах их применения [4]». На данный момент уже проведено большое количество исследований в области ЭБ [5,6]. Понятие ЭБ возникло вследствие эволюции обычных библиотек, сначала простой автоматизации библиотечных процессов компьютерными технологиями, а затем вследствие постепенного перехода на работу с

² В данном случае и в дальнейшем, определяя термины, будет подразумеваться научный контекст. Например, под чисто электронной библиотекой подразумевается чисто электронная научная библиотека.

электронными документами (ЭД). ЭД – документ, носителем которого является электронная среда. Электронная библиотека выполняет следующие *функции*: сбора, хранения, классификации, поиска, выдачи информации и метаинформации. Заметим что сбор и выдачу уже готовой информации.

По аналогии с электронной библиотекой существует понятие электронное издательство (ЭИ) (E-Publishing House) [7]. Как и в случае с библиотекой, это понятие появилось в процессе эволюции издательств. Если обычное издательство выполняло только одну главную функцию – подготовку печатного издания, то в идеологии НИС ЭИ выполняет *функцию* подготовки документа для помещения ее в репозиторий библиотеки, а так же извлечения ее из репозитория и представление в требуемой форме.

Приписка «чисто» в понятиях ЧЭБ и ЧЭИ подчеркивает основную



ориентированность на работу с электронными документами.

Мы подошли к ключевому понятию НИС – электронной научной публикации (ЭНП) (или просто ЭП). ЭНП – это электронный документ, смыслом которого является научная публикация, подготовленный электронным издательством и размещенный в электронной библиотеке. Теперь мы можем определить основной *объект* работы НИС – это электронная научная публикация.

Электронный консалтинг (E-Consulting) (ЭК) – в рамках идеологии НИС, осуществляет *функции* обратной связи читателя с автором, чего так не хватает в бумажных публикациях. Главным элементом ЭК является форум

для общения. Известно, что обсуждение идей часто приводит к рождению новых. Наведение конструктивной критики, получение откликов и рецензий обеспечивает значимое улучшение результатов работ.

Только в комплексе все эти три отдельно развивающиеся технологии (ЭБ, ЭИ, ЭК) позволят создать НИС, обеспечивающую основу жизненного цикла электронной научной публикации.

4 Подсистемы НИС

Построение НИС значительно облегчается за счет использования технологий ЭБ, ЭИ и ЭК. В рамках данной статьи, не представляется возможным даже обозначить все необходимые подсистемы. Рассмотрим лишь наиболее актуальные.

Адаптивное представление. Все электронные публикации это некоторые файлы, представленные в необходимом формате. 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 ЭП 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 получаемый документ может быть адаптирован под физические возможности читателя или программные и аппаратные возможности его системы.

Достоверность, важность информации. Существует понятие рейтинга журнала, так называемый Impact Factor основанный на частоте цитирования его статей [8]. В рамках системы НИС можно ввести два понятия *рейтинг статьи* (РС) и *рейтинг автора* (РА). РА оценивается как суммарный рейтинг всех его статей. РС это кол-во ссылок на эту статью, рейтинг ссылающихся статей и их авторов, кол-во реплик в форуме ассоциированных

с данной статьей и рейтинг авторов написавших эти реплики. В данный момент разрабатывается система подсчета, основанная на множестве этих и других различных факторов, которая позволит более качественно оценить важность и достоверность научной публикации, тем самым, предоставляя читателю возможность выборочного чтения наиболее рейтинговой информации.

Защита интеллектуальной собственности. Соблюдение прав автора самый краеугольный камень электронных публикаций. Однако уже все механизмы и инструменты их обеспечения уже разработаны [9]. Это *трастовые системы управления цифровыми правами* обеспечивающие следование, пользователями системы, некоторых правил. *Расширяемый язык разметки прав XrML* используемый для описания прав трастовых систем. *Идентификатор цифрового объекта (DOI)* – уникальный не изменяющийся идентификатор произведения, обеспечивающий связь между пользователем произведения и обладателем прав на это произведение, присвоенный специально созданным международным фондом DOI. *Самозащищающийся документ* – активный документ, сохраняющий свою конфиденциальность и целостность, принуждающий пользователей следовать правам, связанным с этим документом. *Цифровой водяной знак* позволяющий подтвердить право собственности на цифровое произведение. Адаптация всех этих механизмов к НИС позволит решить техническую сторону проблемы защиты прав автора.

Извлечение знаний. Объем накопленных человечеством знаний огромен. Нередко время для изучения всех материалов, даже по интересующей тематике, ограничено временем человеческой жизни. Нужна хотя бы частичная автоматизация извлечения знаний. Существуют реализованные разработки ассоциативно-статистического подхода анализа текста [11], основанного на результатах нейропсихологических исследований [10], которые установили, что анализ печатного текста, опираясь на зрительное пространственное (а не на линейное слуховое) восприятие, реализуется преимущественно правым полушарием мозга, использующим ассоциативную

статистическую модель. В основе подхода лежит интегральное представление смысла текста в форме ассоциативной семантической сети, в качестве критерия построения которой используется частота совместной встречаемости понятий в предложениях текста. Читатель, исследуя такую сеть, представленную в виде тематического дерева (дерево ключевых и связанных с ними понятий текста), может в значительной степени ускорить процесс исследования текста и выявления требуемой информации. Семантическая сеть так же используется для решения таких задач, как автоматическое реферирование, тематическая классификация и кластеризация текстов, смысловой поиск, автоматическое построение гипертекста и т.д. Существуют уже готовые решения, которые можно легко встроить и использовать в НИС.

К сожалению, все перечисленные подсистемы это пока только отдельно развивающиеся технологии, которые необходимо адаптировать и связать воедино для практического применения в НИС.

5 Заключение

Осознание необходимости построения системы, подобной НИС, пришло давно. До сего времени шел процесс накопления и оттачивания знаний. Наступил момент нового витка в эволюции процессов информационного обмена в научном сообществе.

Не так давно разработанные стандарты платформы XML [12] и реализованные на их основе инструменты, подвели мощнейшую базу для построения интероперабельных открытых систем. Только в 2001-2002 годах, были разработаны реально применимые инструменты для извлечения знаний. К этому же времени были практически отработаны механизмы защиты интеллектуальной собственности, являющимися ключевыми для построения реально используемой НИС. Только 10 января 2002 года был принят

федеральный закон РФ «Об электронной цифровой подписи», являющийся основополагающим в механизмах защиты авторского права.

К настоящему моменту существуют и отработаны все компоненты необходимые для построения качественно новой научной информационной системы.

Множество подобных систем распределенных по центрам научной коммуникации (университеты, институты, академии) позволят создать единую среду научного информационного обмена.

6 Литература

1. А.С. Лопатенко. Современные научные информационные системы. Перспективы использования. 2001. (http://derpi.tuwien.ac.at/~andrei/CRIS_DOC.htm)
2. Э.М. Мирский. Массив публикаций и система научной дисциплины. "Системные исследования", 1977 г.
3. S.M. Griffin NSF/DARPA/NASA Digital Libraries Initiative. D-Lib Magazine, July/August 1998.
4. Report on Workshop "Interoperability, Scaling, and Digital Library Research Agenda", 1995. (<http://www.ccic.gov/pubs/iita-dlw/>)
5. М.Р. Когаловский. Научные коллекции информационных ресурсов в электронных библиотеках. Институт проблем рынка РАН. 1999. (<http://www.dl99.nw.ru/PDF/02.pdf>)
6. С.А. Арнаутов. Роль и место виртуальных цифровых библиотек в Интернете. Институт химической физики им. Н.Н. Семенова РАН. 2001. (http://rcdl2001.krc.karelia.ru/papers/papers/arnautov/arnautov_paper.rtf)
7. П.А. Дмитриев. Проектирование комплексных систем поддержки Электронных Изданий. Вычислительный центр РАН. Москва. 2001. (http://rcdl2001.krc.karelia.ru/papers/papers/dmitriev/complete_text.doc)

8. М.В. Алфимов, А.Н. Либкинд, И.А. Либкинд, В.А. Минин. Информационные потоки в РФФИ: Новый подход к цитированию. (http://intra.rfbr.ru/pub/vestnik/V4_01/1_1.htm)
9. Правовая охрана и защита цифровой интеллектуальной собственности. Институт Открытое общество. (<http://www.intellect.vsu.ru/>)
10. Н.Н. Брагина, Т.А. Доброхотова. Функциональные асимметрии человека. – М.: Медицина, 1981. – 287 с.
11. А.Е. Ермаков, В.В. Плешко. Ассоциативная модель смысла текста в прикладных задачах компьютерного анализа полнотекстовых документов. // Русский язык: исторические судьбы и современность. Международный конгресс. Труды и материалы. – Москва, МГУ, 2001. (http://research.metric.ru/art_ling.asp)
12. М.Р. Когаловский. Стандарты платформы XML и базы данных. Институт проблем рынка РАН. 2001. (http://rcdl2001.krc.karelia.ru/papers/papers/kogalovsky/kogalovsky_paper.rtf)
13. Гребнев А.Н. Система электронных публикаций научных статей //Материалы XL Международной научной студенческой конференции "Студент и научно-технический прогресс": Информационные технологии/Новосиб. гос. ун-т. Новосибирск, 2002.- С. 32-33.
14. Маслов С.Г., Моченов С.В. Методология создания электронных учебных изданий //Материалы XXIII научно-метод. конф. ИжГТУ (19-22 февраля).- Ижевск: Изд-во ИжГТУ, 2001.- С. 29-31.

A.N. Grebnev

This article offers the definition of science information system (SIS) in aspect of science communication. It defines goal, subjects, objects and main functions of SIS. It gives explanation in terms of digital library, publishing house and consulting. It describes more topical logical blocks (subsystems).

Гребнев А.Н. Научные информационные системы // Вестник УдГУ: Математика.- Ижевск: Изд-во УдГУ, 2003. С.99-106.

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>