

В. Г. Шамаев, А. В. Жаров

Электронное издание “Итоги науки и техники” ВИНТИ

Технология, разработанная для создания Русскоязычной БД и Цифровой библиотеки позволила подготовить новый электронный информационный продукт — электронные научно-технические аналитические обзоры. В настоящее время на DVD-дисках выпущены электронные издания “Итоги науки и техники” ВИНТИ по астрономии, физике, математике, выпуски Смирновского сборника МГУ.

Информационное обеспечение науки издавна базировалось не только на проведении конференций, симпозиумов, семинаров и т. п., но и в огромной степени на деятельности научных библиотек. В настоящее время к ним присоединяются электронные библиотеки. В виде локальных или распределенных информационных систем они позволяют накапливать, надежно сохранять и самое главное эффективно использовать разнообразные электронные документы. Эти библиотеки доступны для читателя (пользователя) и через глобальные сети передачи данных. Наполнение таких библиотек — чрезвычайно трудоемкая задача, связанная как с разработкой технологий для подготовки новых электронных продуктов, так и с рутинной работой по переводу в цифровую форму накопленных бумажных массивов.

Начиная с конца 1980-х гг. правительства некоторых государств, а также частные фирмы поддерживали в финансовом плане ряд проектов по оцифровке книг и газет. Всего по этим проектам было отсканировано несколько сотен тысяч книг. В качестве одного из наиболее амбициозных проектов можно назвать “Проект миллиона книг (Million Book Project)” библиотек университета Carnegie Mellon [1]. На конец 2005 г. в рамках этого проекта отсканировано свыше 600 тыс. книг. Большой интерес вызывает начавшийся проект под названием “Google Print Library Project” [2]. Этот проект, возникший в результате сотрудничества известной поисковой системы с пятью крупнейшими библиотеками США — в Мичиганском, Гарвардском, Стэнфордском, Оксфордском университетах, публичной библиотекой Нью-Йорка, — включает десятки миллионов книг. Такие проекты открывают широкий доступ к печатным коллекциям крупных библиотек.

Сканирование и последующее распознавание текстов книг — это только первый этап, хотя и обязательный. Необходимо также создание аннотаций, указателей, ссылок и т. д., т. е. необходима база данных, которая позволит в дальнейшем использовать заложенную, но трудно выявляемую информацию, например, индекс цитирования в научных статьях, ссылки на химические или математические формулы, составление справочников, словарей и тезаурусов и т. д., а также генерировать новую информацию, получаемую в результате сопоставления и поиска в политематической базе данных. Важно также иметь в виду, что наличие изображений, а не просто сканированного текста, как в известных интернет-библиотеках типа lib.ru, lib.aldebaran.ru и др., позволяет сохранить больше информации, чем кажется на первый взгляд. Ведь

на протяжении столетий существования печати издательские и типографские сообщества разработали стандарты, а точнее традиции представления текста на странице, которые способствуют эффективному чтению и усвояемости (деление страниц на текстовые блоки, их размещение на странице и по отношению к иллюстративному материалу, выделение зон текста, наличие колонок, шрифтовые различия и т. п.). Нельзя выкидывать эти элементы оформления, концентрируясь лишь на переводе печатных изображений в текстовую форму в кодировке ASCII, LaTeX, Unicode и др. Если не учитывать эту информацию, то сильно обедняется впечатление от источника. Недаром, для многих гораздо приятнее читать печатную книгу, чем ее суррогат на компьютерном экране. Учет типографских традиций, несомненно, помогает получению дополнительных знаний. Поэтому подготовка полноценных оцифрованных изданий является непростой задачей и проблема получения таких продуктов имеет неоднозначное решение.

Поражаясь отмеченным выше уникальным по своим масштабам проектам, отметим, что особое место в информационном обеспечении научных исследований, да и не только в них, занимает информационная деятельность по составлению обзоров. Вот на это и направлен проект под названием «Электронное издание “Итоги науки и техники”» [3]. По своему значению он, конечно, гораздо шире, чем просто перевод всех серий “Итогов науки и техники” (ИНТ) ВИНТИ или других аналитических обзоров в цифровую форму. Смысл его — сделать издание доступным читателю (пользователю, если говорить на компьютерном жаргоне) в полном объеме, а не только его текстовой части. В книговедении есть подходящий термин — факсимильное издание, т. е. повторение не только содержания, но и вида издания.

Аналитические обзоры, в которых описываются последние достижения (обычно не более чем за 5 последних лет), содержащие критическое осмысление анализируемой информации, чрезвычайно полезны для исследователей. Составить такой обзор может только хорошо знакомый с данной отраслью знания человек, сам являющийся активным научным работником. Его деятельность по написанию обзора — такая же научная работа, как и написание монографии, аккумулирующей результаты многолетних исследований. Нередко такие научные обзоры имеют большое значение для инженеров и других потребителей (аспирантов, студентов, чиновников и т. д.), занимающихся решением практических вопросов, которые связаны с получением и применением современных знаний.

**Количество вышедших томов “Итогов науки и техники”
по физико-математическим наукам за период 1957–1992 гг.**

Физико-математические науки	Годы						Всего
	1957–1964	1965–1969	1970–1974	1975–1979	1980–1984	1985–1992	
Астрономия, Геодезия, Исследование космического пространства	—	9	14	20	25	47	115
Математика	2	15	19	31	32	142	241
Физика	5	7	—	—	10	45	67
Всего:	7	31	33	51	67	234	423

Такие аналитические обзоры под названием “Итоги науки и техники” начали выходить в ВИНТИ с 1957 г. До 1992 г. ВИНТИ ежегодно подготавливал и издавал до 100 томов этих обзоров [4]. В таблице по годам представлено количество томов ИНТ в период 1957–1992 гг. по физико-математическим наукам.*

Мы приводим лишь эту статистику, так как именно эти Итоги и были на первом этапе предметом нашего внимания при разработке технологии подготовки электронного издания ИНТ. Наряду с созданием баз данных и электронных журналов по физико-математическим наукам [5] именно эти издания до последнего времени трудно поддавались переводу в электронную форму из-за обилия сложных математических формул и символов.

Важность этой работы заключалась еще и в том, что в настоящее время только в небольшом количестве крупнейших научных библиотек России хранятся ретромассивы бумажных изданий “Итогов науки и техники” ВИНТИ. Спрос читателей на этот вид литературы значителен и часто превосходит спрос на другие специальные издания. ИНТ по естественным и техническим наукам прежних лет не теряют своей актуальности до настоящего времени. Поэтому большинство специалистов, особенно в начале работы над новой тематикой, обращаются, прежде всего, к вторичной информации, которая кратко по содержанию и достаточно полно по числу цитированной литературы представлена в ИНТ. Кроме того, следует принять во внимание, что в ИНТ ВИНТИ наиболее полно представлены работы советских ученых, составляющие “золотой фонд” мировой науки. Достаточно вспомнить о крупных научных школах в области физики, астрономии, химии, математики, механики, геофизики и т. д. Большинство работ ученых СССР 1950–1980 гг. было опубликовано на русском языке и поэтому слабо представлено за рубежом.

На первый взгляд, эта задача по оцифровке кажется достаточно простой, так как имеются печатные сборники “Итогов науки и техники”. Более того, коллектив, работающий с электронными ИНТ, занимается и выпуском текущих номеров РЖ и ИНТ и при их подготовке, конечно, использует и разработанные им же средства подготовки баз данных реферативной информации. Однако очевидные рассуждения не всегда бывают правильными или не совсем правильными. Например, и это хорошо знают люди, интересовавшиеся даже на любительском уровне астрономией, в 1940–

1950-е гг. возникла и некоторое время просуществовала такая область знания, как астроботаника. Создателям астроботаники и в первую очередь чл.-корр. АН СССР Г. А. Тихову, да и многим астрономам казалось вполне ясным, что у этой науки есть не только вполне определенный предмет изучения — растительные формы жизни во Вселенной, но и конкретные реальные объекты исследования. В те годы многие специалисты были абсолютно уверены в существовании обширных растительных покровов на поверхности Марса. Но, с запуском космических кораблей к Марсу, оказалось, что растений на Марсе нет, и астроботаника Солнечной системы прекратила существование.

Однако отметим, что астроботаника обогатила наши знания новыми методами наблюдений, поставила интересные проблемы в области земной биологии и, тем самым, сыграла положительную роль в истории науки и, в частности, позволила на своем опыте заложить основы современной экзобиологии.

Сейчас на основе современного уровня развития информационных технологий мы видим слабые стороны печатной технологии подготовки информационных изданий и осознаем, что электронная версия, как уже указывалось ранее, это не только отсканированный текст. Бытующее среди непрофессионалов мнение, что информационные технологии в недалеком будущем позволят чуть ли не автоматизировать поиск решений научных задач — результат рекламных акций, не вполне адекватных состоянию информатики как науки. История с астроботаникой и подобные ей случаи призывают нас быть осторожными и побуждают к тщательной проработке как технологических, так и организационных решений. Поэтому мы ограничились первоначальным решением лишь нескольких задач, среди которых создание Русскоязычной базы данных (РубД) и связанной с ней Электронной библиотеки и подготовка выпуска их продуктов (сигнальная информация в электронном виде, электронные ИНТ и др.) [3, 6].

Технология создания РубД и цифровой библиотеки позволила подготовить новый вид электронного информационного издания — всех серий “Итогов науки и техники” ВИНТИ на DVD-носителях [7, 8]. В настоящее время на DVD-дисках выпущены ИНТ по физико-математическим наукам и разработана технология сжатия информации с тем, чтобы можно было использовать и CD-носители [9].

Технология подготовки электронного издания ИНТ включает следующие этапы [10]:

* В дальнейшем выпуск ИНТ в ВИНТИ носил эпизодический характер.

- углубленная разметка каждого тома серии, включая титульные страницы, страницы с опечатками и аннотациями и т. п.;

- ввод библиографических описаний и аннотаций в Русскоязычную базу данных;

- постраничное сканирование источника;

- сборка отдельных изображений страниц в pdf-файлы, соответствующие статьям издания;

- формирование с помощью программы Itogi файлов web-интерфейса;

- подготовка DVD или CD-дисков электронного издания.

Необходимо было также реализовать возможность постоянного пополнения базы данных новыми выпусками серий, а также формирования указателей по всем занесенным в базу данных томам соответствующей серии.

Разработанный web-интерфейс представляет собой систему указателей, созданных в виде html-файлов, обеспечивающих доступ к файлам полнотекстовых изображений выпусков через Интернет и переносимость электронной версии на отдельно стоящую рабочую станцию или на DVD(CD)-носители.

Для генерации файлов web-интерфейса издания используются следующие программные средства:

- программа Itogi, предназначенная для создания html-файлов интерфейса издания;

- программа DisAssembler, предназначенная для подготовки pdf-файлов для помещения на web-сервер в соответствии со специальным идентификатором документа;

- утилита SID2 Editor, предназначенная для проверки и исправления соответствия записи издания и порядка нумерации страниц томов ИИТ.

Основные результаты выполненной работы:

- определены форматы и интерфейсы информационно-поисковой системы "Электронное издание научно-технических аналитических обзоров". База данных реализована на основе СУБД MS SQL Server 2000;

- разработана технология подготовки электронных изданий научно-технических обзоров;

- осуществлен пилотный проект издания "Итоги науки и техники" ВИНТИ в виде демонстрационной версии в Интернете;

- изготовлены DVD или CD-диски с электронными изданиями "Итоги науки и техники" ВИНТИ по астрономии, физике, математике и Смирновского сборника МГУ.

В заключение отметим, что предыдущий проект под названием "Электронный реферативный журнал" [5] неожиданно для авторов получил полноценную жизнь в Интернете на сайте Воронежского университета [11]. Надеемся, что и представляемый проект, а также его продукты послужат

реальной, а не только декларируемой, доступности электронных продуктов широкому кругу научных работников, студентов и аспирантов, специалистов пуждающимся в них.

СПИСОК ЛИТЕРАТУРЫ

1. Denise Troll. Carnegie Mellon University Libraries. Frequently Asked Questions about the Million Book Project. March 20, 2006. <http://www.library.cmu.edu/Libraries/MBP-FAQ.html>.

2. Google Print Library Project. <http://books.google.com>.

3. Шамаев В. Г., Жаров А. В., Горшков А. Б. База данных и Электронная библиотека русскоязычной литературы по физико-математическим наукам // ИИТ. Сер. 1.— 2006.— № 7.— С. 6–13.

4. Черный А. И. Всероссийский институт научной и технической информации.— М.: ВИНТИ, 2005.— 316 с.

5. Шамаев В. Г., Жаров А. В. Электронный реферативный журнал ВИНТИ по физико-математическим наукам // ИИТ. Сер. 1.— 2006.— № 3.— С. 15–25.

6. Шамаев В. Г., Жаров А. В., Батурина О. Н., Горшков А. Б., Лось Е. К., Максимов И. Н., Старцева О. Б. Технология подготовки базы данных по русскоязычным источникам и Электронная библиотека русскоязычной литературы по физико-математическим наукам. <http://www.benran.ru>.

7. Шамаев В. Г., Жаров А. В. Электронное издание "Итоги науки и техники" ВИНТИ по физико-математическим наукам. ВИНТИ.— М., 2005. 8 с.— Рус.— Деп. в ВИНТИ 24.05.2005 № 738-B2005.

8. Шамаев В. Г., Жаров А. В., Батурина О. Н., Горшков А. Б., Максимов И. Н. Описание технологии подготовки электронного издания "Итоги науки и техники" ВИНТИ по физико-математическим наукам. ВИНТИ.— М., 2005. 87 с.— Рус.— Деп. в ВИНТИ 24.05.2005 № 740-B2005.

9. Кувылин Ю. И., Шамаев В. Г. Электронное издание "Итоги науки и техники" ВИНТИ по физико-математическим наукам. 2. Дополнительные аспекты технологии подготовки. ВИНТИ.— М., 2006. 27 с. Библиогр. 4 назв.— Рус.— Деп. в ВИНТИ 26.09.2006, № 1177-B2006.

10. Шамаев В. Г., Жаров А. В., Кувшинова Е. Е., Старцева О. Б. Проект "Электронное издание научно-технических аналитических обзоров". ВИНТИ.— М., 2006. 18 с. Библиогр. 9 назв., Ил. 10.— Рус.— Деп. в ВИНТИ 18.10.2006, № 1246-B2006.

11. Зональная научная библиотека ВГУ. <http://www.lib.vsu.ru/resources/math.phtml>.

Материал принят к опубликованию 05.12.06.