

Электронные библиотеки и проблемы интеграции информационных ресурсов в гетерогенном сетевом пространстве

E-Libraries and Integration of Information Resources within a Heterogeneous Network Environment

Електронні бібліотеки та проблеми інтеграції інформаційних ресурсів у гетерогенному мережевому просторі

О. В. Сунтюренко, Л. Ф. Борисова

Всероссийский институт научной и технической информации Российской Академии наук (ВИНИТИ РАН), Москва, Россия

Oleg Syuntyurenko and Lyudmila Borisova

All-Russian Institute of Scientific and Technical Information (VINITI RAN), Russian Academy of Sciences, Moscow, Russia

О. В. Сунтюренко, Л. Ф. Борисова

Всеросійський інститут наукової та технічної інформації Російської Академії наук (ВІНІТИ РАН), Москва, Росія

Дан анализ проблем и тенденций развития электронных ИР. Показаны сопоставительные данные по контенту национальных сегментов Интернета. Рассматриваются вопросы развития электронных библиотек, а также проблемы интеграции разнородных ИР в сетевом пространстве.

The paper considers the challenges and development trends of digital information resources and presents comparative data on the content of the national segments of the Internet. It also deals with the development of e-libraries and integration of various information resources within a network environment.

Подано аналіз проблем і тенденцій розвитку електронних ІР. Зіставлені показники щодо контенту національних сегментів мережі Інтернет. Розглядаються питання розвитку електронних бібліотек, а також проблеми інтеграції різнорідних ІР у мережевому просторі.

На заседании Совета по развитию информационного общества, которое проходило под председательством Д. А. Медведева 12 февраля 2009 г. отмечалось, что за последние годы ИТ и информационные услуги стали достаточно существенной статьей российского не сырьевого экспорта, достигнув уровня приблизительно в миллиард долларов. С 2000 г. по 2008 г. – 30% в год рост ИКТ. Никакая отрасль в РФ не развивалась столь быстрыми темпами. Рост трафика в российском сегменте интернета с того же 2000 г. – в 183 раза. Тем не менее по ключевым показателям в соответствующих международных рейтингах Россия занимает 70-80 места. Такая неоднозначная и противоречивая картина во многом объясняется медленным ростом контента российского (а также СНГ в целом) сегмента интернет, низкими темпами формирования электронных информационных ресурсов.

Трансформация информации в цифровую форму и создание электронных фондов на основе документов, прежде всего из ретроспективных фондов научно-технических библиотек, связано с переводом в машиночитаемый вид значительных объемов текстовых и графических материалов и, соответственно, со значительным объемом работ и финансовых затрат на их выполнение.

В Библиотеке Конгресса США за последние 12 лет оцифровано 10% фондов, т. е. 11 млн. материалов (стоимость работ – 198 млн. долл.). Для оцифровки всего фонда (134 млн. объектов) потребуется не одно десятилетие. В Национальной библиотеке Франции в 2006 году переведено в цифровой формат и размещено на веб-сайтах 80 тыс. текстовых документов. Стоимость их оцифровки оценивается в 80 тыс. евро. Фонды французской Национальной библиотеки планируют оцифровывать по 150 тыс. объектов ежегодно. Начатая в 1995 году оцифровка библиотечных фондов быстро осуществляется в Германии. Немецкое исследовательское общество планирует с 2008 г. предоставлять от 10 до 20 млн. евро для цифровой обработки фондов научных библиотек. В пределах 5,6 и 7-й рамочных программ ЕС уже реализованы и реализуются целевые программы, научно-исследовательские и технологические проекты, связанные с оцифровкой.

Бюджет только одной программы eContentplus (2005-2008), направленный на производство европейского цифрового контента, составляет 149 млн. евро.

Все это сложно и дорого, но «овчинка стоит выделки», так как цифровая трансформация фондов научных библиотек открывает следующие возможности:

- получать информацию через Интернет независимо от времени и места нахождения пользователя;
- существенно повысить оперативность предоставления пользователям необходимой литературы, документов и данных;
- использовать машиночитаемые копии для сохранения оригинальных документов и создание страховых массивов документов на случай утраты оригиналов;
- производить работу с электронными документами, которая выходит за рамки простого чтения или просмотра изображения (в т. ч. редактировать, соединять, добавлять, вводить подразделы, перестраивать электронные документы, создавать на их основе новые);
- открывать новые формы библиотечного и информационного обслуживания, делать доступными для значительно большего числа пользователей документы, имеющиеся в библиотеках в ограниченном количестве или одном экземпляре.

Как направление работ оцифровка фондов осуществляется в рамках реализации Федеральной целевой программы «Электронная Россия». Электронные библиотеки (ЭБ) и полнотекстовые базы данных – относительно новый информационный ресурс, получивший развитие в последние 10-14 лет –. Как научное направление в информатике – это область исследований и разработок, ориентированных на развитие теории и практики обработки, распространения, хранения, поиска и анализа данных различной природы. Создание и использование ЭБ требует решения многих научных, технологических, методологических, правовых и других вопросов. Технологии электронных библиотек стремительно развиваются. Однако остаются актуальными проблемы семантики, интеграции информации, восприятия и представления разнообразных видов данных.

В прикладном аспекте ЭБ – собрание, в котором все материалы сохраняются в цифровом формате и доступны в сети Интернет. Электронная библиотека содержит не только текстовые материалы, но и переведенные в электронную форму с помощью имитационного и 3D моделирования изображения объектов, моделей, изделий и любых других экспонатов. Полнотекстовые БД и ЭБ создаются на многих уровнях – в федеральных центрах по видам документов, государственных библиотеках, негосударственными организациями. В то же время в нашей стране создание электронных информационных ресурсов идет крайне медленно. По сравнению с США и Китаем эти ресурсы не соответствуют современным требованиям. В табл. 1. сведены сопоставительные данные по периодическим и продолжающимся изданиям США, Китая и России /1/, представленных в сети Интернет (по данным на 2006 г.).

Таблица 1

Страна Объект	США	Китай	Россия
Периодические издания	14000	3200	720
Представление в Интернет	8000	2600	283
Соотношение (в%)	55%	80%	38%

В 1998 году была создана Научная электронная библиотека (НЭБ) при финансовой поддержке Российского фонда фундаментальных исследований. Базы данных НЭБ содержат более 200 Гб полнотекстовой научно-технической информации. Значительную долю в полнотекстовых электронных ресурсах занимают ИР коммерческих структур, которые отчасти заполняют пробелы, существовавшие в ГСНТИ в части правовой, конъюнктурно-экономической, финансовой, справочной и деловой информации. К таким, наиболее крупным системам, относятся прежде всего «Рос-бизнесконсалтинг», «Гарант», «Консультант-Плюс» и «Кодекс».

Основными методологическими проблемами и технологическими задачами развития электронных библиотек являются:

- модели и стандарты представления информации и метаинформации в ЭБ;
- автоматическая классификация информации;
- доступ к распределенным и разнородным коллекциям (интероперабельность, масштабируемость, обнаружение релевантной информации, интеграция метаинформации);
- интерфейсы пользователей, визуализация и анализ данных, вопросы интеллектуальной собственности;
- анализ и обработка естественного языка, изображений, видео- и аудиоданных;
- многоязыковый доступ к данным и обслуживание данных на нескольких языках;
- мобильные технологии и агенты в контексте электронных библиотек;
- методы и средства поиска, каталогизация, индексирование, поддержка целостности и непротиворечивости коллекций, безопасность ЭБ и защита информации /7/.

Доминирующей тенденцией в мировой и европейской системах поставки научно-технической информации является использование сетевых продуктов и услуг /2-6/, а самым прогрессивным способом получения информации выступает режим on-line. Вследствие этого актуальной проблемой является интеграция разнородных информационных ресурсов в гетерогенном сетевом пространстве. Постоянно действующими факторами при формировании единого (виртуального) информационного пространства являются:

- иерархичность информационных систем и ресурсов;
- разнородность ресурсов и программно-технических сред, объединяемых в едином сетевом операционном пространстве;
- распределенность элементов информационной инфраструктуры.

Концепция создания распределенного гетерогенного информационного пространства, объединяющего разнородные ресурсы /6 /, предполагает удовлетворение следующих основных требований к информационным системам:

- способность систем функционировать в условиях информационной и реализационной неоднородности, распределенности и автономности информационных ресурсов ИС;
- обеспечение интероперабельности, повторного использования неоднородных информационных ресурсов в разнообразных применениях;
- возможность объединения систем в более сложные, интегрированные образования, основанные на интероперабельном взаимодействии компонентов;
- осуществлении миграции унаследованных систем в новые системы, соответствующие новым требованиям и технологиям при сохранении их интероперабельности;
- обеспечение более длительного жизненного цикла систем.

С учетом этих требований создание развитой инфраструктуры для представления и обмена метаданными является одним из приоритетных направлений формирования единого информационного пространства и совершенствования современной глобальной сети. В настоящее время многие информационные центры, занимающиеся сбором и распространением метаданных, проявляют активную заинтересованность в организации взаимодействия с целью обмена имеющимися у них фондами. Как правило, в основе такой интеграции фондов лежит выработка стандарта на формат для представления метаданных, одновременно с унификацией массивов нормативно-справочной информации. Однако такой подход применим не для всех типов ресурсов. Появляются новые, быстро эволюционирующие типы ресурсов (например, мульти-медиа ресурсы, интерактивные сервисы сети, электронные модели исследуемых объектов, электронные карты, телеконференции, электронные коллекции и т. п.), разработка стандартов для которых, в силу их динамической природы и новизны, не успевают за темпами развития данных предметных областей. Это – проблема.

Конкретные функции метаданных и их состав в значительной мере зависят от специфики системы и характера конкретных информационных ресурсов.

В самых общих категориях назначение метаданных в прикладной информатике можно охарактеризовать следующим образом:

- а) обеспечивают механизмы интеграции информационных ресурсов из разных источников сведениями о свойствах этих ресурсов;
- б) являются источниками сведений о свойствах и содержании информационных ресурсов для механизмов управления данными в информационных системах;

в) являются источником информации, необходимой для осуществления реинжиниринга информационных систем;

г) обеспечивают представление сведений о системе, ее информационных и других ресурсах для различных приложений и пользователей системы.

Метаданные, так же, как и данные, в информационных системах могут обладать различной степенью динамичности. Они могут оставаться неизменными на протяжении всего жизненного цикла ИС. В других случаях в этот период в предметной

области системы могут происходить такие изменения, которые влекут за собой необходимость изменений системных и/или пользовательских метаданных.

Эта проблема связана с темпоральными свойствами данных и/или метаданных и приводит к эволюции логической схемы баз данных. В таких ситуациях ИС должна быть способна одновременно поддерживать данные в рамках прежней и новой схем. Кроме того, серьезные трудности могут быть связаны с необходимостью поддержки индивидуальности сущностей предметной области, изменяющихся во времени, с обеспечением сопоставимости таких сущностей в разные временные периоды.

В информационных системах используются разнообразные способы и средства представления метаданных в зависимости от применяемых информационных технологий. Это обстоятельство значительно осложняет решение проблем повторного использования и интеграции информационных ресурсов и информационных технологий, мобильности приложений, а также реинжиниринга информационных систем.

Для преодоления указанных проблем активно ведутся разработки стандартов метаданных, ориентированных на различные технологии. В этой области уже существует ряд международных и национальных стандартов.

В заключение можно кратко сформулировать основные задачи и направления работ по созданию единого информационного пространства как в России, так и в СНГ:

- форсированное развитие электронных информационных ресурсов;
- активная поддержка исследований и разработок в области интероперабельных программных средств и систем;
- совершенствование российского сегмента сети Интернет посредством развития инфраструктуры для предоставления и обмена метаданными, с учетом полной гармонизации национальных (государственных) стандартов с международными;
- внедрение единых (гармонизированных) стандартов информационно-библиотечной деятельности;
- государственная поддержка развития сети научно-образовательных телекоммуникаций (трудности внедрения новых телекоммуникационных технологий для научно-образовательных сетей на территории РФ заключаются не в отсутствии know-how, а в проблеме финансирования первоочередных работ по развитию региональных сетей и их средств внешней связи).

Литература

1. Сюнтюрено О. В. Проблемы информационной поддержки инновационной деятельности и научная политика государства //Будущее России. Вызовы и проекты: Экономика. Техника. Инновации. Под ред. Г. Г. Малинецкого, М.:ЛКИ, 2009, с. 264.
2. Борисова Л. Ф., Сюнтюрено О. В. Проблемы информационного обеспечения научно-инновационной и промышленной сферы: новые концептуальные подходы //Научно-техническая информация. Сер. 1. № 4. 2009. С.: 9–12.
3. Сюнтюрено О. В. Информационные ресурсы в экономике, основанной на знаниях //Сборник трудов конференции «Информационное обеспечение фундаментальных и прикладных наук»: Москва, 27 мая 2005. – М.:ВИНИТИ. – С.:86–89.
4. Разработка комплекса мероприятий по развитию информационной инфраструктуры национальной инновационной системы на базе государственной системы научно-технической информации: Отчет по НИР №02. 551. 11. 4011 от 01. 06. 2008 (Роснаука). ВИНТИ, 2008.
5. Борисова Л. Ф., Сюнтюрено О. В. Реферативный банк ВИНТИ РАН: перспективы постобработки информации с использованием методов анализа данных. //Научно-техническая информация. Сер. 1. № 11. 2007. С.: 6–11.

6. Концепция создания распределенного гетерогенного информационного пространства, объединяющего разнородные ресурсы: Отчет по НИР. Гос. Рег. № 01. 2. 006 08818. ВИНТИ, 2005.
7. Sjuntyurenko O. The Problems Providing Information Security: The Case of Information Infrastructure //Studies in Eastern Europe: Technological and Environmental Policy. ed. Gerhard Banse. – Berlin, 2007.