

**Проблемы интеграции разнородных информационных ресурсов
на примере библиотек и музеев**

**Problems of Integration of Diverse Information Resources
(Based on the Example of Libraries and Museums)**

**Проблеми інтеграції різнорідних інформаційних ресурсів
на прикладі бібліотек та музеїв**

Н. А. Мазов

*Институт нефтегазовой геологии и геофизики им. А. А. Трофимука
Сибирского Отделения РАН, Новосибирск, Россия*

О. Л. Жижимов, А. М. Федотов

Институт вычислительных технологий Сибирского Отделения РАН, Новосибирск, Россия

Nikolay Mazov

*Trofimuk Institute of Petroleum Geology and Geophysics,
Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

Oleg Zhizhimov and A. Fedotov

*Institute of Computational Technologies,
Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

М. О. Мазов

*Институт нефтегазовой геологии та геофизики ім. А. О. Трофімука
Сибірського Відділення РАН, Новосибірськ, Росія*

О. Л. Жижимов, А. М. Федотов

Институт обчислювальних технологій Сибірського Відділення РАН, Новосибірськ, Росія

В докладе рассматриваются проблемы интеграции разнородных информационных библиотечной и музейной тематики на основе применения протокола Z39.50 (ISO-23950) для построения распределенной информационной системы. Утверждается, что технологии Z39.50 являются наиболее актуальными для организации доступа к базам данных разнородных информационных коллекций в распределенных информационных системах.

The report deals with the problems of integration of diverse information library and museum items on the basis of Z39.50 (ISO-23950) for a distributed information system. Z39.50 technologies are described as the most actual for the organization of access to the databases of diverse information collections in distributed information systems.

В доповіді розглядаються проблеми інтеграції різнорідної інформації бібліотечної та музейної тематики на основі застосування протоколу Z39.50 (ISO-23950) для побудови розподіленої інформаційної системи. Стверджується, що Z39.50 є найбільш актуальними для організації доступу до баз даних різнорідних інформаційних колекцій в розподілених інформаційних системах.

Предоставление возможности удаленного доступа к информационным ресурсам при работе с библиографическими базами данных, с музейными электронными коллекциями и другими типами информации, возможность интеграции ресурсов на основе использования протокола Z39.50 [1,2] выводят работы, связанные с информационным обслуживанием на другой, более высокий уровень. Внедрение новых сетевых информационных технологий в настоящее время кардинально меняет работу информационных органов, библиотек, архивов, музеев. Об этом свидетельствуют ряд программ как регионального, так и Российского масштаба, направленных на решение проблем интеграции в свете национальных проектов.

Опыт работы в Сибирском отделении Российской академии наук (СО РАН) по данному направлению позволил сделать некоторое обобщение и попытку систематизировать используемые информационные ресурсы, предоставляемые пользователям, основу которых составляют научные сотрудники и студенты Новосибирского государственного университета.

Несмотря на то, что в СО РАН активно ведутся работы в области интеграции информационных ресурсов и создания соответствующих пользовательских интерфейсов, в настоящее время все же

сохраняется сложность при получении пользователем в короткое время полной и систематизированной информации в соответствии со своими информационными потребностями и в едином интерфейсе. Широкое использование всех имеющихся информационных ресурсов СО РАН самими пользователями затруднено из-за разнообразия форм наполнения и представления этой информации. Поэтому разработка технологии функционирования распределенных информационных систем, предназначенных для хранения и отображения информации различного характера и создание специализированных распределенных информационных систем для представления электронных коллекций разнородной научно-технической тематики является на сегодняшний день актуальной задачей.

Если в области унификации доступа к библиотечным информационным ресурсам сегодня более или менее есть какая-то определенность, то с музейными экспонатами дело обстоит совершенно иначе. Электронные музейные экспонаты носят самый разнородный характер: это различные базы данных и идеографическая (описательная) информация; изображения (фотографии, рисунки, чертежи и пр.) и видео- (аудио-) фрагменты; наблюдения и списки данных; модели и модельные данные и др. Эти ресурсы, как правило, принадлежат различным организациям, которые проводят самостоятельную политику в отношении их использования и предоставления публичного доступа к ним.

При продуманном и правильном построении распределенной информационной интегрированной системы, все эти ресурсы могут быть объединены в единое виртуальное информационное пространство, что может позволить пользователям открыть для себя всю информацию, накопленную за долгое время и в разных местах, и быстро получить исчерпывающие ответы на сложные запросы. При этом фрагменты данных, выбранные по запросу, подвергаются определенным процедурам, выполняемым специализированным программным обеспечением, что позволяет представлять их пользователю в адекватной форме.

Структурная и семантическая неоднородность информации, накапливаемой в различных учреждениях СО РАН, требует при интеграции специальных приемов объединения распределенных ресурсов для создания виртуальных организаций – единой динамической среды метаданных распределенных ресурсов [3].

В настоящее время наиболее распространенным способом для реализации подобной системы является WEB, как наиболее простой и эффективный способ организации сетевого доступа к информационным ресурсам различного характера. Однако следует отметить, что технология WEB не позволяет построить распределенную информационную систему со сквозным поиском и едиными для всех систем пользовательскими интерфейсами, поскольку в этой технологии отсутствует такая важная компонента, как глобальная стандартизация на уровне организации данных и форматов их представления. Насколько известно авторам, единственной технологией, содержащей подобную некоммерческую компоненту, апробированную в режиме многолетней промышленной эксплуатации, является технология, основанная на международном стандарте ISO-23950 (Z39.50) [1].

К настоящему моменту авторами отработана технология создания электронных коллекций научно-технической библиографической информации с использованием метаданных и схем данных, основанных на открытых международных стандартах и протоколах [4,5]. В качестве основы при использовании этой технологии для построения распределенных информационных систем используется открытый протокол Z39.50 (ISO-23950), который позволяет унифицировать сетевой доступ к любым базам данных, поскольку в своей основе предполагает абстрагирование от конкретных БД и СУБД. Несмотря на то, что этот протокол создавался по инициативе библиотечного сообщества и для библиотек, сегодня область его применения существенно расширена. Он эффективно используется для доступа к научно-технической информации, к геоинформационным ресурсам, к глобальным базам метаданных. Существует возможность предоставления доступа по протоколу Z39.50 к музейной информации и цифровым коллекциям [6]. Технология построения распределенных систем на основе этого протокола действительно позволяет объединять разнородные ресурсы и обеспечивать к ним унифицированный доступ.

Авторами предложена координация категорий данных и принятых форм записей, применяемых при каталогизации музейных коллекций [5]. На наш взгляд это необходимо, прежде всего, для того, чтобы любой музей при создании электронного каталога своей коллекции был уверен в том, что работу эту не придется переделывать, что его электронный каталог возможно будет в дальнейшем интегрирован в мировое музейное и библиотечное информационное пространство для обмена информацией или обобщения данных, содержащихся в каталогах различных музеев и библиотек.

Правила, которые регламентируют работу с музейной информацией по Z39.50, сформулированы в документах CIMI (Consortium for the Computer Interchange of Museum Information) и составляют профиль CIMI [7,8]. Элементы этого профиля имеют глобальные идентификаторы и являются частью международного стандарта ISO-23950. На основе этих объектов определена модель поиска и извлечения данных с музейной информацией.

Полное описание профиля CIMI достаточно объемно и здесь приводиться не будет. Ниже будут отмечены только некоторые аспекты, которые на наш взгляд характерны для этого профиля. Во-первых, все серверы CIMI обязаны поддерживать запросы типа 1 (RPN) (запросы в обратной польской нотации) с набором поисковых атрибутов CIMI-1, который содержит как атрибуты общего плана (Title, Author, Abstract и т. п.), так и атрибуты специфические для музейных коллекций (repositoryPlace, dateOfOrigin, placeOfOrigin, dateCollected и т. п.). Во-вторых, схема данных и абстрактная структура записи CIMI основана на иерархической модели данных с использованием схем Digital Collections и Dublin Core.

Здесь следует сделать некоторые замечания относительно интеграции библиотечных и музейных ресурсов. Эта интеграция возможна, по-видимому, только на уровне реализации сквозного поиска. При этом в терминах Z39.50 необходимо наличие пересечения множеств поисковых атрибутов различных баз данных. Таким естественным пересечением является множество поисковых атрибутов Dublin Core, которые в Z39.50 существуют как в виде отдельного набора атрибутов (XD-1), так и в виде подмножества BIB-1 (USE 1097-1111). К сожалению, сегодня эти поисковые атрибуты, насколько известно авторам, не использует ни одна библиотечная система России, что затрудняет интеграцию гетерогенных информационных ресурсов.

Тем не менее, некоторые попытки интеграции музейных и библиотечных ресурсов существуют. Пример – появление на сервере библиотечной ассоциации АРБИКОН странички с названием «Музейно-библиотечный портал» (http://www.arbicon.ru/services/index_ml.asp) с возможностью сквозного поиска в библиографических базах данных и базе данных RUSART, предоставленной ОАО «АльтСофт». К сожалению, используемый для этого профиль RUSMARC для решаемой задачи малоэффективен, а другие профили ассоциация АРБИКОН не поддерживает.

Наконец следует отметить, что на сегодняшний день в России существует лишь один программный продукт, позволяющий одновременно поддерживать библиотечный и музейный профили и создавать распределенные информационные системы с целью объединения разнородных информационных ресурсов – серверное программное обеспечение Z39.50 ZooPARK, разработанный СО РАН и успешно функционирующий в различных регионах России [2], в том числе и в упомянутом выше ОАО «АльтСофт».

Литература

1. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50 -1995, July 1995.
2. Жижимов О. Л., Мазов Н. А. Принципы построения распределенных информационных систем на основе протокола Z39.50. – ОИГМ СО РАН, Новосибирск: ИВТ СО РАН. – 2004. – 361 с.
3. Жижимов О. Л., Мазов Н. А., Федотов А. М. Центр доступа к электронным информационным ресурсам СО РАН // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: 13-я междунар. конф. «Крым 2006» (10-18 июня 2006 г., г. Судак): Труды конф. – М.: Изд-во ГПНТБ России, 2006. – [Эл. ресурс].
4. Мазов Н. А., Жижимов О. Л. Метаданные и их роль в распределенных информационных системах на основе использования протокола Z39.50. Лекция // Библиосфера. – 2006. – № 2. – С. 51-60.
5. Мазов Н. А. Содержание электронных музейных коллекций: определение и семантика категорий // Информационные технологии в гуманитарных исследованиях. – Новосибирск: ИАЭ СО РАН, 2006. – № 10. – С. 42-45.
6. Жижимов О. Л., Мазов Н. А. О доступе к информационным ресурсам по культурному наследию по протоколу Z39.50 // EVA'2000. «Электронная конвергенция: новые технологии в музеях, галереях, библиотеках и архивах», 30 октября – 3 ноября 2000 г.: Матер. конф., М., Центр ПИК Минкультуры РФ, 2000, 08-2-1 – 08-2-2.
7. The CIMI Profile Release 1. 0H A Z39.50 Profile for Cultural Heritage Information <http://www.cimi.org/documents/HarmonizedProfile/HarmonProfile1.htm>
8. Мазов Н. А. CIMI: профиль Z3950 для обмена информацией о культурном наследии // Информационные технологии в гуманитарных исследованиях. – Новосибирск: ИАЭ СО РАН, 2006. – № 11. – С. 21-50.