

УДК 004.7

ББК 32.973.26-018.2+32.973.202

МЕТАДАННЫЕ И ИХ РОЛЬ В РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ПРОТОКОЛА Z39.50

Лекция

© Н.А. Мазов, О.Л. Жижимов, 2006

*Объединенный институт геологии, геофизики и минералогии
Сибирского отделения Российской академии наук
630090, г. Новосибирск, пр. Академика В.А. Коптюга, 3*

В последнее время в связи с бурным развитием новых информационных технологий и средств вычислительной техники продолжают накапливаться всё большие объемы информации в электронном виде, хранящейся, как правило, под управлением различных СУБД и в различных форматах. При этом форматы описания и представления данных в различных СУБД усложняют организацию эффективного поиска, а зачастую не позволяют осуществлять прямой доступ к такой информации в распределенной информационной среде, поскольку отсутствуют взаимосвязи между информационными ресурсами и представлением данных. В связи с этим задача интеграции информационных ресурсов на основе использования метаданных, позволяющих обеспечить взаимосвязь между разнородными источниками информации и их эффективным использованием, является наиболее актуальной и выдвигается сегодня на передний план.

Ключевые слова: метаданные; протоколы; Z39.50; технологии; распределенные информационные системы.

В последнее время большое значение уделяется разработке технологий и практике построения распределенных информационных систем, технологиям доступа к распределенным информационным ресурсам (коллекциям, базам данных и т.п.) /1–3/. Это обусловлено тем, что в настоящее время накоплены (и продолжают накапливаться!) огромные разнородные информационные ресурсы, хранящиеся под управлением различных систем управления базами данных (СУБД), имеющих различное функциональное назначение и расположенных на различных информационных серверах. Это различные библиографические реферативные и полнотекстовые базы данных, музейные коллекции, фото-, аудио-, видеоколлекции и др. Как правило, взаимосвязи между такими информационными ресурсами и представлениями данных отсутствуют, а это существенно усложняет организацию эффективного поиска. Обеспечение универсальных способов работы с такими данными, где заранее неизвестно, с какими видами объектов придется работать, унификация их предоставления являются актуальной задачей при интеграции информационных ресурсов, решение которой невозможно без использования метаданных, позволяющих обеспечить необходимое абстрагирование представления данных от конкретных операционных платформ и СУБД,

позволяющих сохранить функциональность данных. Другая, не менее часто возникающая задача – это универсальная обработка похожих атрибутов, представляющих различные виды объектов (например, одинаковая обработка частей документов, в некоторых из которых присутствует «лишнее» поле).

1. Что такое метаданные и определения метаданных

В связи с интенсивным развитием сетевых и информационных технологий, которые предполагают формирование и существование распределенных информационных систем, в которых друг с другом взаимодействуют пользователи с различными уровнями знаний и интересов, а границы между традиционными ролями (например, издатель, автор, библиотека, читатель) размыты, существенно вырос интерес к метаданным (информации об объектах, как физических, так и электронных).

В настоящее время существует большое количество определений метаданных. Так или иначе, все эти определения имеют право на существование. Ниже мы приведем некоторые из них, которые нам удалось обнаружить и которых будет достаточно для того, чтобы понимать по-своему определение метаданных.

Термин «metadata» состоит из двух терминов – «meta» и «data», первый из которых может быть переведен как «абстрактный», а второй – «данные».

Как правило, *метаданные* интерпретируются как *данные о других данных*. Это одно из основных определений, которое принимается в настоящее время.

Ниже приведены еще несколько определений метаданных.

«Метаданные – данные о данных: каталоги, справочники, реестры, базы метаданных, содержащие сведения о составе данных, содержании, статусе, происхождении, местонахождении, качестве, форматах и формах представления, условиях доступа, приобретения и использования, авторских, имущественных и смежных с ними правах на данные и др.».

«Метаданные – информация о документе, понимаемая компьютером».

«Метаданные – данные, которые являются описанием других данных, их характеристик, местонахождения, способов использования».

«Метаданные – это данные о данных, информация об информации, описание контента» /4, 5/.

Как видно из представленных определений, все они имеют право на существование, несмотря на некоторые незначительные отличия.

Запись метаданных состоит из набора атрибутов или элементов, необходимых для описания данного ресурса (объекта). Например, наиболее часто встречающаяся и понятная система метаданных – это библиотечный каталог (традиционный или электронный), содержащий набор записей метаданных с элементами, которые описывают книгу либо другую единицу хранения: автор, заглавие, дата создания или публикации, шифры и номера хранения, определяющие местонахождение единицы на полке, и др. Связь между записью метаданных и ресурсом (объектом), который она описывает, может осуществляться двумя способами:

- элементы могут содержаться в записи, хранящейся отдельно от описываемой единицы, как это происходит в библиотечных каталогах;
- метаданные могут храниться непосредственно в теле ресурса.

Примером внедренных метаданных, которые хранятся вместе с самим ресурсом, является каталогизация в публикациях – CIP (Cataloging In Publication), содержащая данные, печатающиеся на обороте титульного листа книги.

Многие стандарты метаданных, используемые в настоящее время, не предопределяют тот или иной тип связи, оставляя решение при каждой конкретной реализации за разработчиком. Огромный выбор стандартов описания и приложений для электронных ресурсов сможет улучшить нахождение релевантных ресурсов в разнородных информационных ресурсах. Как отмечают лидеры в об-

ласти развития метаданных: «Объединение стандартизованных описательных метаданных с сетевыми объектами обладает потенциальными возможностями существенно улучшить возможность нахождения ресурсов, предоставляя возможность поиска по полям (например, автору, заглавию), пермутационным индексам для нетекстовых объектов, осуществлять доступ к суррогату содержания, значительно отличается от доступа к собственно содержанию ресурса» /5, 6/.

2. Назначение, роль и виды метаданных

Возможности использования метаданных исключительно широки, и роль метаданных для интеграции электронных информационных ресурсов будет постоянно возрастать, поскольку метаданные способны ускорить процесс международного доступа к информации, так как могут быть представлены на языках, отличных от языка самого объекта. Иногда метаданные рассматривают как разновидность традиционной практики библиотечной каталогизации, однако это не так.

Метаданные – понятие исключительно широкое и емкое. В настоящей работе ограничимся применением метаданных, которые ориентированы прежде всего на пользователей и создателей электронных информационных ресурсов в рамках распределенных информационных систем. Применительно к этой области, роль метаданных об электронных ресурсах, прежде всего, состоит в следующем:

- управлении жизненным циклом информационных ресурсов (процессами создания, использования и хранения цифровых документов);
- предоставлении возможностей более быстрого, точного и полного обнаружения необходимых ресурсов;
- обеспечении гибких и разнообразных механизмов отбора и представления информации в соответствии с требованиями пользователя (поисковым запросом).

Учитывая то, что часть задач может решаться и реально решается без участия человека, метаданные подразделяют на предназначенные для использования приложениями и для использования человеком. В английском языке этому подразделению соответствуют термины «machine-readable» и «human-readable». Они отличаются друг от друга областью применения, используемыми подходами и пр.

В настоящее время существуют различные классификации метаданных, которые отличаются между собой главным образом степенью детализации. Наиболее целесообразной, на наш взгляд, является классификация метаданных, предложенная в /5/, на две большие группы:

- 1) административные метаданные,

2) метаданные описания содержания.

Административные метаданные объединяют различные группы и отличаются большим разнообразием. Основная группа административных метаданных предназначена для того, чтобы проводить владельцу ресурса четкую и гибкую политику в отношении информационного объекта, которая включает авторизацию, аутентификацию, управление авторскими правами, доступом, а также служит для идентификации и категоризации объектов в рамках специальной коллекции или организации. Другая группа может представлять собой схемы хранения данных в базах данных, схемы распределенных баз данных и др. Третья группа административных метаданных может использоваться для позиционирования данного информационного ресурса в контексте группы подобных документов, информационно-поисковой системы, предметной области и т.д.

Метаданные описания содержания охватывают описание всех аспектов данного информационного ресурса (объекта) как отдельной сущности. Иногда их дополнительно подразделяют на структурные и описательные.

Метаданные состоят из элементов, объединенных в наборы. Наиболее известным примером набора элементов метаданных является Дублинское ядро (Dublin Core, DC) /18/.

Такие наборы разрабатываются с различными целями (например, для описания различных информационных объектов) различными организациями, которые предпринимают в случае целесообразности усилия по распространению и стандартизации своих разработок. В том случае, если набор элементов метаданных рассматривается и принимается соответствующей уполномоченной организацией, например ISO (International Standard Organisation), то он становится официальным стандартом метаданных.

Поскольку могут существовать и реально существуют различные наборы метаданных, возникает потребность в специальных форматах обмена метаданными между различными информационными системами.

3. Использование метаданных

Традиционно метаданные ассоциируются с «каталожными» системами баз данных или описаниями хранилищ, предоставляющими информацию как о хранящихся данных, так и об ассоциированных с ними процессах. Лишь сравнительно недавно метаданные приобрели большую актуальность в качестве основного компонента в архитектуре корпоративных баз данных и в качестве указателя на любые данные, используемые для идентификации, описания и обнаружения сетевых

электронных ресурсов. Ниже будут описаны различные способы использования метаданных, в соответствии с различными доступными информационными ресурсами.

В корпоративных средах, основанных на взаимоотношениях клиент–сервер с широким разнообразием распределенных систем и приложений, метаданные являются критическим элементом для активизации взаимодействия между системами различных поставщиков. В настоящее время предпринимаются попытки разработки различных стандартов обмена метаданными путем определения расширяемого механизма, позволяющего поставщикам информации обмениваться общими метаданными, так же как и переносить «подходящие» метаданные.

Так, например, одной из наибольших трудностей использования среды Интернета в настоящий момент является широкая неоднородность пространств для поиска, протоколов обмена (таких, как HTTP, FTP и др. /7/), информационных ресурсов и механизмов поиска и индексирования.

Основным типом объектов, доступных в веб-среде, являются электронные документы в формате HTML. Такие документы не имеют дополнительной информации, которая могла бы облегчить их поиск, кроме ключевых слов и полнотекстового содержания, используемого подходящими средствами поиска. А это, в свою очередь, порождает ряд проблем, таких как недостаток точности в конечном результате, отсутствие индексирования информационных мультимедиа-ресурсов и, как следствие, ведет к перегрузке сетей в результате реализаций различных стратегий поиска.

Существует три основных аспекта применения метаданных, наблюдаемых в настоящее время, – это описание ресурсов, производство метаданных и использование метаданных /8/.

Первый аспект касается информации, которая описывается метаданными. Основным вопросом здесь является тип ресурса и назначение дескрипторов для его описания. Второй аспект – производство метаданных. Метаданные представляют собой обобщенные описания данных. Ручная генерация такой информации, несмотря на современные средства ввода, является дорогим процессом. Автоматизация процедуры генерации метаданных должна стать единственно возможным направлением в этих работах. Третий аспект применения метаданных касается их использования и доступа к ним. Этот аспект является особенно важным как механизм локализации ресурса в распределенных информационных сетевых средах, таких, например, как Интернет. Он включает в себя такую информацию, как идентификация ресурса, описание предмета и структуры и т.п., которая является очень важной для доступа к ресурсам.

Применение метаданных, при правильном их понимании и использовании, сулит интересные перспективы, поскольку количество различных электронных ресурсов растет очень быстро. Такие ресурсы, особенно те, которые расположены в среде Интернета, по природе являются динамическими, т.е. часто появляются новые версии, документы часто переименовываются или перемещаются. Очень часто не является очевидным индексирование тех или иных цифровых ресурсов. Это касается, например, веб-страниц, относящихся к объекту поиска или мультимедийных ресурсов.

Другими фундаментальными проблемами, относящимися к применению метаданных, являются следующие /8/:

- Необходима интернационализация стандартов, поскольку большинство из них основаны на английском языке.
- Из-за большого разнообразия стандартов метаданных необходимо учитывать ситуацию описания одного ресурса более чем одним набором атрибутов.
- Необходима поддержка расширений существующих стандартов метаданных для приспособления их к локальной информации и новым типам ресурсов.
- Необходима близкая связь метаданных с описываемыми ресурсами. Совместимость метаданных является очень существенной для использования данных. Следовательно, метаданные должны генерироваться одновременно с описываемым ресурсом, а при изменениях ресурса – необходима модификация метаданных. Однако некоторые специфические типы метаданных могут быть сгенерированы отдельно, такие, например, как отзывы, рецензии и др.
- Метаданные являются данными. Следовательно, для них так же, как и для данных, существуют вопросы хранения, доступа и правильной интерпретации их содержания.

В последние годы различными специалистами за рубежом и в России предприняты значительные усилия в области обсуждения стандартов метаданных, которые имеют целью существенно улучшить результаты поиска по информационным ресурсам в распределенных информационных средах.

4. Стандарты и модели метаданных

В этом разделе мы рассмотрим некоторые из наиболее важных стандартов, описанных в литературе. Поскольку, на наш взгляд, невозможно объять необъятное, мы в основном сфокусируем наше внимание на стандартах, основанных на текстовой информации и проанализируем основную часть существующих подходов. Эти текстовые стандарты различаются между собой по сложности и первичной задаче – некоторые из них предназначены

в основном для каталогизации библиотечных документов, другие были созданы для поиска электронных ресурсов в веб-среде. Существуют также стандарты, связанные с другими специфическими областями применения. В данном разделе мы приведем стандарты для текстовой информации.

4.1. Стандарт метаданных MARC для библиографической каталогизации

Стандарты, объединенные в данной категории, а именно модель семейства MARC, различные ее варианты и диалекты, оказали сильное влияние на другие стандарты в прочих категориях описания ресурсов.

Стандарт MARC (**M**achine **R**eadable **C**atalogue) был создан на основе стандарта обмена ISO-2709 /9/ в конце 60-х гг. прошлого столетия в целях оказания помощи службам каталогизации в обмене каталожными записями. Он использовался в библиотечных автоматизированных службах как основа для манипулирования библиотечными записями – индексирования и отображения. Варианты стандарта MARC, такие как USMARC и UKMARC, а позднее UNIMARC /10, 11/, как основа для создания национальных версий форматов, появились как ответ на существующие конфликты форматов в качестве реализации индивидуальных требований различных библиотек.

Хочется отметить, что хотя изначально стандарт USMARC не предназначался для описания сетевых ресурсов, он стал основой для проекта OCLC. Этот проект развил реализацию стандарта MARC, объединенного с правилами AACR2 (**A**nglo-**A**merican **C**ataloguing **R**ules) /12/ для описания веб-ресурсов различных типов и осуществления доступа к ним.

Любой формат MARC, будь то USMARC или RUSMARC /13/, или любой другой, является трудным для модификации, поскольку он весьма структурирован. Создание нового элемента, например, требует согласования с сообществом USMARC или RUSMARC. Библиографические описания следуют наборам правил AACR2 и ISBD (**I**nternational **S**tandard for **B**ibliographic **D**escription) /14/ для формата USMARC и правилам каталогизации, отраженным в ГОСТ 7.1–84 /15/ для формата RUSMARC. На самом деле, следует сказать, что имеющаяся MARC-модель обслуживает библиотеки и пользователей несовершенно, поскольку в основном она направлена на печатные первоисточники. Основной причиной этого является то, что по существу структура формата MARC является плоской и требует фрагментации иерархически подчиненных компонентов в отдельные дискретные записи. Такая структура записи отражает требования компьютеров начала 70-х гг. XX в., когда системы были жестко привязаны к возможностям

и объемам хранения данных, и в настоящее время уже является несовершенной. Хотя все чаще и чаще отмечается, что стандарт MARC совершенно непригоден для взаимодействия с новыми технологическими требованиями информационно-библиотечных систем, он еще много лет будет оставаться на плаву из-за миллиардов библиотечных записей, функционирующих в информационно-библиотечных системах. По этой причине одной из основных задач, которую приходится сегодня решать, – это различные отображения между стандартом формата MARC и другими форматами.

4.2. Стандарты метаданных для кодирования текста и обмена данными

Стандарты этой категории связаны с использованием SGML (Standard Generalized Marked Language) – стандартного обобщенного языка разметки текстов, метки которого предоставляют структуру и способы доступа к различной библиографической информации для онлайн-систем.

Одним из широко известных является стандарт независимых заголовков TEI (Text Encoding Initiative) /16/. Основой этого стандарта явился профиль, определяющий набор общих правил для представления текстовых материалов в электронном виде, позволяющих осуществлять их обмен и повторное использование. Первоначально проект был нацелен на разработку руководств по подготовке и обмену электронных текстов для научных целей, однако вскоре область его применения вышла за рамки представления научных текстов.

Профиль TEI определяет текстовые возможности в терминах элементов и атрибутов SGML, сгруппированных в наборы меток. Элемент представляет собой текстовую единицу, такую как параграф. В заголовке элемент должен быть единицей, такой как название или автор. Атрибут дает информацию о конкретном вхождении элемента и должен иметь структуру пары «атрибут-значение».

Каждому кодированному TEI тексту должен предшествовать заголовок, описывающий текст. Различные элементы в TEI сгруппированы в наборы тэгов, наборы ядер (элементы, требуемые для всех документов), базовые наборы (наборы элементов, подходящие определенному классу документов – например, книги, статьи из журналов, патенты и т.п.), дополнительные наборы (элементы для специальной обработки текстов для различных классов документов) и дополнительные наборы, содержащие специализированные элементы.

Заголовок TEI составлен следующим образом:

- описание документа (его библиографические характеристики);
- описание кодирования (информация, касающаяся обработки текста и деталей процесса кодирования);

- описание профиля (дополнительная небиблиографическая информация, такая как используемый язык, детальная информация об участниках, классификация предмета и т.п.);

- описание исправлений.

Следует отметить, что профиль TEI ориентирован на описание объектов и не содержит никаких соображений по описанию различных сервисов. Использование независимых заголовков возможно с помощью любого программного обеспечения, поддерживающего SGML. Создание документов HTML является частным случаем использования профиля TEI. Настоящий стандарт кодирования является гибким: использование этого стандарта требует меньшей подготовки, нежели кодирование в MARC.

Различные профили по использованию TEI служат доказательством ограничения возможностей MARC в структурировании небиблиографической информации (такой, как цифровые коллекции и различная иерархическая информация), которая может быть использована для оценки электронных документов. Как и MARC, заголовки TEI предоставляют полную библиографическую информацию, однако в них недостает дополнительной информации о документах, содержащей более детальный анализ электронного текста.

Наряду со стандартом TEI существует еще ряд стандартов, например EAD (Encoding Archival Description) /17/, предназначенный для описания архивных материалов, на которых мы не будем останавливать здесь свое внимание, поскольку они менее используются в информационной практике.

4.3. Стандарты метаданных Dublin Core и GILS

Ниже будет рассмотрен стандарт, известный в информационной практике как Дублинское ядро – (ДЯ) (Dublin Core) /18/. Система ДЯ является результатом первого симпозиума по метаданным, проведенного в 1995 г. OCLC (Online Computer Library Center) / NCSA (National Center for Supercomputer Applications). Целью этого форума была разработка формата записи метаданных для описания сетевой электронной информации без предварительного описания структуры записи. Основной целью этой модели является идентификация и определение синтаксиса независимого набора элементов метаданных для определения ресурсов в веб-среде таким путем, что они могут быть также отображены в более сложные и контролируемые системы, такие как USMARC, UNIMARC и др. В результате был определен набор из 13 основных элементов для описания электронного ресурса, учитывающий следующие характеристики:

- назначение: для описания свойств объекта;
- расширяемость: для возможности включения дополнительного описания;

МАСТЕР-КЛАСС

- факультативность: для указания, что все элементы являются факультативными;
- повторяемость: означает, что все элементы являются повторяемыми;
- модифицируемость: означает то, что каждый определенный в схеме элемент может подразумевать самоопределение, для удовлетворения различных потребностей.

Эти характеристики важны, поскольку они нацелены на автоматическое создание сетевых средств публикаций, содержащих шаблоны элементов метаданных, упрощая тем самым задачу создания записей метаданных. Следует заметить, что спецификации DC еще доопределяются, что дает некоторые новые возможности, однако даже в существующем виде они предоставляют значительную гибкость для механизмов поиска.

В Z39.50 не существует глобального профиля, который бы соответствовал Dublin Core. Однако на DC отображаются, конечно, с потерями информации, все схемы данных Z39.50. Более того, некоторые схемы данных Z39.50, например CIMI, явно содержат элементы DC.

Тем не менее Dublin Core «присутствует» в Z39.50 в виде поисковых атрибутов (табл. 1).

Т а б л и ц а 1

Поисковые USE-атрибуты Bib-1 для Dublin Core

Атрибут	Значение USE Bib-1
DC-Title	1 097
DC-Creator	1 098
DC-Subject	1 099
DC-Description	1 100
DC-Publisher	1 101
DC-Date	1 102
DC-ResourceType	1 103
DC-ResourceIdentifier	1 104
DC-Language	1 105
DC-OtherContributor	1 106
DC-Format	1 107
DC-Source	1 108
DC-Relation	1 109
DC-Coverage	1 110
DC-RightsManagement	1 111

Наличие поисковых атрибутов DC в наборах поисковых атрибутов Z39.50 предоставляет потенциальную возможность включить ресурсы Z39.50

в систему индексации веб-ресурсов и сделать их доступными для поисковых веб-машин. Обратная же задача, т.е. предоставление доступа по протоколу Z39.50 к веб-ресурсам, не является однозначной ввиду отсутствия в Z39.50 схемы данных DC.

Ярким представителем стандартов для глобальной информационной инфраструктуры является GILS (Government Information Locator Service) /19/, главное назначение которого состоит в том, чтобы обеспечить механизм для определения местонахождения полезной информации, производимой различными правительственными учреждениями. Инициатива создания настоящего стандарта принадлежит американскому федеральному правительству, для того чтобы люди могли находить информационные ресурсы всюду по его многим агентствам.

GILS идентифицирует и описывает эти ресурсы, добавляя другие правительственные и коммерческие механизмы информационного распространения. В GILS определено около 70 зарегистрированных атрибутов (так называемых основных GILS элементов), включающих название, создателя, дату публикации, место публикации, язык, реферат (аннотацию), контролируемые предметные индексы, пространственную область и много других.

Так как GILS стандартизован в ANSI Z39.50 /20/, следовательно, возможен доступ к таким записям по протоколу Z39.50. Важным моментом этого стандарта является то, что он должен гарантировать семантическую интероперабельность между различными GILS серверами.

Несмотря на то что рассмотренные стандарты первоначально не были рассчитаны для веб-среды, они были расширены до описания электронных ресурсов. Рассмотренные стандарты (MARC, GILS, DC) имеют больше шансов на успешное использование в веб-среде, ввиду их тесной связи с протоколом Z39.50, который уже позволяет проводить поиск по MARC, GILS и DC метаданным.

В отличие от Dublin Core профиль GILS создан для Z39.50 и жестко определяет структуру описания информационного ресурса. Следствием этого является возможность построения на основе GILS распределенных информационных систем с хорошо развитым поисковым сервисом.

4.4. Collection

Если в первые годы своего существования протокол Z39.50 использовался преимущественно для организации доступа к библиографическим ресурсам, то на сегодняшний день область его применения существенно расширена, он используется для доступа к научно-технической информации, к геоинформационным ресурсам, к глобальным базам метаданных и др. Существует также возможность предоставления доступа по Z39.50 к музейной информации и цифровым коллекциям.

Под цифровыми коллекциями принято понимать упорядоченный набор цифровых объектов, которые могут иметь, а могут и не иметь нецифровых аналогов. Примеры цифровых объектов – оцифрованные аудиозаписи, изображения картин, цифровые фотографии, программы, базы данных и т.п. Подобные объекты и коллекции этих объектов могут быть систематизированы и описаны, т.е. могут быть созданы метаданные для цифровых объектов. Естественно, что они будут существенно отличаться от обычных библиографических описаний. В частности, обязательным элементом описания цифрового объекта является ссылка на сам объект, например по URL.

Для доступа к метаданным цифровых объектов и коллекций существует профиль Digital Collections /21/.

Профиль Digital Collections определяет конформное подмножество Z39.50-1995 для доступа к цифровым коллекциям, организованным посредством метаданных, структура которых описывается в этом профиле. Он также предоставляет семантику для навигации по цифровым коллекциям обнаружения и получения объектов.

Данный профиль предполагает, что сопутствующие профили (совместимые расширения или подмножества рассматриваемого профиля) будут разработаны независимо, расширяя или ограничивая использование этого профиля для определенных приложений или классов информации, например музейных объектов, спутниковых фотографий, географических данных или химических соединений.

Таким образом, данный профиль непосредственно не решает все задачи доступа к цифровым коллекциям, однако предоставляет сопутствующим профилям каркас для этого:

- Цифровые объекты представляются атомарными, т.е. их содержимое является «непрозрачным» и не обрабатывается профилем. Таким образом, профиль предоставляет поиск описательной информации, а не поиск цифровых объектов. Сопутствующие профили могут моделировать содержимое специфических объектов (например, музейных).

- Связанные описания (например, средство поиска, каталоги выставок и т.д.) рассматриваются как «непрозрачные» (их содержимое не обрабатывается профилем), хотя клиенты могут иметь в своем распоряжении дополнительные приложения, дающие возможность их обработки и отображения. Сопутствующие профили могут моделировать содержимое связанных описаний.

- Профиль не моделирует сложные взаимодействия между объектами всех классов. Сопутствующие профили могут делать это для определенных классов объектов.

Хотя в рассматриваемом профиле особое внимание уделено логическому отделению описательной информации от содержимого, он не включает в себя руководства или спецификации, определяющие информацию как описание или содержимое. Профиль также не указывает принципы организации, т.е. пути сбора объектов в коллекции /21–23/.

Несмотря на то что данный профиль адресован на распределенные коллекции, он не нацелен на распределенные базы данных. Различные части коллекции могут управляться различными организациями таким образом, что различные базы данных, соответствующие одной коллекции, могут располагаться на различных серверах, однако индивидуальная база данных не может быть распределена по многим серверам.

Настоящим профилем используется только протокол Z39.50. Профиль, однако, не препятствует наличию мультипротокольных клиентов, шлюзов или цифровых коллекций, в которых одна часть доступна по протоколу Z39.50, другая же – по какому-либо другому протоколу.

В табл. 2 показано, с какими объектами Z39.50 оперирует профиль Digital Collections.

Для программного обеспечения, декларирующего совместимость с профилем Digital Collections, определяется три уровня совместимости как для клиента, так и для сервера:

1. Базовая совместимость – поддержка версии 2 протокола Z39.50-1995, схемы и структуры описательной записи Digital Collections и синтаксиса записи GRS-1.

2. Базовая V3 совместимость – базовая совместимость и дополнительно поддержка версии 3 протокола Z39.50-1995.

3. Расширенная V3 совместимость – поддержка базовой V3 совместимости и дополнительно поддержка:

- параметра CompSpec сервиса Present,

Т а б л и ц а 2

Объекты Z39.50 для Digital Collections

Поисковые атрибуты	Bib-1	1.2.840.10003.3.1
	Collection-1	1.2.840.10003.3.7
Диагностика	Bib-1	1.2.840.10003.4.1
Форматы	GRS-1	1.2.840.10003.5.105
Схемы	Digital Collections Schema	1.2.840.10003.13.3
Наборы меток	TagSet-M	1.2.840.10003.14.1
	TagSet-G	1.2.840.10003.14.2
	Collections TagSet	1.2.840.10003.14.5

МАСТЕР-КЛАСС

- спецификации элементов eSpec,
- набора вариантов variant-1,
- фрагментации,
- Metadata и appliedVariant для GRS-1.

Для структуры записи в этой схеме характерна одна деталь. Поскольку толкование цифрового объекта достаточно широкое, он может быть таким, что для его описания используется другая схема данных. В этом случае в записи присутствует OID другой схемы данных. Примером такого случая является музейная информация.

4.5. CIMI

Правила, которые регламентируют работу с музейной информацией в Z39.50, сформулированы в документах CIMI (Consortium for the Computer Interchange of Museum Information) и составляют профиль CIMI /24/. Элементы этого профиля имеют глобальные идентификаторы и являются частью международного стандарта ISO-23950.

Полное описание профиля CIMI достаточно объемно и здесь приводиться не будет. Ниже будут отмечены только некоторые аспекты, которые характерны для этого профиля.

В табл. 3 показано, с какими объектами Z39.50 оперирует профиль CIMI.

Т а б л и ц а 3

Объекты Z39.50 для CIMI

Поисковые атрибуты	Bib-1	1.2.840.10003.3.1
	CIMI-1	1.2.840.10003.3.8
Диагностика	Bib-1	1.2.840.10003.4.1
Форматы	GRS-1	1.2.840.10003.5.105
	SUTRS	1.2.840.10003.5.101
	Usmarc	1.2.840.10003.5.10
Схемы	Digital Collections Schema	1.2.840.10003.13.3
	CIMI Schema	1.2.840.10003.13.5
Наборы меток	TagSet-M	1.2.840.10003.14.1
	TagSet-G	1.2.840.10003.14.2
	Collections TagSet	1.2.840.10003.14.5
	CIMI TagSet	1.2.840.10003.14.6

На основе этих объектов определена модель поиска и извлечения данных с музейной информацией.

Во-первых, все серверы CIMI обязаны поддерживать запросы типа 1 (RPN) с набором поисковых атрибутов CIMI-1, который содержит как

атрибуты общего вида (Title, Author, Abstract и т.п.), так и атрибуты, специфические для музейных коллекций (repositoryPlace, dateOfOrigin, placeOfOrigin, dateCollected и т.п.). Кроме этого, набор атрибутов CIMI содержит в себе поисковые атрибуты Dublin Core /18/.

Во-вторых, схема данных и абстрактная структура записи CIMI основана на иерархической модели с использованием схем Digital Collections и Dublin Core. Элементы последней явно присутствуют в CIMI на первом уровне вложения. Собственно структура записи CIMI вложена в элемент actualDO записи Digital Collection /21/.

Схема CIMI позволяет представить в стандартном виде описательную информацию о музейных ресурсах, т.е. является схемой метаданных. Связь с изображениями и полными текстами осуществляется через URL в элементе mrObject/rendition/resource. Некоторые аспекты применения схемы данных CIMI можно найти в работах /25–29/.

В ситуации, когда музейное информационное сообщество России только начинает создавать информационные системы с предоставлением доступа к ним через Интернет, очень актуальным является вопрос о выборе правильной технологии и правильной схемы метаданных. Несомненно, предпочтения должны отдаваться решениям, которые опираются на международные стандарты, на технологии, которые прошли проверку временем. Технология Z39.50 здесь является одним из самых привлекательных вариантов.

5. Заключение

При правильном подходе к каталогизации информационных ресурсов получаемые описания могут быть хорошо структурированы и содержать всю основную информацию о первичном объекте. Наиболее характерное свойство метаданных – одинаковая структура вторичных данных, соответствующих первичным объектам одного типа. Именно это свойство удобно использовать при организации поиска в распределенной информационной системе, когда поиск информации производится не по массивам первичной информации, а по массивам их описаний – метаданным. Более того, можно утверждать, что информационные ресурсы одного типа могут быть объединены только на уровне метаданных, так как только метаданные в разных системах могут иметь одинаковую структуру и семантику. Однако, чтобы этого добиться, требуется стандартизация типов метаданных.

Список литературы

1. Жижимов, О. Л. Принципы построения распределенных информационных систем на основе протоко-

- ла Z39.50 / О. Л. Жижимов, Н. А. Мазов. – Новосибирск : ОИГТМ СО РАН, 2004. – 361 с.
2. Жижимов, О. Л. Принципы построения распределенных информационных систем на основе протокола Z39.50. Ч. 1 / О. Л. Жижимов, Н. А. Мазов // Науч.-техн. информ. Сер. 1, Орг. и методика информ. работы. – 2004. – № 5. – С. 1–14.
 3. Жижимов, О. Л. Принципы построения распределенных информационных систем на основе протокола Z39.50. Ч. 2 / О. Л. Жижимов, Н. А. Мазов // Науч.-техн. информ. Сер. 1, Орг. и методика информ. работы. – 2004. – № 11. – С. 12–26.
 4. [http://yandex.ru/yandsearch?text="](http://yandex.ru/yandsearch?text=)метаданные”
 5. Хохлов, Ю. Е. Обзор форматов метаданных [Электронный ресурс] / Ю. Е. Хохлов, С. А. Арнауты // Институт развития информационного общества. Портал «Российские электронные библиотеки». – Режим доступа: http://www.elbib.ru/index.phtml?env_page=methodology/metadata/md_review/md_review.html
 6. Метаданные и их применение / А. Н. Бездушный, А. А. Вежневцев, В. А. Серебряков, А. В. Шкотин // Электронная конвергенция: новые технологии в музеях, галереях, библиотеках и архивах. EVA'2001 : Материалы конф. – М., 2001. – С. 08–3–11 – 08–3–14.
 7. *Carvalho Moura, A. M., de*. A Survey on Metadata for Describing and Retrieving Internet Resources / A. M. Carvalho Moura, de, M. L. M. Campos, M. C. Barreto // World Wide Web. – 1998. – № 1. – С. 221–240.
 8. *Iannella, R.* Metadata: Enabling the Internet. 1997 [Электронный ресурс] / R. Iannella, A. Waugh. – Режим доступа: <http://www.dstc.edu.au/RDU/reports/CAUSE97>.
 9. International Organization for Standardization. Documentation. Format for Bibliographic Information Interchange on Magnetic Tape. – Geneva : ISO, 1981. – 3 p.
 10. Формат USMARC. Краткое описание : в 3 ч. – М. : ГПНТБ России. – 1996. – 3 ч.
 11. Руководство по UNIMARC : рук. по применению междунар. коммуникатив. формата UNIMARC. – М. : ГПНТБ России. – 1992. – 320 с.
 12. AACR. Anglo-American Cataloguing Rules. – [2nd ed.], rev. – 1988.
 13. Российский коммуникативный формат представления библиографических записей в машиночитаемой форме : (Рос. вариант UNIMARC). – СПб., 1998.
 14. ISBD, International Federation of Library Associations and Institutions, ISBD(G). General International Standard Bibliographic Description : Annot. Text. – London : IFLA Intern. Office for UBC, 1997.
 15. Библиографическое описание документа : общ. требования и правила сост. : ГОСТ 7.1–84. – Изд. офиц. – М. : Гос. ком. СССР по стандартам, 1984. – 77 с.
 16. *Burnard, L.* The Text Encoding Initiative Guidelines [Электронный ресурс]. – 1994. – Режим доступа: <http://www.uic.edu/orgs/tei>
 17. *Swetland, A. J. G.* Encoded Archival Description Document Type Definition (DTD) [Электронный ресурс] : final draft 1996. – Режим доступа: <http://scriptorium.lib.duke.edu/people/tom/guidelines/>
 18. *Burnard, L.* A Syntax for Dublin Core Metadata [Электронный ресурс] / L. Burnard, L. Q. Miller, C. M. Sperberg-McQueen // Recommendations from the Second Metadata Workshop. – 1996. – Режим доступа: <http://www.uic.edu/~cmsmcq/tech/metadata.syntax.html>
 19. *Christian, E. J.* GILS What is it? Where's it going? [Электронный ресурс] // D-Lib Magazine. – 1996. – № 1. – Режим доступа: <http://www.dlib.org>
 20. 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.
 21. Z39.50 Profile for Access to Digital Collections [Электронный ресурс] : (Final Draft for Rev.) : May 3, 1996. – Режим доступа: <http://lcweb.loc.gov/z3950/agency/profiles/collections.html>
 22. *Bull, R. A.* Analysis of the Z39.50 Profile for Access to Digital Collections and the Z39.50 Explain Service for UKOLN [Электронный ресурс] // Crossnet Job No. F17. – 29 October 1997. – CSL017-TN-001. – Issue 2. – Режим доступа: <http://www.ukoln.ac.uk/metadata/cld/study/crossnet/zpdc.pdf>
 23. *Moen, W. E.* Resource Discovery Using Z39.50: Promise and Reality [Электронный ресурс] // Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium. – 2001. – 574 p. – Режим доступа: <http://lcweb.loc.gov/catdir/bibcontrol/moenpaper.html>
 24. The CIMI Profile. Release 1.0H. A Z39.50 Profile for Cultural Heritage Information [Электронный ресурс]. – Режим доступа: http://www.cimi.org/public_docs/HarmonizedProfile/HarmonProfile1.htm
 25. *Moen, W. E.* Accessing Distributed Cultural Heritage Information // Comm. of ACM. – 1998. – Vol. 41, N 4. – P. 45–48.
 26. *Moen, W. E.* CIMI's Z39.50 Interoperability Testbed: Search and Retrieval of Distributed Cultural Heritage Information [Электронный ресурс]. – 1998. – Режим доступа: <http://www.unt.edu/wmoen/publications/CIMIACMrev.htm>
 27. *Velegrakis, Y.* On Z39.50 Wrapping and Description Logics / Y. Velegrakis, V. Christophides, P. Constantopoulos // Intern. J. Digit. Libr. – 2000. – N 3. – P. 208–220.
 28. Жижимов, О. Л. О доступе к информационным ресурсам по культурному наследию по протоколу Z39.50 / О. Л. Жижимов, Н. А. Мазов // Электронная конвергенция: новые технологии в музеях, галереях, библиотеках и архивах. EVA'2000 : Материалы конф. – М., 2000. – С. 08–2–1 – 08–2–2.
 29. Жижимов, О. Л. О доступе к информационным ресурсам по культурному наследию по протоколу Z39.50 [Электронный ресурс] / О. Л. Жижимов, Н. А. Мазов // Электронные библиотеки. – 2001. – Т. 4, вып. 4. – Режим доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2001/part4/ZM>

Материал поступил в редакцию 6.02.2006 г.

Сведения об авторах: Мазов Николай Алексеевич – кандидат технических наук, заведующий информационно-библиотечным центром, тел. (383) 333-22-16; e-mail: mazov@uiggt.nsc.ru
 Жижимов Олег Львович – доктор технических наук, заведующий сектором информационных технологий, тел. (383) 333-20-05; e-mail: zhizhim@uiggt.nsc.ru