

Подход к интеграции информационных коллекций в ИСИР РАН

А.Н. Бездушный, А.М. Меденников, В.А.Серебряков
ВЦ РАН, Москва
{bezdushn, meden, serebr}@ccas.ru

С.А. Власова, Н.Е. Каленов
БЕН РАН, Москва
{sav, nek}@ben.ix.ru

Аннотация

Проблема интеграции разнородных информационных коллекций теоретически изучена достаточно полно. Однако реализация теоретических выводов вызывает затруднения. В этой статье описывается подход, разработанный в рамках проекта ИСИР РАН, направленный на практическую реализацию технологии интеграции информационных коллекций при определенных ограничениях. Вводится понятие *общей модели*. Оцениваются различные форматы данных с точки зрения их пригодности в качестве общего формата. Рассматриваются проблемы, возникающие в процессе интеграции коллекций, на примере интеграции системы "Наука России" БЕН РАН.

1 ВВЕДЕНИЕ

В настоящее время в мире накоплены значительные объемы информации. Многие организации сталкиваются с необходимостью предоставления доступа к своим информационным ресурсам в режиме онлайн. Часто для решения этой задачи разрабатывается специальная система, ориентированная на работу с конкретным информационным массивом. Такая система может предоставлять пользователям специализированные интерфейсы доступа к данным, так что пользователь имеет возможность получить наиболее детальную информацию. В то же время, он может столкнуться с проблемами, в частности, с ограничениями на операции поиска и навигации. Эти ограничения могут быть вызваны, например:

- Неполным описанием предметной области. В модели данных могут быть не отражены объекты, их свойства, связи между ними. Часто содержание коллекции представляется как совокупность ресурсов одного типа, например, документы, библиографические описания, хотя, среди свойств этих ресурсов в той

или иной мере присутствуют описания других сущностей, например, авторы документов или публикаций, сведения об организациях. При этом, исходя из имеющихся в коллекции значений этих свойств, часто невозможно установить, описывают ли они одну и ту же сущность. С другой стороны, даже будучи отнесенными к одному типу, ресурсы могут существенно различаться, например, книги, журналы и статьи. Это затрудняет навигацию по элементам коллекции и поиск с ограничениями на эти связи. В результате, пользователям приходится прикладывать значительные усилия для получения подробной информации о таких объектах. Отсутствие связей между элементами коллекции, в то время как в реальном мире они имеют место (например, между журналом и содержащимися в нем статьями в библиографической базе данных) также снижает ценность информационной коллекции.

- Отсутствием связей между коллекциями. Одни и те же сущности могут быть описаны в разных коллекциях, с разных точек зрения, но пользователи не могут получить объединенное представление этой информации. Так, данные о человеке могут содержаться в базах данных сотрудников двух организаций, а также в авторитетных файлах библиотек. Получить эту информацию в обобщенном виде невозможно.

Указанные ограничения снимаются путем интеграции коллекций в единое информационное пространство. Имеются всевозможные подходы различной сложности и ориентации к проблеме интеграции неоднородных данных: системы неоднородных баз данных [1, 2], онтологии [3, 4], слабоструктурированные данные [5, 6] и т.п. Однако в этих подходах рассматриваются слишком общие случаи, так что их практическая реализация затруднена.

Эта работа, ведущаяся в рамках проекта ИСИР РАН, нацелена на создание работающей технологии, позволяющей практически выполнять интеграцию разнородных информационных массивов при определенных ограничениях.

©Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

2 ОБЗОР ФОРМАТОВ ИНФОРМАЦИОННОГО ОБМЕНА

Сущностью или объектом будем называть объект реального мира. Ресурсом будем считать условную единицу информации, имеющую внутреннюю структуру, которая может иметь как достаточно тривиальной, так и довольно сложной. Ресурсом может быть все, что обладает индивидуальностью, при этом он может представлять некоторую сущность или быть исключительно "виртуальным". Ресурс имеет метаданные и, возможно, содержание. Метаданные – это структурированные сведения о ресурсе, представляющие его свойства (или атрибуты). Это данные, на основе которых осуществляется поиск ресурсов, вывод результатов поиска, управление ресурсами, взаимодействие с ними и т.п. К метаданным относится уникальный идентификатор ресурса, отличающий его от любых других ресурсов в соответствующей среде. Каждый ресурс относится к определенному типу, который определяет некоторую совокупность сведений, внутренне присущих ресурсам этого типа. Между ресурсами могут быть установлены отношения.

Используя эти понятия, мы можем построить модель данных, основанную на концепции взаимосвязанных типизированных ресурсов, которую назовем общей моделью. Процесс интеграции коллекции включает в себя три компонента:

- обмен данными (физическая передача информации)
- преобразование данных
- связывание данных

Вопросы физического обмена данными выходят за рамки статьи. Рассмотрим более подробно остальные компоненты.

Для хранения информации используются разнообразные форматы. Модели данных, лежащие в основе коллекций, также различны. Подход ИСИР РАН состоит в том, чтобы, используя общий формат для представления информации, отобразить модель данных коллекции на общую модель.

Существует достаточно много стандартов и предложений по форматам данных для обмена. В библиотечной среде широко используются MARC (Machine Readable Cataloguing) форматы (USMARC [7], UNIMARC [8], RUSMARC [9]), GILS [10] и т.д. Они позволяют детально каталогизировать информационные ресурсы, но предназначены для определенной предметной области. Форматы CDF (Common Data Format) [11] и HDF (Hierarchical Data Format) [12] основаны на специфических моделях данных: первый использует представление информации в виде многомерного массива, второй – в виде иерархической структуры. Формат SOIF (Summary Object Interchange Format) [13], широко используется в Интернет для обмена метаинформацией о ресурсах, однако он не имеет концепций структуризации и вложения элементов, как и средств поддержки связей между ресурсами.

Интересной технологией описания ресурсов является методика RDF [14], которая предлагается консорциумом W3C в качестве стандарта для определения и обработки метаданных Web-ресурсов. Основная цель создания RDF

состояла в том, чтобы определить механизм описания ресурсов, который не делал бы никаких предположений относительно специфики предметной области, но был бы удобен для описания и обработки сведений о любой области. Примечательной стороной RDF является то, что он позволяет сделать утверждения не только о ресурсах, но и о самих утверждениях.

Семантика используемых в RDF описаниях слов, задается с помощью ссылок на так называемые RDF-схемы [15], которые можно рассматривать как своего рода словари или пространства имен. Схема определяет термины, которые могут быть использованы в RDF-утверждениях, позволяет специфицировать их значения, используя механизм формирования типов, чтобы обеспечить автоматическую обработку метаданных. Для многих стандартных форматов разработаны такие схемы. Однако большинство из них сохраняют структуру моделей данных, лежащих в основе их исходного формата, что делает использование этих пространств имен с RDF не совсем удобным из-за ограничений, налагаемых на RDF-утверждения этой структурой. В связи с этим, наиболее широко с RDF используется схема Dublin Core Metadata Set (DC) [16].

Dublin Core – это международная и междисциплинарная попытка определить набор элементов описания электронных информационных ресурсов, который бы ориентирован на их обнаружение и простую каталогизацию. Разработка формата следовала следующим принципам:

- Формат должен быть прост в использовании, как для непрофессиональных пользователей, так и для программных приложений.
- Необходимо иметь механизм расширения описательных возможностей формата, позволяющий сохранить обратную совместимость с исходным форматом.
- Все элементы должны иметь ясные достаточно общие определения, которые могут быть модифицированы, изменены с помощью необязательных квалификаторов.
- Все элементы формата должны быть не обязательными и повторяться неограниченное число раз.

Элемент – это некоторая характеристика или атрибут ресурса. Каждый элемент, определенный в DC, имеет некоторое семантическое значение. Он может определять допустимые значения, типы ресурсов, которые он может описывать, его отношение с другими элементами. Значения элементов необходимо различать по видам описываемого содержимого, то есть в некотором смысле указывать семантику значений. Такое разделение значений обеспечивается именованием элементов.

Дополнительная систематизация значений осуществляется с помощью квалификаторов [17]. Квалификатор элемента рафинирует, уточняет семантику элемента, чтобы точнее указать отношение элемента к ресурсу. Квалификатор элемента не изменяет определение элемента, не модифицирует его значение. Квалификатор значения определяет способ, в соответствии с которым значение закодировано. Обычно это указывается ссылкой на

список разрешенных термов, описание формата, правил разбора.

Такое определение модели метаданных и приведенных квалификаторов позволяет использовать формат DC в его стандартной простой форме и для сложных локальных описаний, за счет введения собственных значений классификаторов элементов. В настоящее время стандартизирован только набор элементов DC из 15 элементов. Однако работа над стандартом продолжается. Перечень квалификаторов пока не полностью определен.

Концепция RDF довольно проста: это, в сущности, бинарная модель данных, хотя, как уже упоминалось, стандарт предусматривает возможность построения более сложных конструкций (утверждения об утверждениях). Тем не менее, значительная часть информационных коллекций основана на бинарных моделях. Система понятий ИСИР позволяет создавать общие бинарные модели большой сложности. Кроме того, этот стандарт получает все большее распространение в качестве средства обмена метаинформацией. Это привело нас к выбору RDF как формата обмена данными.

3 МЕТОДИКА ИНТЕГРАЦИИ КОЛЛЕКЦИЙ

В процессе преобразования данных необходимо решить ряд проблем.

1. Требуется создать общую модель, которая в заданной предметной области могла бы инкорпорировать достаточно большой класс электронных коллекций. Она должна содержать основные сущности и их атрибуты из предметной области, а также основные связи между ними. Кроме того, модель должна быть такой, чтобы к ней можно было добавить новые элементы минимальными усилиями. Например, ИСИР РАН содержит ресурсы следующих типов: Организация, Подразделение, Персона, Проект, Публикация. Это дает замкнутое множество взаимосвязанных между собой ресурсов. Все типы ресурсов имеют собственные атрибуты, специфичные для каждого типа: тип Организации, научное звание Персоны, язык Публикации. Кроме того, имеются атрибуты, общие для нескольких типов ресурсов: адрес, телефон, комментарий. Также имеется возможность динамического расширения множества типов атрибутов без изменения структуры хранения данных. Таким образом, модель ИСИР РАН дает достаточно хорошее описание предметной области, при этом предоставляя возможности для его расширения.

Особое внимание было уделено разработке модели ресурса Публикация. Из многочисленных форматов, предназначенных для представления библиографических сведений, видно, что имеется довольно много типов публикаций, каждый из которых имеет собственный набор основных атрибутов. Кроме того, существуют десятки характеристик публикаций, используемых достаточно редко или только с определенными типами публикаций. С другой стороны, с расширением потока электронных изданий возникает необходимость представления специфичных для них свойств. Это заставило нас реализовать модель публикации на основе концепции Dublin Core. В этой модели публикация может иметь атрибуты, соответствующие

пятнадцати элементам DC. При необходимости семантика любого атрибута может быть специфицирована более точно с помощью квалификаторов, представленных в виде элементов словаря. Структура словаря позволяет определять иерархические системы квалификаторов.

2. Необходимо отобразить атрибуты коллекции на атрибуты общей модели. Эта задача может быть довольно нетривиальной. Во-первых, элемент коллекции может быть отображен на несколько ресурсов общей модели. Например, в массиве MARC-записей каждый элемент может быть отображен в три взаимосвязанных ресурса: Публикация, Персона и Организация. Во-вторых, для каждого атрибута интегрируемой коллекции необходимо определить атрибут общей модели, наиболее точно соответствующего его семантике. При отсутствии такого "семантически близкого" атрибута может потребоваться расширение схемы. В некоторых случаях это может потребовать большой работы, в результате чего оптимальным решением будет отказ от использования в общей модели значений такого атрибута.

3. Стандарт RDF позволяет создать несколько различных конструкций, описывающих одно и то же множество ресурсов. Часть различий, обусловленных использованием допускаемых стандартом сокращений, пропадает при преобразовании данных из синтаксиса XML в направленный граф ресурсов (DLG - Direct Labeled Graph). Но и после этого сохраняется неоднозначность, выражающаяся в существовании нескольких различных графов, эквивалентных с точки зрения общей модели. Здесь возникает задача отображения графа ресурсов в структуру, используемую для хранения данных. В частности, в ИСИР РАН, использующей в качестве хранилища реляционную базу данных, задача состоит в отображении этого графа в команды СУБД (SQL-выражения). Сложность этой проблемы была частично снята введением более строгих, чем предусмотрено стандартом, требований к синтаксису RDF-описаний, что позволило уменьшить количество возможных подграфов и упростить их отображение во внутреннюю структуру.

4. Использование пространств имен с RDF ставит перед нами две проблемы. Во-первых, существуют стандартные пространства имен, которые необходимо поддерживать. В ИСИР РАН поддерживаются пространство Dublin Core и vCard [18]. Первое, как уже упоминалось, предназначено для описания документов, публикаций ит.п. Схема vCard разработана для представления информации о персоналиях. DC отображается в модель ИСИР РАН довольно просто, так как ресурс Публикация разрабатывался с использованием идей, заложенных в Dublin Core. Поддержка пространства имен vCard реализуется сложнее, поскольку семантика и синтаксис его элементов и атрибутов ресурсов Персона, Организация и Подразделение, используемых в ИСИР РАН, имеют существенные отличия. На данном этапе в общую модель преобразуется минимальный набор элементов vCard (FN, N, ADR, TEL, EMAIL).

Во-вторых, часто стандартных пространств имен недостаточно для выражения семантики всех необходимых атрибутов, или же это требует сложного преобразования исходных данных. В этом случае приходится создавать специальное пространство имен для введения необходимых семантических элементов. Так, при работе с систе-

Таблица 1: Элементы схемы x-ben.ixex.ru. Квалификаторы элементов.

dcq.Title.Alternative	Альтернативное (дополнительное, параллельное и т.п.) заглавие
dcq.Description.BibDesc	Библиографическое описание публикации по ГОСТ
dcq.Description.Pages	Информация о номерах или количестве страниц
dcq.Description.Issue	Том и номер журнала
dcq.Relation.PartOf	Ссылка на ресурс, содержащий текущий как физическую часть
dcq.Relation.Serial	Для выпуска журнала – сериальное издание, частью которого является выпуск
dcq.Relation.References	Ресурс, на который есть ссылка (напр., цитирование) из текущего

Таблица 2: Элементы схемы x-ben.ixex.ru. Элементы словарей.

dct.Type.a	публикация аналитического уровня (статья)
dct.Type.m	публикация монографического уровня (книга, том журнала)
dct.Type.s	публикация уровня сериального издания (журнал, серия)

Таблица 3: Элементы схемы x-isir.ras.ru. Типы ресурсов.

publication	публикация
person	физическое лицо
organization	организация

Таблица 4: Элементы схемы x-isir.ras.ru. Виды квалификаторов.

eq	квалификатор элемента
vq	квалификатор значения

Таблица 5: Элементы схемы x-isir.ras.ru. Свойства ресурса "person".

name	полное имя или имя с инициалами
employer	организация, в которой работает
speciality	специальность ВАК

Таблица 6: Элементы схемы x-isir.ras.ru. Свойства ресурса "organization".

orgname	полное название организации
orgabbr	сокращенное название организации
address	почтовый адрес
zipcode	почтовый индекс
city	город

мой "Наука России" было определено два пространства имен. Одно (с префиксом x-isir.ras.ru:) содержит элементы, представляющие семантику атрибутов модели ИСИР РАН. Другое (x-ben.ixex.ru:) состоит из согласованного с БЕН РАН множества элементов, используемых в качестве квалификаторов Dublin Core и предназначенных для точного отображения семантики атрибутов базы данных "Наука России". В нем также определены множества элементов, являющихся словарями значений перечисленных типов. Реализация выполнена таким образом, что любому атрибуту общей модели может быть сопоставлено несколько элементов из одного или из разных пространств имен, что позволяет сопоставлять собственным пространствам имен стандартные по мере разработки последних.

Таким образом, имея методику синтаксического и структурного преобразования интегрируемой коллекции к унифицированному виду, мы можем включить ее данные в общее информационное пространство. И здесь встает проблема семантической интероперабельности. Для каждого ресурса, полученного из данных коллекции, в общем пространстве может уже существовать "семантически эквивалентный" ресурс (термин "семантически эквивалентный" описывает наиболее сильное из четырех ви-

дов "семантического сходства" ресурсов, введенных в [1]), причем он может быть создан и на основе данных другой коллекции и, следовательно, иметь другие свойства. Это усложняется тем обстоятельством, что ресурсы коллекции обычно уникально идентифицированы только в пределах этой коллекции или ее части. Задача, стоящая на этапе связывания, состоит в том, чтобы для каждого ресурса, добавляемого в пространство общей модели, найти семантически эквивалентный ресурс, если таковой имеется, на основе сравнения их атрибутов. Ниже описана методика, которая, по нашему убеждению, сможет в определенной степени решить эту проблему.

Для каждого типа ресурса задается список, состоящий из наборов атрибутов этого типа ресурса. В процессе связывания ресурса последовательно выполняется поиск среди ресурсов одного с ним типа с условием совпадения значений соответствующих атрибутов из очередного набора. Результатом каждого поиска может быть ноль, один или более ресурсов, удовлетворяющих указанным условиям. Если результат поиска – пустое множество, мы можем остановиться, считая, что "семантически эквивалентного" ресурса не существует, или повторить поиск со следующим набором атрибутов. В случае, когда результатом является единственный ресурс, можно сделать предполо-

```

(1) <?xml version="1.0"?>
(2) <rdf:RDF
(3) xml:lang="ru"
(4) xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
(5) xmlns:dc="http://purl.org/dc/elements/1.0/"
(6) xmlns:isir="x-isir.ras.ru:">

(7) <isir:organization rdf:about="x-ben.irex.ru:org.045">
(8) <isir:orgname>Вычислительный центр РАН</isir:orgname>
(9) <isir:orgabbr>ВЦ РАН</isir:orgabbr>
(10) <isir:address zipcode="117967" city="Москва" rdf:value="ул. Вавилова, 40"/>
(11) </isir:organization>

(12) <isir:organization rdf:about="x-ben.irex.ru:org.0305">
(13) <isir:orgname>Библиотека по естественным наукам РАН</isir:orgname>
(14) <isir:orgabbr>БЕН РАН</isir:orgabbr>
(15) <isir:address zipcode="119890" city="Москва" rdf:value="ул. Знаменка, 11"/>
(16) </isir:organization>

(17) <isir:person rdf:about="x-ben.irex.ru:per.P1">
(18) <isir:name>Алексеев Н.Г.</isir:name>
(19) <isir:employer rdf:resource="x-ben.irex.ru:org.0305"/>
(20) </isir:person>

(21) <isir:publication rdf:about="x-ben.irex.ru:pub.S2">
(22) <dc:Type rdf:resource="x-ben.irex.ru:dct.Type.s"/>
(23) <dc:Title>Библиотека</dc:Title>
(24) </isir:publication>

(25) <isir:publication rdf:about="x-ben.irex.ru:pub.M219965">
(26) <dc:Type rdf:resource="x-ben.irex.ru:dct.Type.m"/>
(27) <dc:Relation rdf:parseType="Resource">
(28) <isir:eq rdf:resource="x-ben.irex.ru:dcq.Relation.Serial"/>
(29) <rdf:value rdf:resource="x-ben.irex.ru:pub.S2"/>
(30) </dc:Relation>
(31) <dc:Description rdf:parseType="Resource">
(32) <isir:eq rdf:resource="x-ben.irex.ru:dcq.Description.Issue"/>
(33) <rdf:value>(5)</rdf:value>
(34) </dc:Description>
(35) <dc>Date>1996</dc>Date>
(36) </isir:publication>

(37) <isir:publication rdf:about="x-ben.irex.ru:pub.A2">
(38) <dc:Title>Ветеран войны - ветеран культуры: [К 75-летию директора БЕН
    РАН А.Г. Захарова].</dc:Title>
(39) <dc:Creator rdf:resource="x-ben.irex.ru:per.P1"/>
(40) <dc:Relation rdf:parseType="Resource">
(41) <isir:eq rdf:resource="x-ben.irex.ru:dcq.Relation.PartOf"/>
(42) <rdf:value rdf:resource="x-ben.irex.ru:pub.M219965"/>
(43) </dc:Relation>
(44) <dc:Description rdf:parseType="Resource">
(45) <isir:eq rdf:resource="x-ben.irex.ru:dcq.Description.Pages"/>
(46) <rdf:value>34</rdf:value>
(47) </dc:Description>
(48) </isir:publication>
(49) </rdf:RDF>

```

Рис. 1: Пример RDF-описания

жение, что семантически эквивалентный ресурс найден.

Однако это предположение может оказаться ошибочным. Неверные исходные данные, отсутствие значения в используемом для поиска атрибуте и тому подобные ситуации достаточно часто встречаются в информационных коллекциях. Это приводит нас к необходимости введения авторитетного лица: человека, которому предоставлено право и возможности корректировать результаты процесса автоматизированного обмена метаинформацией. Для обеспечения эффективности его работы необходимым условием является возможность отложенной обработки спорных ситуаций, т.е. возможность проведения интеграции вручную в любое время после окончания работы автоматизированного процесса.

В системе, где данные поступают из разных источников, необходим механизм защиты от изменения или уничтожения данных при преобразовании атрибутов разных коллекций, имеющих одинаковую семантику, в атрибуты ресурсов общей модели. С этой целью каждой коллекции, данные которой интегрируются в систему, присваивается идентификатор. При создании нового значения атрибута ресурса, полученного из некоторой коллекции, оно помечается идентификатором этой коллекции. Если рассматриваемый атрибут уже имеет значение, но помеченное идентификатором другой коллекции, оно не может быть изменено или уничтожено.

Описанная технология была применена для интеграции данных информационной системы "Наука России", поддерживаемой БЕН РАН, в ИСИР РАН. Текущая версия ИСИР поддерживает такие типы ресурсов, как Организация, Подразделение, Персона, Проект, Публикация, а также Рубрикаторы, например рубрикаторы ВАК и РФФИ. Система "Наука России" обеспечивает поддержку четырех взаимосвязанных баз данных, содержащих сведения об организациях, персоналиях (в частности, авторах публикаций), публикациях на аналитическом и монографическом уровнях, изданиях на сводном уровне (журналах, сборниках и т.п.). В рамках комплекса реализована возможность задания пристрастной библиографии (библиографических ссылок).

На рис. 1 приводится фрагмент RDF-описания, представляющего информацию, содержащуюся в базе данных "Наука России". В строках с 7 по 11 приводится информация об организации (точнее, о ресурсе типа Организация). Значение атрибута `rdf:about` представляет собой внутренний идентификатор ресурса системы "Наука России", записанный в виде URI. Для задания атрибутов ресурса (строки 8–10) используются элементы пространства имен `isir`. В строке 17 начинается описание ресурса типа Персона. Элемент `isir:employer` в строке 19 указывает идентификатор организации, в которой человек работает. Строки 21–48 содержат описания трех ресурсов типа Публикация. В системе "Наука России" эти данные хранятся в виде двух записей в разных подбазах: одна находится в базе источников и описывает сериальное издание в целом (полное название, краткое название, вид издания), а другая хранится в базе публикаций и ее поля содержат сведения как о статье, так и о выпуске журнала, в котором статья издана. Система также поддерживает связь между статьей и источником. Модель ИСИР РАН для публикаций предусматривает в таких случаях создание трех ресурсов: для сериального издания в целом, для

отдельного выпуска серии и для статьи, причем последние два содержат указатель на включающий их ресурс. На рис. 1 строки 21–24 являются описанием сериального издания, что определяется значением атрибута `dc:Type` (22), а отдельный выпуск описывается строками 25–36. Строки 27–30 представляют собой элемент пространства имен Dublin Core `dc:Relation` с квалификатором, указывающим на вхождение описываемой публикации в серию. Элемент `rdf:value` (29) содержит идентификатор ресурса, представляющего сериальное издание. Прочие элементы представляют информацию, относящуюся к выпуску: номер выпуска (31–34) и год издания (35). Статья описывается строками 37–48 и содержит заглавие (38), идентификатор автора (39), идентификатор выпуска серии (40–43), номера страниц в выпуске (44–47).

4 ЗАКЛЮЧЕНИЕ

В результате проведенной работы ИСИР РАН имеет возможность предоставить научному сообществу информацию о научных публикациях сотрудников организаций РАН с поддержкой необходимых связей. Представленная технология достигает поставленных целей и может быть использована как основа для разработки методов более тесной интеграции электронных библиотек.

Список литературы

- [1] A. Sheth (ed). Semantic issues in multidatabase systems. Special Issue. *SIGMOD Record*, 1991.
- [2] A. Sheth, J. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys* 22(3), 1990.
- [3] L. Kalinichenko. Emerging semantic-based interoperable information system technology. *Computers as our best partners*, 1994
- [4] D. Briukhov, S. Shumilov. Ontology Specification and Integration Facilities in a Semantic Interoperation Framework. *Proc. of the International Workshop ADBIS'95*, Springer, 1995
- [5] P. Buneman. Semistructured data. *Proc. of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, Tucson, Arizona, 1997.
- [6] S. Abiteboul. Querying semi-structured data. *Proc. of the Int. Conf. on Database Theory*, Delphi, Greece, 1997.
- [7] Форматы USMARC. Краткое описание. В 3-х ч. / Пер. с англ.; ГПНТБ России. - М., 1996.
- [8] Руководство по UNIMARC / Пер. с англ.; ГПНТБ России. - М., 1992.
- [9] Руководство по применению RUSMARC, 1997
- [10] The GILS standard. <http://www.usgs.gov/gils/>
- [11] The CDF standard. <http://nssdc.gsfc.nasa.gov/cdf/>

- [12] The HDF standard. <http://hdf.ncsa.uiuc.edu/>
- [13] Harvest User Manual, Appendix B, <http://harvest.cs.colorado.edu/Harvest/brokers/soifhelp.html>
- [14] Resource Description Framework (RDF) Model and Syntax, W3C Recommendation, 1999. <http://www.w3.org/TR/REC-rdf-syntax>
- [15] Resource Description Framework (RDF) Schema Specification. W3C Candidate Recommendation, 2000. <http://www.w3.org/TR/rdf-schema>
- [16] Dublin Core Metadata Initiative. http://purl.org/metadata/dublin_core/
- [17] Qualified Dublin Core Metadata for Simple Resource Discovery. <http://www.mailbase.ac.uk/lists/dc-datamodel/files/rfc3.html>
- [18] F. Dawson, T. Howes. vCard MIME Directory Profile, 1998. <http://www.imc.org/rfc2426>