Обзор некоторых направлений интеграции гетерогенных ресурсов в электронных библиотеках

© Новицкий А.В.

Институт программных систем НАН Украины
alex@zu.edu.ua

Аннотация

Работа посвящена интеграции информации. Сделан обзор текущих проектов по интеграции информации в электронных библиотеках. Проблемы, которые возникают при интеграции данных, связны с тем, что ресурсы описаны метаданными различным способом с различной семантикой. Семантическое аннотирование ресурсов это будущее Интернета и электронных библиотек. С недавнего времени существуют ряд технологий Semantic Web, которые способны автоматически решать проблему интеграции аннотированных веб-документов. В роботе сделано обзор текущих проектов по семантической интеграции в электронных библиотеках, а также показано пример преобразования семантической аннотации RDFa веб-документа в RDF для ЭБ.

1 Введение

В информационной интеграции можно выделить следующие проблемы: интеграция схемы [19], хранилищ данных, интеграция данных (также известна как интеграция информации предприятия, EII enterprise information integration) и интеграция каталога.

Подход к интеграции данных с использованием онтологий называется интеграция данных на основе онтологий.

В общем, есть ряд шагов, которые необходимо выполнять при интеграции информационных систем с использованием онтологий. К ним относятся:

- интерпретация запроса в терминологию общей онтологии;
- выявление соответствия между семантически связанными сущностями в локальной и общей онтологии;
- перевод соответствующих данных из локальных информационных источников (участвующих в обработке запроса пользователя) в формализм представления знаний системы интеграции информационных систем;
- согласования результатов, полученных из различных локальных информационных источников, а именно выявление и устранение, например избытка, дублирование и др.

Семантическая гетерогенность [13], как правило, отличается от синтаксической гетерогенности и структурной гетерогенности в семействе баз данных [3, 11, 10, 12].

Синтаксическая неоднородность связана с неоднородностью форматов данных. Стандартизация форматов данных принимается в качестве подхода к решению проблем синтаксической неоднородности. Например, XML используется в качестве стандартного формата для всех видов доступных Web данных.

Структурные неоднородности связаны с различными моделями данных, структур данных или схем, например, реляционных и объектно-ориентированные модели базы данных. Примером решения проблемы структурной неоднородности является использование RDF, который основан на синтаксисе XML и обеспечивает унифицированный способ структуры источников информации.

Следует обратить внимание, несмотря на то, что в электронной библиотеке информация может быть представлена в различных видах, семантика этой информации представляется с помощью текстовых метаданных, соответственно мы будем сосредотачиваться на интеграции семантических метаданных.

Когда два информационных источника смоделированы в одном и том же формате данных с применением одной и той же модели данных, как и раньше, могут возникать проблемы семантической неоднородности [20]:

- семантические конфликты. Различные разработчики моделей не испытывают точно такой же набор объектов реального мира, но вместо этого они представляют наборы которые пересекаются (включение или перекрытие элементов набора). Например, "Студент" объект класса может возникнуть в

Труды 11[ii] Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009.

одной схеме, в то время как более ограниченный объект класса "Студенты специальности информатика" находится в другой схеме. При интеграции двух схем класс "Студент специальности информатика" будет интегрирован как подкласс класса "Студент".

- описательные конфликты. Описательные конфликты относятся к конфликтам именования вследствие омонимов и синонимов.
- структурные конфликты. Структурные конфликты отличаются от структурной неоднородности. Даже если два разработчика моделей, используют одну и ту же модель данных, они могут выбирать различные конструкторы для представления объектов реального мира. Например, в объектно-ориентированной модели, когда разработчик описывает компонент объекта типа О, он оказывается между выбором, создания нового типа объекта или добавить атрибут к О.

Каждый домен использует локальные онтологии, которые являются результатом концептуализации домена. Поскольку процесс концептуализации не является однозначным, то это порождает гетерогенность источников. Для того чтобы их объединить, необходимо сделать больше, чем простой механизм маркировки соответствия объектов, классов или содержания. На самом деле, часто возникает ситуация, когда понятия не совсем совпадают, поскольку они могут иметь различия по свойствам в видовой или родовой классификации [18]. В целом можно выделить несколько видов сопоставления онтологий:

- расширение: предусматривает определение онтологии домена, связывая некоторые понятия между двумя выходными онтологиями. Две концептуальные модели дополняют друг друга, например концепты первой онтологии уточняются во второй через дополнительные атрибуты, которые не указаны в первой.
- гармонизация: предполагает семантическую эквивалентность между доменом и прикладными онтологиями, касается одного и того же онтологического обязательства. В этом случае текущий домен можно рассматривать как специализация в другом домене, который является более общим или расположен на абстрактно-формальном уровне.
- выравнивание: предполагает обобщение онтологии домена через общие понятия и аксиомы. Обе модели имеют (много/несколько) общих совместных концептов.

Для электронных библиотек решения проблемы семантической гетерогенности можно решать на двух уровнях: на уровне метаданных и на уровне контента. Для того чтобы показать текущее состояние дел, сделаем краткий обзор проектов по интеграции электронных библиотек. Следует обратить внимание, что рассмотрены различные аспекты проблематики интеграции данных. Такими аспектами могут быть особенности поискового механизма или архитектурные особенности. Тем не менее, данный обзор предоставляет, в некотором смысле, общий взгляд на текущее состояние проблемы.

Цель обзора - прояснить особенности интеграции информации в семантической среде. Следует обратить внимание, что если получиться представление контента (метаданных) в модели RDF, то проблему интеграции для контента (метаданных) можно считать решенной, или решение проблемы будет преведено к задаче мапинга онтологий.

В разделе 2 будет попытка сделать обзор проектов семантической интерпарабельности с позиций принципов интеграции и поисковых механизмов, затронуты и сопутствующие вопросы.

В разделе 3 приводится пример, который показывает, насколько простым может быть решение интеграции контента (метаданных) если изначально наш контент имеет семантическую аннотацию.

2 Краткий обзор семантической интерпарабельности в Европейских проектах.

2.1 Проект SWHi

Рассмотрим европейские проекты по внедрению Semantic Web в электронные библиотеки. Онтология SWHi [6] разработана для электронной библиотеки с точки зрения, когда наши основные источники данных в репозитории описаны метаданными. Эти метаданные отображаются и хранятся в онтологии, которая базируется на онтологии схемы. Для обогащения онтологии, также добавляют новую, связанную информацию из выбранных веб-документов.

Поиск в этой системе реализован в двух формах, простой и сложной. Система использует SeRQL как язык запросов к RDF [2]. Процессор генерирования запросов SeRQL сталкивается по крайней мере с двумя проблемами. Во-первых, он не знает, в каком классе или свойстве могут быть найдены слова. Чтобы избежать этой проблемы, прикладное программное обеспечение Semantic Web, такое как OpenAcademia [16], требует от пользователей вводить ключевые слова в соответствующее поле (автор, название или год) в ее расширенный поисковый интерфейс. Во-вторых, существуют некоторые ограничения в подстроках соответствия SeRQL при использовании символа общности '*'. Эту проблему можно решить с помощью информационно-поисковых программ, таких как Lucene. Другие семантические системы, такие как KIM, используют машину поиска Lucene для индексации и поиска семантически аннотированных документов. Одновременно OWLIR и Swoogle используют индексно-поисковую Haircut для

индексирования и поиска RDF документов, используя механизм N-грамм в качестве терминов индексации.

Помимо самого поиска, важным также является вопрос представления результатов поиска. Одним из направлений является визуализация поиска. В Semantic Web визуализация становится все более важной. Существуют случаи сложных взаимоотношений между ресурсами, которые не могут быть представлены с помощью простого списка. Кроме того, как правило, отражается только небольшое количество результатов поиска. К документам находящимся в хвосте результата поиска, скорее всего, никогда не будут обращаться.

Общая архитектура системы SWHi состоит из трех уровней: система управления знаниями (Kms), семантических веб-приложений (SWA), и уровень пользовательских интерфейсов (Web UI).

Очевидно, онтологии играют центральную роль в SWHi. Для развития SWHi онтологии, повторно используют имеющиеся онтологические ресурсы для структурирования и сохранение исторической информации, а именно: PROTON базовая онтология, таксономия предметной классификации NewsBank/Readex, Дублинское Ядро и словарь FOAF Vocabulary. Эти онтологии сохраняется с использованием Sesame2.

Экземпляры для этой онтологии брались с данных ранней Американской истории 1639-1800. Метаданные состояли из 36305 записей, которые подробно описаны (название, автор, дата публикации и т.д.) в MARC21. В будущем планируется расширить источники данных из других исторических электронных журналов, включив их полные тексты.

2.2 Проект eCulture

eCulture это семантическая поисковая система, которая позволяет одновременно искать в нескольких коллекциях учреждений культурного наследия [15]. Работает путем переноса коллекций в RDF, связывая объекты коллекций как экземпляры классов через общедоступные словари, тем самым создавая большой RDF граф.

Основным механизмом поиска является использование Prolog [21]. Запросы прикладной логики выражаются как Prolog цели на необработанных данных RDF и/или модулях суждения RDFS/OWL.

В eCulture разработана методология портирования культурных хранилищ для Semantic Web и RDF. Эта методология основана на том , что мы можем рассчитывать, как правило, на два типа данных:

-метаданные, которые описывают культурные объекты и фотографии;

-локальные словари, которые используются в некоторых метаданных.

Основное внимание в проекте уделено конвертации XML в RDF. Обращая внимание на то, что в отличие от ранее предложенных подходов, где целевая модель RDF следует из источников данных XML, в eCulture такой подход неприемлем. Трансформация XML в RDF основана на правилах свойств, правилах очередности, замены значений и других.

2.3 Проект IPISAR (Image Preservation, Information Systems, Access and Research)

Проект IPISAR исследует распространение, изучение и рациональное использование культурного наследия, а также попытки представить решения общих проблем в этих областях в рамках Semantic Web (SW) [14]. В рамках проекта предложено ряд идей, которые возможно, упростят интеграцию информации.

В проекте разработано приложение «Pescador», которое будет хранить каталогизированные данные в устойчивых тройных хранилищах (чьи функции будут такими же, что и реляционные базы данных в традиционных системах).

Pescador использует модель SW для каталогизации, где каждому формату записи будет соответствовать отдельный вид структурированного графа, в который включена специальная лексика и правила, с ссылкой на специализированную прикладную логику.

Одним из направлений Pescador является обеспечение интерфейсов программирования для пользователей, которые выполняют дизайн, моделирование и программирование каталогов. Для достижения этой цели была предложена семантическая компонентная архитектура (SCA). То есть, адаптация компонентов архитектуры в соответствии с принципами SW, в которых данные, структура и правила прикладной логики тесно связаны между собой. SCA должен координировать подключаемые "компоненты", которые "обернуты" оболочкой и могли бы взаимодействовать со следующими типами: схемами; ограничениями; правилами вывода; онтологиями; определениями путей; программным кодом; спецификациями вывода; информацией о конфигурациях Abox; ссылками к внешним источникам данных.

SCA должна включать средства определения модели пути. Предварительный обзор существующих механизмов определения путей показывает, что SPARQLeR [9] расширение SPARQL [4] может быть лучшим кандидатом для адаптации SCA. SPARQLeR предназначена для запросов по семантический ассоциации. Запрос в SPARQLeR сосредоточен на построении шаблонных путей, включая в себя неориентированные и направленные, пути направления которых, задаются определенными свойствами.

2.4 Проект EPOCH и AMA для библиотеки культурного наследия

EPOCH представляет собой сеть из более ста европейских культурных институтов, объединение

их привело к повышению качества и эффективности использования информационно-коммуникационных технологий для культурного наследия.

При интеграции ЕБ возникает ряд проблем.

Первая проблема заключается в том, что каждый справочник имеет свою поисковую систему и использует свою грамматику метаданных для описания и индексации данных, в частности, она никогда не будет работать на других системах. Ни одна из этих систем метаданных не может проанализировать всю информацию на веб-сайте, если мы не будем делать их доступными через машинно-читаемую форму с использованием RDF [17].

Вторая проблема касается непосредственно информации: огромное количество различных форматов, используемых для индексирования данных, является большим препятствием на пути к интеграции, и должны быть серьезно проанализированы. Даже если мы ограничиваем наши усилия исключительно для архивов культурного наследия, (например, базы данных музеев и коллекций, археологические раскопки, отчеты, доклады и другие неструктурированные данные), мы вынуждены признать, что информация, также является гетерогенной.

Чтобы создать единый концептуальный слой, семантическая информация должна быть взята из базы данных, HTML-страниц, описательных текстов, метаданных и представлена в стандартном формате, с целью получения концептуального содержания информации, создав концептуальный мапинг[1]. Как только концептуальный слой для данных и метаданных готов семантическая информация будет храниться в контейнере, основанном на RDF и онтологии.

Для упрощения и доступности процесса отображения в проекте AMA было разработано программное обеспечение AMA Mapping Tool. Этот инструмент способствует сопоставлению различных моделей данных с разной структурой, в том числе и робота с неструктурированными данными (текстовое описание). Этот мапинг основывается на известной онтологии CIDOC CRM.

Для неструктурированных документов используется ПО AMA TextTool.

Большая часть информации для наполнения CIDOC CRM-онтологии получается из текстов вручную. Для этого в рамках проекта разработано ПО AMA TextTool, которое предназначено для полуавтоматического кодирования археологических текстов в CIDOC-CRM. Даное ПО работает на понятиях и методиках компьютерной лингвистики. AMA TextTool реализует KWIC (ключевое слово в контексте). ПО используется для поиска слова, фразы или шаблонов слов и, возможно, XML-разметки в тексте. Пользователи могут затем проанализировать текст и знаки, которые представлены в следствие KWIC и отметить согласованные элементы. Система затем вставляет соответствующие отметки в тексты файла (ов).

2.5 Мапинг данных культурного наследия в CIDOC-CRM.

На данный момент несколько описательных стандартов уже отражено в CIDOC-CRM. Например, MIDAS стандарт данных Великобритании для информации об исторической среде, разработанной в интересах Форума информационных стандартов в области охраны наследия (FISH).

В введении к справочнику CIDOC CRM его авторы отмечают, что "поскольку прогнозируемая сфера применения CRM является подмножеством реального мира, и поэтому потенциально бесконечна", то модель была разработана для расширения посредством связей с совместимыми внешними типами иерархий.

В этом смысле "совместимость расширения с CRM означает, что данные, структурированные в соответствии с расширением, должны также оставаться правильными, как экземпляры класса CRM ".

В документации к CRM-CIDOC описано целый ряд процедур, которые можно использовать для расширения, придерживаясь выше поставленных требований:

- существующие классы высокого уровня могут быть расширены через подкласс или динамически с использованием типа иерархии.
- существующие свойства высокого уровня могут быть расширены с помощью структурированных подсвойств, а в некоторых случаях, динамично, с использованием атрибутов свойств, позволяющих подтипы.
- дополнительная информация, которая выходит за рамки семантики формально определенной в CRM, и может быть записана как неструктурированные данные, используя E1.CRM_Entity. P3.has_note: E62.String.

Начальные и целевые структуры, возможно, не всегда совпадают: в этом случае новая модель (источник), которая согласовывается с CIDOC CRM-иерархией будет создана путем явной декларации некоторых понятий, которые не явные в источнике, в соответствии с аксиомами/путями, описывающих структуру и иерархию CIDOC-CRM. Это своего рода процесс анисоморфизм[2] (anisomorphic), который изменяет очевидную структуру первоначального источника.

[1] метод для представления знаний в виде графов

[2] различия в семантический сфере применения терминов, относящихся к реальной жизни: например, английский и русский языки являются анисоморфизм в том, что касается терминологии цвета. Английский определяет светло-голубой (light blue) и темно-синего цвета (navy blue), как оттенки одного цвета, но русский трактует как не связные оттенки различных цветов

Это показывает, что отображение не является простым вопросом, или линейным процессом и требует дисциплинарной компетенции, а также глубокое понимание неявных предположений о модели исходного источника.

В рамках выполнения проекта было осуществлено гармонизацию с CIDOC-CRM доменных онтологий AMICO, DC, EAD, FRBR, TEI. Также было осуществлено расширение и специализация CIDOC-CRM с дополнительными прикладными онтологиями X3D и MPEG7, а также осуществлено отображение к CIDOC-CRM онтологии задач MIDAS, English Heritage, ICCD, PERSEUS. Для осуществления отображения использовалось ПО AMA (Archaeological Mapping Tool) [7].

3. Семантическая аннотация

Для научных исследований предложенные решения для интеграции библиотек, несомненно, важны и интеграция результатов научных экспериментов со знаниями, которые представлены в электронных библиотеках есть перспективным направлением. Опорой в этом направлении мы считаем технологию Semantic Web, а именно аннотация контента.

Как следует с обзора, обязательным этапом интеграции информации в электронной библиотеке есть перенос коллекции в RDF. В случаи структурированных кратких описательных метаданных (например, в рамках ДЯ) этот процесс возможно автоматизировать. Но для возможности автоматически анализировать содержания документа, таких аннотаций явно недостаточно. Поэтому в последнее время большое внимание уделяется более подробному раскрытию смысла контента через аннотации. Другими словами для анализа научных данных интеграция схем метаданных является не достаточной.

При анализе научных данных необходимо подавлять разного типа гетерогенность. Интеграция позволит объединить основные сведения из различных электронных архивов и других научных источников, которые могут быть пересмотрены и в которых возможно осуществить поиск, как в единой целой электронной библиотеке.

Хотя с момента появления Semantic Web прошло уже около 10 лет, тем не менее, должной популярности в широких массах эта технология не набрала. Виной тому множество рекомендаций и стандартов, которые существуют в данном направлении. С одной стороны нет удобных приложений, которые работали бы с RDF, с другой стороны отсутствуют RDF данные и онтологии.

Следующим этапом мы считаем более глубокое проникновение семантических технологий в электронные библиотеки, тому есть несколько причин. Во-первых, сейчас уже создано достаточное количество онтологий в различных предметных областях, например, Basic Formal Ontology [http://www.ifomis.org/bfo/], CIDOC Conceptual Reference Model [http://cidoc.ics.forth.gr/], Open Biomedical Ontologies [http://www.obofoundry.org/] и т.д. Во-вторых, разработано ряд приложений, которые способствуют внедрению Semantic Web на практике. Важным этапом на пути интеграции информации в Semantic Web есть принятия в качестве рекомендации языка запросов SPARQL (W3C Recommendation, January 15, 2008) и рекомендации по повторному использованию RDF-данных в XHTML - RDFa (W3C Recommendation, October 18, 2008).

В электронных библиотеках очень хорошая среда для наполнения Semantic Web посредством семантического аннотирования. Электронная библиотека хороша тем, что данные в ней структурированы с помощью метаданных. Однако метаданные хоть и представляются в машиночитаемом формате, но не дают полного представления об контенте информационного ресурса которые они описывают.

Семантическая разметка или аннотирование представляет собой явное описание семантики контента ресурса при помощи понятий семантической модели (онтологии или словаря). Такое явное описание семантики выполняется указанием четкого соответствия между определенной частью контента ресурса и его семантикой, описанной в семантической модели. Аннотирование при этом базируется на RDF.

Сегодняшние Web-ресурсы разрабатываются по большей части для использования людьми. Несмотря на постепенное появление в сети данных, предназначенных для машинного восприятия, эти данные в основном распространяются отдельным файлом в определенном формате. Притом соответствие машинной версии человеческому представлению весьма ограничено. Как следствие, Web-браузеры могут обеспечить пользователей лишь минимальной поддержкой в анализе и обработке сетевых данных, ведь браузеры только представляют информацию. Технология RDFa [18] позволяет сопроводить графические данные машиночитаемыми подсказками с помощью набора XHTML-атрибутов. RDFa — это способ выражения RDF-данных в XHTML, в рамках которого данные, предназначенные для человека, используются повторно.

Примером использования RDFa может служить закладывания фрагмента кода, который описывает название и автора статьи, которая расположена в электронной библиотеке. При описании используется схема метаданных Дублинского Ядра (xmlns:dc=http://purl.org/dc/elements/1.1/).

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<head profile="http://www.w3.org/2003/g/data-view">
<title> Доповідь про http://oai.org.ua</title>
</head>
<body>
  <h1>Ресурс http://oai.org.ua</h1>
```

```
        <dl about="http://eprints.zu.edu.ua/2648/">
        <dt>Назва доповіді</dt>
        <dd property="dc:title">Інтеграція наукових
електронних бібліотек України: всеукраїнський портал
збору та пошуку метаданих http://oai.org.ua</dd>
        <dt>Автор</dt>
<dd property="dc:creator">Новицький, О.В.</dd>
        </dl>
 </body>
</html>
```

Рис. 1 Фрагмент XHTML кода ЭБ с разметкой RDFa.

Сам по себе механизм RDFa был бы малоинтересен, хоть определяет семантику контента. Необходимым условием является возможность извлечение семантической аннотации со страниц. Такой механизм к счастью разработан и носит названия GRDDL - **G**leaning **R**esource **D**escriptions from **D**ialects of **L**anguages (http://www.w3.org/TR/grddl/).

При помощи GRDDL возможно однообразно извлекать микроформатированный контент. Спецификация GRDDL определяет разметку на основе существующих стандартов для объявления о том, что XML документ включает в себя данные совместимые с RDF, а также ссылку на алгоритм (как правило, представленный в XSLT), для извлечения данных из документа.

Разметки содержат определения пространства имен общего назначения для XML-документов, а также ссылку на профиль отношений для использования в валидных XHTML документов.

Ниже представлен фрагмент XHTML кода в соответствии с GRDDL.

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<head profile="http://www.w3.org/2003/g/data-view">
<link rel="transformation" href="RDFaRDF.xsl"/>
<title> Доповідь про http://oai.org.ua</title>
</head>
<body>
  <h1>Ресурс http://oai.org.ua</h1>
        <dl about="http://eprints.zu.edu.ua/2648/">
        <dt>Назва доповіді</dt>
        <dd property="dc:title">Інтеграція наукових
електронних бібліотек України: всеукраїнський портал
збору та пошуку метаданих http://oai.org.ua</dd>
        <dt>Автор</dt>
        <dd property="dc:creator">Новицький, О.В.</dd>
        </dl>
 </body>
</html>
```

Рис. 2 Фрагмент XHTML кода с разметкой RDFa и GRDDL

При обработке преобразования данного фрагмента средствами XSLT будет получена модель данных и представлена в RDF с помощью XML Рис. 3.

```
<rdf:RDF xmlns:h="http://www.w3.org/1999/xhtml"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
        <rdf:Description rdf:about="">
<transformation xmlns="http://www.w3.org/1999/xhtml"
rdf:resource="RDFa2RDFXML.xsl"/>
        </rdf:Description>
        <rdf:Description
rdf:about="http://eprints.zu.edu.ua/2648/">
        <dc:title
xmlns:dc="http://purl.org/dc/elements/1.1/">Інтеграція
наукових електронних бібліотек України: всеукраїнський
портал збору та пошуку метаданих http://oai.org.ua</dc:title>
        </rdf:Description>
        <rdf:Description
rdf:about="http://eprints.zu.edu.ua/2648/">
        <dc:creator
xmlns:dc="http://purl.org/dc/elements/1.1/">Новицький,
О.В.</dc:creator>
        </rdf:Description>
</rdf:RDF>
```

Рис. 3 RDF представлен с помощью XML

Стоит обратит внимание, что GRDDL имеет возможность преобразования разметки RDFa для которой, например, используется схема данных Дублинского Ядра, непосредственно в другие схемы метаданных, таких как CIDOC-CRM.

Такой подход применим только к веб-документам, но возможно получится применение данной технологии к мультимедиа форматам.

Еще одним применением данного подхода может быть процесс который предложен в [8].

Пример автоматического внесения документов (с возможностью распределенности) и построения индексов. Идея заключается в GRDDL обработке источников документов и извлечения встроенного RDFa для подключения в хранилища RDF. Далее SPARQL запросы выбирали бы с этого хранилища соотвествующие результаты, которые были бы представлены в виде автоматически генерируемой веб-страницы Рис. 4.

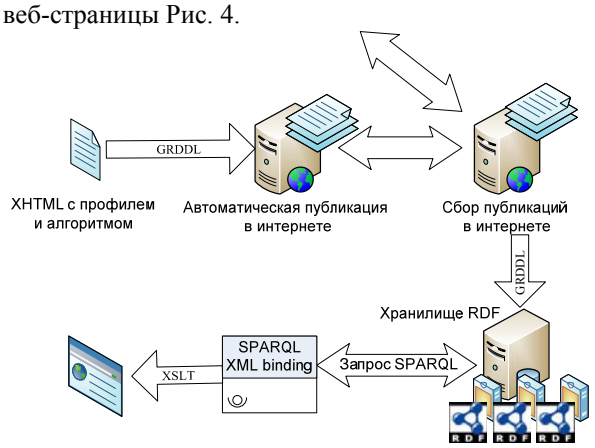

Рис. 4 Пример RDFa и GRDDL

Литература

[1] *Mapping, Embedding and Extending: Pathways to Semantic Interoperability The Case of Numismatic Collections.* . **Andrea D' Andrea, Franco Niccolucci.** Tenerife : б.н., 2008. Semantic Interoperability in the European Digital Library, Proceedings of the First International Workshop. стр. 63-75.

[2] *SeRQL: A Second Generation RDF Query Language.* **Broekstra J, Kampman A.** Proc SWAD-Europe Workshop on Semantic Web Storage and Retrieval 2003.

[3] *Answering XML queries over heterogeneous data sources.* **Daniela Florescu, Ioana Manolescu, and Donald Kossmann.** б.м. : Morgan Kaufmann. 27th International Conference on Very Large Data Bases (VLDB 2001). стр. 241-250.

[4] **Eric Prud'hommeaux Andy Seaborne.** SPARQL Query Language for RDF. *W3C.* [В Интернете] 2008 г. http://www.w3.org/TR/rdf-sparql-query.

[5] **Ismail Fahmi, Junte Zhang, Henk Ellermann, Gosse Bouma.** SWHi System Description: A Case Study in Information Retrieval, Inference, and Visualization in the Semantic Web. *The Semantic Web: Research and Applications, 4th European Semantic Web Conference.* Innsbruck, Austria : Springer, 2007, стр. 769-778.

[6] **Felicetti A. M. Ioannides, D.Arnold, F. Niccolucci, K. Mania (eds.).** *MAD – Management of Archaeological Data.* Budapest : б.н., 2006,. стр. 124 – 131, The e-volution of Information Communication Technology in Cultural Heritage – Project papers.

[7] *Semantic Maps and Digital Islands: Semantic Web technologies for the future of Cultural Heritage Digital Libraries.* **A. Felicetti, H. Mara.** Tenerife, Spain : б.н., 2008. SIEDL 2008: Semantic Interoperability in the European Digital Library. стр. 51-62.

[8] **Gandon, Fabien.** Digital library example. *Institut National de Recherche en Informatique et en Automatique / Centre de recherche Sophia Antipolis - Mŭditerranŭe.* [В Интернете] 2009 г. http://www-sop.inria.fr/acacia/personnel/Fabien.Gandon/tmp/grddl/rdfaprimer/PrimerRDFaSection.html.

[9] *SPARQLeR: Extended Sparql for Semantic Association Discovery.* **Krys Kochut, Maciej Janik.** 2007. 4th European Semantic Web Conference (ESWC2007). http://www.eswc2007.org/pdf/eswc07-kochut.pdf.

[10] *Data integration: a theoretical perspective.* **Lenzerini, Maurizio.** New York : ACM Press, 2002. 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2002).

[11] *Locating data sources in large distributed systems.* **Leonidas Galanis, Yuan Wang, Shawn R. Jeffery, and David J. DeWitt.** б.м. : Morgan Kaufmann. 29th International Conference on Very Large Data Bases (VLDB 2003). стр. 874–885.

[12] *Combining artificial intelligence and database for data integration.* **Levy., Alon Y.** б.м. : Berlin/Heidelberg, LNCS 1600, Springer. In Artificial Intelligence Today: Recent Trends and Developments. стр. 249–268.

[13] **Lin, Yun.** *Semantic Annotation for Process Models: Facilitating Process Knowledge Management via Semantic Interoperability.* б.м. : Department of Computer and Information Science Norwegian University of Science and Technology. 7491.

[14] *Modular, Best-Practice Solutions for a Semantic Web-Based Digital Library Application* . **Martinez, Andrew Russell Green and Jose Antonio Villarreal.** Tenerife, Spain : б.н., 2008. Proceedings of the Workshop on Ontologies: Reasoning and Modularity (WORM-08).

[15] *Porting Cultural Repositories to the Semantic Web.* **Omelayenko, B.** Tenerife, Spain : б.н., 2008. Proceedings of the First Workshop on Semantic Interoperability in the European Digital Library (SIEDL-2008). стр. 14-25.

[16] **OpenAcademia.** [В Интернете] www.openacademia.org.

[17] **RDF Core Working Group.** Resource Description Framework (RDF). *Resource Description Framework* . [В Интернете] W3C. http://www.w3.org/RDF/.

[18] RDFa Primer. Bridging the Human and Data Webs. **W3C Working Group.** [В Интернете] W3C. http://www.w3.org/TR/xhtml-rdfa-primer/

[19] **Shvaiko, Pavel.** *Iterative schema-based semantic matching.* Informatica e Telecomunicazioni. Trento : University of Trento, 2006. Technical Report DIT-06-102.

[20] *Model independent assertions for integration of heterogeneous schemas.* **Stefano Spaccapietra, Christine Parent, and Yann Dupont.** 1, 1992 г., VLDB Journal, Т. 1, стр. 81–126.

[21] **Wielemaker, J., Hildebrand, M., Ossenbruggen, J.R. Van.** Using Prolog as the fundament for applications on the semantic web (2008). *Proceedings of the 2nd Workshop on Applicatiions of Logic Programming and to the web, Semantic Web and Semantic Web Services.* Porto, Portugal : б.н., 2007.

A review of some of the integration of heterogeneous resources in digital libraries

O. Novytskyi

This paper is devoted to the integration of information. An overview of current projects for the integration of information in digital libraries. Problems that arise when integrating data, coupled with the fact that the resources described in the metadata different way with different semantics. Semantic annotation of resources is the future of the Internet and digital libraries. Recently, there are a number of technologies Semantic Web, which can automatically solve the problem of integration of annotated web documents. In the robot made a review of current projects on semantic integration in digital libraries, and shows an example of the transformation of semantic web annotations RDFa document to the RDF for the DL.