

Организация обмена RDF-документами в распределенной информационной системе*

Р.К. Азизов, А.И. Привезенцев[†], А.З. Фазлиев[‡]

Аннотация. В работе описан подход к организации обмена RDF-документами в распределенной информационной системе. Программная реализация обмена метаданными выполнена средствами Java. Обсуждается возможность создания типового веб-сервиса для обмена метаданными на основе предложенного подхода.

Введение

Построение информационных систем (ИС) предполагает наличие информационного ресурса имеющего в своей структуре данные и метаданные [1]. Важность метаданных существенно возрастает в распределенных информационных системах. Дело в том, что зачастую авторы ресурсов не дают доступ к данным, но могут позволять работу с метаданными, в ситуациях, когда объемы данных значительны и их передача коммерчески не выгодна, метаданные приобретают еще большее значение. В структурированных информационных системах, в основе которых, как правило, лежат базы данных, обмен метаданными осуществляется средствами БД. В слабоструктурированных ИС, к числу которых относятся информационные системы с доступом по сети Интернет, метаданные описываются в рамках подходов DC, GILS, PRISM и т. д. [2]. Последние два года консорциум W3C подготавливал рекомендацию для описания метаданных (RDF), которая была принята в феврале 2004 г. [3-5].

В нашей работе рассматривается подход для обмена метаданными описываемыми средствами RDF в рамках распределенной информационной системы. Суть подхода состоит в сведении всех RDF-описаний информационных ресурсов распределенной ИС в единый реестр и организации обмена на основе pull-технологии.

1. Модель

Распределенная информационная система в нашей работе рассматривается как набор сайтов доступных в сети Интернет. Информационные ресурсы этих сайтов представлены html- и xml-документами и соответствующими RDF-документами, содержащими метаданные. Соответствующие XML- и RDF-схемы являются неотъемлемыми компонентами сайта. XML-документы формируются с помощью HTML-форм для ввода данных. RDF-документы, содержащий метаданные создаются либо автоматически при формировании типовых XML-документов или вручную при создании HTML-документов. Важной особенно-

*Работа выполнена при финансовой поддержке РФФИ, грант 02-07-90139 и интеграционного проекта СО РАН № 132.

[†]Томский государственный университет систем управления и радиоэлектроники.

[‡]Институт оптики атмосферы СО РАН, Томск.

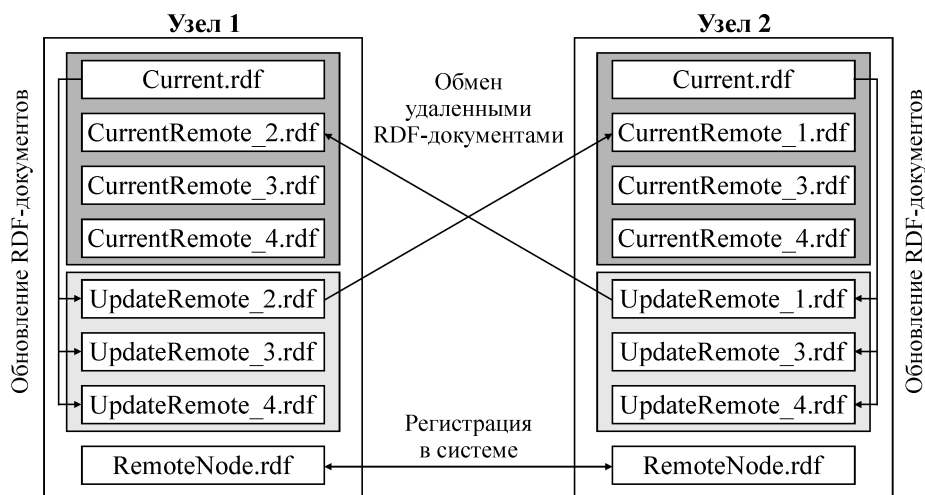


Рис. 1. Структура реестра RDF-документов

стью является тот факт, что RDF-схема может формироваться самостоятельно разработчиками сайтов в рамках той предметной области на которую ориентирована распределенная ИС. Часть этой схемы могут быть и типовые схемы метаданных (CIMI, DC, GILS и т. д.). RDF-описание всех документов сайта сводятся в единый реестр документов данного сайта. Фактически реестр представляет собой описание коллекции типовых XML-документов. Задача, решенная нам, заключалась в организации обмена метаданными между реестрами метаданных каждого сайта распределенной информационной системы. На рис. 1 представлена структура реестров и схема их обновлений. Метаданные разных сайтов имеют атрибут принадлежности (CurrentRemote-n.rdf) по которому в общем реестре их можно выделять. При создании в ИС метаданных они заносятся в Current.rdf и все UpdateRemote.rdf части реестра предназначенные для фиксации обновлений. При запросе на обновления от удаленных узлов метаданные передаются соответствующим узлам и заносятся в части реестра CurrentRemote-n.rdf. После проведения обмена часть реестра UpdateRemote.rdf становится пустой. Список узлов, которым разрешено обмениваться метаданными и атрибуты узлов, помещен в часть реестра RemoteNode.rdf.

В рамках выбранной нами модели цикл создания информационного ресурса в распределенной ИС состоит из следующих этапов: ввод документа осуществляется через HTML-формы, при этом формируются XML-документ (после проверки на соответствие и достоверность) и RDF-документ. Последний заносится в реестр текущего сайта и реестры обновлений. По мере обращения к узлу эти обновления передаются всем узлам распределенной ИС.

2. Реализация сервиса обмена метаданными

Реализация описанной выше модели требует создания ряда приложений, обеспечивающих функционирование сервиса обмена метаданными. Эти приложения реализуют функции:

1. Создание и редактирование XML-документа.
2. Создание и редактирование RDF-документа.
3. Формирование реестра метаданных.
4. Просмотр метаданных.
5. Обмен метаданными.
6. Регистрация узлов (сайтов).

Для реализации информационной системы обмена метаданными была разработана описанная выше модель и применены следующие технологии. Для создания и редактирования XML документа нами применялась объектная модель документа (DOM) и язык XSL Transformations (XSLT) [6]. Поддержка этих технологий реализована на языке программирования Java. Поддержка DOM реализована в стандартных библиотеках языка, а для XSLT существует множество сред исполнения XSLT преобразований (Saxon, Xalan и т. д.), наш выбор остановился на Xalan [7]. Для создания и редактирования RDF документа применен пакет Jena [8] реализованный на Java и упрощающий работу с рекомендациями RDF консорциума W3C. Для формирования реестра метаданных и его просмотра использована технология сервлетов Java, работающих под веб контейнером Tomcat [9] использующего описанные выше пакеты. И для удаленного обмена метаданными мы использовали пакет Axis [10] – реализация рекомендации SOAP [11–13] и рабочего проекта WSDL консорциума W3C [14, 15].

На рис. 2 представлен вариант организации обмена метаданными между сайтами распределенной ИС с применением технологий веб-сервисов.

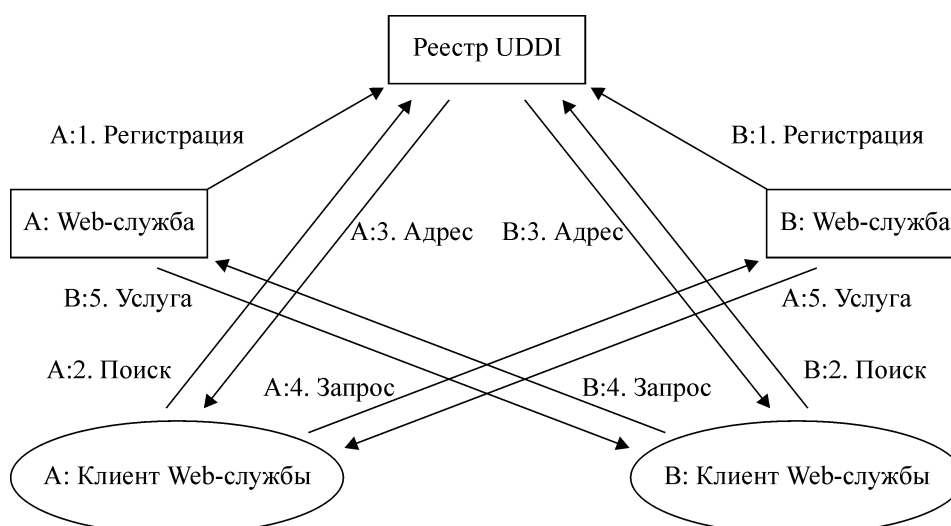


Рис. 2. Организация обмена метаданными с помощью веб-сервисов

Заключение

В работе представлен подход по созданию и поддержке единого реестра RDF-документов в распределенной информационной системе. Программная реализация сервиса обмена метаданными выполнена в рамках технологий Java.

Список литературы

- [1] Коголовский М.П. Перспективные технологии информационных систем. – М.: ДМК Пресс, 2003. – 288 с.
- [2] RDF Primer, W3C Recommendation 10 February 2004. – <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [3] RDF Semantics W3C Recommendation 10 February 2004. – <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>.
- [4] RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004. – <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>.
- [5] GRDF/XML Syntax Specification (Revised) W3C Recommendation 10 February 2004. – <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>.
- [6] XSL Transformations (XSLT) Version 1.0 W3C Recommendation 16 November 1999. – <http://www.w3.org/TR/1999/REC-xslt-19991116/>.
- [7] Xalan-Java version 2.6.0. – <http://xml.apache.org/xalan-j/>.
- [8] Jena 2 – A Semantic Web Framework. – <http://www.hpl.hp.com/semweb/>.
- [9] Apache Tomcat. – <http://jakarta.apache.org/tomcat/>.
- [10] WebServices – Axis. – <http://ws.apache.org/axis/>.
- [11] SOAP Version 1.2 Part 0: Primer. – <http://www.w3.org/TR/2003/REC-soap12-part0-20030624/>.
- [12] SOAP Version 1.2 Part 1: Messaging Framework. – <http://www.w3.org/TR/2003/REC-soap12-part1-20030624/>.
- [13] SOAP Version 1.2 Part 2: Adjuncts. – <http://www.w3.org/TR/2003/REC-soap12-part2-20030624/>.
- [14] Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language W3C Working Draft 26 March 2004. – <http://www.w3.org/TR/2004/WD-wsdl20-20040326/>.
- [15] Web Services Description Language (WSDL) Version 2.0 Part 2: Message Exchange Patterns W3C Working Draft 26 March 2004. – <http://www.w3.org/TR/2004/WD-wsdl20-patterns-20040326/>.