

Інформаційні технології в технічних системах

УДК 004.738.52

В.Н. Анохин, А.И. Бец

Харьковский национальный экономический университет, Харьков

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ ОПИСАНИЯ РЕСУРСОВ RDF В СЕМАНТИЧЕСКОЙ ПАУТИНЕ

Практически параллельно с работами по стандартизации XML консорциум W3C сформулировал новое понятие – Semantic Web – то, каким он видит будущее глобальной сети, и инициировал исследования в этом направлении. В основе предполагаемого развития интернета лежит способность машин не только читать, но и понимать содержание Интернет-ресурсов. На пути к осуществлению поставленной цели необходимо решить несколько задач. С одной стороны программы должны понимать язык соответствующей предметной области, с другой – должны уметь сопоставлять связанные термины различных предметных областей. Технология описания ресурсов – Resource Description Framework (RDF) – является тем самым средством, на которое WWW Консорциум возлагает надежды на решение указанных задач, связанных с описанием семантики.

Ключевые слова: технология описания ресурсов (Resource Description Framework - RDF), Semantic Web (семантическая паутина), всемирная сеть, глобальная сеть, связь, ресурс.

Вступление

На наших глазах технология World Wide Web (WWW) пережила уже два этапа развития: вначале предоставляя информацию в виде вручную сгенерированных HTML-страниц, затем появилась возможность генерировать страницы из баз данных, хранящихся на Web-серверах. Однако оба эти этапа были ориентированы на обработку информации человеком-специалистом и читателей – пользователей компьютеров. Чтобы противостоять неуклонному и неуправляемому расширению «всемирной паутины», которая стала угрожающе запутываться, ведущими учеными и разработчиками был выдвинут лозунг «превратить информационное Интернет-пространство в пространство знаний». Web следующего, третьего поколения, должна обеспечивать машинную обработку информации с учетом ее семантики (machine accessible semantics).

Новую парадигму Web ее создатели и идеологи назвали Semantic Web, делая тем самым акцент на требовании представления семантики в Web, понятного для программных агентов. Росту популярности и широкому распространению технологий Semantic Web способствует разработка и внедрение WWW-консорциумом (W3C) ряда инициатив, связанных со стандартизацией синтаксической и семантической разметки электронных документов, особенно технологий XML и RDF, поддерживающих семантическую совместимость [1].

The Resource Description Framework (RDF) является языком общего назначения для представ-

ления информации в Web [2]. RDF применяется для того, чтобы предоставить информацию о ресурсах в терминах простых свойств и значений этих свойств.

Основная часть

Благодаря семантической паутине компьютерные программы могут использовать возрастающие объемы структурированных данных, распределено и децентрализованно рассеянные по сети в настоящее время. RDF представляет собой абстрактную модель, обеспечивающую способ разбиения знаний на дискретные части.

Главное отличие Semantic Web от Web состоит в том, что каждая страница семантической сети содержит информацию на двух языках: на обычном, понятном человеку и показываемом браузером, и на специальном, информация на котором скрыта от человеческих глаз, но понятна интеллектуальным программам-агентам, роботам. Этот специальный язык описывает представленный на странице материал с помощью тэгов и атрибутов, понятных машинам.

В основе семантической сети лежат три принципа: агрегация, безопасность и логика. Агрегация означает совместное использование данных. Подобно тому, как гипертекст является неотъемлемой частью WWW и благодаря этому «всё можно связать со всем», в Semantic Web при решении поставленной задачи могут быть использованы любые данные. Для этих данных будет создана соответствующая семантическая информация (онтологии), позволяющая использовать их надлежащим обра-

зом. В основу безпеки, забезпечуючої довіру до семантичної мережі, покладено цифрові підписи, які можуть використовуватися агентами і комп'ютерами для перевірки того, що інформація отримана з надійного джерела, наприклад від якого-то публічного сервісу або персонального агента іншого надійного користувача. Логіка – це набір правил описання інформаційної

структури даних, протоколи і мови описання сторінок. Саме логіка надає семантичній мережі правила виводу для проведення висновків і методики вибору тактик виконання операцій з даними, щоб отримати відповіді на запити [3]. Ключовим аспектом технології семантичної мережі також є багаторівнева структура, частиною якою і є технологія RDF (рис.1).

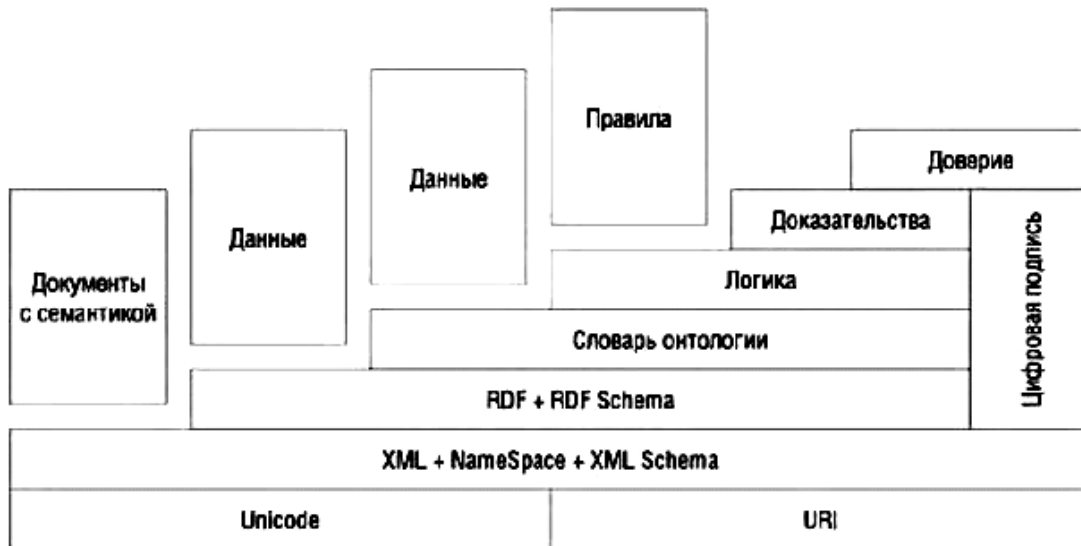


Рис.1. Архитектура семантической сети

Принцип побудови відносин між мережевими ресурсами в специфікації RDF передбачає наявність трьох компонентів – об'єкта, атрибута і значення (аналогічних класическій схемі "підлежаче – сказуване – доповнення"). Кожному елементу даної ланочки (триплету) присвоюється ідентифікатор (URI), з допомогою якого при вказанні одного з ланочок можна автоматично відновити всю ланочку в цілому. З триплетів мови RDF можуть формуватися мережі взаємопов'язаних об'єктів.

Оскільки RDF використовує URI-ідентифікатори для кодування інформації в документі, це забезпечує те, що кожне поняття буде прив'язано до єдиного визначення, яке можна знайти в Мережі.

Базовий будівельний блок в RDF – трійка "об'єкт – атрибут – значення" часто записують у вигляді A(O,V), де O – об'єкт, A – атрибут з значенням V. RDF дозволяє змінювати місцями об'єкти і значення [4].

RDF – це універсальний метод розділення знання на маленькі частини, відповідно до деяких правил, які враховують семантику (смысл) цих частин. Суть в тому, що такий метод повинен бути достатньо простим, щоб з його допомогою можна було описати будь-який факт, і достатньо структурованим, щоб представити

факт в такій формі, в якій комп'ютерні програми зможуть виконувати корисні дії з знаннями, вираженими в форматі RDF.

Інформаційні ресурси описуються за допомогою сукупності термінів. Семантика, зміст самих термінів і словників термінів, фіксується з допомогою глобальних універсальних імен URI. В RDF передбачено словник термінів, названий RDF-схемою, який визначає, які терміни можуть бути використані в RDF-утвердженнях про властивості ресурсів, представляє ієрархію понять предметної області, описує характеристики кожного з термінів.

Модель даних сама по собі всього лише скелет. Для того щоб описання отримало якийсь зміст, необхідно використати словники, які задаються за допомогою додаткової технології – RDF Schema, яка грає для RDF ту ж роль, що і схема для XML.

Під словником слід розуміти сукупність ресурсів, використовуваних для описання властивостей інших ресурсів; класів ресурсів, які можуть бути описані за допомогою заданих властивостей; і обмеження, налагаються на їх значення або набори допустимих значень. При цьому класи можуть складатися з відношеннями "підклас" і аналогічно властивості можуть бути пов'язані відношеннями "підвласність" [2, 5].

Реальное значение RDF невозможно оценить, пока он используется для внутренних целей отдельно взятого приложения. Польза от внедрения RDF будет тогда, когда он станет средством межпрограммного взаимодействия, обмена данными, когда машины получают способность комбинировать информацию, полученную из различных источников, тем самым, получая какую-то новую информацию. Чем больше приложений в Интернете смогут работать с данными, тем выше станет их ценность [6]. Одним из наиболее ярких примеров существующего общеизвестного словаря является Dublin Core – набор атрибутов для метаописания ресурсов с целью каталогизации. [7].

Таким образом, видно, что с помощью RDF можно достаточно эффективно описывать различные взаимоотношения между различными объектами.

Для использования возможностей языка RDF в аннотировании Интернет-ресурсов, представленных в различных XML-форматах, комитетом W3C ведется разработка стандарта RDF/A. Стандарт RDF/A описывает способ, с помощью которого RDF-данные могут помещаться внутри XHTML-файлов. В результате использования данного стандарта появляется возможность организации эффективного поиска по содержимому различных Интернет-ресурсов, т.к. их содержимое доступно для понимания программным службам. Его основные преимущества заключается в том что, описав данные с использованием RDFa, агент пользователя, обрабатывая веб - страницу, может сообщать дополнительную информацию или делать определенные действия, например, сообщать информацию о правах на документ или добавлять события в личный календарь пользователя. Большой плюс использования RDF также в том, что информация, описанная в теле документа, не дублируется. В частности с помощью RDF/a можно разметить информацию о персоне в основной части веб-страницы, что помогает Google лучше находить и отображать ее в результатах поиска. Для примера сделем небольшую аннотацию своей контактной информации:

```
<div xmlns:v="http://rdf.data-vocabulary.org/#"
  typeof="v:Person">
  Мое имя <span property="v:name">Александр
  Бец</span>,
  но друзья зовут меня <span
  property="v:nickname">Alex</span>.
  Вот моя главная страница:
  <a href="http://www.rdfa.ucoz.com"
  rel="v:url">www.rdfa.ucoz.com</a>.
  Я живу в
  <span rel="v:address">
```

```
<span typeof="v:Address">
  <span property="v:locality">Харьков</span>,
  <span property="v:region">Украина</span>
</span>
</span>
и работаю <span property="v:title"> аналитиком
данных</span>
в <span property="v:affiliation">DEV-PRO</span>.
</div>
```

Пример начинается с объявления пространства имен с помощью xmlns. Этот атрибут указывает на пространство имен, содержащее словарь (список сущностей и их компонентов). Объявление пространства имен

```
xmlns:v="http://rdf.data-vocabulary.org/#"
```

можно использовать каждый раз при разметке на странице личных данных, отзывов, товаров или мест. Кроме того, typeof="v:Person" в первой строке указывает, что размеченное содержание представляет сущность "человек". Перед именем свойства typeof ставится префикс v: (typeof="v:Person"). Каждое свойство человека (например, имя или псевдоним) помечается с помощью property. Перед именем свойства ставится префикс v: (). Для указания URL используется rel вместо property следующим образом: www.rdfa.ucoz.com . Префикс v. rel служит для указания связи между сущностями, в данном случае между сущностью "человек" и сущностью "веб-страница". В общем случае RDFa использует простые атрибуты в тегах XHTML (обычно или <div>) для задания кратких и описательных имен сущностей и их свойств. В результате проведенных исследований, часть которых приведена в данной статье, можна сделать вывод, что информация представления в формате RDFa положительно влияет на результаты поиска в системе Google. Также зааннотированные сущности отображаются под результатами поиска страницы, что дает возможность быть более уверенным в релевантности запрошенной информации.

Выводы

Технология описания ресурсов – Resource Description Framework (RDF) – является тем самым средством, на которое WWW Консорциум возлагает надежды на решение указанных задач, связанных с описанием семантики. К сожалению RDF по прежнему не обрел широкого распространения. Причин этому несколько. С одной стороны в отличие от языка XML RDF не сразу получил широкую программную поддержку. С другой – RDF не имел такой первоочередной нацеленности на электронную коммерцию и до сих пор остается

преимущественно в области интересов исследователей. Третья причина видится в том, что текущий синтаксис RDF вызывает многочисленные споры и нарекания со стороны потенциальных пользователей данной технологии. По их мнению предложенные формы записи сложны и громоздки, что делает описание ресурсов не слишком удобным для применения.

Тем не менее, RDF представляется весьма интересной технологией и является важным шагом к созданию Semantic Web. Он продолжает привлекать внимание исследователей и, по мнению многих из них, несет в себе большой потенциал. Семантическая паутина не отменяет полезности Web-сервисов, а дополняет их. Однако, в то время, как задача большинства современных Web-сервисов - обеспечить коммуникацию между приложениями, семантика и использование технологи RDF решает более сложную проблему - повысить ценность информации, которую можно найти в сети Internet. Можно сказать, что это средство еще не известно широким массам разработчиков или только осознается ими, и есть надежда, что в будущем RDF займет предназначенное ему место.

Список литературы

1. Н.В. Рябова, С.С. Щербак Развитие технологий Semantic Web: обработка RDF-графов на основе XSLT [Электронный ресурс]. - Режим доступа: http://shcherbak.net/rdf_xslt_tech.

2. Стин Декер, Сергей Мельник, Франк ван Хермелен, Semantic Web: роли XML и RDF [Электронный ресурс]. - Режим доступа: http://www.publish.ru/os/1072570/text/180411/_p3.html.

3. Даниил Кальченко Семантическая сеть [Электронный ресурс]. - Режим доступа: <http://www.compress.ru/article.aspx?id=12195&iid=468>.

4. Дмитрий Ландэ Семантический веб: от идеи - к технологии [Электронный ресурс]. - Режим доступа: <http://poiskbook.kiev.ua/sw.html>.

5. Dan Brickley, R.V. Guha RDF Vocabulary Description Language 1.0: RDF Schema [Электронный ресурс]. - Режим доступа: <http://www.w3.org/TR/2002/WD-rdf-schema-20021112>.

6. Graham Klyne, Jeremy Carroll Resource Description Framework (RDF): Concepts and Abstract Data Model [Электронный ресурс]. - Режим доступа: <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829>.

7. Diane Hillmann Using Dublin Core [Электронный ресурс]. - Режим доступа: <http://dublincore.org/documents/usageguide>.

Поступила в редколлегию 26.03.2011

Рецензент: канд. экон. наук, проф. И.А. Золотарева, Харьковский национальный экономический университет, Харьков.

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ ОПИСУ РЕСУРСІВ RDF У СЕМАНТИЧНІЙ ПАВУТИНІ

В.М. Анохін, О.І. Бець

Практично паралельно з роботами по стандартизації XML консорціум W3C сформулював нове поняття - Semantic Web - те, яким він бачить майбутнє глобальної мережі, і ініціював дослідження в цьому напрямку. В основі передбачуваного розвитку інтернету лежить здатність машин не тільки читати, а й розуміти зміст Інтернет-ресурсів. На шляху до здійснення поставленої мети необхідно вирішити декілька завдань. З одного боку програми повинні розуміти мову соотвєтствующей предметної області, з іншого - повинні вміти зіставляти пов'язані терміни різних предметних областей. Технологія опису ресурсів - Resource Description Framework (RDF) - є тим самим засобом, на яке WWW Консорціум покладає надії на вирішення зазначених завдань, пов'язаних з описом семантики.

Ключові слова: технологія опису ресурсів (Resource Description Framework - RDF), Semantic Web (семантична павутина), всесвітня мережа, глобальна мережа, зв'язок, ресурс.

THE USE OF RESOURCES DESCRIPTION TECHNOLOGY RDF IN THE SEMANTIC WEB

V.N. Anohin, A.I. Bets

Almost in parallel with work on the standardization of XML W3C consortium has formulated a new concept - a PR Semantis - how he sees the future of the global network and has initiated research in this direction. At the core of the proposed development of the Internet is the ability of machines not only read but also to understand the content of Internet resources. On the way to the fulfillment of the objectives necessary to solve several problems. On the one hand, the program must understand the language of the subject domain, on the other - should be able to compare the related terms of different subject areas. Technology Resource Description - Resource Description Framework (RDF) - is the very means by which WWW Consortium hopes to address these challenges with a description of the semantics.

Keywords: Technology Resource Description (Resource Description Framework - RDF), Semantic Web (Semantic Web), a global network, global network, communications and resource.