

УДК 004.738.52

SEMANTIC WEB КАК НОВАЯ МОДЕЛЬ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ИНТЕРНЕТ

Ф.И. Андон, И.Ю. Гришанова, В.А. Резниченко

Институт программных систем НАН Украины,
03680 Киев, проспект Академика Глушкова, 40.
Тел.: (044) 526 5139, e-mail: reznich@isofts.kiev.ua

Описаны базовые концепции и архитектура Semantic Web, а также положение дел по разработке данного проекта по состоянию на конец 2007 года. Выделены проблемы, которые стоят перед мировым сообществом для дальнейшего развития Semantic Web.

Basic concepts and architecture of Semantic Web is described. State of the art concerning development of the project up to the end of 2007 year is outlined. The problems of future development of Semantic Web are noted.

Введение

Феномен World Wide Web стал возможным только благодаря практическому использованию набора широко распространенных стандартов на разных уровнях, что обеспечило интероперабельность данных. Современная тенденция развития Интернета заключается в переходе от документов, „читаемых компьютером” (machine readable) к документам, которые „понимаемы компьютером” (machine understandable).

Web разрабатывался как информационное пространство, полезное не только для коммуникации человека с человеком, но и как пространство, в котором смогут эффективно сотрудничать и компьютеры. Одно из главных препятствий на пути к этому состоит в том, что большая часть информации в Web предназначена для ее понимания человеком. Очевидно, что такая структура данных не может быть понятной для просматривающего веб-робота. Подход Semantic Web базируется на разработке языков для выражения информации в форме, пригодной для машинной обработки.

Идея Semantic Web была предложена в 1998 году Тимом Бернерсом-Ли (Tim Berners-Lee), который является изобретателем WWW, URI, HTTP и HTML.

Semantic Web представляет собой сеть информационных узлов, которые связаны друг с другом таким образом, чтобы имеющаяся информация могла легко обрабатываться компьютером. Его можно рассматривать как эффективный способ представления данных во Всемирной паутине, или как глобально связанную базу данных. Данный проект предлагает реализацию полной системы по автоматизированному созданию и хранению семантического ядра контента, предоставленного во Всемирной паутине.

Проект Semantic Web – это попытка собрать все устоявшиеся идеи и сделать так, чтобы они смогли работать вместе внутри сети Интернет. Для достижения этой цели используются стандарты, которые разработаны не только консорциумом W3C, но и другими организациями. Цель проекта – разрешить взаимодействовать этим стандартам между собой внутри децентрализованной системы без вмешательства человека.

Проект Semantic Web [1], начатый в 2001 году, на данный момент находится в стадии активной разработки, старается интегрировать в себя все уже имеющиеся на данный момент подходы, с целью создать действительно универсальное средство семантического поиска информации [2, 3]. Большое внимание отводится архитектуре и модели распределенной среды [4], архитектуре метаданных [5 – 8]. Как сказано в определении, которое представлено на домашней странице проекта – «Semantic Web является абстрактным представлением данных во Всемирной паутине, которое базируется на стандартах RDF и других стандартах, имеющих распространение. Проект разрабатывается Консорциумом W3C в содружестве с большим количеством исследователей, ученых и промышленных партнеров» [9].

«Semantic Web – это расширения текущего Web, в котором информация предоставляется с хорошо определенным значением, которое лучше разрешит компьютерам и людям работать вместе. ... Его идея в том, чтобы иметь данные в Web, определенные и связанные между собой таким образом, чтобы их можно было использовать для более эффективного исследования, автоматизации, интеграции и повторного использования в разных приложениях... эти данные могут быть общедоступными и обрабатываемыми автоматическими средствами так же, как и людьми» [2].

В рамках данного проекта задействованы такие передовые технологии, как агентно-ориентированный подход в программировании [10] – проект DAML+OIL (DARPA Agent Markup Language + The Ontology Inference Layer) [11 – 14], онтологии [15, 16], XML [17 – 19], RDF [20 – 22], и др. В настоящее время распространяется использование Web-агентов (в упрощенном виде веб-сервисов), которые разрабатываются как для частных задач, так и для создания ядра Semantic Web [23 – 28].

© Ф.И. Андон, И.Ю. Гришанова, В.А. Резниченко, 2008

ISSN 1727-4907. Проблеми програмування. 2008. № 2-3. Спеціальний випуск

417

Как указал профессор Джон Сова, – Semantic Web – много-дисциплинарная тема, которая объединяет теории и методы трех областей:

- логика – формальные структуры и правила логического вывода;
- онтологии – описание типов сущностей, которые относятся к Предметной области;
- теория моделей.

Интернет – это сеть компьютеров, объединенных каналами и использующие протоколы (TCP/IP) для связи между собой. Web – это сеть сайтов, использующих гиперссылки для переходов между страницами [29]. Традиционный Web базируется на языке разметки документов HTML. HTML-страница описывает форму представления информации в Web-браузере, а этот язык тяжело подвергается автоматическому содержательному анализу. Автоматизировать даже такие тривиальные задачи, как поиск людей, проектов, программ в Интернете невозможно. Следующий этап развития Интернет – Semantic Web – представляет собой переход на новый уровень представления данных – уровень знаний и автоматизированной обработки. Технология Semantic Web разрешит компьютеру интерпретировать информацию, представленную в Web, наравне с людьми, для чего разработана графовая модель описания ресурсов RDF (Resource Description Framework).

В общем виде Semantic Web (по Тиму Бернерсу-Ли) – это:

- интероперабельность данных между программными приложениями и организациями;
- набор интероперабельных стандартов для обмена знаниями;
- архитектура для взаимосвязанных сообществ и словарей [30].

1. Архитектура Semantic Web

С точки зрения архитектуры Semantic Web можно рассматривать как три яруса (рис. 1):

- базис, который состоит из уникальной глобальной идентификации ресурса, метаданных для декларирования фактов о ресурсах, и общего языка для выражения метаданных и знаний, который реализован с помощью онтологий для общедоступного понимания и общего словаря метаданных и правил для добавления новых метаданных и знаний;
- базовый сервис, например логический вывод и запросы к метаданным и онтологиям, разъяснение таких выводов, управление доверием (trust), агенты, поисковые системы, серверы онтологий;
- сервисы приложений, например сервис агентства путешествий.



Рис. 1. Три яруса сети Semantic Web

Технологии, которые задействованы в разработке Semantic Web:

- семантический поиск;
- вопросно-ответные системы;
- агенты;
- объединение знаний (интеграция баз данных);
- всепроникающие вычисления (ubiquitous/pervasive computing) [29].

В 1998 году Тим Бернерс-Ли предложил следующий логический план построения Semantic Web [31]:

- синтаксис для представления знаний, который использует ссылку на онтологии (RDF);
- язык описания онтологий (OWL);
- язык описания веб-сервисов (WSDL, OWL-S);
- инструменты чтения/разработки документов Semantic Web (Jena, Haystack, Protege);
- язык запросов к знаниям, которые записаны в RDF (SPARQL);
- логический вывод знаний (находится на этапе обсуждения);
- семантическая поисковая система (например, SHOE).

Базовая модель Semantic Web (пирог Тима) в редакции 2006 года показана на рис. 2 [32].

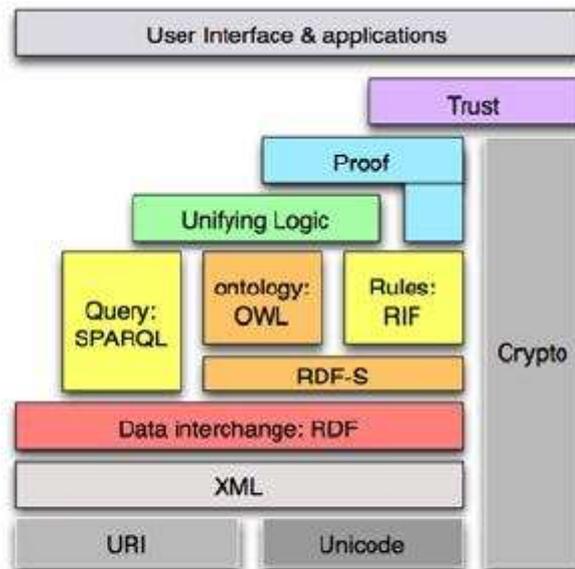


Рис. 2. Базовая модель Semantic Web в редакции 2006 года

Фундаментальными основами Semantic Web являются:

- графовая модель представления полуструктурированных данных (OEM, Lore);
- формальная логика (логика первого порядка, базы знаний, фреймы);
- архитектура WWW (URI/IRI, Unicode, XML, HTTP);
- криптография с открытым ключом.

Рассмотрим структуру базовой модели Semantic Web более детально.

2. URI – универсальный идентификатор ресурсов

В Web для идентификации элементов используются "Унифицированные идентификаторы ресурсов", или сокращенно URI (Uniform Resource Identifier). URI можно присвоить чему угодно, и если эта сущность имеет URI, то о ней можно говорить, что она находится "в Web": это может быть человек, книга, абстрактная концепция, т.е. все, что имеет название.

URI является базисом Web. «URI – это компактная строка символов, которая используется для идентификации абстрактного или физического ресурса» [33].

Одной из форм URI есть URL (Uniform Resource Locator), унифицированный указатель ресурса. URL это адрес, по которому загружается Web-страница.

Также необходимо указать, что в начальной базовой модели в нижнем ярусе было указано еще и базовое кодирование – т.е. общий для всех принцип кодирования всех возможных символов многих языков – кодовая таблица UNICODE.

За синтаксисом URI следит комитет IETF. Документ, который опубликованный этим комитетом RFC 2396 является общей спецификацией URI. Консорциум W3C поддерживает список схем URI.

В 2005 году на смену URI был предложен интернационализированный идентификатор ресурса – Internationalized Resource Identifiers (IRI), идентифицирующий абстрактный или физический ресурс на любом языке мира. URI могут содержать только латинские символы и знаки препинания из набора символов US-ASCII (в общей сложности около 60 символов). Для обеспечения принципов интернационализма, сохранения «читабельности» для человека, в IRI было предложено, что эти идентификаторы могут содержать любые символы Юникода (Unicode/ISO10646) в чистом виде, без всякого кодирования. IRI не ущемляют права вторых языков и ведут к более высокой степени равноправия пользователей Интернет. В будущем идентификаторы IRI призваны заменить URI.

3. Документы: расширяемый язык разметки (XML)

XML[34] (eXtensible Markup Language) представляет собой очень простой и при этом мощный, и гибкий текстовый формат для описания документов произвольной структуры. XML был разработан и утвержден в качестве стандарта в 1998 г Консорциумом W3C для упрощения реализации, а также для обеспечения интероперабельности между SGML и HTML. Он является подклассом языка SGML, однако более прост для понимания и обработки.

Функции XML следующие:

- представление синтаксиса для других языков разметки;
- семантическая разметка Web-страниц. XML-представление может использоваться на Web-странице вместе с таблицей стилей XSL, что определяет корректный вывод на экран разных элементов;

- единый формат обмена данных. XML-представление может передаваться между двумя применениями как объект данных.

Язык XML разрешает каждому создавать свой собственный формат документов и потом писать документы в этом формате. Эти форматы документов могут включать разметку, которая уточняет содержание контента документа. Документ с разметкой может "читаться" компьютером.

4. Утверждения: Общая схема описания ресурсов RDF

Для описания предметной области ресурсов предложен стандарт RDF (Resource Description Framework) [35 – 42], принятый в 1999 году консорциумом W3C и поддержанный многими ведущими производителями ПО, и поставщиками контента. Начальное назначение RDF было в описании XML-ресурсов с разных точек зрения. RDF представляет собой модель описания метаданных. Этот язык использует XML-синтаксис.

В то время, как модель данных XML является графом с обозначенными вершинами и не обозначенными дугами (т.е. без связей), модель данных RDF является графом с обозначенными как вершинами, так и дугами, который разрешает определять связи между сущностями.

Модель Resource Description Framework имеет своей целью стандартизировать определение и использование метаданных, которые описывают ресурсы Web. Однако, RDF также хорошо подходит и для представления данных [43].

Стандарт RDF (Resource Description Framework) включает две основные части – собственно способ описания ресурсов, а также способ задачи схем, по которым ресурс описывается.

Первая часть RDF [44] определяет простую модель для описания объекта, который рассматривается в качестве ресурса, как связей между ресурсами в терминах поименованных свойств и значений.

Вторая (RDF Schema – RDFS) [45, 46] служит для задачи структуры предметной области и аналогична диаграмме классов в UML.

На RDF можно описывать как структуру ресурса, так и связанную с ним предметную область.

RDF описывает ресурсы в виде ориентированного размеченного графа – каждый ресурс может иметь свойства, которые в свою очередь также могут быть ресурсами или их коллекциями.

Базовый строительный блок в RDF – это тройка «объект – атрибут – значение», который часто записывают в виде $A(O,V)$, т.е. «объект O имеет атрибут A со значением V ». Такую связь можно также представить как ребро с меткой A , которое объединяет два узла, O и V : $[O] - A \rightarrow [V]$. Такая нотация довольно полезна, поскольку RDF разрешает менять местами объекты и значения. Таким образом, каждый объект может играть роль значения, которое в графическом представлении отвечает цепочке из двух ребер с метками.

Кроме всего вышесказанного, RDF допускает форму представления, в которой любое выражение RDF в тройке может быть объектом или значением, т.е. графы могут быть как вложенными, так и линейными. В Web это разрешает, например, выражать сомнение или согласие с выражениями, созданными другими людьми.

Главная цель RDF – предложить базовую модель данных «объект – атрибут – значение» для метаданных. Кроме этой семантики, которая описана в стандарте лишь неформально, RDF не содержит каких-либо четких правил, ориентированных на моделирование данных. Также, как XML Schema используется для определения словаря, RDF Schema разрешает разработчикам определять конкретный словарь для данных RDF (такой, как *authorOf*) и указывать виды объектов, к которым могут применяться эти атрибуты. Другими словами, механизм RDF Schema предоставляет базовую систему типов для моделей RDF.

Таким образом, RDF предоставляет возможность формулировать утверждения в виде, пригодном для обработки компьютером и это является основой Semantic Web.

5. Метаданные

В базовой модели Semantic Web, вышепредставленной, предложенной Тимом Бернерсом-Ли, явно не выделено наличие средств описания метаданных. Тем не менее, в своих работах, например, [30, 31], а также в работах других ученых указывается на важность включения в концепцию Semantic Web понятия метаданных.

Метаданные это данные о данных. Более точно, это данные, предназначенные для идентификации, описания или локализации (местоположения) информационных ресурсов, не зависимо от физической природы ресурса.

Было разработано множество схем описания метаданных, среди которых следует упомянуть следующие:

- Topic Maps (XMT) [47] – стандарт ISO (ISO/IEC 13250:2003) для представления и обмена знаниями с точки зрения поиска информации.

- Text Encoding Initiative (TEI) [48] – международный проект по разработке нормативов для разметки (marking up) электронных текстов, таких как романы, пьесы, стихи; главным образом для поддержки исследований в гуманитарной сфере.

- Metadata Encoding and Transmission Standard (METS) [49] – стандарт кодирования и передачи метаданных, был разработан для удовлетворения потребности в стандартной структуре данных для описания сложных цифровых библиотечных объектов.

- Metadata Object Description Schema (MODS) [50] – схема метаданных описания объектов, которая была выведена из MARC 21, и предназначена для перенесения отобранных данных из существующих записей метаданных MARC 21 или для создания оригинальной записи описания ресурса.

- Encoded Archival Description (EAD) [51] – закодированное архивное описание, было разработано как способ разметки данных, которые содержатся в поисковых средствах, для того, чтобы они находились и показывались в оперативном режиме.

- Learning Object Metadata (LOM) [52] – стандарт IEEE 1484.12.1-2002 метаданных объектов учебного процесса для повторного использования ресурсов учебного характера, таких как компьютерного и дистанционного обучения.

- Online Information Exchange (ONIX) [53] – международный стандарт схемы метаданных, который разработан издателями книжной промышленности Соединенных Штатов и Европы.

Однако, базовыми для Semantic Web в данный момент признаются стандарты Dublin Core, FOAF, SIOC и DOAP [54].

FOAF (Friend-Of-A-Friend) [55 – 57] – это формат машинно-обрабатываемых страниц, описывающих персональную информацию о людях и их деятельности (фотографии, календари, блоги и прочее) в формате XML.

SIOC (Semantically-Interlinked Online Communities) [58] – документы, описывающие онлайн-сообщества. SIOC обеспечивает взаимосвязь таких средств обсуждения информации, как блоги, форумы и почтовые рассылки между собой.

Description of a Project Description of a Project (DOAP) [59] – документы, описывающие в сети проекты с открытым исходным кодом.

Среди данных стандартов выделяется Dublin Core [60], как один из базовых стандартов для представления данных об информационных ресурсах в Semantic Web. Dublin Core [61, 62] – набор элементов (свойств) для описания документов, который первоначально был разработан в марте 1995 года. Цель Dublin Core – обеспечение минимального набора элементов описания, которые оказывают содействие внедрению описания и автоматической индексации документоподобных сетевых объектов по принципу, подобному карточкам библиотечного каталога. Набор метаданных Dublin Core предназначался для использования средствами исследования ресурсов Интернета, такими как веб-кроулеры поисковых систем, а также предполагалось, чтобы Dublin Core был достаточно простым набором для понимания и использования широким кругом авторов и случайных публикаторов, которые размещают информацию в Интернете. Элементы Dublin Core широко используются в документировании Интернет-ресурсов. На данный момент элементы Dublin Core определены в Dublin Core Metadata Element Set, Version 1.1: Reference Description [63].

Расширять сам набор элементов можно как самостоятельно, так и с использованием уже имеющихся стандартов. Например, для описания людей и организаций (которые выступают в качестве элементов метаданных Dublin Core: Creator, Publisher или Contributor) можно применить стандарт для электронных бизнес-карт (vCard [64]). Общие соображения по этому поводу даются в [65], а конкретное предложение предоставляется в [66 – 68].

Как отмечается в официальном описании RDF, метаданные могут быть встроенными (embedded) в сам ресурс, например, в HTML страницы [69] или документы, например, MsWord (это простейший подход для описания страниц), а могут сохраняться и обновляться независимо от ресурсов. Многие из производителей программного обеспечения уже выпускают ряд продуктов, которые автоматически формируют некоторый небольшой блок RDF-описания внутри документа. Второй подход является более универсальным, так как в этом случае метаданные могут быть созданы для любого ресурса. В настоящее время уже начат проект на базе Open Directory [70] (поисковая система Google) по автоматическому созданию репозитория RDF-описаний ресурсов Интернет.

В случае размещения метаданных отдельно от ресурса, сами метаданные преимущественно сохраняются (и передаются) в формате XML. При этом максимально используются возможности модели RDF и обеспечивается свободный обмен информацией (interoperability). Обмен метаданными сводится к пересылке RDF/XML-файлов (т.е. текстовых файлов в формате XML или просто ссылок на эти файлы), т.е. может быть полностью автоматизирован.

6. Простое моделирование данных: схема RDF

Первым "пластом" Semantic Web над только что обсужденным синтаксисом является простая модель типизации данных. Схема и онтология – это средства для описания содержания и связи между терминами.

На основе RDF 23 января 2003 был предложен рабочий проект RDF Vocabulary Description Language 1.0: RDF Schema [71]. Схема RDF была разработана как простая модель типизации данных для RDF. Как указывается в документе, RDF является языком общего применения для представления информации в Интернет. Данная спецификация описывает как использовать RDF для описания RDF-словарей. Она определяет базовый словарь, предназначенный для этих целей и принятые соглашения, которые могут быть использованы при создании приложений Semantic Web для поддержки более сложных словарей RDF-описаний. Язык описания словаря RDF определяет классы и свойства, которые могут быть использованы для описания других

классов и свойств, а также производить некоторые более сложные вещи, такие, как создание диапазонов и областей для свойств.

Три наиболее важных понятия, которые дает нам RDF и схема RDF – это "Ресурс" (rdfs:Resource), "Класс" (rdfs:Class) и "Свойство" (rdfs:Property). Эти понятия являются "классами" в том понимании, что этим классам могут принадлежать термины.

Как уже было указано, RDF Schema определяется в терминах базовой информационной модели RDF – структуры графа, который описывает ресурсы и свойства. Все словари RDF используют некоторую базовую структуру: они описывают классы ресурсов и типы связей между ресурсами. Эта общность позволяет использовать разнородные словари, созданные для машинной обработки, и отвечает требованиям по созданию метаданных, в которых утверждения могут быть получены из множества разнородных децентрализованных словарей, созданных различными сообществами по разным принципам и разными методами.

Описание с помощью RDF не ограничивается только описанием документов Интернет. Этот стандарт довольно универсальный и гибкий для того, чтобы описывать большинство типов структурированных данных. Например, в RDF естественно выражаются диаграммы сущность-связь, которые широко применяются для проектирования баз данных. Описание семантики ресурса на RDF может быть как «внешним», когда описывается ресурс в целом, так и «внутренним», когда описывается внутренняя структура ресурса – будь-то база данных, XML-документ, или целый сайт.

Важной особенностью стандарта RDF, как и лежащего в его основе XML, является расширяемость. На RDF можно задать структуру описания источника, используя и расширяя встроенные понятия RDF-схем, такие как классы, свойства, типы, коллекции. Модель схемы RDF включает наследование; наследоваться могут как классы, так и свойства.

Кроме описания структуры, RDF разрешает оперировать утверждениями. Выражение «ресурс R1 как свойство P имеет ресурс R2» можно проинтерпретировать и как предикат P(R1, R2), а потом использовать это утверждение как объект других утверждений. Такая интерпретация разрешает описывать с помощью RDF концептуальную информацию.

Таким образом, RDF целиком подходит на роль универсального языка описания семантики ресурсов и взаимосвязей между ними.

Однако, как утверждают сами авторы стандарта, RDF имеет и ряд отсутствующих свойств, которые они указывают как следующие:

- невозможность указания мощности множества значений свойства, например, что «Человек имеет только одного биологического отца»;
- невозможность указания того, что представленное свойство (например, hasAncestor – имеет предка, прототип) является транзитивным, например, что «если A hasAncestor B, и B hasAncestor C, тогда A hasAncestor C»;
- невозможность указания того, что два разных класса, определенных в разных схемах, фактически представляют одно и то же понятие;
- невозможность указания того, что два разных экземпляра (instances), определенные отдельно, фактически представляют один и самый субъект;
- невозможность определения новых классов в терминах операций (например, объединение и пересечение) над другими классами.

7. Онтологии

Онтологии, в общем виде определяются как совместно используемые формальные концепции конкретных предметных областей, они дают общее представление о понятиях, информацией из которых могут обмениваться люди и приложения. Они разрешают концептуализировать домен фиксированием сущностей (entities) и связей в домене. Указание в каких связях принимает участие сущность частично разрешает понять и ее значение (содержание), поскольку это предоставляет возможность видеть, где данная сущность входит в отношения с другим доменом.

Онтологии основываются на математическом аппарате формальной логики (descriptive logic, DL), малое подмножество которого охвачено RDF-схемой. DL является подмножеством логики первого порядка, которое вычислимо.

Дополнительные возможности, вышеуказанные, в дополнении к имеющимся в RDF, является целью онтологических языков, таких, как DAML+OIL [72, 73] и OWL [74, 75]. Данные два языка основаны на RDF и RDF Schema. Цель данных языков – обеспечение ресурсов дополнительной машинно-обрабатываемой семантикой, т.е. они направлены на обеспечение машинного представления ресурсов в форме, которая более соответствует их оригиналу из реального мира.

Разметка документов Semantic Web с помощью онтологических терминов позволит производить автоматическую обработку их контента. Таким образом, онтологии определяются как ключевая технология для развития Semantic Web.

Онтологии в состоянии сыграть критически важную роль в организации обработки знаний на базе Web, их общего использования и обмена ими между приложениями.

Язык DAML (DARPA Agent Markup Language 2000 год) был разработан агентством передовых оборонных исследовательских проектов (Defense Advanced Research Projects Agency) как расширение XML и RDF. Последняя версия языка DAML+OIL обеспечивает большой набор конструкций для создания онтологий и разметки информации таким образом, чтобы компьютеры были способны их прочитать и понять. В этой связи необходимо также упомянуть еще одну разработку DARPA – язык DAML-S – Semantic Markup for Web Services.

DAML+OIL является языком семантической разметки для Веб-ресурсов. Он основывается на ранних стандартах W3C таких, как RDF и RDF Schema, и расширяет эти языки более полными примитивами моделирования. DAML+OIL обеспечивает примитивы моделирования, которые по обыкновению используются в языках, основанных на фреймах. Онтология DAML+OIL (или база знаний, knowledge base) есть коллекция RDF – троек. Онтология, как правило, содержит иерархию понятий предметной области и описывает важные свойства каждого понятия с помощью механизма «атрибут – значение». Связи между понятиями могут быть описаны с помощью дополнительных логических утверждений.

Язык OWL. Наиболее развитым языком представления онтологий в настоящее время является OWL (Web Ontology Language), который расширяет возможности XML, RDF, и RDF Schema. Этот язык основан на DAML+OIL. Проблемы, которые возникли в DAML+OIL, были вызваны постоянным изменением ядра спецификаций RDF, на котором основан DAML+OIL.

Как указывается в основном рабочем проекте, OWL почти полностью похож на DAML+OIL. Основные и существенные отличия от DAML+OIL состоят в следующем:

- устранение некоторых ограничений;
- способность прямо указывать, что свойство может быть симметричным;
- устранение некоторых неиспользуемых конструкций DAML+OIL, особенно ограничение с дополнительными компонентами.

Существует также несколько маловажных расхождений, которые включают в себя некоторые изменения имен некоторых конструкций, однако основная цель, преследуемая при создании OWL, заключалась в том, чтобы максимально корректно сохранить имена DAML+OIL.

Онтология OWL является последовательностью аксиом и фактов с добавлением ссылок на другие онтологии, которые считаются включенными в онтологию. Онтологии OWL являются Web-документами и на них можно ссылаться. Онтологии также имеют не связанную с логикой компоненту (пока еще не определенную), что может быть использовано для записи авторства, и другая не связанная с логикой информация, ассоциированная с онтологией. Фактически это словарь, который расширяет набор терминов, определенных в RDFS.

Онтологии включают информацию о классах, свойствах и частных случаях, каждый из которых может иметь идентификатор ID, который является ссылкой URI.

OWL имеет три модификации:

- OWL Lite (простой);
- OWL DL (с полной разрешимостью);
- OWL Full с полной выразительной мощностью).

Каждая из этих модификаций (кроме Lite) является расширением предыдущей. Как следствие: любая OWL Lite онтология является OWL DL онтологией, а любая OWL DL онтология является OWL Full онтологией.

Главные характеристики языка веб-онтологий OWL:

- OWL использует синтаксис XML;
- OWL имеет инструкции для представления дерева классов;
- OWL имеет инструкции для указания принадлежности индивидов классам;
- OWL имеет систему описания свойств: область определения, область значений;
- OWL может задавать характеристики свойств: симметричность, транзитивность, функциональность;
- OWL имеет инструкции для указания эквивалентности (склеивание) классов.

8. Языки запросов к RDF хранилищам

Говоря о языках запросов, фактически речь идет о интеграции разных языков (информационно-поисковых, баз данных, манипулирования данными, обмена данными и т.п.) в единый язык запросов Web. При этом все специалисты едины во мнении, что это должен быть декларативный язык, построенный на модели неполноструктурированных данных (semistructured).

Документ "XML-QL: A Query Language for XML" [76] был подготовлен к семинару W3C по поисковым языкам, который прошел в конце 1998 года и явился далеко не единственной попыткой обобщения такого рода.

В настоящее время появилось несколько языков запросов к XML-источникам данных: XQL (1998) [77], XML QL (1998) [78 – 80]. Поиск в XML-документе состоит в нахождении элементов, которые удовлетворяют условиям запроса, с последующим преобразованием найденных элементов в структуру, заданную в запросе.

Язык запросов к RDF-источникам данных (RDF Query) предложен в 1998 [81 – 85] и в данное время имеет уже практическую реализацию в проекте Sesame [86].

В 2006 году консорциум W3C начал разработку языка запросов к RDF и OWL-хранилищам – SPARQL Query Language for RDF, который сейчас имеет статус рекомендованного кандидата (candidate recommendation) [87].

SPARQL – язык запросов, который базируется на паттернах графов.

SPARQL одновременно является как языком запросов, так и протоколом доступа к данным, является одним из ключевых компонент приложений Web 2.0: в качестве стандарта для поддержки гибкой модели данных он дает общий механизм запросов для всех приложений Web 2.

9. Логический вывод

Принцип "логического вывода" очень простой: это возможность выводить новые данные из данных, которые уже есть. В математическом смысле, выполнение запроса является одной из форм логического вывода (например, возможность вывести из массы данных некоторый результат поиска). Логический вывод является одним из ведущих принципов Semantic Web, так как он разрешает очень легко создавать SW-приложения [88].

Для того, чтобы Semantic Web стал довольно выразительным и смог помогать людям в разных ситуациях, возникает необходимость построения мощного логического языка, который поддерживает логический вывод. Дискуссии относительно методов, и даже возможности выполнения этой задачи, до сих пор ведутся очень активно; обращается внимание на то, что в RDF недостаточны возможности квантификации, и что эта область определена недостаточно хорошо. Проблемы логики предикатов подробно рассмотрены в базовой монографии Джона Сова (John Sowa's) «Математические предпосылки (логика предикатов)» – «Mathematical Background (Predicate Logic)» [89].

Rule Interchange Format (RIF) – формат обмена правилами. Цель этого разрабатываемого консорциумом W3C стандарта [90] – определение формата, который бы разрешил транслировать правила между разными языками правил и благодаря этому обеспечить обмен правилами между системами, основанными на правилах.

Системы, основанные на правилах, получили широкое распространение в информационных технологиях. К их числу относятся, например, экспертные системы и системы дедуктивных баз данных. Разработки технологий Semantic Web обеспечивают новую среду использования таких систем. Поэтому консорциум W3C уделяет отдельное внимание этой области. Спецификация RIF может рассматриваться как составная часть комплекса стандартов Semantic Web.

В настоящее время рабочей группой, организованной при консорциуме для разработки этого стандарта, подготовлен и обсуждается рабочий проект документа, который систематизирует случаи использования RIF и требования к этому языку. Важнейшее требование к создаваемому стандарту – обеспечение возможности его использования не только при текущем состоянии технологий, основанных на правилах, но и его гибкости, достаточной для обеспечения его использования в процессе их эволюции.

Рабочий проект документа, который описывает случаи использования, даст возможность определить функциональные требования к RIF и на этой основе разработать адекватные спецификации языка.

Правила вывода новых фактов SWRL. Благодаря дополнению OWL языком RuleML [91] (подмножество Datalog) в виде словаря SWRL (A Semantic Web Rule Language) [92] появилась возможность использовать дизъюнкты Хорна (Horn-like rules) для явного указания способа вывода новых фактов из RDF-утверждений. Пока словарь SWRL находится в стадии стандартизации [93].

Хотя работы над этим уровнем Semantic Web продолжают, однако в нашем распоряжении есть уже достаточный набор средств для построения Semantic Web: утверждение, цитирование (материализация) в RDF, классы, свойства, области, документирование в схеме RDF, непересекающиеся классы, свойства однозначности и уникальности, типы данных, инверсии, эквивалентности, списки и прочее.

10. Доверие и доказательство

Следующий шаг в разработке Semantic Web – доверие и доказательство. Об этом уровне написано очень мало, что является недопустимым, так как в будущем он будет очень важным.

Для обеспечения целостности и непротиворечивости информации, представленной в Semantic Web, важно обеспечить связь приложений Semantic Web с контекстом, а также механизмы проверки доказательства и цифровых подписей.

Приложения Semantic Web будут учитывать контекст в целом для того, чтобы сообщать пользователям, могут ли они доверять предоставленным данным. Если пользователь получает поток RDF-данных от другого пользователя о прочитанной им книге и о его оценке этой книги, то он должен знать, кто этот человек, и можно ли доверять этой информации. Более того, пользователь может потом воспользоваться этой информацией, не сомневаясь в ее источнике. Далее пользователь оставляет на свое собственное усмотрение насколько ему верить полученному критическому отклику о книге.

Необходимо помнить и о том, что над разделяемыми контекстами работают также и группы людей. Если какая-то группа разрабатывает в Semantic Web информационную службу для художников, каталогизируя людей, их имена и места, где находятся картины этих людей, то доверие пользователя к этой группе зависит от того, насколько он доверяет людям, которые принимают участие в этой группе.

В связи с этим в Semantic Web для определения источника информации предлагается использовать цифровые подписи.

Цифровые подписи это есть небольшие фрагменты кода, которые можно использовать для однозначной проверки того, кто написал тот или другой документ. Основанная на работах по математике и криптографии, цифровая подпись является доказательством того, что документ или утверждение написал (или с ним согласен) определенный человек. Разработчики Semantic Web планируют, что каждый пользователь или агент все свои RDF-утверждения будет подписывать персональной уникальной цифровой подписью.

Еще одним аспектом доверительности информации является проверка истинности. Язык проверки истинности это просто язык, который позволит проконтролировать, является ли утверждение правдивым. Реализация языка проверки обычно составляется из списка "элементов" логического вывода, которые используются для получения искомой информации, а также для последующей проверки информации о доверии для каждого из этих элементов.

11. Агенты и сервисы

Ведущую роль в Semantic Web должны сыграть программные агенты. При вышеописанной архитектуре информационного пространства, предполагается, что агенты, обладающие интеллектуальными способностями, смогут выполнять поставленные им пользователями цели и задачи самостоятельно. Например, по поиску необходимой информации, подбору и выбору оптимальных вариантов и т.п. Это в перспективе мобильные, интеллектуальные агенты, способные к целеполаганию, планированию, совместному взаимодействию с другими агентами для достижения цели, имеющими знания как о себе, так и о внешнем мире. Для достижения поставленных задач они должны иметь возможность пользоваться некоторыми стандартными наборами услуг, представленными в Web в качестве веб-сервисов.

Веб-сервис – это программная система, предоставляющая некоторую услугу и обеспечивающая взаимодействие по сети. Обычно это веб-ресурс, характеризующийся абстрактным набором функциональных возможностей, которые в нем реализуются. Функционально веб-сервис может являться агентом, а может быть обычной программой.

Определение веб-сервиса, данное в википедии следующее: это «программная система, идентифицируемая строкой URI, чьи общедоступные интерфейсы определены на языке XML. Описание этой программной системы может быть найдено другими программными системами, которые могут взаимодействовать с ней согласно этому описанию посредством сообщений, основанных на XML, и передаваемых с помощью интернет-протоколов.»

Архитектура веб-сервисов базируется на компонентном подходе, т.е. сервис должен быть достаточно автономным, а также может состоять из нескольких сервисов, подбираемых динамически для выполнения конкретной задачи в соответствии с различными критериями.

Немаловажным аспектом при выборе сервиса является его доступность. Интернет представляет собой динамичную среду, и вопрос доступности ресурса или сервиса является очень актуальным. При проектировании композиции сервисов очень важно учитывать данный аспект.

Задача построения новых сервисов из уже имеющихся поднимает проблему синтеза сервисов.

Для того, чтобы воспользоваться услугами, должна быть возможность их обнаружения, механизм получения информации о том, какие услуги они предоставляют, как к ним обращаться, формат сообщений. Решением этой задачи стало создание каталогов услуг с помощью стандартных методов доступа. Сервисы должны быть описаны в стандартных терминах, а информация о том, как к ним обращаться и другая имеющаяся информация должна кодироваться стандартным способом.

Технология веб-сервисов базируется на следующих открытых XML-стандартах:

- SOAP (Simple Object Access Protocol) [94 – 100] — XML-протокол для удаленного вызова методов веб-сервисов;
- UDDI (Universal Description, Discovery and Integration) [101] — описывает модель данных, предназначенную для каталогизации и обнаружения услуг, предоставляемых веб-сервисами;
- WSDL (Web Services Description Language) [102] — язык описания интерфейсов веб-сервисов.

Формирующиеся дополнения к ним, например, WSCoordination/WS-Transaction (транзакции), WSSecurity (безопасность), WS-Routing (маршрутизации сообщений) и т.д., призваны расширить возможности этой платформы в удовлетворении требований задач интеграции приложений. В рамках инициативы WS-I разрабатываются примеры прикладных решений, предложения и дополнительные требования, призванные гарантировать совместимость решений разных поставщиков. Это сулит широкие возможности по интеграции различных информационных систем в рамках единого согласованного набора спецификаций.

Во многих случаях интеграция информационных ресурсов требует комбинирования обращений более чем к одному веб-сервису для реализации пользовательского запроса. Таким образом, веб-сервисы должны иметь возможность поддерживать взаимодействие с другими приложениями в дополнение к стандартным процедурам обработки данных. Более того, процесс предоставления агрегированной распределенной информации может включать в себя разбиение на набор взаимосвязанных этапов обработки данных, взаимодействие ряда веб-сервисов, вмешательство людей в процесс обработки пользовательских запросов и другие элементы прикладной логики. Поэтому процесс сбора и интеграции гетерогенных данных может представлять собой

логически сложную композицию обращений к хранилищам информационных сущностей посредством интерфейсов веб-сервисов — определять автоматизированный поток обработки данных.

Для описания композиций веб-сервисов на данный момент различными ассоциациями предлагается ряд стандартов. Среди них можно отметить следующие языки описания автоматизированных потоков работ, участниками которых являются веб-сервисы:

- WSFL (Web Services Flow Language) — позволяет определять композиции веб-сервисов в виде графовой модели рабочего процесса;
- BPML (Business Process Modeling Language) — определяет блочную модель композиции веб-сервисов;
- BPEL4WS (Business Process Execution Language For Web-Services) — представляет собой гибрид блочной и графовой моделей описания взаимодействий веб-сервисов.

Эти языки позволяют описывать композиции веб-сервисов, что позволяет определять сложные, распределенные процессы по извлечению, обработке и интеграции информации.

Для решения таких сложных распределенных задач особенно хорошо подходит мультиагентная технология.

Как уже было вышеописано, для выполнения конкретных задач веб-сервисы должны обмениваться сообщениями, сообщать информацию о себе и предоставляемых услугах в виде, удобном как для машинной обработки, так и доступном для понимания человеком. Для решения этой задачи консорциумом были предложены языки метаописаний сервисов WSDL, а также онтологический язык веб-сервисов OWL-S [103]. В настоящее время консорциумом предложен проект языка моделирования сервисов – Service Modeling Language (SML) [104].

Наиболее часто используемое определение агента состоит в том, что программный агент это программная сущность, которая функционирует продолжительно и автономно в конкретном окружении, часто – вместе с другими агентами. Агенты могут быть специализированные, они должны уметь общаться с другими агентами с целью обнаружения сервисов, продуктов, информации или других агентов. Сервисы, представленные в сети, могут быть реализованы как агенты. Возникает проблема создания архитектуры для взаимодействия агентов, где бы агенты могли описывать свои цели с использованием заранее определенных словарей, где возможно было бы производить поиск и подбор необходимых сервисов и информационных ресурсов, а также использовать многие другие возможности.

12. Практическая реализация Semantic Web

Технология Semantic Web на данное время успешно решает следующие задачи:

- независимость данных от приложений;
- семантическая интеграция данных;
- создание основы для повсеместного использования компьютерных агентов (сервисов).

Формирование Semantic Web станет возможным только при условии обеспечения более высокого уровня интероперабельности. Однако уже сейчас сделано много практических шагов по реализации данного проекта. Новый проект на базе поисковой системы Google недавно предоставил свои ресурсы для запросов агентам на выполнение поисковых функций и проверки правописания [105]. Также представляет интерес новый проект по автоматическому созданию RDF-описаний и хранилища метаданных, создаваемый на базе Open Directory [70] поисковым механизмом Google [106]. Кроме того, необходимо также отметить и проект консорциума W3C SWAD-Europe [107], который занимается проблемой связи хранилищ семантических данных с используемыми реляционными системами баз данных, особенно лицензированных как Free Software / Open Source (FS/OS).

В настоящее время необходимо констатировать, что общий объем мета-информации достиг уже критической массы и неуклонно растет. На сентябрь 2006 года пространства имен OWL были использованы в 113 000 документах Semantic Web (это 8% общего объема), пространство имен RDFS объявлено в 677 000 (47%). Owl:Class является наиболее используемым термом из пространства имен OWL, он используется в 1 800 000 высказываниях из 68 000 документов. В августе 2007 года в сети насчитывалось более 2 миллиардов RDF-троек [32, 54, 108, 109].

Интерес к использованию данной информации также постоянно повышается. На март 2006 года [108] из анализа запросов поисковой системы Google видно, что обычными рядовыми пользователями было произведено 2 120 000 запросов к типу „RDF filetype:rdf” и 13 600 “ontology filetype:owl”. Такие цифры говорят о популяризации идей Semantic Web и дает возможность уже реально начинать использовать данную мета-информацию в прикладной сфере.

Дальнейшему развитию Semantic Web оказывает содействие наличие свободно распространяемых систем для разработки приложений Semantic Web:

- Jena Framework (Java);
- Drive RDF Parser (C#).

В настоящее время уже существуют:

- библиотеки для интерпретации стека языков RDF для всех популярных языков программирования (Jena, Redland, RDFLib);
- редакторы онтологий (Protege);
- системы рассуждений над онтологиями (Racer, KAON, FACT);

- семантические хранилища (Sesame, Kowari, YARS);
- семантические браузеры (Simile, Piggy Bank, Gnowsis, Haystack);
- поисковики семантических данных (Swoogle);
- конверторы из разных форматов представления данных в/из RDF/XML (Aperture, RDFizers, D2R);
- прикладные программы (Bibster, FOAF Explorer).

Также необходимо указать и существующие коммерческие продукты: Adobe's XMP – инструмент для создания метаописаний о файлах; Oracle's 10.2 Database – уже имеет встроенную поддержку модели RDF; Tucana's Knowledge Discovery Suite – платформа для интеграции информации применений (Enterprise Information Integration, EII)

На последней VI международной конференции по Semantic Web – Sixth International Semantic Web Conference, которая проходила 11-15 ноября 2007 г. в Корею [109], обозначено следующее положение дел в направлении распространения Semantic Web:

- обозначился резкий рост и возникновение компаний, использующих технологию Semantic Web (Joost, Radar Networks, MetaWeb, Siderean, SandPiper, SiberLogic, Ontology Works, Intellidimension, Intellisophic, TopQuadrant, Data Grid, etc.);
- произошло вовлечение крупных поставщиков ПО – Adobe, Cisco, HP, Microsoft, Nokia, Oracle, Sun, Vodaphone;
- активно развиваются правительственные программы – в США, Объединенной Европе, Японии, Корею, Китае;
- сильно возрос такой важный рынок, как медико-фармацевтический – создана специальная группа при консорциуме Health Care and Life Sciences Interest Group at W3C;
- появилось много инструментов с открытым кодом – Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Onto(xxx). Wilbur.

На этой конференции Semantic Web рассматривался как коллекция всех формальных, машиннообработываемых, доступных в Web, основанных на онтологиях утверждений (семантических метаданных) о веб-ресурсах и прочих сущностях мироздания, выраженных на языке представления знаний, основанном на синтаксисе XML (например, OWL, DAML, DAML+OIL, RDF, etc.). Необходимо констатировать, что в Web уже представлено достаточно большое количество такой информации. Все больше встает проблема ее обработки, объединения, выравнивания, выявления связей.

С 2003 года ежегодно проводится всемирный конкурс Semantic Web Challenge [110], призванный собрать самые последние наработки и показать миру состояние дел по практической реализации идей Semantic Web. При этом был сформулирован следующий перечень минимальных критериев, определяющих понятие «приложение Semantic Web».

Во-первых, приложение должно использовать информационные источники, которые:

- географически распределены;
- имеют различных владельцев, что предполагает отсутствие контроля за их развитием;
- являются гетерогенными (синтаксически, структурно, и семантически);
- содержат данные реального мира, т.е. источники должны быть больше, чем игрушечные примеры.

Во-вторых, приложение должно воспринимать открытый мир; это значит, что оно знает, что информация никогда не бывает полной и постоянно меняется.

В-третьих, приложение должно использовать некоторое формальное описание значения данных.

Помимо этих минимальных критериев, были определены несколько желательных качеств. Приложение должно использовать источники данных в других целях или по-другому, чем первоначально было намечено. Оно также должно использовать контент мультимедийных документов. Пользователи должны быть в состоянии получить доступ к приложению на множестве языков или с других, отличных от PC, устройств. Приложение должно использовать как статические, так и динамические знания, например, комбинация статических онтологий и динамических технологических процессов. Наконец, приложение должно быть масштабируемым (в терминах количества используемых данных и совместно работающих распределенных компонент).

Итоги состязания между представленными проектами ежегодно подводятся на Всемирной конференции по Semantic Web, где обсуждаются научные решения и проблемы, возникшие на данном этапе развития Semantic Web. На последней VI конференции 2007 г. в Корею было выделено 2 поколения приложений Semantic Web [111]. Первое поколение – Семантически привязанные приложения Semantic Web – Semantically Closed SW Applications. Эти приложения используют единую онтологию, очень привязаны к семантическим ресурсам, ограничены в интерактивности. Такие приложения предоставляют однородное представление гетерогенных источников данных и очень ограниченно используют существующие в Semantic Web данные. Существующие на данный момент приложения Semantic Web более похожи на традиционные системы, ориентированные на знания.

В настоящее время встает задача создания приложений второго поколения. Второе поколение приложений Semantic Web должны использовать весь огромный запас уже накопленной семантики. Приложения Semantic Web 2-го поколения должны быть способны использовать:

- множество онтологий;

- быть открытыми для семантических ресурсов;
- быть открытыми для работы с пользователем (user interaction).

В идеале они также должны уметь использовать не только данные Semantic Web, но и другие форматы данных, например, фолксономии и т.п., следовательно должны иметь мощные механизмы по автоматическому извлечению информации.

Также на этой конференции было показано, как Semantic Web предлагает решение проблемы объединения данных, а также практические результаты этой работы.

Результаты VI конференции по Semantic Web показали, что:

- большинство из событий, которые были предположены, свершились, или свершаются в данный момент, темпы этого движения ускоряются;
- некоторые достижения происходят быстрее, чем планировалось ранее (массовый рост RDF-хранилищ, представление рассуждений, наличие онтологий – но очень плохо связанных);
- некоторые планы пока слабо реализуются, но движение в этих направлениях продолжается (публичные источники информации RDF, OWL, зарождение «всепроникающих» вычислений);
- слабое развитие технологии агентов [108].

Заключение

Semantic Web – это динамичная, постоянно развивающаяся концепция, а не набор комплексных, работающих систем.

С точки зрения машинной обработки данных – "Semantic Web – это идея хранения данных в Web таким образом, чтобы они были определены и связаны для дальнейшей возможности автоматизированной обработки, интеграции и повторного использования их в различных приложениях." [9]

С точки зрения интеллектуальных агентов «целью Semantic Web является сделать существующий Web более машинночитаемым с тем, чтобы иметь возможность использовать интеллектуальных агентов для поиска и обработки соответствующей информации.» [112]

С точки зрения распределенных баз данных «концепция Semantic Web заключается в «... обеспечении достаточной гибкости для возможности представления всех баз данных и правил логики таким образом, чтобы связать их все вместе...» [9] "Простое описание Semantic Web заключается в том, что он представляет собой попытку реализовать машинную обработку данных...В частности, трансформировать обработку информации обеспечением общего принципа, по которому данные могут быть получены, связаны вместе и поняты. Перевод Web от типа «большой книги с гиперссылками» к большой связанной базе данных"[112].

С точки зрения автоматизированной инфраструктуры – «Semantic Web является инфраструктурой, а не приложением» [113].

С точки зрения обслуживания человеческих потребностей – идея Semantic Web заключается в освобождении человека от обременительных рутинных задач по добыче, поиску, учету и индексированию информации, содержащейся в Web. «Semantic Web – это видение следующего поколения Интернет, который позволит веб-приложениям автоматически собирать веб-документы из различных источников, учитывать и обрабатывать информацию, а также взаимодействовать с другими приложениями для выполнения сложных задач" [114].

С точки зрения улучшения аннотирования – «идея Semantic Web состоит в обеспечении существующего Web аннотациями, выраженными в машиннообрабатываемой форме и связанными между собой» [115].

С точки зрения улучшения поиска – реализация поиска не только по ключевым словам, но и по контенту.

С точки зрения веб-сервисов – «Semantic Web должен обеспечить доступ не только к статичным документам, содержащим полезную информацию, но и к сервисам, которые предоставляют полезные услуги» [116].

Таким образом, задачи Semantic Web, а равным образом и его проблемы заключаются в следующем:

- индексация и поиск информации;
- разработка и поддержка метаданных;
- разработка и поддержка методов аннотирования;
- представление Web в виде большой, интероперабельной базы данных;
- организация машинной добычи данных;
- обнаружение (discovery) и предоставление веб-ориентированных сервисов;
- исследования в области интеллектуальных программных агентов.

Дополнительная библиография по представленной тематике приведена в [117].

1. W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity>
2. SemanticWeb organization. – <http://www.semanticWeb.org/>
3. Getting into RDF "Semantic Web using N3", Tim Berners-Lee – <http://www.w3.org/2000/10/swap/Primer.html>
4. "Web Architecture: Describing and Exchanging Data", Berners-Lee, Connolly, Swick, W3C Note 7 June 1999. – <http://www.w3.org/1999/04/WebData>
5. Metadata Architecture, W3C Design Issues. – <http://www.w3.org/DesignIssues/Metadata>
6. RDF and Metadata, Tim Bray, June 09, 1998. – <http://www.xml.com/xml/pub/98/06/rdf.html>
7. The Power of Metadata, book chapter by Rael Dornfest, Dan Brickley. – <http://www.openp2p.com/pub/a/p2p/2001/01/18/metadata.html>

8. *Web Metadata: A Matter of Semantics* by Ora Lassila, IEEE Internet Computing, July-August 1998. – <http://computer.org/internet/ic1998/w4030abs.htm>
9. *W3C, The Semantic Web Home Page.* – <http://w3.org/sw/>
10. *AgentWeb*, resource guide and newsfeed covering Agent-related technologies. – <http://agents.umbc.edu/>
11. *A Model-Theoretic Semantics for DAML+OIL*, W3C Note 18 December 2001. – <http://www.w3.org/TR/daml+oil-model>
12. *An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL*, W3C Note 18 December 2001. – <http://www.w3.org/TR/daml+oil-axioms>
13. *DAML+OIL* (March 2001) Reference Description, W3C Note 18 December 2001. – <http://www.w3.org/TR/2001/NOTE-daml+oil-reference-20011218>
14. *XML Schema, RDF Schema & DAML Comparison.* – <http://www.isi.edu/expect/Web/semanticWeb/comparison.html>
15. *W3C Web Ontology.* – <http://www.w3.org/2001/sw/WebOnt/>
16. *Requirements for a Web Ontology Language*, W3C Working Draft. – <http://www.w3.org/TR/Webont-req/>
17. *SemanticWeb: роль XML и RDF/ С. Декер, С. Мельник, Ф. ван Хермелен, Д. Фенсел, М. Клейн, Д. Брукстра, М. Эрдманн, Я. Хорроке // Открытые системы. 2001 - № 9.* – <http://www.osr.ru/os/2001/09/041.htm>.
18. *Distributed XML: the role played by XML in the next-generation Web*, Edd Dumbill. – <http://www.xml.com/pub/2000/09/06/distributed.html>
19. *XML and the Web*, by Tim Berners-Lee, XML World 2000, Boston 2000/09/06. – <http://www.w3.org/2000/Talks/0906-xmlWeb-tbl/>
20. *An Introduction to the Resource Description Framework* by Eric Miller, D-Lib Magazine, May 1998. – <http://www.dlib.org/dlib/may98/miller/05miller.html>
21. *Putting RDF to Work*, Edd Dumbill. – <http://www.xml.com/pub/2000/08/09/rdfdb/index.html>
22. *RDF tutorial*, Pierre-Antoine Champin (for developers). – <http://www710.univ-lyon1.fr/~champin/rdf-tutorial/>
23. *W3C Web Service's Home Page.* – <http://www.w3.org/2002/ws/>
24. *Web Services Architecture*, W3C Working Draft 14 November 2002. – <http://www.w3.org/TR/ws-arch/>
25. *Web Services Architecture Requirements*, W3C Working Draft 14 November 2002. – <http://www.w3.org/TR/wsa-reqs>
26. *Web Services Architecture Usage Scenarios*, W3C Working Draft 30 July 2002. – <http://www.w3.org/TR/ws-arch-scenarios/>
27. *Web Services Description Requirements*, W3C Working Draft 28 October 2002. – <http://www.w3.org/TR/ws-desc-reqs/>
28. *Web Services Glossary*, W3C Working Draft 14 November 2002. – <http://www.w3.org/TR/ws-gloss/>
29. *Лущиц Ю., Семантический Веб, лекция, 2006.* – <http://logic.pdmi.ras.ru/~yura/internet.html>
30. *The Semantic Web.* By Tim Berners-Lee, James Hendler and Ora Lassila. Scientific American, May 17, 2001. – <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>
31. *The Semantic Web Roadmap*, Tim Berners-Lee, 1998. – <http://www.w3.org/DesignIssues/Semantic.html>
32. *State of The SemanticWeb*, Ivan Herman, Stavanger, Norway, 2007.
33. *Semantic Web for Developers.* – <http://logicerror.com/semanticWeb-Webdev>
34. *Extensible Markup Language (XML) 1.0*, W3C Recommendation 10.02.1998. – <http://www.w3.org/TR/1998/REC-xml-19980210>
35. *RDF/XML Syntax Specification (Revised)*, W3C Working Draft 25 March 2002. – <http://www.w3.org/TR/rdf-syntax-grammar/>
36. *RDF Model Theory*, W3C Working Draft 29 April 2002. – <http://www.w3.org/TR/rdf-ml/>
37. *RDF Semantics*, W3C Working Draft 23 January 2003. – <http://www.w3.org/TR/2003/WD-rdf-ml-20030123/>
38. *RDF Primer*, W3C Working Draft 11 November 2002. – <http://www.w3.org/TR/rdf-primer/>
39. *RDF Test Cases*, W3C Working Draft 12 November 2002. – <http://www.w3.org/TR/rdf-testcases>
40. *RDF Tutorial*, W3C. – <http://www.w3.org/TR/rdf-tutorial>
41. *Resource Description Framework (RDF): Concepts and Abstract Data Model*, W3C Working Draft 29 August 2002. – <http://www.w3.org/TR/rdf-concepts/>
42. *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendation 22 February 1999. – <http://www.w3.org/TR/REC-rdf-syntax/>
43. *Using RDF to model multimedia content – slide "Relation with MPEG-7".* – <http://www.w3.org/Architecture/1998/06/Workshop/paper29/slides/slide13-0.html>
44. *RDF syntax*, W3C Recommendation. – <http://www.w3.org/TR/PR-rdf-syntax>
45. *RDF Schema*, W3C Working Draft. – <http://www.w3.org/TR/PR-rdf-schema>
46. *RDF Vocabulary Description Language 1.0: RDF Schema*, W3C Working Draft 23 January 2003. – <http://www.w3.org/TR/2003/WD-rdf-schema-20030123/>
47. *Topic Maps (XMT).* – <http://www.topicmaps.org/>
48. *Text Encoding Initiative.* – <http://www.tei-c.org/>
49. *Metadata Encoding and Transmission Standard.* – <http://www.loc.gov/standards/mets/>
50. *Metadata Object Description Schema (MODS).* – <http://www.loc.gov/standards/mods>
51. *Encoded Archival Description (EAD).* – <http://www.loc.gov/ead>
52. *Learning Object Metadata (LOM).* – <http://www.ltsc.ieee.org/wg12/>
53. *Online Information Exchange (ONIX).* – <http://www.editeur.org/onix.html>
54. *Introduction to the Semantic Web*, Ivan Herman, W3C, International Conference on Dublin Core and Metadata Applications, Singapore, 2007-08-31. – <http://www.w3.org/2007/Talks/0831-Singapore-IH/>
55. *The Friend of a Friend (FOAF) project.* – <http://www.foaf-project.org/>
56. *FOAF Vocabulary Specification.* – <http://www.xmlns.com/foaf/0.1/>
57. *FOAF Vocabulary Specification.* – <http://www.xmlns.com/foaf/spec/>
58. *Semantically-Interlinked Online Communities.* – <http://www.sioc-project.org/>
59. *Description of a Project Description of a Project (DOAP) vocabulary.* – <http://www.usefulinc.com/doap/>
60. *RFC2413, Dublin Core Metadata for Resource Discovery.* – <http://www.faqs.org/rfcs/rfc2413.html>
61. *"DublinCore Qualifiers/Substructure".* – <http://www.loc.gov/marc/dcqualif.html>
62. *"DublinCore qualifiers".* – <http://www.roads.lut.ac.uk/Metadata/DC-Qualifiers.html>
63. *Dublin Core Element Set, Version 1.1 – Reference Description.* – <http://www.dublincore.org/documents/1999/07/02/dces/>
64. *vCard.* – <http://www.imc.org/pdi/>
65. *Names in Dublin Core*, Diane I. Hillmann. – <http://purl.org/dc/documents/notes/notes-hillmann-19981027.htm>
66. *"Guidance on expressing the Dublin Core within the Resource Description Framework (RDF)".* – <http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/>
67. *Representing vCard v3.0 in RDF*, Renato Iannella. – <http://www.dstc.edu.au/RDU/RDF/draft-iannella-vcard-rdf-00.txt>
68. *ROADS.* – <http://ukoln.bath.ac.uk/roads/>
69. *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendation 22 February 1999. – <http://www.w3.org/TR/REC-rdf-syntax/>
70. *Open Directory Project.* – <http://dmoz.org/>
71. *RDF Vocabulary Description Language 1.0: RDF Schema*, W3C Working Draft 23 January 2003. – <http://www.w3.org/TR/2003/WD-rdf-schema-20030123/>
72. *DAML+OIL Project Homepage.* – <http://www.w3.org/TR/daml+oil-reference>

73. *DAML+OIL Primer*. – <http://www.w3.org/TR/rdf-primer/#ref-damloil>
74. *Язык OWL*. – <http://www.w3.org/TR/owl-ref/>
75. *OWL, Primer*. – <http://www.w3.org/TR/rdf-primer/#ref-owl>
76. *XML-QL: A Query Language for XML*. Submission to the World Wide Web Consortium 19.08.1998. – www.w3.org/TR/NOTE-xml-ql/
77. *XQL Tutorial (XML Query Language)*, Jonathan Robie. – <http://www.metalab.unc.edu/xql/xql-tutorial.html>
78. *XML-QL : A Query Language for XML User's Guide Version 0.9*. – <http://www.research.att.com/~mff/xmlql/doc/>
79. *Home of the W3C's XML Query working group*. – <http://www.w3.org/XML/Query>
80. *A Query Language for XML*. Alin Deutsch, Mary Fernandez, Daniela Florescu. University of Pennsylvania, Philadelphia. – <http://www8.org/w8-papers/1c-xml/query/query.html>
81. *RDF Query Language (RQL)*. – <http://139.91.183.30:9090/RDF/VRP/index.html/RQL/index.html>
82. *The RDF Query Rules, W3C*. – <http://www.w3.org/2001/11/13-RDF-Query-Rules/>
83. *The RDF Query Language (RQL), W3C*. – <http://139.91.183.30:9090/RDF/RQL/>
84. *RDF Query Specification, December 3, 1998*. – <http://www.w3.org/TandS/QL/QL98/pp/rdfquery.html>
85. *TRIPLE HomePage*. – <http://triple.semanticWeb.org/>
86. *Sesame, storage and querying middleware system for RDF and RDF Schema*. – <http://sesame.aidadministrator.nl/>
87. *SPARQL Query Language for RDF W3C Candidate Recommendation 14 June 2007*. – <http://www.w3.org/TR/rdf-sparql-query/>
88. *The Semantic Web In Breadth*, Aaron Swartz. – <http://logicerror.com/semanticWeb-long>
89. *Математические предпосылки (логика предикатов) – Mathematical Background (Predicate Logic)*, Джон Сова (John Sowa's). – <http://www.jfsowa.com/logic/math.htm>
90. *RIF: Use Cases and Requirements, W3C Working Draft 10 July 2006*. – <http://www.w3.org/TR/2006/WD-rif-ucr-20060710/>
91. *RuleML*. – <http://www.ruleml.org/>
92. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission 21 May 2004*. – <http://www.w3.org/Submission/SWRL/>
93. *Презентация доклада «Семантический Веб: текущее состояние исследований и перспективные направления», Уланов Д., ИСП РАН, 03.02.2006.* – http://dulanov.wordpress.com/2006/02/02/prezentatsiya_o_proekte_semanticheskii_veb/
94. *SOAP Version 1.2 Part 0: Primer, W3C Candidate Recommendation 19 December 2002*. – <http://www.w3.org/TR/2002/CR-soap12-part0-20021219/>
95. *SOAP Version 1.2 Part 1: Messaging Framework, W3C Candidate Recommendation 19 December 2002*. – <http://www.w3.org/TR/2002/CR-soap12-part1-20021219/>
96. *SOAP Version 1.2 Part 2: Adjuncts, W3C Candidate Recommendation 19 December 2002*. – <http://www.w3.org/TR/2002/CR-soap12-part2-20021219/>
97. *SOAP Version 1.2 Specification Assertions and Test Collection, W3C Working Draft 26 June 2002*. – <http://www.w3.org/TR/soap12-testcollection>
98. *SOAP Version 1.2 Usage Scenarios, W3C Working Draft 26 June 2002*. – <http://www.w3.org/TR/xmlp-scenarios/>
99. *SOAP 1.2 Attachment Feature, W3C Working Draft 24 September 2002*. – <http://www.w3.org/TR/soap12-af/>
100. *SOAP Version 1.2 Email Binding, W3C Note 3 July 2002*. – <http://www.w3.org/TR/soap12-email>
101. *Universal Description, Discovery, and Integration (UDDI) OASIS Standard*. – <http://www.uddi.org>
102. *Web Services Description Language (WSDL) Version 1.2, W3C Working Draft 24 January 2003*. – <http://www.w3.org/TR/2003/WD-wsdl12-20030124/>
103. *Semantic Markup for Web Services, W3C Member Submission 22 November 2004* <http://www.w3.org/Submission/OWL-S/>
104. *Service Modeling Language, Version 1.1 W3C Working Draft 3 March 2008*, <http://www.w3.org/TR/sml/>
105. *Open Directory Project, RDF dumps*. – <http://dmoz.org/rdf.html>
106. *Google Search Engine*. – <http://google.com>
107. *SWAD-Europe: Mapping Semantic Web Data with RDBMSes, W3C Semantic Web Advanced Development for Europe (SWAD-Europe), 2003-01-23*. – http://www.w3.org/2001/sw/Europe/reports/scalable_rdbms_mapping_report/
108. *Introduction and Overview to the Semantic Web*, James A. Hendler, Rensselaer Polytechnic Institute, The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, 11-15 ноября 2007. – http://videlectures.net/iswc07_hendler_ios/
109. *The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, 11-15 November 2007, Busan, Korea*. – <http://iswc2007.semanticweb.org/main/default.asp>
110. *Semantic Web Challenge Homepage*. – <http://challenge.semanticWeb.org/>
111. *Enrico Motta, The Open University, Semantic Web Applications, The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference, 11-15 ноября 2007*. – http://videlectures.net/iswc07_motta_swa/
112. *Sean B. Palmer, The Semantic Web: An Introduction, 2001-09*. – <http://infomesh.net/2001/swintro>
113. *Semantic Web As "Perfection Seeking:" A View from Drug Terminology*, Tuttle M., Brown S., Campbell K., Carter J., Keck K., Lincoln M., Nelson S., Stonebraker M., 2001.
114. *Semantic Web Modeling and Programming with XDD*, Anutariya, Wuwongse, Akama, Wattanapailin, In Proceedings of SWWS'2001.
115. *Towards a principled approach to semantic interoperability*, Euzenat, IJCAI 2001, Workshop on ontology and information sharing, 2001, Seattle (WA US)
116. *"Explorer's Guide to the Semantic Web"*, Thomas B. – Passin; June2004. – 304 p.
117. *Библиография по тематике Semantic Web, Type of content, Class blog*. – <http://typecontent.net/blog/wp-content/uploads/2007/02/semanticbibliography.pdf>