

О «Semantic Web» и работе над средствами описания семантики в Сети

(Ниже представлена информация о из учебника:

Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д.

[Онтологии и тезаурусы: модели, инструменты, приложения](#)

БИНОМ. Лаборатория знаний, Интернет-университет информационных технологий - ИНТУИТ.ру, 2009)

Идея Семантической Сети (Semantic Web) впервые была провозглашена в 2001 году Тимом Бернерсом-Ли (создателем World Wide Web). Однако она не является новой ни для автора, ни для web-сообщества в целом. Суть ее состоит в автоматизации "интеллектуальных" задач обработки значения (в семантическом смысле) тех или иных ресурсов, имеющих в Сети. Обработкой и обменом информации должны заниматься не люди, а специальные интеллектуальные агенты (программы, размещенные в Сети). Но для того, чтобы взаимодействовать между собой, агенты должны иметь общее (разделяемое всеми) формальное представление значения для любого ресурса. Именно для цели представления общей, явной и формальной спецификации значения в Semantic Web используются онтологии.

На рис. представлена диаграмма, называемая иногда стеком (или даже "слоеным пирогом") Semantic Web.

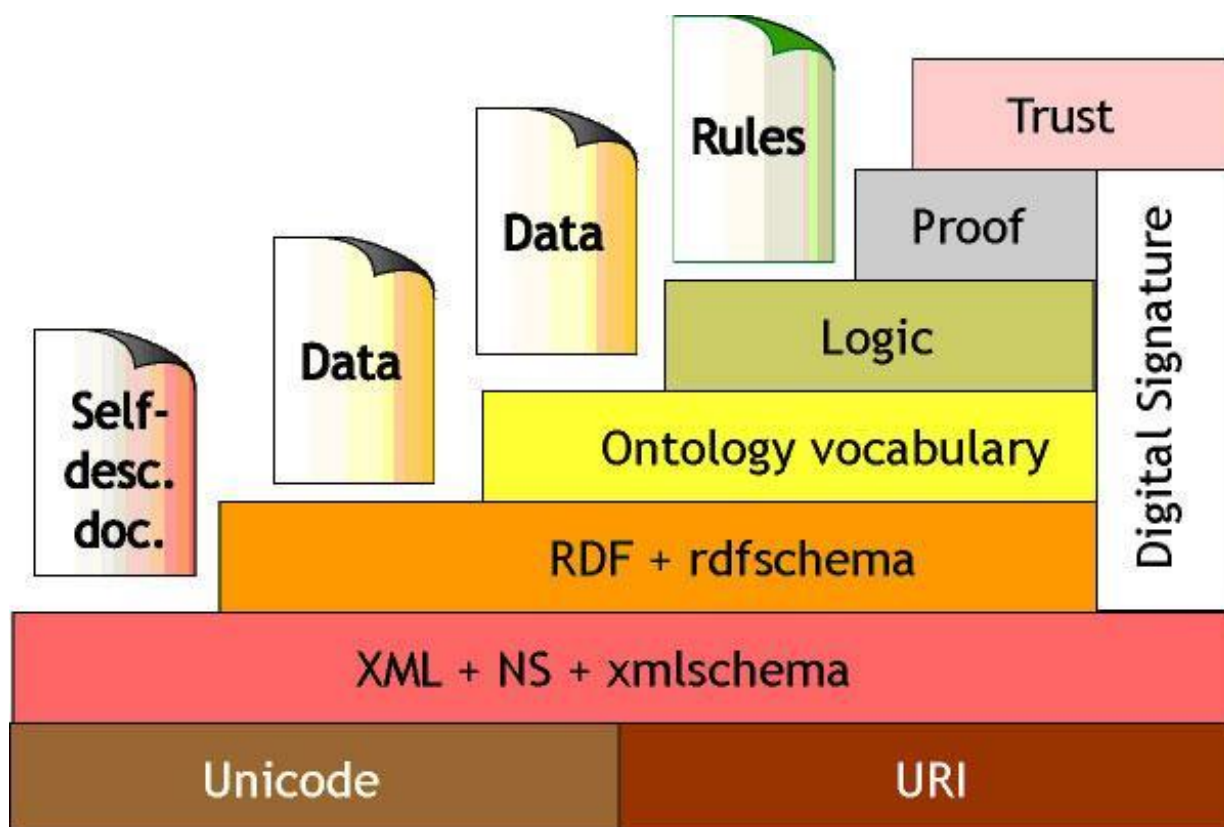


Рис. Диаграмма Semantic Web

За время, прошедшее с момента первой публикации о Semantic Web, был разработан целый ряд стандартов и рекомендаций, реализовано множество проектов. Но, несмотря на отдельные успехи, до сих пор (и это признает сам Т. Бернерс-Ли) нельзя сказать, что идея Semantic Web реализована на практике.

Все основные уровни диаграммы описаны ниже.

Работа над средствами описания семантики в Сети началась задолго до публикации 2001 года. В 1997 году консорциум W3C определил спецификацию RDF (Resource Description Framework). RDF предоставляет простой, но мощный язык описания ресурсов, основанный на триплетях (triple-based) "Субъект-Предикат-Объект" и спецификации URI. В 1999 году RDF получает статус рекомендации. Этот шаг в направлении улучшения функциональности и обеспечения

интероперабельности (т.е. возможности обмениваться данными несмотря на их разнородность) в Сети считается одним из важнейших. Концептуально RDF дает минимальный уровень для представления знаний в Сети. Спецификация RDF опирается на ранние стандарты, лежащие в основе Web:

- Unicode служит для представления символов алфавитов различных языков,
- URI используется для определения уникальных идентификаторов ресурсов,
- XML и XML Schema — для структурирования и обмена информацией и для хранения RDF (XML синтаксис RDF).

Кроме RDF был разработан язык описания структурированных словарей для RDF — RDF Schema (RDFS). Он предоставляет минимальный набор средств для спецификации онтологий. RDFS получил статус рекомендации W3C в 2004 году. Однако препятствием для Semantic Web стало то, что документов, написанных на языке RDF/RDFS, было относительно мало. В период с 2001 по 2004 годы шла интенсивная работа по созданию программных средств для обработки и автоматической генерации RDF-документов.

Результатом в 2004 году стал язык GRDDL (Gleaning Resource Descriptions form Dialects of Languages). Его назначение состоит в предоставлении средств для извлечения RDF-триплетов из XML и XHTML данных (в особенности это относится к документам, автоматически генерируемым из закрытых баз данных). Развивалось и программное обеспечение для Semantic Web. В области создания библиотек классов и построения логических выводов над RDF-графами была создана библиотека Jena Framework, в области создания модулей расширения для браузеров — Simile для Firefox. В области создания визуальных сред редактирования большое число редакторов онтологий стали поддерживать RDF.

В 2004 году статус рекомендации получил язык OWL (Web Ontology Language). Он имеет 3 диалекта (3 множества структурных единиц), используемых в зависимости от требуемой выразительной мощности. OWL фактически является надстройкой над RDF/RDFS и поддерживает эффективное представление онтологий в терминах классов и свойств, обеспечение простых логических проверок целостности онтологий и связывание онтологий друг с другом (импорт внешних определений). Многие формализмы описания знаний могут быть отображены на формализм OWL (два из его диалектов — OWL Lite и OWL DL — соответствуют двум дескриптивным логикам, имеющим разную выразительную силу). Большое число создаваемых в настоящее время онтологий кодируются на OWL; уже существующие онтологии транслируются в него.

На этом работа по обеспечению Semantic Web необходимыми стандартами не остановилась.

В 2005 году началась работа над форматом обмена правилами — RIF (Rule Interchange Format). Его назначение — соединить в одном стандарте несколько формализмов для описания правил (по которым может осуществляться нетривиальный логический вывод): логику клауз Хорна, логики высших порядков, продукционные модели и т.п.

Язык SPARQL — язык запросов к RDF-хранилищам — в январе 2008 года приобрел статус официальной рекомендации Консорциума W3C. Синтаксически он очень похож на SQL. Он уже широко используется разработчиками информационных систем.

Уровням "Ontology vocabulary" и "Logic" соответствуют OWL и RIF. Уровень "Trust" на данный момент остается незатронутым никакими стандартами. Здесь и возникает одно из существенных препятствий к реализации всей идеи: поддержка автоматической проверки корректности и правдивости информации. В самом деле, у многих поставщиков семантических описаний может возникнуть соблазн "обмануть" программу-агента, предоставив информацию, не соответствующую действительности, либо навязчивую рекламу, как это в настоящее время проделывается с поисковыми машинами, спам-фильтрами и т.п.

Еще одним камнем преткновения для создания Semantic Web является фактическое отсутствие работающих интеллектуальных агентов. Не всякая программа, обрабатывающая RDF, является агентом Semantic Web, точно так же как и не всякая программа, написанная на ПРОЛОГЕ, является приложением в области искусственного интеллекта.

Завершая раздел, нужно подчеркнуть, что Семантическая Сеть продолжает развиваться — появляются новые стандарты.

Новый шаг — начало разработки формата обмена правилами RIF, построенными над онтологиями, и определение требований и области его применения. Появилось множество свободно распространяемых библиотек для разработки приложений "под Semantic Web". Главными задачами, стоящими перед сообществом Семантической Сети, остаются создание новых онтологий и согласование существующих.