

Лекция № 8

Информационные технологии. Семантический Веб «Алгоритмы для Интернета»

*Кафедра прикладной математики
и кибернетики СибГУТИ*

Юрий Иванович Молородов

yumo@ict.sbras.ru

Содержание курса

1. Основы распределенных вычислительных систем
2. Системы переноса кода
3. Агентные технологии
4. Технологии одноранговых вычислений
5. Объектные и компонентные распределенные системы
6. Сервис-ориентированные технологии и веб-сервисы
7. Грид-технологии

Введение

Существующая сеть WWW представляет собой гигантское количество информации в формате, приспособленном для человеческого восприятия. Пользователь может переходить с одной ссылки на другую, давать запросы различным поисковым системам или же находить сайты, просто вводя их адреса.

И хотя веб-страницы весьма привлекательны для человека, для компьютерной программы же, обрабатывающей их содержимое, они не более чем строчки из случайных символов.

Введение

Всемирная паутина представляет собой информационный ресурс с практически неограниченным потенциалом.

Однако, этот потенциал является относительно неиспользованными, поскольку он является сложным для машиной обработки, а интегрировать эту информацию необходимо.

В последнее время исследователи приступили к изучению возможности привлечения Веб-содержимого в целях создания Semantic Web.

Вместо того чтобы полагаться на обработку естественного языка, для извлечения смысла из имеющихся документов, такой подход требует, чтобы авторы описывали документы, используя знания языка представления.

Портал

Сайт (совокупность сайтов), обеспечивающий удовлетворение основных потребностей пользователей путем реализации услуг в следующих областях: информация, бизнес, общение, а также предоставления инструментария, необходимого пользователю для продвижения собственного контента в рамках портала.

Т. о. портал должен охватывать 4 основных типа услуг:

- информационные (помогают найти и получить информацию);
- бизнес-функции (ориентация непосредственно на продажу товаров и услуг);
- инструментарий пользователя (помогает создавать и продвигать свой "контент" в Сети, прежде всего бесплатные хостинг и e-mail, рейтинги, баннеры, "анонсировщики" и др.);
- Коммуникация (обеспечение общения).

Введение

Компьютерная программа не способна, загрузив произвольный документ, будь то веб-страница или какой-то файл, понять его содержание. Она, конечно, может сделать некие догадки, основываясь на HTML- или XML-тэгах, но всё равно требуется человек-программист, который должен разобраться в них и понять смысл, или семантику, каждого из тэгов.

С точки зрения компьютера, существующая Сеть WWW — это полная неразбериха.

К счастью, выход есть: это Семантическая Сеть.

Semantic Web

Семантическая паутина (*Semantic Web*) — это направление развития Всемирной паутины, целью которого является представление информации в виде, пригодном для машинной обработки.

В обычной Паутине, основанной на HTML-страницах, информация заложена в тексте страниц и извлекается человеком с помощью браузера.

Семантическая паутина предполагает запись информации в виде семантической сети с помощью онтологий.

Т.о. программа-клиент может непосредственно извлекать из паутины факты и делать из них логические заключения.

Semantic Web

Семантическая паутина работает параллельно с обычной Паутиной и на её основе, используя протокол HTTP и идентификаторы ресурсов URI (*Uniform Resource Identifier*). Термин «семантическая паутина» был впервые введён сэром Тимом Бернерсом-Ли (изобретателем Всемирной паутины) в мае 2001 года в журнале «Scientific American», и называется им «следующим шагом в развитии Всемирной паутины». Эта концепция была принята и продвигается Консорциумом Всемирной паутины.

Онтологии используются в процессе программирования как форма представления знаний о реальном мире или его части. Основные сферы применения — моделирование бизнес-процессов, семантическая паутина, искусственный интеллект.

Semantic Web

В интервью корреспондентам Times Online Сэр Тимоти Бернерс-Ли, рассказал о том, какие возможности мог бы предоставить пользователям интернет в случае перехода к семантической сети (Semantic Web).

Проект Semantic Web предполагает создание специализированной системы, в которой компьютеры смогут взаимодействовать друг с другом без участия человека, а приложения научатся распознавать информацию.

Для этого в каждый документ или веб-страницу будут добавляться метаданные - сведения о том, где, когда, кем был создан файл, как он отформатирован, для чего предназначен и т.д.

Semantic Web

По словам Бернерса-Ли, семантическая сеть позволит создавать веб-приложения, по функциональности превосходящие любой из современных сервисов. Он отмечает, что, например, компания Google нашла очень эффективный способ поиска страниц в интернете, но возможности этого поисковика не идут ни в какое сравнение с тем, чего можно было бы добиться в "сети будущего".

По словам Бернерса-Ли, компания, которая первой сможет реализовать возможности семантической сети, сможет без особых усилий отнять у Google корону славы. Аналогично, сверхпопулярные сейчас социальные сети Facebook и MySpace, в перспективе могут быть просто-напросто вытеснены сервисами, связывающими не только людей, а буквально всё в Сети.

Semantic Web

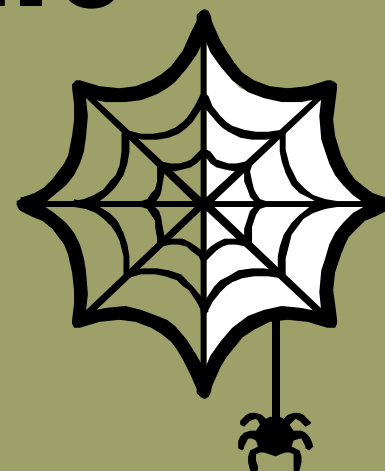
С целью иллюстрации возможностей семантической сети Бернерс-Ли приводит следующий пример.

Допустим, перечень банковских чеков и календарь могут "говорить на одном языке" и обмениваться информацией друг с другом.

В этом случае владелец счета сможет запросто совместить перечень с календарем, чтобы получить представление о том, когда и какие суммы денег были потрачены.

Концепция Web 2.0

Internet как *платформа*



Пользователи становятся активными участниками работы сети.

Они наполняют сеть информацией, сами её организуют, сами определяют параметры поиска, способы обработки...

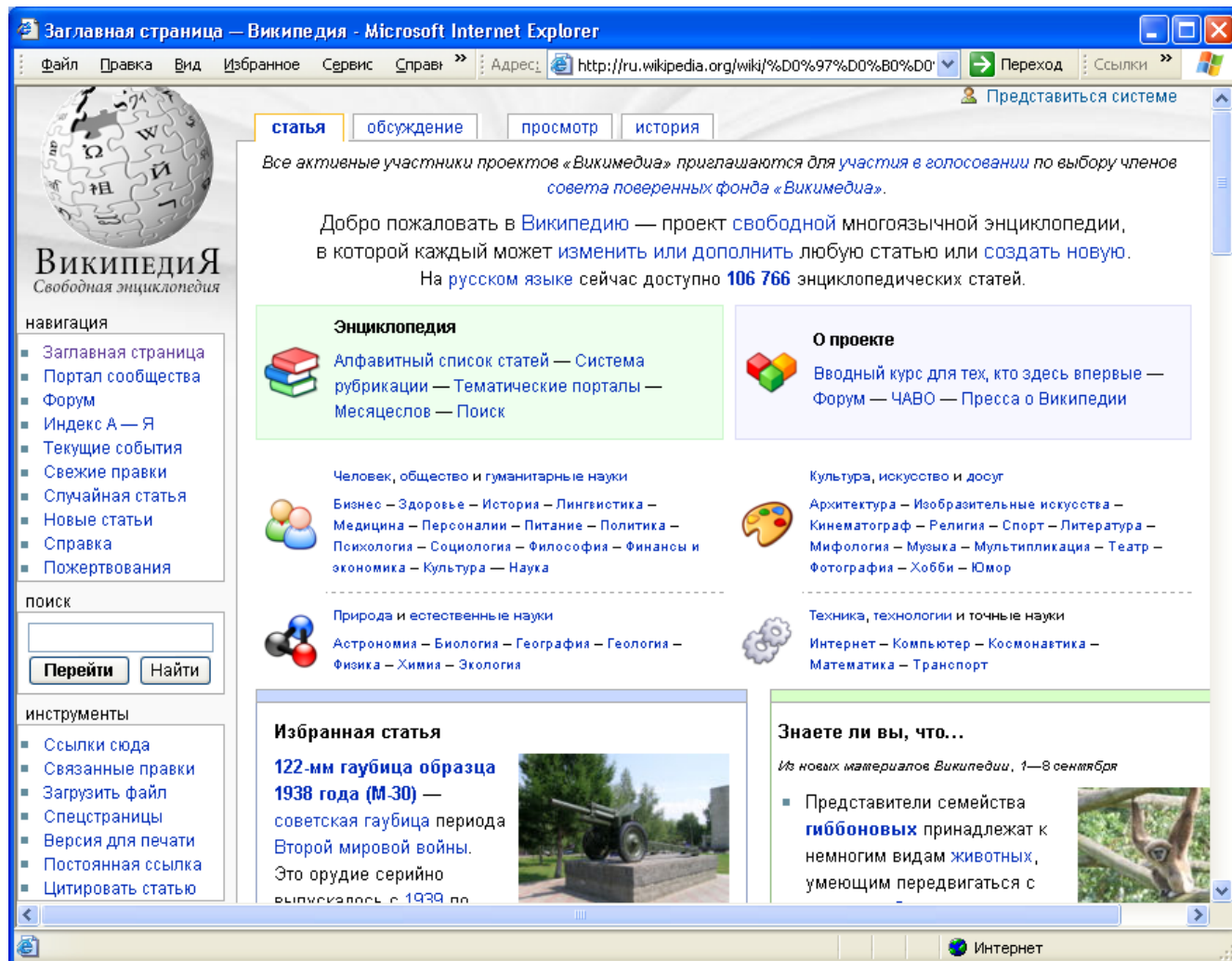
Основные свойства Web 2.0 – это простота и удобство в занесении информации, её поиске и совместном использовании.

Мы из пользователей превращаемся в соавторов.

Примеры элементов Web 2.0

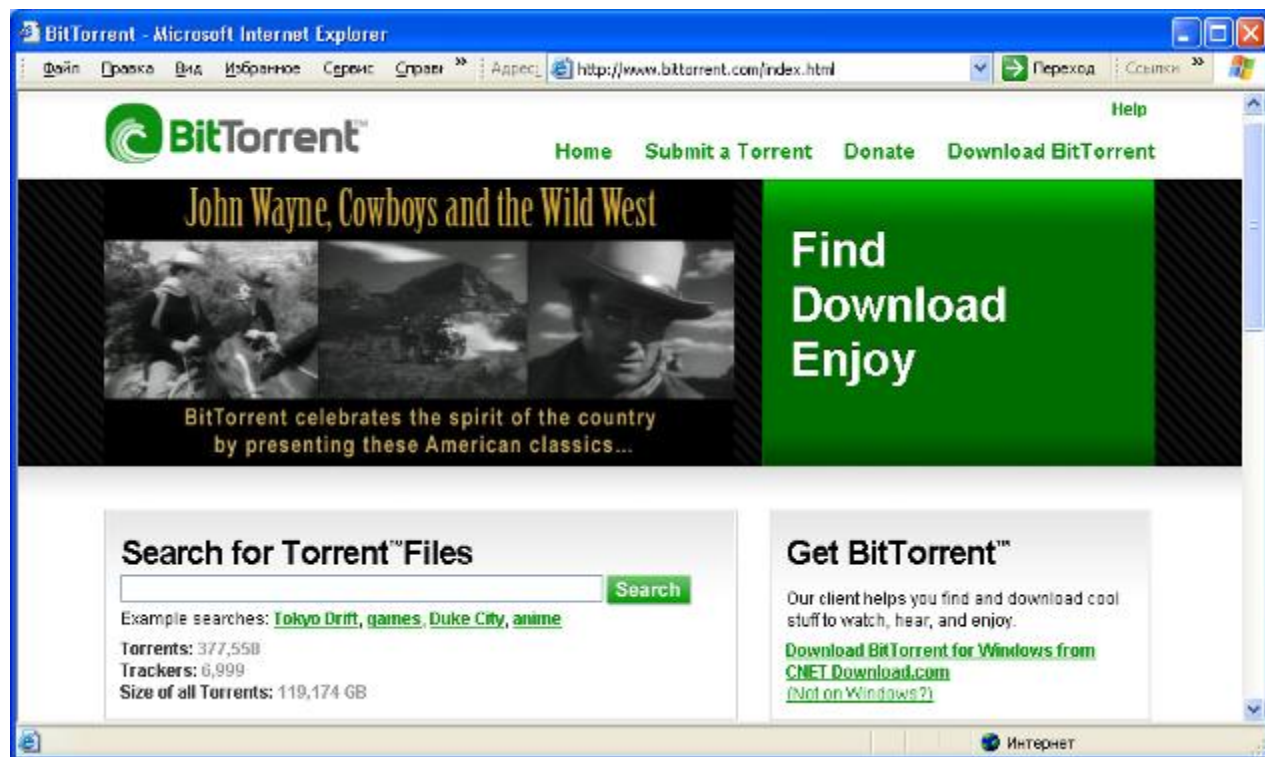
Wikipedia — энциклопедия, заполняемая совместными усилиями энтузиастов.

Они не только публикуют свои статьи, но и редактируют, статьи созданные другими авторами — *распределенная, совместная работа авторов в сети.*

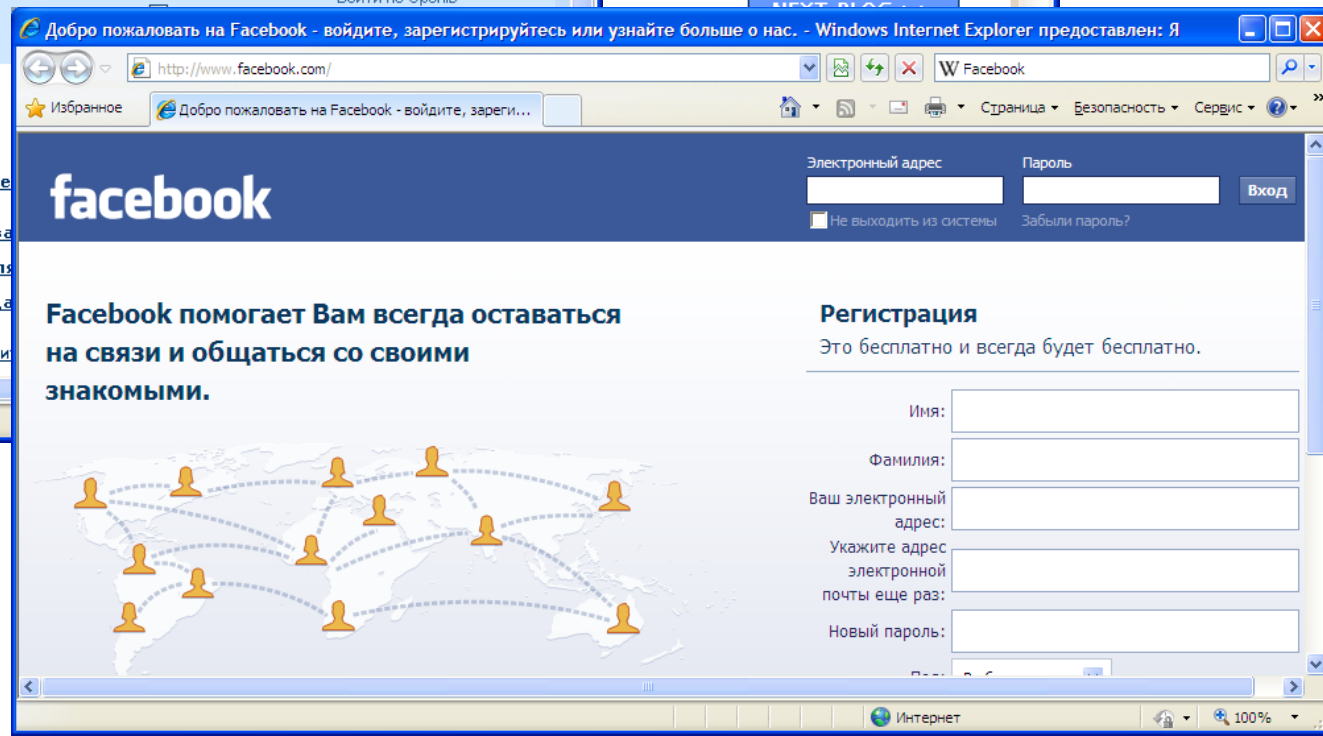
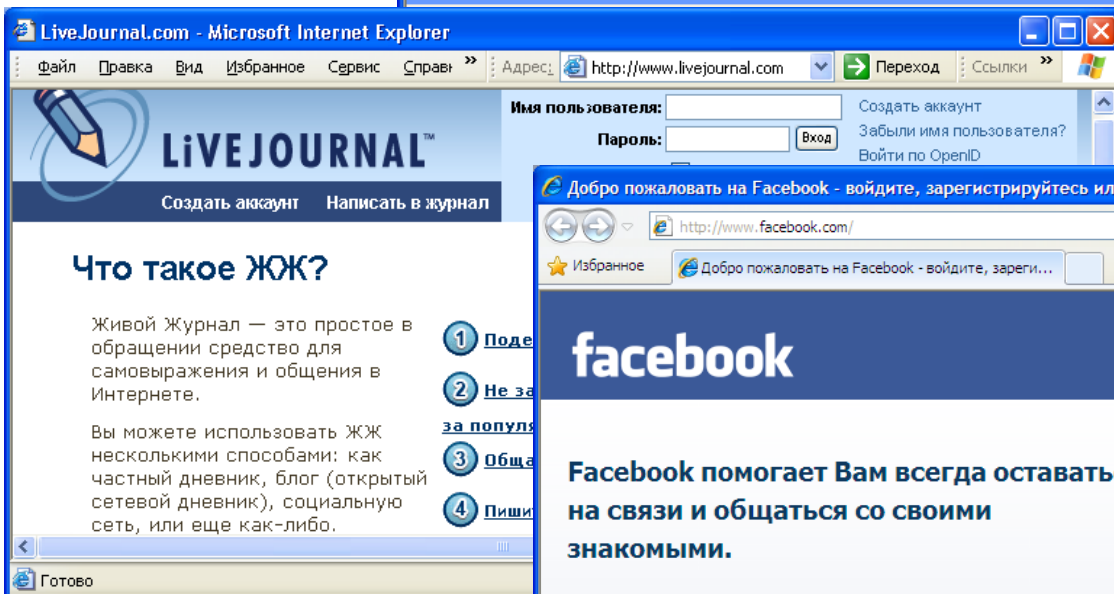
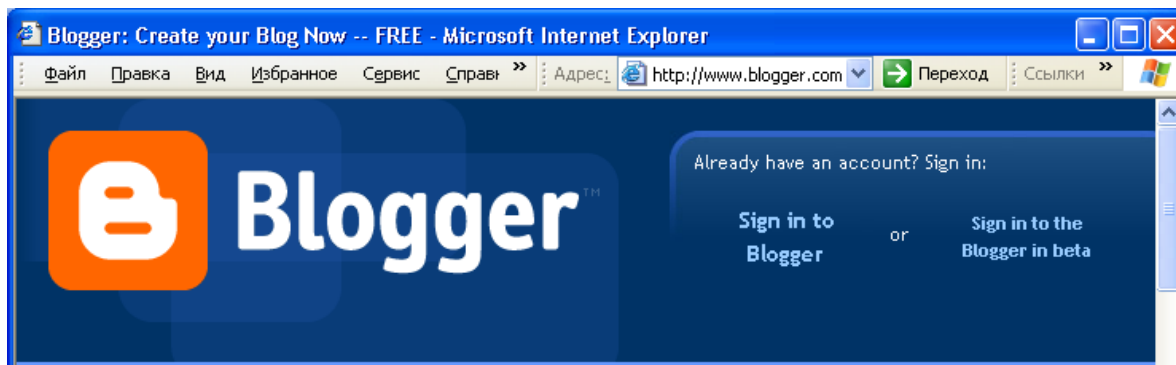


Пиринговая (**Peer to Peer** – «равный к равному») сеть **BitTorrent** – пользователи не только делятся информацией друг с другом, но и выделяют друг другу ресурсы своего компьютера – **распределенная сеть равноправных пользователей** (*децентрализация Web 2.0*). Каждый участник и клиент и сервер.

Централизованного хранилища данных нет, нагрузка сети распределяется равномерно. Чем больше пользователей будет отдавать конкретный файл, тем быстрее его получают те, у кого его еще нет. Значит качество сети улучшается по мере роста числа пользователей.



Социальные сети — Блоги, Facebook, Видео-хостинг ...



Видео-хостинг

Размещение
видео-роликов и
вещание в
реальном времени
с собственной
Web-камеры

YouTube - Broadcast Yourself. - Windows Internet Explorer предоставлен: Яндекс

http://www.youtube.com/

Избранное YouTube - Broadcast Yourself.

Поиск | Просмотр | Добавить видео | Создать аккаунт | Войти

Добро пожаловать на YouTube!
Предполагаемый язык (выбран автоматически): Русский
Чтобы изменить параметры языка, воспользуйтесь ссылками в нижней части страницы.
Нажмите "ОК", чтобы выбрать эту настройку, или "Отмена", чтобы просматривать сайт на английском языке.
OK Отмена
Show message in English

Присоединитесь к самому большому в мире видеосообществу!
Создать аккаунт > У вас уже есть аккаунт? Войти

Рекомендованные вам Подробнее

Персональные рекомендации видео
Чтобы начать получать персональные рекомендации, достаточно просмотреть несколько видео, а затем вернуться на эту страницу. Здесь появятся предложения YouTube, подобранные специально для вас.

Самые популярные

Кино и ТВ

Katy Perry's Cleavage Traumatizes Mot...
Просмотров: 521446
sxephil

Музыка

4MINUTE'Б Heart to Heart'Б
Просмотров: 210627
4minuteofficial

Новости и политика

New Video of an entire town being was...
Просмотров: 249938
pablomaron

Фильмы и анимация

Black Snake Moan and the NEXT 7!!! (D...
Просмотров: 134902
PhilipDeFranco

Спорт

Shannon Brown switch-hand dunk agains...
Просмотров: 227905
getbangedon

Обучение и стиль

Mature / Hooded eyes make-up tutorial
Просмотров: 80097
pixiwoo

Okami - Episode 46
13:38
We gain the ability to control tornadoes, and then head out to sea ...
chuggaaconroy Просмотров: 244716 От партнеров

Тенденции

Amazing Dancer - Norske Talenter 2011 - Tord (...)
от пользователя henrik117
Просмотров: 1209527

Britney Spears - Till The World Ends - LIVE @ R...
от пользователя jsleiman06
Просмотров: 772932

Britney Spears - Big Fat Bass Full Live from V...
от пользователя 08ngin01
Просмотров: 518229

Видео от партнеров

Интернет 100%

Web-сервисы

– отдельные программы многократного использования, которые представляют свои функции другим приложениям через Internet. Любое приложение может использовать службы независимо от языка их реализации и устройств на которых они установлены.

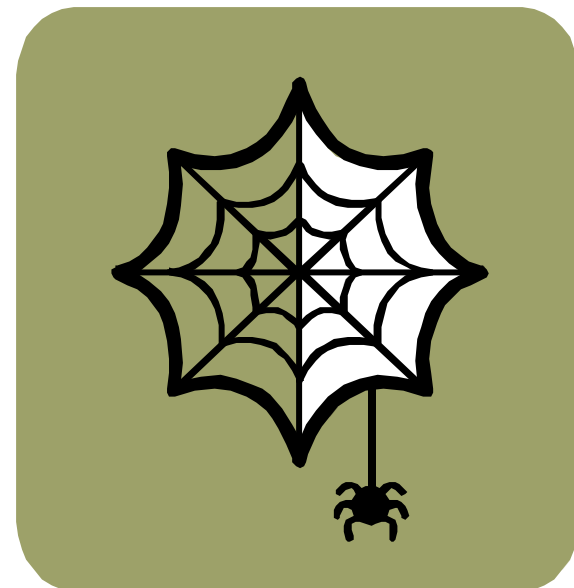
Сервис ориентированная архитектура — мечта индустрии программирования о замене «кустарного» локального кодирования программ «от и до» на «промышленную» сборку приложений из «стандартных комплектующих» (расположенных в Internet), как это делается в автомобильной, или других традиционных отраслях промышленности.

RSS-ленты (Web-каналы) – оперативная доставка динамического Web.

Концепция Web 3.0

Web как *семантическая паутина*

СЕМАНТИКА - анализ отношения между языковыми выражениями и миром, реальным или воображаемым.



Попытка человечества из «визуальных» HTML-страниц получить «смысл», что позволит автоматизировать процесс извлечения новых понятий и знаний из Internet, позволит автоматизировать анализ и прогнозирование событий на основе Internet-данных и др.

Онтологии язык **OWL** (Web Ontology Language)

В семантической паутине предполагается повсеместное использование, во-первых, универсальных идентификаторов ресурсов (**URI**), а во-вторых — *онтологий* и языков описания *метаданных* (**OWL**) вместо или в дополнение к традиционному HTML – языку *визуализации* Internet-страниц.

Онтология (в информатике) — это попытка всеобъемлющей и детальной формализации некоторой области знаний с помощью *концептуальной схемы*. Формальная семантика OWL описывает, как получить логические следствия, имея онтологию, т.е. получить факты, которые не представлены в онтологии буквально, но *следуют* из ее семантики.

Корпоративные порталы знаний

– среда для выявления, хранения, классификации, обработки и распространения знаний корпорации на основе различных методов семантики.

На жестких дисках LAN, MAN – огромные массивы документов. Необходимо из информации извлекать *знания*, т.е. кроме **базы данных** информации необходимо добавлять **онтологии** (метаданные о базе данных) =>

база знаний

Характерные задачи, решаемые в базе знаний:

- поиск в новый понятий, сбор информации об объектах, определение связей объектов,
- прогнозирование ситуаций, анализ событий, многофакторные динамические исследования,
- рекомендации по принятию решений,
- кластеризация и систематизация документов,
- автоматическое реферирование, перевод,
- проведение интеллектуально и нечеткого поиска,
- «машинные» ответы на вопросы,

Проблемы Web 3.0

На этапе обработки и вывода информации появляется ряд проблем:

1. появляется необходимость дополнительных затрат для создания семантической версии каждого сайта, что делает технологию значительно менее доступной;
2. отсутствие гарантии адекватного описывания веб-мастерами собственных ресурсов (аналогично с историей использования тега «keywords»);
3. невозможность принятия единого формата описания свойств ресурсов в условиях существующей конкуренции из-за корпоративно-рекламной политики создателей ресурса и наличия широкого поля для манипуляций описательными механизмами.

Semantic Web

Главная трудность при построении семантической сети заключается в том, чтобы найти способ унифицированного представления всех типов данных. Для этого предлагается использовать три основные составляющие - механизм описания ресурсов RDF (Resource Definition Framework), язык онтологий OWL (Web Ontology Language) и язык запросов SPARQL.

Консорциум World Wide Web (W3C) обнародовал и спецификации SPARQL. Таким образом, эра семантической сети может наступить быстрее, чем кажется.

Semantic Web

сферы применения — моделирование бизнес-процессов, семантическая паутина, искусственный интеллект.

1. История и мотивация

Семантический Веб — новая концепция развития Веба и сети Интернет, принятая и продвигаемая W3C (World Wide Web Consortium).

Эта организация разрабатывает и внедряет технологические стандарты для Всемирной паутины. Когда сейчас слово «*Интернет*» употребляется в обиходе, то чаще всего имеется в виду Всемирная паутина и доступная через нее информация, а не сама физическая сеть компьютеров.

Интернет — всемирная система объединенных компьютерных сетей, построенная на использовании протоколов (TCP/IP) для связи и маршрутизации пакетов данных.

1. История и мотивация

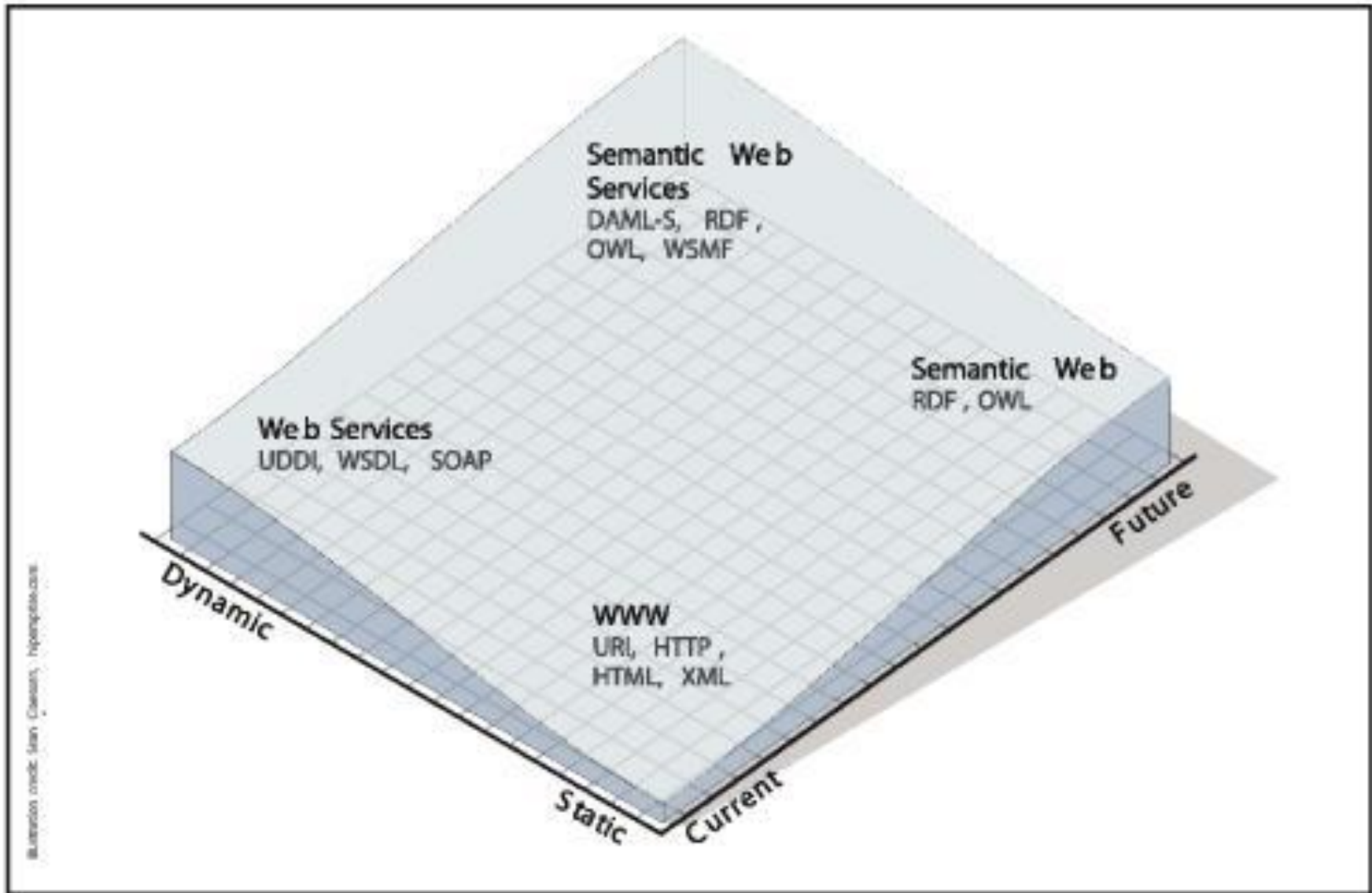


Рис.1. Эволюция сети WWW.

1. История и мотивация

Веб — глобальное информационное пространство, основанное на физической инфраструктуре Интернета и протоколе передачи данных HTTP. При его создании в 1989 Тимом Бернерсом-Ли предполагалась не только человеческая взаимосвязь, но и участие компьютеров в осмысленном манипулировании информацией.

1. История и мотивация

Основной помехой для этого является тот факт, что большинство данных в Веб_е хранится в форме, рассчитанной на восприятие человеком. Таким образом, даже при учете того, что информация получена из какой-либо базы, ее структура не очевидна роботу.

Компьютер способен умело воспроизвести веб-страницу, но, вообще говоря, у него нет надежного способа для извлечения семантики документа.

Семантический Веб как раз имеет своей целью наверстать это упущение.

1. История и мотивация

Семантический Веб — это не какая-то отдельная сеть, а расширение уже существующей, такое, что в ней информация снабжена точно определенным значением, что позволит человеку и машине успешней взаимодействовать. Первые этапы на пути к «вплетению» Семантической Сети в структуру имеющейся Сети уже осуществляются полным ходом. В ближайшем будущем данные разработки возвестят о новых значительных функциональных возможностях, когда машины станут намного лучше обрабатывать «понимать» те данные, которые сейчас они просто показывают на экране.

1.1. Сценарии будущего

Агент — программа, работающая без непосредственного управления со стороны человека или другого постоянного контроля, созданная для достижения целей, поставленных перед ней пользователем.

Обычно агенты собирают, фильтруют и обрабатывают информацию, найденную в Сети, иногда с дополнительной помощью со стороны других агентов. Семантический Веб привнесет структуру в смысловое содержание веб-страниц, тем самым создав среду, в которой агенты, переходя со страницы на страницу, смогут без особого труда выполнять запросы пользователя.

1.1. Сценарии будущего

При условии существования подобных программ возможны следующие запросы человека своему агенту:

- Закажи для меня эту книгу в ближайшей библиотеке.
- Посмотри на расписание электричек и мое расписание и выбери билеты в театр, чтобы я мог успеть после работы.
- Скажи мне, какое вино нужно купить к каждому из блюд в этом меню. И кстати, я не люблю Агдам.
- Микроволновка, сходи на сайт производителя и загрузи оптимальные параметры подогрева.

1.1. Сценарии будущего

Можно выделить несколько вариантов использования Семантического Веба.

1. ***Семантический поиск.***
2. ***Объединение знаний (интеграция баз данных).***
3. ***Всепроникающие вычисления (ubiquitous computing).***

1. История и мотивация

Семантический поиск. Поисковая система сможет выдавать только те сайты, где упоминается в точности искомое понятие, а не произвольные страницы, в тексте которых встретилось данное многозначное ключевое слово. Сегодняшние поисковые системы зачастую выдают бесчисленное множество совершенно не относящихся к запросу страниц, обрекая пользователя на длительный ручной отбор материала.

1. История и мотивация

Например, если вы ввели для поиска слово «cook», то компьютеру совершенно непонятно, имеете ли вы в виду повара, хотите ли найти информацию о рецептах приготовления пищи, или же вам нужно какое-то место, человек или компания или еще что-либо, в чьем имени или названии встречается слово «cook».

Вся проблема в том, что для компьютера слово «cook» не имеет четкого смысла, или другими словами, семантического содержания.

1. История и мотивация

Объединение знаний (интеграция баз данных).

Мы уже говорили, что отсутствие семантики ведет к тому, что компьютеру не ясно, как поступить с информацией из базы данных даже при условии того, что он знает названия всех столбцов полученной таблицы. Семантический Веб, именуя всякое понятие просто с помощью URI-идентификатора, даст возможность каждому выражать новые понятия, которые он изобретает, с минимальными усилиями. Его универсальный логический язык позволит постепенно связать все эти понятия в универсальную Сеть.

1. История и мотивация

Эта структура сделает знания и достижения человечества доступными для анализа программными агентами и предложит нам новый класс средств, с помощью которых мы сможем вместе жить, работать и учиться. В частности, работа вопросо-ответных систем станет значительно эффективней.

1. История и мотивация

Всепроникающие вычисления (*ubiquitous computing*). На следующем этапе своего развития Семантическая Сеть уже вырвется из виртуальной области и расширит сферу своего влияния на физический мир, поскольку URI-идентификаторы могут указывать на что угодно, в частности, и на физические объекты, такие как сотовый телефон или телевизор. Эти устройства смогут рекламировать свои функциональные возможности (что они умеют делать и каким образом ими можно управлять) практически так же, как это делают программные агенты.

1. История и мотивация

Например, так называемая домашняя автоматизация требует сейчас тщательной настройки всех устройств для их совместной работы.

Семантическое же описание возможностей и порядка функционирования этих устройств позволит достичь той же автоматизации, но уже с минимальным вмешательством человека.

1.2. Хронология

- 1994: Создание W3C. Консорциум разработал стандарты: HTML, URL, XML, HTTP, PNG, SVG, CSS.
 - 1998: Тим Бернерс-Ли публикует план Семантического Веба (Semantic Web Roadmap).
 - 1999: W3C создает группы по проектированию Семантического Веба, публикуется первая версия RDF.
 - 2000: Американские военные начинают исследования по описанию онтологий (DAML+OIL project).
 - 2001: В журнале Scientific American публикуется описание Семантического Веба.
 - 2004: Выпущена новая версия RDF, представлен язык описания онтологий OWL.
 - 2006: Представлена версия языка запросов SPARQL (candidate recommendation).
- В разработке проекта участвуют: HP, Mozilla, IBM, MIT, Stanford, ...

2. Архитектура Семантического Веба

Что нужно сделать, чтобы можно было осуществить вышеприведенные сценарии будущего? Будем отталкиваться от того, что было разработано при создании обычного Веба: протокол передачи данных, язык разметки страниц и браузер.

Таким образом, для реализации нужно следующее:

1. Разработать язык огромной выразительной силы, на котором можно описать все знания человека и который был бы понятен компьютерам.
2. Перевести все сайты на этот язык.
3. Написать программы, работающие со знаниями на этом языке (обработка запросов, логический вывод).

Первые два пункта представляются невыполнимыми.

Рассмотрим более тонкое решение, предложенное Тимом Бернерсом-Ли.

07.11.2016

2.1. Общая архитектура

В отличие от подхода, использующего искусственный интеллект для обучения компьютеров поведению людей, его идея заключается в поэтапной и распределенной разработке языка, способного выражать информацию в понятной машине форме. Т.о. цель Семантического Веба — создание языка, на котором можно будет описать как данные, так и правила рассуждений об этих данных, так что правила вывода, существующие в какой-либо системе представления знаний, можно будет экспортировать в Веб. Чтобы определить язык, необходимо задать его синтаксис и семантику.

Синтаксис — набор правил построения фраз языка, позволяющий определить корректные предложения в этом языке. Основным инструментом синтаксиса является наличие правил проверки, позволяющих судить о том, удовлетворяет ли текст синтаксису или нет.

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка. Примером языка с синтаксисом, но без семантики, является XML.

Примером семантики без синтаксиса — человеческая речь.

Поэтому программам так трудно разобратся, где что.

Тим Бернерс-Ли предложил отдельно разрабатывать синтаксис и семантику языка описания всех знаний человечества:

- RDF (Resource Description Framework) — язык, отвечающий за синтаксис документов Семантического Веба. В нем широко используются ссылки на онтологии для определения смысла слов.
- OWL (Ontology Web Language) — язык описания онтологий.

Онтология — описание классов объектов, их свойств и взаимоотношений для какой-то предметной области (домена).

Т.о. план Тима Бернерса-Ли требует последовательно разработать:

- 1.** Синтаксис для представления знаний, использующий ссылки на онтологии (RDF).
- 2.** Язык описания онтологий (OWL).
- 3.** Язык описания веб-сервисов (WSDL, OWL-S).

В настоящее время уже существует множество автоматизированных веб-сервисов безо всякой семантики. Но у других программ, таких как агенты, нет никакого способа разыскать в сети подобную программу, выполняющую ту или иную функцию.

Этот процесс, называемый обнаружением сервисов, станет возможным лишь после появления единого языка, который может описывать сервисы, чтобы агенты могли «понимать», что может делать данный сервис и каким образом им пользоваться.

Сервисы и агенты могут рекламировать выполняемые ими функции, например, занося подобные описания в справочники, подобные «Желтым Страницам».

4. Инструменты чтения и разработки документов Семантического Веба (Jena, Haystack, Protege).
Главный минус концепции Семантического Веба — сложность внедрения. Формат RDF был разработан людьми с академическим образованием и изначально не был рассчитан на применение рядовыми пользователями Интернета. Даже многим веб-мастерам и программистам бывает сложно освоить RDF и OWL.

Несмотря на это, Тим Бернерс-Ли утверждает, что в будущем никаких специальных знаний для создания страниц не потребуется.

Язык запросов SPARQL

5. Язык запросов к знаниям, записанным в RDF - SPARQL. SPARQL — новый язык запросов для быстрого доступа к данным RDF. Технология SPARQL позволяет извлекать данные из распределенных источников и может служить в качестве средства интеграции разнородной информации. Используя обычный протокол и язык SPARQL, приложения могут анализировать RDF-описания ресурсов и получать из сети нужную информацию.

Язык запросов SPARQL

В спецификации **SPARQL** не накладываются ограничения на формат данных. Благодаря этому становится возможным взаимодействие между ресурсами различного типа.

Тимоти Бернерс-Ли, автор идеи World Wide Web и руководитель консорциума W3C, характеризует значимость **SPARQL** следующим образом:

"Пытаться использовать семантическую сеть без **SPARQL** - это все равно, что работать с реляционной базой данных без языка структурированных запросов SQL".

Иными словами **SPARQL** превращает доступ к данным в некое подобие веб-сервиса.

SPARQL Query Language for RDF (SPARQL)

Язык запросов SPARQL для RDF.

SPARQL – проект нового стандарта платформы XML. Он определяет язык запросов информационных ресурсов в среде Веб в терминах их описания средствами стандарта RDF.

Создание этого языка завершит деятельность консорциума W3C по формированию многоуровневой информационной архитектуры семантического Веб. В сочетании со спецификациями RDF язык SPARQL будет определять полнофункциональную модель данных семантического уровня представления информационных ресурсов Веб.

Имеющиеся в настоящее время средства семантического описания информационных ресурсов Веб (стандарты RDF и OWL) не обладают какими-либо операционными возможностями, адекватными этому уровню представления информационных ресурсов Веб. Стандарт *SPARQL* восполняет этот пробел, позволяя извлекать данные по запросу из RDF-спецификации, а также конструировать новые компоненты RDF-спецификации.

Разработчики стандарта *SPARQL* рассматривают этот язык также как средство интеграции информационных ресурсов из нескольких неоднородных источников. Наряду с разработкой собственно языка *SPARQL* в консорциуме W3C ведутся работы по созданию стандарта протокола *SPARQL* для RDF и стандарта, определяющего XML-формат представления результатов обработки *SPARQL*-запросов.

Ключевым аспектом технологий Семантического Веба является их многоуровневая архитектура, которая показана ниже на рисунке.

Верхние уровни зависят от нижних, а они, в свою очередь, могут использоваться независимо.

В самом низу находится Unicode — стандарт кодирования символов. Она позволяет представить знаки практически всех письменных языков.

Рядом с ним находится URI, гарантирующий то, что каждое понятие, используемое в документе — это не просто слово, а нечто, привязанное к единому определению, которое каждый желающий может найти в Сети.

Иметь подобный механизм в самом низу необходимо для того, чтобы предоставить каждому человеку универсальный способ описания ресурсов.

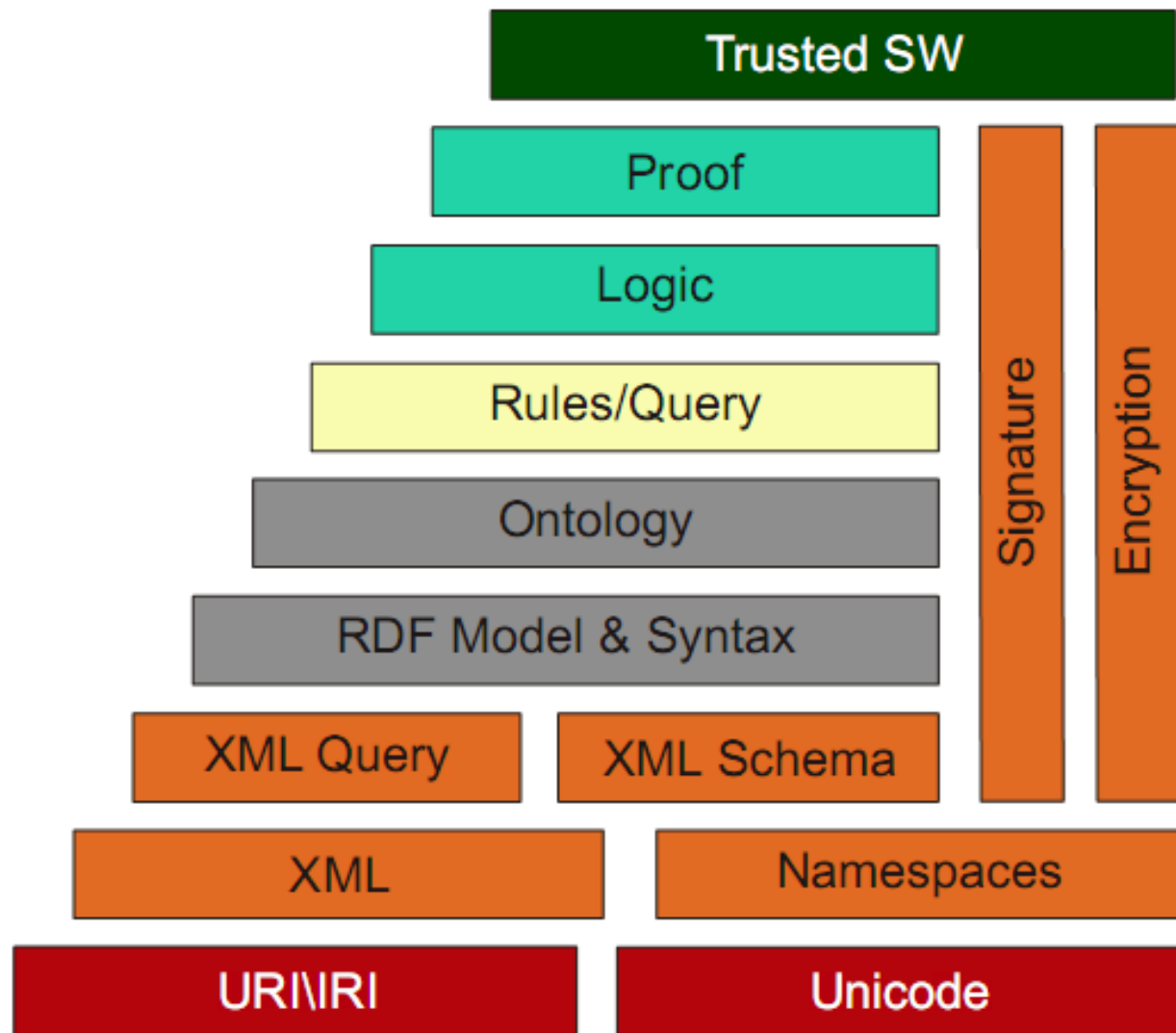


Рис. 2. Пирог уровней Семантического Веба.
(Тим Бернерс-Ли, 2000)

XML используется в качестве синтаксиса документов. Совместно с пространством имен может использоваться для выражения информации и обмена ею между программами.

XML Schema используется для выражения набора правил, которым должен удовлетворять XML-документ, чтобы быть признанным корректным.

Протокол *HTTP* и язык *HTML* лежат где-то в области этого уровня, но воспринимаемые человеком документы и их передача в значительной степени отличаются от материала, ориентированного преимущественно на машины. Поэтому на этой диаграмме они не показаны.

Пока пропустим *RDF* и *OWL*. О них поговорим дальше. Логический вывод (*Logic*) используется для обеспечения связности и корректности информации и для получения новых данных.

Доказательства (*Proof*) отслеживают и объясняют шаги логического вывода. Заслуживающий доверия Семантический Веб (*Trusted SW*) — средства, выполняющие аутентификацию, проверку достоверности информации, надежности сервисов и агентов.

2.2. RDF: синтаксис Семантического Веба

Язык *XML* позволяет пользователям создавать документы произвольной структуры. Но сам язык ничего не говорит о том, что означает эта структура. Смысл выражается посредством языка *RDF*, который кодирует его с помощью деревьев глубины три (Notation 3). Здесь каждое дерево состоит из **субъекта** (подлежащее), **предиката (свойства)** (сказуемое) и **объекта** (дополнение).

Объект можно назвать функцией свойства от субъекта. Например, утверждение «Небо голубого цвета» в RDF-терминологии можно представить следующим образом: субъект — «небо», свойство — «иметь цвет», объект — «голубой». Для идентификации субъектов, свойств и объектов используются [Uniform Resource Identifiers \(URI\)](#).

В начале любого **RDF** документа идет список ссылок на онтологии.

Таким образом, каждая вершина может быть задана строкой или ссылкой на объект из какой-либо онтологии.

Вершины могут иметь дополнительные уточняющие квалификаторы.

Частным случаем формата **RDF** является формат **RSS**.

RSS — семейство XML-форматов, предназначенных для описания лент новостей, анонсов статей, изменений в блогах и т.п.

Информация из различных источников, представленная в формате RSS, может быть собрана, обработана и представлена пользователю в удобном для него виде специальными программами-агрегаторами.

Под **RSS** может пониматься:

Rich Site Summary (RSS 0.9х) — обогащённая сводка сайта;

RDF Site Summary (RSS 0.9 и 1.0) — сводка сайта с применением инфраструктуры описания ресурсов;

Really Simple Syndication (RSS 2.x) — очень простое приобретение информации.

Обычно с помощью **RSS 2.0** даётся краткое описание новой информации, появившейся на сайте, и ссылка на её полную версию.

Интернет-ресурс в формате RSS называется **RSS-каналом**, **RSS-лентой** или **RSS-фидом**.

Пример документа в формате RDF

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
xmlns:edu="http://www.example.org/">
  <rdf:Description rdf:about="http://www.princeton.edu">
    <geo:lat>40.35</geo:lat>
    <geo:long>-74.66</geo:long>
    <edu:hasDept
      rdf:resource="http://www.cs.princeton.edu"
      dc:title="Department of Computer Science"/>
  </rdf:Description>
</rdf:RDF>
```

Пример документа в формате RDF

Верхний тэг означает, что все его содержимое — *RDF-описание*. Его семантическую часть страницы, понимает компьютер.

Атрибут *xmlns* является ссылкой на пространство имен. После двоеточия следует название, а в кавычках *URI*, *rdf* — базовое пространство имен, а три остальные — дополнительные.

Первое из них, *dc*, будет рассмотрено в дальнейшем, второе, *geo*, отвечает географической онтологии, а третье — образовательной.

Тэг *rdf:Description* определяет тройное предложение, его атрибут *rdf:about* — **субъект**. Далее следуют три **свойства**: «широта», «долгота» и «иметь факультет». Объекты в первых двух случаях заданы не ссылкой, а строкой. В третьем случае у тэга *edu:hasDept* есть атрибут *rdf:resource*, определяющий **объект**. Далее, используя пространство имен Dublin Core, задается название объекта.

OWL

W3C Язык Web-онтологий OWL 2 (**OWL**) является Semantic Web языком. Он предназначен для представления знаний об отдельных лицах, группах лиц и отношений между людьми.

OWL представляет собой вычислительный логико-ориентированный язык. Знание, выраженное в OWL может быть обосновано с помощью компьютерных программ либо для проверки согласованности этих знаний, либо для того, чтобы сделать неявное знание явным. **OWL** документы, известные как онтологии, могут быть опубликованы в World Wide Web и могут ссылаться или на них могут быть ссылки из других OWL онтологий.

OWL является частью семантического стека Web технологий консорциума W3C, которая включает RDF и SPARQL.

OWL 2/1

Язык OWL 2 при отсутствии каких бы то ни было ограничений называют OWL 2 Full. Подмножество OWL 2 DL предназначен для пользователей, которым нужна максимальная выразительность при сохранении полноты вычислений (все логические заключения, подразумеваемые той или иной онтологией, будут гарантированно вычислимыми) и разрешаемости (все вычисления завершатся за определенное время). OWL DL так назван из-за его соответствия дескрипционной логике — дисциплине, в которой разработаны логики, составляющие формальную основу OWL.

OWL 2/2

- **OWL 2 Full** предназначен для пользователей, которым нужна максимальная выразительность и синтаксическая свобода [RDF](#) без гарантий вычисления. Например, в OWL Full класс может рассматриваться одновременно как собрание индивидов и как один индивид в своём собственном значении. OWL Full позволяет строить такие онтологии, которые расширяют состав предопределённого (RDF или OWL) словаря. Маловероятно, что какое-либо программное обеспечение будет в состоянии осуществлять полную поддержку каждой особенности OWL Full.
- В первой версии языка также присутствовало подмножество **OWL Lite**, призванное ограничить выразительность языка и повысить скорость алгоритмов. В новой версии стандарта OWL Lite отсутствует.

2.3. OWL: язык описания онтологий

Основные компоненты OWL включают **классы, свойства и индивидуальные элементы**.

Класс — это концепция в онтологии. Классы являются основными блоками OWL и обычно образуют таксономическую иерархию (т.е. систему подкласс-надкласс).

OWL поддерживает шесть основных способов описания классов. Самый простой — класс с именем (*named*). Другие типы: классы пересечений (*intersection*), объединений (*union*), дополнений (*complement*), ограничений (*restrictions*) и классы перечислений (*enumerated*).

2.3. OWL: язык описания онтологий

Индивидуальные элементы — это элементы классов. В RDF они будут объектами и субъектами.

Мир классов и индивидов был бы совершенно неинтересным, если бы мы могли только определять *таксономии*.

Свойства позволяют нам утверждать общие факты о членах классов и особые факты об индивидах. Они включают две основные категории:

свойства-объекты, которые связывают индивидуальные элементы между собой и **свойства-значения** (datatype properties), которые связывают индивидуальные элементы со значениями типов данных.

Для определения типов данных OWL использует схему XML.

3. Проекты Семантического Веба

3.1. Дублинское ядро (Dublin Core)

Оно появилось раньше RDF. Теперь это просто аннотации (метаданные) к любым объектам, записанным на RDF с помощью онтологии DC. Цель DC — установить единый формат метаданных для облегчения поиска по автору, названию, году выпуска и т. д.

Множество метаданных Дублинского ядра (Dublin Core Metadata Element Set (DCMES)) состоит из 15 элементов: ***Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights.***

Dublin Core Metadata Element Set "simple Dublin Core" <http://purl.org/dc/elements/1.1/>.

Contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type.

Квалифицированный Dublin Core включал три дополнительных элемента (*аудитория, происхождение и правообладатель*), а также группу уточнений элементов (также называемых *квалификаторами*), которые могли бы уточнить семантику элементов таким образом, чтобы быть полезными в обнаружение ресурсов.

DC

С 2012 года оба (простой и квалифицированный DC) были включены в состав DCMI метаданных как единый набор терминов с использованием Resource Description Framework (RDF). Полный набор элементов находится в пространстве имен <http://purl.org/dc/terms/>. Поскольку определение терминов часто содержит области определения и значений, которые могут не быть совместимы с до-RDF определениями, используемыми для исходных 15 элементов Dublin Core, сохранилось и старое пространство имен.

2.3. OWL: язык описания онтологий

3.2. Проект Друг друга (*Friend of a Friend, FOAF*)

Этот проект позволяет описывать отношение знакомства с помощью RDF. Любой его участник может идентифицировать себя уникальным образом с помощью URI (например, `mailto-`адресом электронной почты, адресом блога, и т. п.). Он может создать свой профиль, используя predetermined для FOAF отношения на языке RDF, и перечислить идентификаторы людей, которых этот участник знает. Это описание может обрабатываться автоматически. На его основе можно строить сети доверия, анализировать структуру социальных групп, и т. д. Иными словами проект FOAF — онтология характеристик личностей и человеческих взаимоотношений. Цель — снабдить домашние страницы и профили машинно-понимаемыми описаниями, объединив все социальные сети в одну глобальную базу.

2.3. OWL: язык описания онтологий

3.3. *Haystack*

В проекте *Haystack* разрабатывается индивидуальная система управления информацией. Два подпроекта имеют отношения к Семантическому Вебу.

Piggy-Bank (Re:Search) сохраняет историю навигации в RDF формате, позволяя проводить «умный» поиск по материалам уже просмотренных страниц.

Haystack может использоваться как семантический браузер по документам, снабженным RDF-описаниями, а также производить обработку таких страниц..

2.3. OWL: язык описания онтологий

3.4. Другие проекты

- Поисковая система SHOE
(<http://www.cs.umd.edu/projects/plus/SHOE/search/>):
поиск в Семантическом Вебе.
- Jena (<http://jena.sourceforge.net>): среда разработки приложений для Семантического Веба, включает исполнитель SPARQL-запросов.
- Simile (<http://simile.mit.edu>): Семантический Веб для электронных библиотек.
- Protege (<http://protege.stanford.edu>): редактор онтологий из Стэнфорда.

Выводы

Семантический Веб — снабжение Интернет страниц описаниями, которые понятны компьютерам.

Описания пишутся на языке RDF со ссылками на онтологии, построенные с помощью языка OWL.

Отдельные подпроекты Семантического Веба имеют самостоятельное значение: FOAF, DC, RSS.

В полную силу Семантический Веб будет реализован тогда, когда люди создадут множество программ, которые, знакомясь с содержимым Сети из различных источников, обрабатывают полученную информацию и обмениваются результатами с другими программами.

Выводы

Эффективность таких программных агентов будет расти экспоненциально по мере увеличения количества доступного машинно-воспринимаемого веб-контента и автоматизированных сервисов (включая других агентов).

Выводы

Общее определение целей будущего веба выглядит очень заманчиво и амбициозно. Семантический веб - это не технология, а идеология. Для Бернерса-Ли Семантический веб - это следующая ступень эволюции Интернета.

Нынешний веб нацелен исключительно на предоставление информации людям и может эффективно читаться только людьми. "Вместо этого Семантический веб разрабатывает языки для выражения информации в форме, доступной для машинной обработки". То есть Семантический веб - это веб не для людей, а для машин, это комплекс технологий, позволяющих машинам лучше разбираться в Интернете, что, в свою очередь, сделает веб более полезным для людей.

Источники

[1] Тим Бернерс-Ли, Джеймс Хендлер и Ора Лассила.
Семантический Веб

http://ezolin.pisem.net/logic/semantic_web_rus.html

[2] Joshua Tauberer. What Is RDF?

<http://www.xml.com/pub/a/2001/01/24/rdf.html>

[3] Рекомендация W3C. Перевод Дмитрия Щербины.
OWL, язык веб-онтологий. Руководство

http://sherdim.rsu.ru/pts/semantic_web/REC-owl-guide-20040210_ru.html

[4] А.В. Манцивода. Система метаописаний Dublin Core

<http://teacode.com/concept/eor/dc.html>

[5] Страница курса

<http://logic.pdmi.ras.ru/yura/internet.html08ia.pdf>

Заключение

**Семантический Веб позволит машинам
ПОНИМАТЬ семантику документов и данных,
но не человеческую речь или его сочинения.**

**Т.Бернерс-Ли,
Дж.Хендлер,
О.Лассила
Семантический Веб, 2001**

Лекция окончена !

Благодарю за внимание !