

Семантический Веб
Лекция N 8 курса
“Алгоритмы для Интернета”

Юрий Лифшиц

ПОМИ РАН - СПбГУ ИТМО

Осень 2006

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

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

- 1 История и мотивация
 - Сценарии будущего
 - Хронология

- 1 История и мотивация
 - Сценарии будущего
 - Хронология
- 2 Архитектура семантического Веба
 - Общая архитектура
 - RDF: синтаксис семантического Веба
 - OWL: язык описания онтологий

- 1 История и мотивация
 - Сценарии будущего
 - Хронология
- 2 Архитектура семантического Веба
 - Общая архитектура
 - RDF: синтаксис семантического Веба
 - OWL: язык описания онтологий
- 3 Проекты семантического Веба

Часть I

Как мы представляем себе всемирную сеть будущего?

История усилий по разработке сети нового поколения —
Семантического Веба

В чем разница между понятиями **Интернет** и **Веб (WWW)**?

В чем разница между понятиями **Интернет** и **Веб (WWW)**?

Интернет: сеть **компьютеров**, объединенных каналами и использующих протоколы (TCP/IP) для связи. Изобретатели: Vint Cerf и Robert Kahn

Веб: сеть **сайтов**, использующих гиперссылки для переходов от страницы к странице. Изобретатель: Tim Berners-Lee



Человек — своему веб-агенту:

- Закажи для меня эту книгу в ближайшей библиотеке

Человек — своему веб-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

Человек — своему веб-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
- Скажите мне, какое вино мне нужно купить к каждому из блюд в этом меню. И, кстати, я не люблю Сотерн

Человек — своему веб-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
- Скажите мне, какое вино мне нужно купить к каждому из блюд в этом меню. И, кстати, я не люблю Сотерн
- Микроволновка, сходи на сайт производителя продукта и загрузи оптимальные параметры подогрева

- Семантический поиск
- Вопросо-ответные системы
- Агенты в семантическом Вебе
- Объединение знаний (интеграция баз данных)
- Всепроникающие вычисления (ubiquitous computing)

Хронология

- **1994:** Создание W3C. Они разработали стандарты: HTML, URL, XML, HTTP, PNG, SVG, CSS
- **1998:** Тим Бернерс-Ли публикует план семантического Веба (Semantic Web Road map)
- **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)

Хронология

- **1994:** Создание W3C. Они разработали стандарты: HTML, URL, XML, HTTP, PNG, SVG, CSS
- **1998:** Тим Бернерс-Ли публикует план семантического Веба (Semantic Web Road map)
- **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)

Участвуют: HP, Mozilla, IBM, MIT, Stanford, ...

Часть II

Из каких частей состоит инфраструктура семантического Веба?

Как устроен синтаксис его страниц?

Как определяется семантика его страниц?

Неформальный план

Что нам нужно сделать, чтобы можно было осуществить вышеприведенные сценарии будущего?

Неформальный план

Что нам нужно сделать, чтобы можно было осуществить вышеприведенные сценарии будущего?

Наивный план:

- 1 Разработать МЕГА-язык, на котором можно описать все знания человечества и который был бы понятен компьютерам
- 2 Под угрозой смерти заставить всех веб-мастеров снабжать сайты переводом на этот язык
- 3 Написать программы, работающие со знаниями на этом языке (обработка запросов, логический вывод)

Неформальный план

Что нам нужно сделать, чтобы можно было осуществить вышеприведенные сценарии будущего?

Наивный план:

- 1 Разработать МЕГА-язык, на котором можно описать все знания человечества и который был бы понятен компьютерам
- 2 Под угрозой смерти заставить всех веб-мастеров снабжать сайты переводом на этот язык
- 3 Написать программы, работающие со знаниями на этом языке (обработка запросов, логический вывод)

Тим Бернерс-Ли предложил более тонкое решение для первого шага

Что такое синтаксис и что такое семантика?

Что такое синтаксис и что такое семантика?

Синтаксис — набор правил построения фраз языка, позволяющий определить, осмысленные предложения в этом языке

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка

Что такое синтаксис и что такое семантика?

Синтаксис — набор правил построения фраз языка, позволяющий определить, осмысленные предложения в этом языке

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка

Синтаксис без семантики:

Что такое синтаксис и что такое семантика?

Синтаксис — набор правил построения фраз языка, позволяющий определить, осмысленные предложения в этом языке

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка

Синтаксис без семантики: XML

Что такое синтаксис и что такое семантика?

Синтаксис — набор правил построения фраз языка, позволяющий определить, осмысленные предложения в этом языке

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка

Синтаксис без семантики: XML

Семантика без синтаксиса:

Что такое синтаксис и что такое семантика?

Синтаксис — набор правил построения фраз языка, позволяющий определить, осмысленные предложения в этом языке

Семантика — система правил истолкования отдельных языковых конструкций. Семантика определяет смысловое значение предложений языка

Синтаксис без семантики: XML

Семантика без синтаксиса: человеческая речь

RDF и OWL

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

RDF (Resource Description Framework) — синтаксис документов семантического Веба. Использует ссылки на **ОНТОЛОГИИ**

OWL (Ontology Web Language) — язык описания онтологий

RDF и OWL

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

RDF (Resource Description Framework) — синтаксис документов семантического Веба. Использует ссылки на **онтологии**

OWL (Ontology Web Language) — язык описания онтологий

Онтология — описание классов объектов, их свойств и взаимоотношений для какой-то предметной области (например, магазин игрушек)

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)

Нужно последовательно разработать:

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

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)
- 3 Язык описания веб-сервисов (начато: WSDL, OWL-S)
- 4 Инструменты чтения/разработки документов семантического Веба (начато: Jena, Haystack, Protege)

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)
- 3 Язык описания веб-сервисов (начато: WSDL, OWL-S)
- 4 Инструменты чтения/разработки документов семантического Веба (начато: Jena, Haystack, Protege)
- 5 Язык запросов к знаниям, записанным в RDF (начато: SPARQL)

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)
- 3 Язык описания веб-сервисов (начато: WSDL, OWL-S)
- 4 Инструменты чтения/разработки документов семантического Веба (начато: Jena, Haystack, Protege)
- 5 Язык запросов к знаниям, записанным в RDF (начато: SPARQL)
- 6 Логический вывод знаний (не сделано)

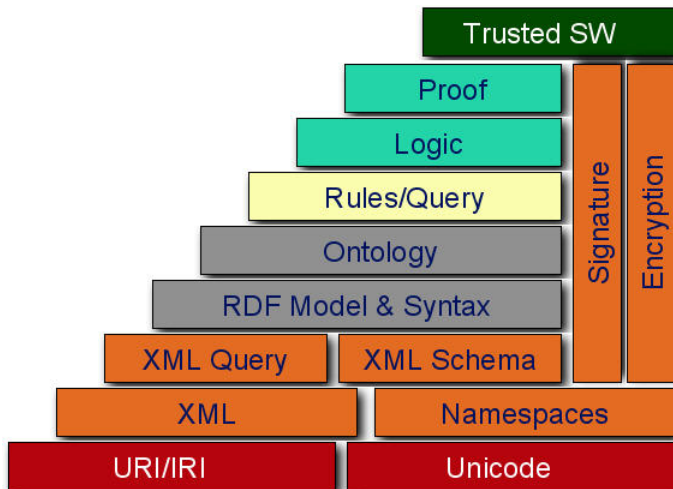
Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)
- 3 Язык описания веб-сервисов (начато: WSDL, OWL-S)
- 4 Инструменты чтения/разработки документов семантического Веба (начато: Jena, Haystack, Protege)
- 5 Язык запросов к знаниям, записанным в RDF (начато: SPARQL)
- 6 Логический вывод знаний (не сделано)
- 7 Семантическая поисковая система (начато: SHOE)

Нужно последовательно разработать:

- 1 Синтаксис для представления знаний, использующий ссылки на онтологии (сделано: RDF)
- 2 Язык описания онтологий (сделано: OWL)
- 3 Язык описания веб-сервисов (начато: WSDL, OWL-S)
- 4 Инструменты чтения/разработки документов семантического Веба (начато: Jena, Haystack, Protege)
- 5 Язык запросов к знаниям, записанным в RDF (начато: SPARQL)
- 6 Логический вывод знаний (не сделано)
- 7 Семантическая поисковая система (начато: SHOE)
- 8 Агенты семантического Веба (не сделано)

Пирог Тима Бернерса-Ли



RDF: синтаксис семантического Веба

- RDF является уточнением XML
- В начале RDF документа идет список ссылок на онтологии (namespaces)
- Основная часть RDF-документа состоит из деревьев глубины 3
- Верхняя вершина — субъект, средний уровень — свойства, нижний уровень — объект
- Каждая вершина может быть задана строкой или ссылкой на объект из одного из namespaces
- Вершины могут также иметь уточняющие квалификаторы

RDF: синтаксис семантического Веба

- RDF является уточнением XML
- В начале RDF документа идет список ссылок на онтологии (namespaces)
- Основная часть RDF-документа состоит из деревьев глубины 3
- Верхняя вершина — субъект, средний уровень — свойства, нижний уровень — объект
- Каждая вершина может быть задана строкой или ссылкой на объект из одного из namespaces
- Вершины могут также иметь уточняющие квалификаторы

RSS (изначально RDF Site Summary): возник как подмножество RDF

Пример RDF документа

```
<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>
```

- OWL использует синтаксис XML
- Есть команды для задания дерева классов
- Команды для задания индивидов принадлежности классам
- Система описания свойств: область определения, область значений
- Характеристики свойств: симметричность, транзитивность, функциональность
- Команды для склеивания (эквивалентности) классов

Часть III

Что уже сделано для реализации семантического Веба?

Где на все это можно посмотреть?

Дублинское ядро (Dublin Core) — появился раньше RDF, но теперь это просто аннотации (метаданные) к любым объектам, записанные на RDF с помощью онтологии DC

Цель DC — установить единый формат метаданных, для облегчения поиска по автору, названию, году выпуска, и т.д.

Дублинское ядро (Dublin Core) — появился раньше RDF, но теперь это просто аннотации (метаданные) к любым объектам, записанные на RDF с помощью онтологии DC

Цель DC — установить единый формат метаданных, для облегчения поиска по автору, названию, году выпуска, и т.д.

Где находится Дублин?

```
<?xml version="1.0"?> <metadata
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/">
  <dc:title>
    Automated Theorem Proving
  </dc:title>
  <dc:creator>
    Mantsivoda Andrei
  </dc:creator>
  <dc:subject xsi:type=dcterms:UDC>
    681.3
  </dc:subject>
  <dc:date>
    2004-04-04
  </dc:date>
  <dc:type>
    Article
  </dc:type>
  <dc:identifier>
    http://andrei.baikal.ru/atp2004
  </dc:identifier>
</metadata>
```

Друг друга (Friend of a Friend, FOAF) — онтология характеристик личностей и человеческих взаимоотношений

Цель — снабдить домашние страницы и профили машинно-понимаемыми описаниями, объединив все социальные сети в одну глобальную базу.

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-описанием, и производить обработку (например сортировку товаров по цене) таких страниц

[Поисковая система SHOE](#): поиск в семантическом Вебе.

[Jena](#): среда разработки приложений для семантического Веба, включает исполнитель SPARQL-запросов

[Simile](#): семантический Веб для электронных библиотек

[Protege](#): редактор онтологий из Стэнфорда

Сегодня мы узнали:

- Семантический Веб — снабжение интернет-страниц описаниями, которые понятны компьютерам

Сегодня мы узнали:

- Семантический Веб — снабжение интернет-страниц описаниями, которые понятны компьютерам
- Описания пишутся на языке RDF со ссылками на онтологии, построенные на языке OWL

Сегодня мы узнали:

- Семантический Веб — снабжение интернет-страниц описаниями, которые понятны компьютерам
- Описания пишутся на языке RDF со ссылками на онтологии, построенные на языке OWL
- Отдельные подпроекты семантического Веба имеют самостоятельное значение: FOAF, DC, RSS

Сегодня мы узнали:

- Семантический Веб — снабжение интернет-страниц описаниями, которые понятны компьютерам
- Описания пишутся на языке RDF со ссылками на онтологии, построенные на языке OWL
- Отдельные подпроекты семантического Веба имеют самостоятельное значение: FOAF, DC, RSS


Сегодня мы узнали:

- Семантический Веб — снабжение интернет-страниц описаниями, которые понятны компьютерам
- Описания пишутся на языке RDF со ссылками на онтологии, построенные на языке OWL
- Отдельные подпроекты семантического Веба имеют самостоятельное значение: FOAF, DC, RSS


Вопросы?


Страница курса <http://logic.pdmi.ras.ru/~yura/internet.html>

Использованные материалы:

 Тим Бернерс-Ли, Джеймс Хендлер и Ора Лассила
Семантический Веб
http://ezolin.pisem.net/logic/semantic_web_rus.html

 Joshua Tauberer
What Is RDF?
<http://www.xml.com/pub/a/2001/01/24/rdf.html>

 Рекомендация W3C, перевод Дмитрия Щербины
OWL, язык веб-онтологий. Руководство
http://sherdim.rsu.ru/pts/semantic_web/REC-owl-guide-20040210_ru.html

 А.В. Манцивода
Система метаописаний Dublin Core
<http://teacode.com/concept/eor/dc.html>