What is Semantic Web?
What is the Semantic Web?

- “An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”
  - Sir Tim Berners-Lee et al., Scientific American, 2001: tinyurl.com/i59p
- “…allowing the Web to reach its full potential…” with far-reaching consequences
- “The next generation of the Web”
Semantic Web

- Tim Berners-Lee has a vision of a Semantic Web which
  - has machine-understandable **semantics** of information, and
  - millions of small specialized **reasoning** services that provide support in automated task achievement based on the accessible information
The Semantic Web in essence

- The word “semantic” stands for “the meaning of”:
  - The Beatles were a popular band from Liverpool; Lennon was a member of the Beatles; "Hey Jude" was recorded by the Beatles

- The Semantic Web is a Web that is able to describe things in a way that computers can process
Metadata and Semantics

Types of Metadata and Annotations

**Ontology**
(Example: Anatomy, Diagnostics, ...)

**Semantic Metadata**
(Example ontology-driven metadata:
Region: Upper Abdomen
Organ: Liver
Pathological_Structure: Abscess, Abscess located in Liver)

**Structural Metadata**
(document structure: DTDs, XSL
clustering and similarity processing: concept extraction)

**Syntactic Metadata**
(language, format, document length, creation date, source,
audio bit rate, encryption, affiliation, date last reviewed, authorization, ...)

**Data**
(Structured, semi-structured and unstructured)
Ontology

- The semantic Web is essentially based on ontologies
  - ontologies are *formal* and *consensual* specifications of conceptualizations…
  - providing a *shared and common* understanding of a domain that can be communicated across people and application systems
A very simple Ontology

Ontologies describe concepts and their Relations.
Semantic Web - Language tower

Tim Berners-Lee
Keynote Speech in 2005
What is Semantic Web for?

- Integrating - trying to solve the problem of data and service integration
- Searching - Providing better communication between human and computers by adding machine-processable semantics to data.
Semantic Integration

- Top-Down approach: Building up different domain ontologies for better data integration and communication within the domain:
  - PapiNet.org: Vocabulary for Paper Industry
  - BPMI.org: Vocabulary for exchanging Business Process Models
  - XML-HR: Vocabularies for human resources (HR)
  - DMTF: Distributed Management Task Force: Vocabularies for managing enterprises
Semantic Integration

- CRM
- Marketing
- Sales

Semantics
Syntax
Software
Hardware

Plumbing (networking: tcp/ip, corba, rmi, http, soap)
## Semantic Differences: Example

### Marketing

<table>
<thead>
<tr>
<th>Person</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P#</strong></td>
<td>76798</td>
<td></td>
</tr>
<tr>
<td><strong>Name</strong></td>
<td>de Bruijn</td>
<td></td>
</tr>
<tr>
<td><strong>FName</strong></td>
<td>Jos</td>
<td></td>
</tr>
<tr>
<td><strong>DName</strong></td>
<td>Jos de Bruijn</td>
<td></td>
</tr>
<tr>
<td><strong>BDate</strong></td>
<td>1979-06-23</td>
<td></td>
</tr>
<tr>
<td><strong>LSale</strong></td>
<td>2001-04-07</td>
<td></td>
</tr>
</tbody>
</table>

### Sales

<table>
<thead>
<tr>
<th>Customer</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CustNr</strong></td>
<td>43526</td>
<td></td>
</tr>
<tr>
<td><strong>Name</strong></td>
<td>Jos Debruijn</td>
<td></td>
</tr>
<tr>
<td><strong>Surname</strong></td>
<td>Debruijn</td>
<td></td>
</tr>
<tr>
<td><strong>Initials</strong></td>
<td>J</td>
<td></td>
</tr>
<tr>
<td><strong>BDate</strong></td>
<td>1979-06-23</td>
<td></td>
</tr>
</tbody>
</table>

### Syntax Semantics

1. distinct  equivalent
2. equivalent  distinct
3. equivalent  equivalent
4. distinct  distinct

[Diagram showing the comparison between marketing and sales records]
Information Integration Patterns (1): Ad Hoc Integration

\[ O(n^2) \]
Information Integration Patterns (2): Global Integration

O(n)

Global Model

CRM

App4

App5

Marketing

Sales

Not silver bullet!
Semantic Searching
Semantic Searching

Text Search Results

Semantic Search Augmentation
Semantic Web: Past
The current (syntactic / structural) Web
Was the Web meant to be more?
How to realize Tim’s vision

● Another chance for “Artificial Intelligence (AI)”?
  ● Knowledge Representation (representing semantics)
  ● Logic Programming (reasoning semantics)

● Decisions for:
  ● Background logic for semantic web language (RDF, OWL)
    ● Description Logic
    ● DAPAR + EU = DAML+OIL (in 2001)
AI Influence

- Too much AI
  - Ontologies are too heavy
    - Too many axioms, complicated rules, concepts and relationships
  - Things are too formal
    - Too many formal logic, logic reasoning,
    - Knowledge base, expert system
Ontology or Oncology?

- Struggle to form a community
  - Lay the foundation for the birth of the Semantic Web community
  - Now it is continued as KnowledgeWeb (http://knowledgeweb.semanticweb.org/semanticportal/sewView/frames.html)
Slowly we found our stand

- Stand in the scientific community
  - Own international conference (ISWC, ESWC, ASWC)
  - Own journal (JoDS)
  - A research field -- topics in many other major conferences
  - Education

- Chances in Industry
  - Semantic Technology (http://www.semantic-conference.com/)
  - Europe
Semantic Web: Now
Social Web – Web 2.0

- The term Web 2.0 was made popular by Tim O’Reilly:
  - http://en.wikipedia.org/wiki/Web_2.0
    - “Web 2.0 … has … come to refer to what some people describe as a second phase of architecture and application development for the World Wide Web.”
- The Web where “ordinary” users can meet, collaborate, and share using social software applications on the Web (tagged content, social bookmarking, AJAX, etc.)
- Popular examples include:
  - Bebo, del.icio.us, digg, Flickr, Google Maps, Skype, Technorati, orkut, 43 Things, Wikipedia…
Social Networks
A move from the Web to a “social Web”

“On the Internet, nobody knows you’re a dog.” (A dog, sitting at a computer terminal, talking to another dog.)

ID: 22230, Published in The New Yorker July 5, 1993

“I had my own blog for a while, but I decided to go back to just pointless, incessant barking.” (One dog talking to another.)

ID: 121304, Published in The New Yorker September 12, 2005
When did Web 2.0 appear?
Features / principles of Web 2.0

1. The Web as platform
2. Harnessing collective intelligence
3. Data is the next “Intel Inside”
4. Rich user experiences
Web 2.0 meme cloud
W3C SWEO Linking Open Data Project

- Project aims to
  - Publish existing open license datasets as linked data on the web
  - Interlink things between different data sources
  - Develop clients and applications that consume linked data from the web
Bubbles in May 2007

Over 500M RDF triples
Around 120K RDF links between data sources
Bubbles in April 2008

>2B RDF triples
Around 3M RDF links
Organization participating in the LOD community

- **Academic**
  - MIT, Univ Southampton, DERI, Open Univ, Univ London, Univ Hannover, Penn State Univ, Univ Leipzig, Univ Karlsruhe, Joanneum (AT), Free Univ Berlin, Cyc, SouthEast Univ (CN), …

- **Commercial**
  - BBC, OpenLink, Talis, Zitgist, Garlik, Mondeca, Renault, Boad Interactive
What LOD can bring?

- It will lift current document web up to a data web
- LOD browsers can let you navigate between different data sources by following RDF links.
- It can drill down to the lower granularity of the information
  - allowing you for more fine search on the web
  - making the question-answer search on the Web possible
  - meshing up different data through RDF links
  - Making the built-on-top application easier
DBTune (http://dbtune.org)
DBTune
Industry pick-up

Semantic Arts
Celtx
empolis
Cambridge Semantics
Talis
Semantic Insights
Digital Harbor
Cycorp
Smartlogic
Ontotext
Aduna
Semantic Universe
Industry pick-up

- Semantic Technology Conferences – The major industrial conference in semantic web area.
  - Attendance include major IT giants (Google, Yahoo, IBM, Oracle, Intel, Vulcan)
- Rader Networks Company raised $18M to implement semantics
Semantic Web: Future
Metaweb ≡ social semantic information spaces
1+1>2

- Semantic forums
- Semantic blogs
- Semantic wikis
- Semantic social nets
- Semantic desktop

Semantic Web + social software
The path to Web 3.0 (the Semantic Web)

- The Semantic Web effort is mainly towards producing standards and recommendations that will **interlink data and applications**
- The Web 2.0 is about **providing user applications**
- Not mutually exclusive:
  - [http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html](http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html)
  - With a little effort, many Web 2.0 applications can and do use Semantic Web technologies to great benefit
Document Web vs. Data Web

- **Document Web**
  - Glued by hyperlinks
  - Data are HTML pages
  - Query result is HTML pages, which cannot be further processed
  - Data are just interlinked, but not integrated
  - Data access through different APIs

- **Data Web**
  - Glued by RDF links
  - Data are RDF triples
  - Query result is RDF triples which can be easily further processed (e.g., web services)
  - Data are interlinked and integrated, and links are typed
  - Data access through a single and standardized access mechanism (maybe it will be called LOD API?)
Web 3.0

Social Web + Semantic Web → Next generation Web

Document Web → Data Web → Service Web
Will Google 2.0 be Semantic?

- Google could be superseded, says web inventor:
  - TimesOnline:
    - http://technology.timesonline.co.uk/tol/news/ttech_and_web/article3532832.ece

- Google 2.0 embraces Semantic Web
  - Government Computer News:
    - http://www.gcn.com/online/vol1_no1/44290-1.html#
Thanks

- Contact
- Ying Ding
- LI029
- (812) 855 5388
- dingying@indiana.edu