Semantic Web

Lecture I – 12.10.2009
Introduction
Dieter Fensel and Ioan Toma
## Where are we?

<table>
<thead>
<tr>
<th>#</th>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.10.2009</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>12.10.2009</td>
<td><strong>Semantic Web architecture</strong></td>
</tr>
<tr>
<td>3</td>
<td>09.11.2009</td>
<td>RDF and RDFs</td>
</tr>
<tr>
<td>4</td>
<td>09.11.2009</td>
<td>Web of hypertext (RDFa, Microformats) and Web of data</td>
</tr>
<tr>
<td>5</td>
<td>23.11.2009</td>
<td>Semantic annotations</td>
</tr>
<tr>
<td>6</td>
<td>23.11.2009</td>
<td>Repositories and SPARQL</td>
</tr>
<tr>
<td>7</td>
<td>07.12.2009</td>
<td>OWL</td>
</tr>
<tr>
<td>8</td>
<td>07.12.2009</td>
<td>RIF</td>
</tr>
<tr>
<td>9</td>
<td>11.01.2010</td>
<td>Web-scale reasoning</td>
</tr>
<tr>
<td>10</td>
<td>11.01.2010</td>
<td>Social Semantic Web</td>
</tr>
<tr>
<td>11</td>
<td>11.01.2010</td>
<td>Ontologies and the Semantic Web</td>
</tr>
<tr>
<td>12</td>
<td>25.01.2010</td>
<td>SWS</td>
</tr>
<tr>
<td>13</td>
<td>25.01.2010</td>
<td>Tools</td>
</tr>
<tr>
<td>14</td>
<td>25.01.2010</td>
<td>Applications</td>
</tr>
<tr>
<td>15</td>
<td>TBD</td>
<td>Exam</td>
</tr>
</tbody>
</table>
Course Organization

• The lecturers are:
  Dieter Fensel (dieter.fensel@sti2.at)
  Ioan Toma (ioan.toma@sti2.at)

• The tutors are:
  Srdjan Komazec (srdjan.komazec@sti2.at)

• Lectures and Tutorials every two weeks. (Check lecture and tutorial page for dates)
Course material

- Web site:
  
  http://www.sti-innsbruck.at/teaching/courses/ws200910/details/?title=semantic-web
  
  - Slides available online before each lecture

- Mailing list:
  
  https://lists.sti2.at/mailman/listinfo/sw2009
Examination

- Exam grade:

<table>
<thead>
<tr>
<th>score</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>75-100</td>
<td>1</td>
</tr>
<tr>
<td>65-74.9</td>
<td>2</td>
</tr>
<tr>
<td>55-64.9</td>
<td>3</td>
</tr>
<tr>
<td>45-54.9</td>
<td>4</td>
</tr>
<tr>
<td>0-44.9</td>
<td>5</td>
</tr>
</tbody>
</table>

- You can get up to 25 points if you perform very well in the tutorials. These points count for the final exam grade.
Agenda

1. Motivation
2. Technical solution
   1. Introduction
   2. Semantic Web – architecture and languages
   3. Semantic Web - data
   4. Semantic Web - processes
3. Illustration
4. Extensions
5. Summary
6. References
MOTIVATION
Today Web

- The current Web represents information using
  - natural language (English, German, Italian, …)
  - graphics, multimedia, page layout
- Humans can process this easily
  - can deduce facts from partial information
  - can create mental associations
  - are used to various sensory information

However they can do this only if there is a small amount of information that is available to them
• Tasks often require to combine data on the Web
  – hotel and travel information may come from different sites
  – searches in different digital libraries
• Again, humans combine this information easily
  – even if different terminologies are used!
• Problems with existing services and applications
However...

- Machines are ignorant!
  - partial information is unusable
  - difficult to make sense from, e.g., an image
  - drawing analogies automatically is difficult
  - difficult to combine information automatically
  - …
How to improve current Web?

- Increasing automatic linking among data
- Increasing recall and precision in search
- Increasing automation in data integration
- Increasing automation in the service life cycle

- Adding semantics to data and services is the solution!
Approaches to semantics

• Statistics + Linguistics
  – mathematical algorithms
  – extract info from text
  – no understanding of the content

• Semantic Web
  – smarter applications
  – share & link data – Web of Data
  – more expressive queries
The KIM Platform

• A statistics and linguistic platform

• It offers:

  services and infrastructure for:

  – (semi-) automatic semantic annotation and
  – ontology population
  – semantic indexing and retrieval of content
  – query and navigation over the formal knowledge

• Based on Information Extraction technology
XYZ announced profits in Q3, planning to build a $120M plant in Bulgaria, and more and more and more and more and more and more text.
The KIM Platform includes:

- **Ontologies** (PROTON + KIMSO + KIMLO) and KIM World KB
- **KIM Server** – with a set of APIs for remote access and integration
- **Front-ends**: Web-UI and plug-in for Internet Explorer.
KIM is based on the following open-source platforms:

- **GATE** – the most popular NLP and IE platform in the world, developed at the University of Sheffield. Ontotext is its biggest co-developer. [www.gate.ac.uk](http://www.gate.ac.uk) and [www.ontotext.com/gate](http://www.ontotext.com/gate)

- **OWLIM** – OWL repository, compliant with Sesame RDF database from Aduna B.V. [www.ontotext.com/owlim](http://www.ontotext.com/owlim)

- **Lucene** – an open-source IR engine by Apache. [jakarta.apache.org/lucene](http://jakarta.apache.org/lucene)
TECHNICAL SOLUTION
INTRODUCTION TO SEMANTIC WEB
The Vision

More than 2 billion users
more than 50 billion pages

Static

WWW
URI, HTML, HTTP
Serious problems in

- information finding,
- information extracting,
- information representing,
- information interpreting and
- and information maintaining.

Static

WWW
URI, HTML, HTTP

Semantic Web
RDF, RDF(S), OWL
What is the Semantic Web?

• “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

• “…allowing the Web to reach its full potential…” with far-reaching consequences

• “The next generation of the Web”
What is the Semantic Web?

• The next generation of the WWW

• Information has machine-processable and machine-understandable semantics

• Not a separate Web but an augmentation of the current one

• The backbone of Semantic Web are ontologies
Ontology definition

- Unambiguous terminology definitions
- Conceptual model of a domain (ontological theory)
- Formal, explicit specification of a shared conceptualization
  - Machine-readability with computational semantics
  - Commonly accepted understanding

Gruber, “Toward principles for the design of ontologies used or knowledge sharing?” , Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6, 1995
• “An ontology is an explicit specification of a conceptualization”
• Ontologies are the modeling foundations to Semantic Web
  – They provide the well-defined meaning for information
An ontology is:
• A conceptualization
  – An ontology is a model of the most relevant concepts of a phenomenon from the real world
• Explicit
  – The model explicitly states the type of the concepts, the relationships between them and the constraints on their use
• Formal
  – The ontology has to be machine readable (the use of the natural language is excluded)
• Shared
  – The knowledge contained in the ontology is consensual, i.e. it has been accepted by a group of people.

Ontology example

Concept
conceptual entity of the domain

Property
attribute describing a concept

Relation
relationship between concepts or properties

Axiom
coherency description between Concepts / Properties / Relations via logical expressions

holds(Professor, Lecture) => Lecture.topic = Professor.researchField
Types of ontologies

describe very general concepts like space, time, event, which are independent of a particular problem or domain.

describe the vocabulary related to a generic domain by specializing the concepts introduced in the top-level ontology.

describe the vocabulary related to a generic task or activity by specializing the top-level ontologies.

the most specific ontologies. Concepts in application ontologies often correspond to roles played by domain entities while performing a certain activity.

[Guarino, 98] Formal Ontology in Information Systems
http://www.loa-cnr.it/Papers/FOIS98.pdf
Types of ontologies - examples

• Top Level/Upper ontologies:
  – Cyc, DOLCE, SUMO, DublinCore

• Domain ontologies:
  – medicine, telecom ontologies, etc.

• Task ontologies:
  – diagnosing, selling, scheduling ontologies

• Application ontologies:
  – Cell Cycle Ontology (CCO)
The Semantic Web is about...

• Web Data Annotation
  – connecting (syntactic) Web objects, like text chunks, images, ... to their semantic notion (e.g., this image is about Innsbruck, Dieter Fensel is a professor)

• Data Linking on the Web (Web of Data)
  – global networking of knowledge through URI, RDF, and SPARQL (e.g., connecting my calendar with my rss feeds, my pictures, ...)

• Data Integration over the Web
  – seamless integration of data based on different conceptual models (e.g., integrating data coming from my two favorite book sellers)
Annotating Hydrogen

With an atomic mass of 1.00794 g/mole (u), hydrogen is the lightest element. It is also the most abundant, constituting roughly 75% of the universe's elemental mass. In the Universe, NASA Website URL: http://www.ontoprise.de/, hydrogen is mainly composed of hydrogen in its elemental state. Elemental hydrogen is relatively rare on Earth, and is industrially produced from fossil fuels, coal, and natural gas. After which most pure hydrogen is used "captively" meaning locally at the production site, with the largest markets being the production of ammonia production (mostly for the fertilizer market). However, hydrogen can easily be produced from water using the process of electrolysis.

The most common naturally occurring oxidant of hydrogen is a single electron and no hydrogen. In metal compounds, it can take an either a positive charge (becoming a cation) or a negative charge (becoming an anion) known as a hydride. Hydrogen can form compounds with most elements and is present in metals and most organic compounds. It plays a particular role in acid-base chemistry, in which the atom has several between suitable molecules. As the only neutral atom for acidity, study of the energetics and bonding of the hydronium ion is also significant. The properties of hydrogen include:

| Property | Value
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>1.00794 g/mole (u)</td>
</tr>
<tr>
<td>Density</td>
<td>0.07143 kg/m^3</td>
</tr>
<tr>
<td>Melting point</td>
<td>-252.87°C</td>
</tr>
<tr>
<td>Boiling point</td>
<td>-252.7°C</td>
</tr>
</tbody>
</table>

Discovery of Hz

Hydrogen gas, H₂, was first artificially produced and formally described by Henry Cavendish (also known as the "father" of the science of electricity) via the mixing of metals with strong acids. He was unaware that the flammable gas produced by this chemical reaction was a new chemical element. In 1671, Robert Boyle rediscovered and described the reaction between iron filings and dilute acids, which results in the production of hydrogen gas. Iron could be used in the reaction, but hydrogen was not recognized as a distinct substance. By identifying the gas from a metal-acid reaction as "inflammable air", and further finding that the gas produces water when burned, Cavendish had stumbled on hydrogen when experimenting with acids and mercury. Although he wrongly assumed that hydrogen was a liberated component of the mercury rather than the "air", he was still able to accurately describe several key properties of hydrogen. He is usually given credit for its discovery as an element. In 1783, Antoine Lavoisier gave the element the name of hydrogen when he (with Laplace) reproduced Cavendish's finding that water is produced when hydrogen is burned. Lavoisier's name for the gas won out.

Role in history of quantum theory

Because of its relatively simple atomic structure, consisting only of a proton and an electron, the hydrogen atom,
Data Linking on the Web

http://linkeddata.org/
Data Linking on the Web

• **Linked Open Data statistics:**
  – data sets: 108
  – total number of triples: 4,712,896,432
  – total number of links between data sets: 142,375,048
Data linking on the Web principles

- Use URIs as names for things
  - anything, not just documents
  - you are not your homepage
  - information resources and non-information resources
- Use HTTP URIs
  - globally unique names, distributed ownership
  - allows people to look up those names
- Provide useful information in RDF
  - when someone looks up a URI
- Include RDF links to other URIs
  - to enable discovery of related information
Data Integration over the Web

http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial

Same URI = Same resource
SEMANTIC WEB – ARCHITECTURE AND LANGUAGES
• Things are denoted by URIs
• Use them to denote things
• Serve useful information at them
• Dereference them
Semantic Web Architecture

- Give important concepts URIs
- Each URI identifies one concept
- Share these symbols between many languages
- Support URI lookup
Semantic Web - Data

Topics covered in the course:
- SparQL
- Logic framework
- OWL
- Rules
- DLP bit of OWL/Rul
- RDF Schema
- RDF Core
- XML
- Namespaces
- URI
- Unicode

Trust
- Signature
- Encryption
• Uniform Resource Identifier (URI) is the dual of URL on Semantic Web
  – it’s purpose is to identify resources

• eXtensible Markup Language (XML) is a markup language used to structure information
  – fundament of data representation on the Semantic Web
  – tags do not convey semantic information
• Resource Description Framework (RDF) is the dual of HTML in the Semantic Web
  – simple way to describe resources on the Web
  – sort of simple ontology language (RDF-S)
  – based on triples (subject; predicate; object)
  – serialization is XML based

• Ontology Web Language (OWL) a layered language based on DL
  – more complex ontology language
  – overcome some RDF(S) limitations
• **SPARQL**
  - Query language for RDF triples
  - A protocol for querying RDF data over the Web

• **Rule languages (e.g. SWRL)**
  - Extend basic predicates in ontology languages with proprietary predicates
  - Based on different logics
    - Description Logic
    - Logic Programming
SEMANTIC WEB - DATA
- URIs are used to identify resources, not just things that exist on the Web, e.g. Sir Tim Berners-Lee

- RDF is used to make statements about resources in the form of triples
  \(<\text{entity, property, value}>\)

- With RDFS, resources can belong to classes (my Mercedes belongs to the class of cars) and classes can be subclasses or superclasses of other classes (vehicles are a superclass of cars, cabriolets are a subclass of cars)
KIM Browser Plugin
Web content is annotated using ontologies
Content can be searched and browsed intelligently
Derereferencable URI

Disco Hyperdata Browser
navigating the Semantic Web as an unbound set of data sources
Faceted DBLP uses the keywords provided in metadata annotations to automatically create light-weight topic categorization.
43% of businesses resort to manual processes and/or new software when integrating information for reporting.
Semantic Web Data

Semantic Broker

Existing legacy systems “wrapped” in semantic technologies

Declarative definition of Business Rules

Billing
Sales
CRM
Order Processing
Marketing
Inventory

Reasoning enables inference of new facts from existing data sources

Based on lightweight, open standards from W3C
SEMANTIC WEB - PROCESSES
Processes

• The Web is moving from static data to dynamic functionality
  – Web services: a piece of software available over the Internet, using standardized XML messaging systems over the SOAP protocol
  – Mashups: The compounding of two or more pieces of web functionality to create powerful web applications
Semantic Web - Processes

- Web services and mashups are limited by their syntactic nature
- As the amount of services on the Web increases it will be harder to find Web services in order to use them in mashups
- The current amount of human effort required to build applications is not sustainable at a Web scale
The addition of semantics to form Semantic Web Services and Semantically Enabled Service-oriented Architectures can enable the automation of many of these currently human intensive tasks – Service Discovery, Adaptation, Ranking, Mediation, Invocation

Frameworks:
- **OWL-S**: WS Description Ontology (Profile, Service Model, Grounding)
- **WSMO**: Ontologies, Goals, Web Services, Mediators
- **SWSF**: Process-based Description Model & Language for WS
- **SAWSDL (WSDL-S)**: Semantic annotation of WSDL descriptions
Semantic Web - Processes

Conceptual Model for SWS

Formal Language for WSMO

Ontology & Rule Language for the Semantic Web

Execution Environment for SWS

More about in Semantic Web Services lecture
Semantic Web uptake

- Major companies offer Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Microsoft, Dow Jones, ...
Others are using it (or consider using it) as part of their own operations: Novartis, Boeing, Pfizer, Telefónica, …
Some of the names of active participants in W3C SW related groups: ILOG, HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Pfizer, Sun, Eli Lilly, …
Example 1
Find the right experts at NASA

- Expertise locator for nearly 20,000 NASA civil servants using RDF integration techniques over 6 or 7 geographically distributed databases, data sources,

From Kendall Clark, Clark & Parsia, LLC
Example II
Vodafone live!

- Integrate various vendors’ product descriptions via RDF
  - ring tones, games, wallpapers
  - manage complexity of handsets, binary formats
- A portal is created to offer appropriate content
- Significant increase in content download after the introduction

From Kevin Smith, Vodafone Group R&D
More Examples

  - Cultural Heritage
  - Health Care
  - Life Sciences
  - eCommerce
  - B2B integration
  - eTourism
  - ...
Case study: BT Research and Venturing

Integration with Semantic Mediation

- The complexity of supply chains has increased, they involve many players of differing size and function.
- Support for “Operational Support Systems (OSS)” integration using semantic descriptions of system interfaces and messages.
- Internet Service Providers integrate their OSS-s with those of BT (via a gateway).
- Integration of heterogeneous OSS systems of partners.
- The approach reduces costs and time-to-market; ontologies allow for a reuse of services.

Courtesy of Alistair Duke, BT, (SWEO Use Case)
Cloud computing

- Grid Computing
  - solving large problems with parallel computing

- Utility Computing
  - Offering computing resources as a metered service

- Software as a service
  - Network-based subscription to applications

- Cloud Computing
  - Next generation internet computing
  - Next generation data centers
Including semantic technologies in Cloud Computing will enable:

- Flexible, dynamically scalable and virtualized data layer as part of the cloud
- Accurate search and acquire various data from the Internet,
Extending the mobile and sensors networks with Semantic technologies, Semantic Web will enable:

- **Interoperability at the level of sensors data and protocols**
- **More precise search for mobile capabilities and sensors with desired capability**

http://www.opengeospatial.org/projects/groups/sensorweb
SUMMARY
Summary

• Semantic Web is not a *replacement* of the current Web, it’s an *evolution* of it
• Semantic Web is about:
  – annotation of data on the Web
  – data linking on the Web
  – data Integration over the Web
• Semantic Web aims at *automating* tasks currently carried out by humans
• Semantic Web is becoming *real* (maybe not as we originally envisioned it, but it is)
References

- RDF Primer: http://www.w3.org/TR/rdf-primer/
- RDF Semantics: http://www.w3.org/TR/rdf-mt/
<table>
<thead>
<tr>
<th>#</th>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.10.2009</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>12.10.2009</td>
<td>Semantic Web architecture</td>
</tr>
<tr>
<td>3</td>
<td><strong>09.11.2009</strong></td>
<td>RDF and RDFs</td>
</tr>
<tr>
<td>4</td>
<td><strong>09.11.2009</strong></td>
<td>Web of hypertext (RDFa, Microformats) and Web of data</td>
</tr>
<tr>
<td>5</td>
<td>23.11.2009</td>
<td>Semantic annotations</td>
</tr>
<tr>
<td>6</td>
<td>23.11.2009</td>
<td>Repositories and SPARQL</td>
</tr>
<tr>
<td>7</td>
<td>07.12.2009</td>
<td>OWL</td>
</tr>
<tr>
<td>8</td>
<td>07.12.2009</td>
<td>RIF</td>
</tr>
<tr>
<td>9</td>
<td>11.01.2010</td>
<td>Web-scale reasoning</td>
</tr>
<tr>
<td>10</td>
<td>11.01.2010</td>
<td>Social Semantic Web</td>
</tr>
<tr>
<td>11</td>
<td>11.01.2010</td>
<td>Ontologies and the Semantic Web</td>
</tr>
<tr>
<td>12</td>
<td>25.01.2010</td>
<td>SWS</td>
</tr>
<tr>
<td>13</td>
<td>25.01.2010</td>
<td>Tools</td>
</tr>
<tr>
<td>14</td>
<td>25.01.2010</td>
<td>Applications</td>
</tr>
<tr>
<td>15</td>
<td>TBD</td>
<td>Exam</td>
</tr>
</tbody>
</table>
Questions?