Summer Academy 2009

Semantic Web

Prof. Dr. Steffen Staab
Dr. Maciej Janik
Organizational Issues

Contact:

Prof. Dr. Steffen Staab  staab@uni-koblenz.de
Dr. Maciej Janik       janik@uni-koblenz.de

Web site:
http://www.uni-koblenz-landau.de/koblenz/fb4/studying/summer-academy/courses2009/semantic_web
http://www.uni-koblenz-landau.de/koblenz/fb4/institute/IFI/AGStaab/Teaching/SS09/sw09/

Language:
English

Lectures:
Prof. Steffen Staab, Dr. Maciej Janik

Exercises:
Dr. Maciej Janik, Noam Bercovici
## Timetable

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>8:30 - 10:00</td>
<td>K208</td>
<td>K208</td>
<td>K208</td>
<td>F414</td>
</tr>
<tr>
<td>Exercise</td>
<td>10:15 - 12:00</td>
<td>F112</td>
<td>F113</td>
<td>F113</td>
<td>A024</td>
</tr>
<tr>
<td>Lecture</td>
<td>16:00 - 17:30</td>
<td>K207</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise</td>
<td>18:00 - 19:30</td>
<td>F112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td>8:30 - 10:00</td>
<td>D238</td>
<td>A308</td>
<td>A308</td>
<td>A308</td>
</tr>
<tr>
<td>Exercise</td>
<td>10:15 - 12:00</td>
<td>F112</td>
<td>F113</td>
<td>F112</td>
<td>A024</td>
</tr>
<tr>
<td>Lecture</td>
<td>14:00 - 15:30</td>
<td></td>
<td></td>
<td></td>
<td>D238</td>
</tr>
<tr>
<td>Exercise</td>
<td>16:00 - 17:30</td>
<td></td>
<td></td>
<td></td>
<td>F112</td>
</tr>
<tr>
<td>Lecture</td>
<td>8:30 - 10:00</td>
<td>A308</td>
<td>D238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise</td>
<td>10:15 - 12:00</td>
<td>F112</td>
<td>F113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXAM</td>
<td>12:15 - 13:45</td>
<td></td>
<td></td>
<td></td>
<td>D238</td>
</tr>
</tbody>
</table>
Semantic Web Books

- H. Stuckenschmidt
  **Ontologien**: Konzepte, Technologien und Anwendungen

- P. Hitzler, M. Krötzsch, S. Rudolph, Y. Sure
  **Semantic Web**
Lectures overview

- Semantic Web foundations
  - XML / RDF
  - RDFS / OWL
  - Ontology engineering
  - SPARQL, Graphs

- Applied Semantic Web
  - User Interfaces
  - Linked Open Data
  - Extraction (Wikipedia)
  - Semantic Search
  - Semantic Social Web
Semantic Web in a nutshell

Some history
Overview of technologies and applications
...a short history of the Web ...

- ~1989

- Researchers
  - Aalta
  - …
  - Zyström

- Aalta
  - Phone 789
  - Friend
  - Phone 981

- Zyström
World Wide Web

WWW :=

Hypertext &

Internet &

Social Phenomenon
1 Billion users later …

- Sir Tim Berners-Lee
...the Web history continues ...

~1995

Phone Directory CERN
Aalta 789
...Zyström 981

<HTML>
Aalta
Hobby ...
Phone 789

<HTTP>
Zyström
Friend
Phone 981
...but there is problem with the Web ...

~1995
…have a closer look …

That’s how computer “see” webpages
...short history of the Semantic Web ...

~1995
What exactly do we see here?

1. **Language for (meta-) data**

2. **Language for schema definition for (meta-) data**

3. **Exchange of data and schema via Internet**

4. **Many people and applications**
Semantic Web :=

Semantic Web Data & Ontologies & Internet & Social phenomenon

1. Language for (meta-) data
2. Language for schema definition for (meta-) data
3. Exchange of data and schema via Internet
4. Many people and applications
Agenda

- Foundations of Semantic Web
  - Linked data
  - Ontologies
  - Query languages

- Applications
  - eScience
  - Semantic Desktop
  - Tagster

- Semantic Web 2.0
  - RDF Mash-Ups
  - SemaPlorer
  - Networked RDF Graphs
  - How to do all this?
Semantic Web Foundations
Semantic Web Building Blocks

(Previously: Semantic Web Layer Cake)

RDF

RDF/XML Syntax Specification (Revised)
W3C Recommendation

RDF Vocabulary Description Language 1.0: RDF Schema
W3C Recommendation

RDF Primer
W3C Recommendation

Resource Description Framework (RDF): Concepts and Abstract Syntax
W3C Recommendation

RDF Semantics
W3C Recommendation

RDF Test Cases
W3C Recommendation
Semantic Web in a nut shell

Ontology

Person
- rdfs:subClass of Person
- rdfs:Domain
- rdfs:Range

Employee
- rdfs:subClass of Employee

PostDoc
- rdfs:subClass of PostDoc

Professor
- rdfs:subClass of Professor

cooperatesWith
- rdfs:Ranger

Steffen Staab
- rdfs:ID = "person_sst"
- swrc:name = "Steffen Staab"
- swrc:cooperatesWith rdf:resource = "http://www.uni-koblenz.de/~staab/#person_sst"

Siegfried Handschuh
- rdfs:ID = "person_sha"
- swrc:name = "Siegfried Handschuh"
- swrc:cooperatesWith rdf:resource = "http://www.uni-koblenz.de/~staab/#person_sst"

Web site

http://www.deri.ie/~sha

URL

http://www.uni-koblenz.de/~staab

Research:
- Semantic Web
- Knowledge Management
- Natural Language
Resource Description Framework – RDF

- URI – uniform resource identifiers
- XML – underlying syntax
- Linked data
- Semantics: axiomatization

Not supported:
- Identity,
- Cardinality,
- Negation, etc.
OWL – Web Ontology Language

OWL - Web Ontology Language Overview

OWL Web Ontology Language Guide

OWL Web Ontology Language Reference

OWL Web Ontology Language Semantics and Abstract Syntax

OWL Web Ontology Language Test Cases

OWL Web Ontology Language Use Cases and Requirements

OWL Web Ontology Language XML Presentation Syntax. W3C Note 11 June 2003
Taxonomy := segmentation, classification and order of elements in classification system
Thesaurus

- Terminology about specific field
- Taxonomy plus additional, fixed relationships (synonym, relatedTo, antonym)
- Widespread in libraries

Object
  - Person
  - Topic
  - Document
  - Student
  - Researcher
  - Semantics
  - PhD Student
  - Doctoral Student

• Synonym
• Similar
• Topics (nodes), relationships and their realization (in documents)
• ISO-Standards in SGML and XML
• Typical for navigation and visualization
• Application in book publishing (digital book index)
• Languages: F-Logic, OWL (W3C recommendation)
• Originated from the knowledge representation
Example: Ontologies in Biology

Tim Berners-Lee, ISWC November 2005,
http://www.w3.org/2005/Talks/1110-iswc-tbl/#(14)
Web Ontology Language (OWL)

- Description logic
  - Allows definition of concepts with names
  - Consists of concepts and roles (T-Box), e.g. Concept(Employee), Role(advises),
  - and individuals filling specific role (A-Box), e.g. Prof(Steffen)

- Equivalence and subsumption
  \( \text{Prof} \equiv \text{Professor}, \text{advises} \lor \text{responsibleFor} \)

- Algebraic properties of defined roles (inversion)
  \( \text{advises}^{-1} \equiv \text{isAdvisedBy} \)

- Complex definitions
  \( \text{CSProf} \equiv \text{Professor} \cup \text{ComputerScientist}, \)
  \( \text{HuBi} \equiv \text{MedicalScientist} \cup \neg \text{MedicalDoctor} \)
  \( \text{Prof} \lor \text{Lecturer} \lor \exists_{\geq 2} \text{teaches.Course} \lor \exists_{\leq 4} \text{teaches.Course} \)
Reasoning in OWL

OWL: Subset of DL
Semantics: model theory

T-Box
- Consistency (achievable)
  - Is there a model (interpretation) I for ontology O?
- Subsumption
  - Calculate Taxonomy
  - Does concept C subsumes concept D in all models (interpretation) I?

A-Box
- Extension of concepts and relationships
- Instantiation
  - Is i an instance of C?
- Filling roles
  - Find all pairs \((x,y)\in R\)

All problems reduce to consistency checking:
- e.g. \(C \lor D\) if \(\neg D \cup C\) inconsistent in O
SPARQL query language

SPARQL Query Language for RDF
W3C recommendation 15 January 2008
Example

- Data:
  <http://example.org/book/book1>
  <http://purl.org/dc/elements/1.1/title>
  "SPARQL Tutorial"

- Query:
  SELECT ?title
  WHERE
  {
  }

Query Result:

<table>
<thead>
<tr>
<th>title</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;SPARQL Tutorial&quot;</td>
</tr>
</tbody>
</table>
Constructing graphs

- **Data:**
  
  ```
  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
  _:a foaf:givenname "Alice" .
  _:a foaf:family_name "Hacker" .
  _:b foaf:firstname "Bob" .
  _:b foaf:surname "Hacker" .
  ```

- **Query:**
  
  ```
  PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> CONSTRUCT {
  ?x vcard:N _:v .
  _:v vcard:givenName ?gname .
  _:v vcard:familyName ?fname
  }
  WHERE {
  { ?x foaf:firstname ?gname }
  UNION
  { ?x foaf:givenname ?gname } .
  { ?x foaf:surname ?fname } .
  { ?x foaf:family_name ?fname } .
  }
  ```

- **Result:**
  
  ```
  @prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
  _:v1 vcard:N _:x .
  _:x vcard:givenName "Alice" .
  _:x vcard:familyName "Hacker" .
  _:v2 vcard:N _:z .
  _:z vcard:givenName "Bob" .
  _:z vcard:familyName "Hacker" .
  ```
Applications of Semantic Web
Semi-automatic generation of metadata

The system has found the value "Landhotel & Ferienpark Leonorenwald" for the Concept Hotel. Do you agree?
Used ontologies and common vocabularies:

- InChI identifier
- SMILES string
- CML (Chemical Markup)

Open Biomedical Ontologies:
- Gene Ontology
- Sequence Ontology
- Cell Ontology
Semantic Web @ Koblenz

- Ontologies & Semantic Web
- Multimedia
- Personal information management
- Information retrieval & mining
- Web Services
- Peer-to-Peer
- Web2.0
Standard file system

2344 entries in my main work folder C:/work

No management system like Subversion / BSCW / Wiki

The problem:
I need to find a presentation of the Semantic Desktop I talked about with Thomas for the next lecture
Instant Messaging und Email

“I need information about Semantic Desktop?”

“Here is a good paper …”

“.. it could be also interesting for you.

:B dns:plays xco:SenderRole
:A dns:plays xco:RecipientRole
:infObject3 dns:plays xco:TransferredFile
:x-cosim.pdf dns:realizes : infObject3
:A dns:plays xco:SenderRole
:C dns:plays xco:RecipientRole
:infObject3 dns:plays xco:Attachment
:infObject4 dns:plays xco:Emailbody
RDF Browser

Information Object: http://localhost:8080/openrdf/default#message-18049b9b:112ae55057:-7f8d
Realized by File: file:/home/franz/tmp/x-cosim.pdf
Sent by: address:A@kater.uni-koblenz.de
Sent to: address:C@kater.uni-koblenz.de
Sent at: 2007-05-21T15:13:29.288+02:00
Transferred as: http://isweb.uni-koblenz.de/Research/x-cosim/com-module.owl#Attachment
Part of Conversation: http://localhost:8080/openrdf/default#conversation-18049b9b:112ae55057:-7f92

Information Object: http://localhost:8080/openrdf/default#message-18049b9b:112ae55057:-7fa8
Realized by File: file:/home/franz/tmp/x-cosim.pdf
Sent by: address:A@kater.uni-koblenz.de
Sent to: address:C@kater.uni-koblenz.de
Transferred as: http://isweb.uni-koblenz.de/Research/x-cosim/com-module.owl#TransferedFile
Part of Conversation: http://localhost:8080/openrdf/default#conversation-18049b9b:112ae55057:-7f65
Semantic Desktop

K-Cap

Indexing Ontologies with Semantics-enhanced Keywords
Erkki Huovinen, Murat Cetin, Sergio De Maio, Paul Lewis, Nigel Shadbolt
A Methodology for Asynchronous Multi-User Editing of Semantic Web Ontologies
Julian Baidach, Allen Bierlen
A Framework for Evaluating Semantic Metadata
Yugang Li, Vincent Uher, Enrico Valletta
Ontologies & Semantic Web

- Personal information management
- Multimedia
- Information retrieval & mining
- Web Services
- Peer-to-Peer
- Web2.0
Tagster: local annotation with Tags

- browse personal data with common filebrowser view
- select new files + folders to tag
- type in tags to assign to selected files

Tagster window:
- Browse and Tag tabs
- Search in: music folder
- Files listed: Ane Brun, Audioslave, Bush, Disturbed, E la luna, Evanescence, Filter, Foo Fighters, K's Choice, Lacuna Coil, Linkin Park, Lunik, Mando Diao
- Tags: rock, alternative, metal, favorite

Dateiename: turbid "Filter" "Foo Fighters" "Linkin Park" "Metallica" "Millencolin" "P.O.D." "Placebo"
Dateityp: Alle Dateien
Tags: rock alternative metal favorite
Tagster: Peer-to-Peer Tagging

- browse all available tags (includes automatically extracted tags, e.g. path+file)
- search by tag
- browse displayed resources (files)
- select additional tags
- modify tags directly
Semantic Web 2.0

Interlinking applications and integrating knowledge from different sources
Mashing up means creating new content by combining and reusing the existing components and content.

Examples:
- Merging multiple news feeds into one
- Find “apartment in Mountain View” and present result in Google Maps
Mashups and Semantic Web

- **Mashup most popular model:**
  - *Hack-and-Hope*

- **Disadvantages:**
  - Screen-Scraping
  - No agreement on data model

- It may be different sometimes:
  - Google Web Service
  - Amazon Web Service

- **Semantic Web most popular model:**
  - *Crawl-Integrate-and-Reason*

- **Disadvantages:**
  - Data is *outdated*
  - Data is not declarative, but only has implicit semantics
  - Scalability problems
  - **Access rights:** not all is allowed to be copied / published
  - **Provenance** – data sources can be blurred
FOAFer

Please enter a FOAF-Resource: [http://www.uni-koblenz.de/~sschenk/foaf.rdf]

go

Name: Simon Schenk
Givenname: Simon
Family Name: Schenk
MBOX_SHA1SUM: 14c0da3c553cc8f7866a1

knows:
Steffen Staab
Thomas Franz
Carsten Saathoff

DBLP
Simon Schenk
List of publications from the DBLP Bibliography Server - FAQ
Coauthor Index - Ask others: ACM DL - ACM Guide - CiteSeer - CSB - Google

2006

2005

Coauthor Index
1 Thomas Franz
2 Olaf Gehrke
3 Carsten Saathoff
4 Sergei Sirov
5 Steffen Staab

ISWeb - Information Systems & Semantic Web
Steffen Staab staab@uni-koblenz.de

<isweb>
Declarative, dynamic semantic mashups can be very powerful.
Solutions with RDF Mashups

- **Mashup most popular model:**
  - *Hack-and-Hope*

- **Disadvantages:**
  - *Screen-Scraping*
  - No agreement on data model

- It may be different sometimes:
  - *Google Web Service*
  - *Amazon Web Service*

**Improvements:**
- Reuse of data instead of Screen-scraping
- More current views
- Data integration using graph definition
- Possible to evaluate data sources and clients
- Networked, dynamic Semantic Web

- **Semantic Web most popular model:**
  - *Crawl-Integrate-and-Reason*

- **Disadvantages:**
  - Data is outdated
  - Data is not declarative, but only has implicit semantics
  - Scalability problems
  - *Access rights*: not all is allowed to be copied / published
  - *Provenance* – data sources can be blurred

---

ISWeb - Information Systems & Semantic Web
Steffen Staab
staab@uni-koblenz.de
Semantic Web
54 of 62
Sema Plorer

Context: Sheffield

Search
- Sheffield
- Locations
  - Sheffield
  - Sheffield, Illinois
  - Sheffield station
  - Sheffield, Tasmania
  - Sheffield Manor

Persons
- Sally Sheffield
- Johnny Sheffield
- David Sheffield
- Bill Sheffield
- Jeremy Sheffield

Tags
- Sheffield

Wordnet
- Sheffield

Place Object
Type Event

Sheffield
Sheffield is a city and metrop...

Sights
- Sheffield Cathedral
- Orchard Square
- Fargate
- Kelham Island Museum
- Cathedral Church of St Marie, She

Locations
- Sheffield
- Attercliffe
Where does knowledge come from?

Based on application created for Rolls-Royse

Is the information reliable?
Report knowledge

Original metadata

```<isweb>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@baseURI http://www.problem.com/sampleDocument

:G1 {
  T5678 hasProblem overheating.
  overheating causedBy vibration.
}

:G2 {
  T5678 hasProblem burnedEdges.
  burnedEdges causesProblem wrongAirflow.
  overheating causesProblem vibration
}

:G3 {
  :G1 source <http://example/report01.doc> .
  :G1 agent Bob.
  :G1 extractor textAnalyzer.
  :G1 accuracy "0.9".
  :G1 timestamp "5/5/2007"
}

:G4 {
  :G2 source <http://example/image01.jpg> .
  :G2 agent Mary.
  :G2 extractor imageAnalyzer.
  :G2 accuracy "0.6" .
  :G2 timestamp "6/6/2006"
}
</isweb>
Knowledge extracted from picture

Original metadata

```turtle
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@baseURI http://www.problem.com/sampleDocument

:G1 {
  T5678 hasProblem overheating.
  overheating causedBy vibration.
}

:G2 {
  T5678 hasProblem burnedEdges.
  burnedEdges causesProblem wrongAirflow.
  overheating causesProblem vibration
}

:G3 {
  :G1 source <http://example/report01.doc> .
  :G1 agent Bob.
  :G1 extractor textAnalyzer.
  :G1 accuracy "0.9".
  :G1 timestamp "5/5/2007"
}

:G4 {
  :G2 source <http://example/image01.jpg> .
  :G2 agent Mary.
  :G2 extractor imageAnalyzer.
  :G2 accuracy "0.6".
  :G2 timestamp "6/6/2006"
}
```
CONSTRUCT ( ?Y causeOf ?Z )
WHERE
{
  {
    { ?X hasProblem ?Y . } AND
    { ?Y causesProblem ?Z . }
  }
  FILTER { ?X = "T5678" }
Algebra of SPARQL

:G1 {
  T5678 hasProblem overheating.
  overheating causedBy vibration.
}

S1="T5678"

Simultaneous execution

\[ \land \]
\[ \mathsf{min} \]

Result: problems are with overheating caused by vibrations.

Original information can be found in Report01.doc, the accuracy of result is 0.9
Conclusion
WWW Vs. Semantic Web

WWW :=

Hypertext &
Internet &
Social Phenomenon

Semantic Web :=

Semantic Web Language/Data &
Ontologies &
Internet &
Social Phenomenon

Without Social Phenomenon
= Intranet

New important paradigms, but not that great impact …

Without Social Phenomenon
= Semantic Data Integration