Semantic Web – State of the Art

Pascal Hitzler
Kno.e.sis Center
Wright State University, Dayton, OH
http://www.knoesis.org/pascal/
New Book

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies
Chapman & Hall/CRC, 2010

Grab a flyer!

http://www.semantic-web-book.org
Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

语义Web技术基础
Tsinghua University Press (清华大学出版社)，2012, to appear

Translators:
Yong Yu, Haofeng Wang, Guilin Qi (俞勇，王昊奋，漆桂林)

http://www.semantic-web-book.org
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics
• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them
• Some current work
The current (World Wide) Web

- Immensely successful.
- Huge amounts of data.
- Syntax standards for transfer of structured data.

BUT:

- Content/knowledge cannot be accessed by machines. Meaning (semantics) of transferred data is not accessible.
Examples

• Find that landmark article on data integration written by an Indian researcher in the 1990s.
  [If you manage this without knowing the answer, let me know how you did it.]

• Which car is called a “duck” in German?
  [This needs some intelligent integration of content from different websites plus background knowledge.]
Another example

“Identify congress members, who have voted “No” on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts.”

In principle, all the required knowledge is on the Web – most of it even in machine-readable form.

However, without automated processing and reasoning we cannot obtain a useful answer.
Very brief history of the Semantic Web

- 1990s: W3C metadata activity (lead to RDF(S))
- USA: DAML-Programme 2000-2005 approx. $90M.
- Many large scale EU projects since 2002 and ongoing. FP6/FP7
- Major IT companies and venture capital now investing.
Semantic Technologies in the US

• Funding available e.g. via
  – NIH
  – NSF
  – DoD, DoE, AFRL
  – IARPA, DARPA
  – ...

• Considerable industrial take-up
  – Annual Semantic Technology Conference in CA
    Taylored towards industry
  – Major IT players (Oracle, IBM, HP, ...) invest
  – Major government contractors (BBN, Lockheed, ...)
  – Venture capital (e.g. Vulcan, Inc.).
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics
• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them
• Some current work
Basic Idea of the Semantic Web

Ontology represents Schema knowledge e.g. every publication has an author

Data e.g. on Websites

DL Rules Krötzsch, Rudolph, Hitzler ECAI 2008
Basic Idea of the Semantic Web

e.g. every publication has an author

DL Rules
Krötzsch, Rudolph, Hitzler
ECAI 2008
Basic Idea of the Semantic Web

Ontology represents
Schema knowledge
\[\text{e.g. every publication has an author}\]

Data e.g. on Websites

DL Rules
Krötzsch, Rudolph, Hitzler
ECAI 2008
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics

• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them

• Some current work
What Is Semantic Web Semantics?

• Opinions Differ. Here’s my take.

• Semantic Web requires a shareable, declarative and *computable* semantics.

• I.e., the semantics must be a formal entity which is clearly defined and automatically computable.

• Ontology languages provide this by means of their formal semantics.

• Semantic Web Semantics is given by a relation – the *logical consequence* relation.

• Note: This is considerably more than saying that the semantics of an ontology is the set of its logical consequences!
In other words

We capture the meaning of information

not by specifying its meaning (which is impossible)
but by specifying

how information interacts with other information.

We describe the meaning indirectly through its effects.
Simple Logical Reasoning

If I ask for soccer team members, I also want to get the goalkeepers listed ...

If I ask for cities, I also want all capitals listed ...

inheritance reasoning
Less Simple Reasoning

What was again the name of that Russian researcher who worked on resolution-based calculi for EL?

Which car is called „duck“ in German?

What is "Käuzchen" in English?

Answering requires merging of knowledge from many websites and using background knowledge.
SNOMED CT

- SNOMED CT: commercial ontology, medical domain ca. 300,000 axioms

- InjuryOfFinger \( \rightarrow \) Injury \( \cup \) 9site.Finger_s
- InjuryOfHand \( \rightarrow \) Injury \( \cup \) 9site.Hand_s
- Finger_s \( \lor \) Hand_p
- Hand_p \( \lor \) Hand_s \( \cup \) 9part.Hand_E

- Reasoning has been used e.g. for
  - classification (computing the hidden taxonomy)
    e.g., InjuryOfFinger \( \lor \) InjuryOfHand
  - bug finding
So what happened?

- In 2004, two W3C Recommendations were completed:
  - RDF + RDF Schema with formal model-theoretic semantics
  - OWL with formal model-theoretic semantics

- OWL 2 update emerged 2009.
- RDF update is being discussed right now.
Ontology languages

- Of central importance for the realisation of Semantic Technologies are suitable representation languages.
- Meaning (semantics) provided via logic and deduction algorithms.
- Scalability is a challenge.

Expressivity

OWL DL  OWL EL  RDF

Scalability

Language standards recommended by W3C
Ontology Example

Declaration of classes

rdfs:Class

x:Professor

x:Employee

x:Tutor

x:PhD-Student

x:Student

subClass

instantiation

Declaration of properties

x:email

rdfs:Literal

rdfs:domain

rdfs:range

x:advises

x:Student

x:supervises

x:Employee

rdfs:subPropertyOf

rdfs:domain

rdfs:range

x:responsible_for

rdfs:subPropertyOf

schema knowledge
PhDStudent v 9advisedBy.Professor

rules
responsible_for(y,x) \AE Professor(y) ! Employee(x)
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics

• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them

• Some current work
Basic Idea of the Semantic Web

Ontology represents schema knowledge

e.g. every publication has an author

Data e.g. on Websites

DL Rules
Krötzsch, Rudolph, Hitzler
ECAI 2008

fact knowledge
Currently it's looking like this

Data e.g. on Websites

facts only. (almost) no schema knowledge

DL Rules
Krötzsch, Rudolph, Hitzler
ECAI 2008
Mashups
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics
• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them
• Some current work
Example: GeoNames

<table>
<thead>
<tr>
<th>ID</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,518,403</td>
<td>P.PPL</td>
<td>populated place</td>
</tr>
<tr>
<td>48,483</td>
<td>P.PPLX</td>
<td>section of populated place</td>
</tr>
<tr>
<td>39,336</td>
<td>P.PPLL</td>
<td>populated locality</td>
</tr>
<tr>
<td>13,306</td>
<td>P.PPLQ</td>
<td>abandoned populated place</td>
</tr>
<tr>
<td>2,684</td>
<td>P.PPLA4</td>
<td>seat of a fourth-order administrative division</td>
</tr>
<tr>
<td>2,028</td>
<td>P.PPLA</td>
<td>seat of a first-order administrative division</td>
</tr>
<tr>
<td>1,847</td>
<td>P.PPLW</td>
<td>destroyed populated place</td>
</tr>
<tr>
<td>1,006</td>
<td>P.PPLF</td>
<td>farm village</td>
</tr>
<tr>
<td>930</td>
<td>P.PPLA3</td>
<td>seat of a third-order administrative division</td>
</tr>
<tr>
<td>695</td>
<td>P.PPLA2</td>
<td>seat of a second-order administrative division</td>
</tr>
<tr>
<td>253</td>
<td>P.PPLS</td>
<td>populated places</td>
</tr>
<tr>
<td>249</td>
<td>P.STLMT</td>
<td>israeli settlement</td>
</tr>
<tr>
<td>235</td>
<td>P.PPLC</td>
<td>capital of a political entity</td>
</tr>
<tr>
<td>57</td>
<td>P.</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>P.PPLR</td>
<td>religious populated place</td>
</tr>
<tr>
<td>6</td>
<td>P.PPLG</td>
<td>seat of government of a political entity</td>
</tr>
<tr>
<td>2,629,547</td>
<td>Total for P</td>
<td></td>
</tr>
</tbody>
</table>
Example: GovTrack

“Nancy Pelosi voted in favor of the Health Care Bill.”

Vote: 2009-887

Votes: 2009-887/+ vote:votedBy

Aye rdfs:label

people/P000197 name

Nancy Pelosi
dc:title

On Passage: H R 3962 Affordable Health Care for America Act
dc:title

H.R. 3962: Affordable Health Care for America Act
vote:hasAction

Bills:h3962

vote:vote
“Identify congress members, who have voted “No” on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts.”

In principle, all the knowledge is there:

- GovTrack
- GeoNames
- DBPedia
- US Census

But even with LoD we cannot answer this query.
Example querying LoD

“Identify congress members, who have voted “No” on pro-environmental legislation in the past four years, with high-pollution industry in their congressional districts.”

Some missing puzzle pieces:

• Where is the data?
  - GovTrack
  - GeoNames
  - US Census

requires intimate knowledge of the LoD data sets
Example querying LoD

“Identify congress members, who have voted “No” on pro \textit{environmental legislation} in the past four years, with \textit{high-pollution industry} in their congressional districts.”

Some missing puzzle pieces:

- Where is the data?
  (smart federation needed)

- Missing background (schema) knowledge.
  (enhancements of the LoD cloud)

- Crucial info still hidden in texts.
  (ontology learning from texts)

- Added reasoning capabilities (e.g., spatial).
  (new ontology language features)
Don’t get me wrong

Linked Open Data is great, useful, cool, and a very important step.

But we need to make use of the added value of formal semantics in order to advance towards the Semantic Web vision!
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics
• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them
• Some current work
To leverage LoD, we require **schema knowledge**

- **application-type driven** (reusable for same kind of application)
- **less messy than LoD** (as required by application)
- **overarching several LoD datasets** (as required by application)
Schema on top of the LoD cloud
Contents

• What is Semantic Web?
  – Limitations of the current World Wide Web
  – The basic Semantic Web idea
  – Semantic Web Semantics
• Semantic Data Web (state of the art)
  – its limitations
  – and how to overcome them
• Some current work
LOQuS – Querying Linked Open Data

Work in progress.

- Schema creation for
  - query federation
  - utilizing background knowledge
  - compilation of LOD knowledge into reason-able form

- Reasoning algorithm (on suitable language) for very efficient data-intensive reasoning

Linked Open Data

Traditional Web content

Schema
### Table 4. Results of various systems for LOD Schema Alignment. Legends: Prec=Precision, Rec=Recall, M=Music Ontology, B=BBC Program Ontology, F=FOAF Ontology, D=DBpedia Ontology, G=Geonames Ontology, S=SIOC Ontology, W=Semantic Web Conference Ontology, A=AKT Portal Ontology, err=System Error, NA=Not Available

<table>
<thead>
<tr>
<th>Test</th>
<th>Alignment API</th>
<th>OMViaUO</th>
<th>RiMoM</th>
<th>S-Match</th>
<th>AROMA</th>
<th>BLOOMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec</td>
<td>Rec</td>
<td>Prec</td>
<td>Rec</td>
<td>Prec</td>
<td>Rec</td>
</tr>
<tr>
<td>M,B</td>
<td>0.4</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>M,D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>F,D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>G,D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>S,F</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.2</td>
<td>0.52</td>
<td>0.11</td>
</tr>
<tr>
<td>W,A</td>
<td>0.12</td>
<td>0.05</td>
<td>0.16</td>
<td>0.03</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>W,D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.07</td>
<td>0.01</td>
<td>0.17</td>
<td>0.0</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Jain, Hitzler et al, ISWC2010
Table 1. Results on the oriented matching track. Results for RiMOM and AROMA have been taken from the OAEI 2009 website. Legends: Prec=Precision, A-API=Alignment API, OMV=OMViaUO, NaN=division by zero, likely due to empty alignment.

<table>
<thead>
<tr>
<th>Test</th>
<th>A-API</th>
<th>OMV</th>
<th>S-Match</th>
<th>AROMA</th>
<th>RiMoM</th>
<th>BLOOMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec</td>
<td>Rec</td>
<td>Prec</td>
<td>Rec</td>
<td>Prec</td>
<td>Rec</td>
</tr>
<tr>
<td>1XX</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.06</td>
<td>0.01</td>
<td>0.71</td>
</tr>
<tr>
<td>2XX</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>0.30</td>
</tr>
<tr>
<td>3XX</td>
<td>0.01</td>
<td>0.03</td>
<td>0.02</td>
<td>0.047</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.

2. **Construction of the BLOOMS forest** $T_C$ for each class name $C$, using information from Wikipedia.

3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.

4. **Post-processing** of the results with the help of the Alignment API and a reasoner.
**BLOOMS trees**

**Fig. 1.** BLOOMS trees for Jazz Festival with sense Jazz Festival and for Event with sense Event. To save space, some categories are not expanded to level 4.
1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.

2. **Construction of the BLOOMS forest** $T_C$ for each class name $C$, using information from Wikipedia.

3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.

4. **Post-processing** of the results with the help of the Alignment API and a reasoner.
1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.

2. **Construction of the BLOOMS forest** $T_C$ for each class name $C$, using information from Wikipedia.

3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.

4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

) **We’re currently evaluating the LOQuS querying approach while utilizing BLOOMS.**
Reasoning: useful scalable languages

- OWL 2: complexity > exponential
- ELP: complexity = polynomial [ISWC2008]
- OWL 2 EL and RL: complexity = polynomial
- hybrid ELP: data complexity = polynomial [ECAI2008]
Thanks!

Collaborators on the covered topics:

Kno.e.sis: Prateek Jain, Adila Alfa Krisnadhi, Frederick Maier, Raghava Mutharaju, Amit Sheth

Accenture: Kunal Verma, Peter Z. Yeh

Karlsruhe: Sebastian Rudolph

Oxford: Markus Krötzsch

Lisboa: Matthias Knorr, Jose J. Alferes

http://www.semantic-web-book.org
http://www.semantic-web-journal.net
References


References


References


• Shasha Huang, Qingguo Li, Pascal Hitzler, Reasoning with Inconsistencies in Hybrid MKNF Knowledge Bases. Logic Journal of the IGPL. To appear.

• Frederick Maier, Yue Ma, Pascal Hitzler, Paraconsistent OWL and Related Logics. Semantic Web journal. To appear.
References


• Matthias Knorr, Jose Julio Alferes, Pascal Hitzler, Local Closed-World Reasoning with Description Logics under the Well-founded Semantics. Artificial Intelligence 175(9-10), 2011, 1528-1554.


References


References


References


References
