Towards Reasoning Pragmatics: State of the Art and Vision for the Semantic Web

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

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies
Chapman & Hall/CRC, 2009

Grab a flyer!

http://www.semantic-web-book.org
Remember?

Tim Berners-Lee, James Hendler and Ora Lassila
The Semantic Web
Scientific American, May 17, 2001

• Talks explicitly about knowledge representation and logic as required ingredient.
“The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.”
Formal Semantics

• RDF as of 2001 had no formal semantics.

• What actually is semantics? What is formal semantics?
Syntax and Semantics

Syntax: character strings without meaning
Semantics: meaning of the character strings

IF cond(A,B) THEN display(_354)

Show pixel set „_354“ on screen if „A“ is of type „B“.

meaning, e.g., „in the world“
Semantics of Programming Languages

Syntax

FUNCTION f(n:natural):natural;
BEGIN
IF n=0 THEN f:=1
ELSE f:=n*f(n-1);
END;

Procedural Semantics

What happens at program execution

Formal Semantics

\[ f : n \mapsto n! \]

Intended Semantics

computing factorial

What happens at program execution
So ...

• Semantics tells us more about something than meets the eye.

• Semantics gives access to **implicit knowledge**.

• Semantics helps to focus on the implicit knowledge, and abstracts from concrete representations.
  – [there’s always more than one way to code something]
Semantics for the Semantic Web

- Need semantics for data (not for programs).
- How to define semantics? How to encode data?

  - Logic-based formalisms
  - Model-theoretic semantics
Semantics of Logic

Syntax

$\forall x (p(x) \lor q(x))$

Intended Semantics

All humans are mortal

Model-theoretic semantics

provability in a calculus

Proof-theoretic semantics

logical consequence

2
So what happened?

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

• The hype started a bit earlier, actually.
Very brief history of the Semantic Web

- Idea as old as 1989.
- 1990s: W3C metadata activity (lead to RDF(S))
- SciAm article: 2001

- USA: DAML-Programme 2000-2005 approx. 70M€.
- Many large scale EU projects since 2002 and ongoing. → FP6

- Now funding mostly application oriented (EU FP7, US NIH)
Basic Idea of the Semantic Web

Ontology represents schema knowledge and mediates implicit knowledge. For example, "every publication has an author."
Basic Idea of the Semantic Web

e.g. “every publication has an author”
Basic Idea of the Semantic Web

Ontology
- represents Schema knowledge
- mediates implicit knowledge
  e.g. "every publication has an author"

Data e.g. on Websites

DL Rules
Krötzsch, Rudolph, Hitzler
ECAI 2008
“The Semantic Web is about two things. It is about **common formats for integration and combination of data** drawn from diverse sources, where the original Web mainly concentrated on the interchange of documents. It is also about **language for recording how the data relates to real world objects.**”
The new buzzword: **Linked Data**

Five Aspects mentioned:

- **Linked Data:**
  “The Semantic Web is a Web of data”

- **Vocabularies:**
  OWL, SKOS – “enrich data with additional meaning”

- **Query:**
  “If the Semantic Web is viewed as a global database ...”

- **Inference:**
  “discovering new relationships”

- **Vertical Applications:**
  “innovation adoption through Semantic Web technology”
Linked Open Data

But where is the semantics?
Example: GeoNames

<table>
<thead>
<tr>
<th>Populated Place Features (city, village,...)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.PPL</td>
</tr>
<tr>
<td>P.PPLX</td>
</tr>
<tr>
<td>P.PPLL</td>
</tr>
<tr>
<td>P.PPLQ</td>
</tr>
<tr>
<td>P.PPLA4</td>
</tr>
<tr>
<td>P.PPLA</td>
</tr>
<tr>
<td>P.PPLW</td>
</tr>
<tr>
<td>P.PPLF</td>
</tr>
<tr>
<td>P.PPLA3</td>
</tr>
<tr>
<td>P.PPLA2</td>
</tr>
<tr>
<td>P.PPLS</td>
</tr>
<tr>
<td>P.STLM</td>
</tr>
<tr>
<td>P.PPLC</td>
</tr>
<tr>
<td>P.PPLR</td>
</tr>
<tr>
<td>P.PPLG</td>
</tr>
<tr>
<td>2,629,547</td>
</tr>
</tbody>
</table>
"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
“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? (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)
How to get there

- Schema ontologies
  - made for specific purposes (e.g., querying)
  - spanning several LoD datasets
  - incorporating schema knowledge hidden in the LoD datasets
  - including additional background knowledge needed for design purpose

- Added reasoning capabilities extending OWL as needed.
  - rules
  - extended datatypes
  - spatial and temporal reasoning etc.

- Making use of ontology lifecycle state-of-the-art tools
  - ontology evaluation
  - ontology learning from texts
  - ontology evolution etc.
From no semantics to low semantics

1. Take a no-semantics or low-semantics solution.
   E.g., naive LoD querying using SPARQL.

2. Identify where added value could be obtained by formal semantics.
   E.g., by using schema knowledge as query entry points; by using
   schema knowledge to get better answers.

3. Identify (or develop!) ontology language which has the required
   features (→ really interesting research!).
   E.g., spatial reasoning.

4. Realize application and publish (additional) data as LoD data.

Important: Keep it simple, stupid!
A little semantics can go a long way.
Take-Home Message

Metadata without formal semantics is simply more data.

Get added value
from using formal semantics / access to implicit knowledge.

Lift your applications *carefully* to the use of deeper semantics.

... and thanks to Prateek Jain for the LoD querying example ...
Thanks!
References


References


References


References
