Ontologies in a Data-driven World

Pascal Hitzler
DaSe Lab for Data Semantics
Wright State University
http://www.pascal-hitzler.de
Textbook

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)

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), 2013.

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

http://www.semantic-web-book.org
Semantic Web journal

- EiCs: Pascal Hitzler
  Krzysztof Janowicz

- New journal with significant uptake.

- We very much welcome contributions at the “rim” of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.

- Non-standard (open & transparent) review process.

- http://www.semantic-web-journal.net/
Ontologies?
A Basic Idea of the Semantic Web

- Person 1 exchanges symbols with Person 2, agreeing on the "Duck" concept.
- Agent 1 and Agent 2 exchange symbols, agreeing on the Ontology description.
- The concept of "Duck" is part of a specific domain, e.g., Animals.
A Basic Idea of the Semantic Web

Reconciling OWL and Rules
Knorr, Hitzler, Maier
ECAI 2012

Ontology represents
general domain knowledge

Data e.g. on Websites

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

Reconciling OWL and Rules
Knorr, Hitzler, Maier
ECAI 2012

e.g. every publication has an author
Ontology represents general domain knowledge

- e.g. every publication has an author

Reconciling OWL and Rules
Knorr, Hitzler, Maier
ECAI 2012

Data e.g. on Websites
The ontology hype

- Large, well-thought-out ontologies (foundational/domain/etc).
- Networked, interlinked ontologies

“You just have to get your formal definitions right, and a lot of the rest will just fall into place.”
The ontology hype

• “You just have to get your formal definitions right, and a lot of the rest will just fall into place.”

  – This does not even work for
    • scientists
    • wanting to share and reuse scientific data
    • through well-kept data repositories

  – So how is this supposed to work for the web at large?
Multiple perspectives

- Try to find a universal definition for
  - Forest
  - Mountain
  - City
  - River
  - Etc.

- The stronger our ontological commitments, the more we lose reusability.

- We need to accept that conceptualizations are often very local, resulting in “micro-ontologies”.
Multiple perspectives

Two ontologies.
Left: transportation domain
Right: agriculture domain

We cannot simply equate a:Canal and b:Canal!
The well-done ontologies

- Brittle
- Expensive
- Sometimes unintuitive
- Unwieldy
- Single-perspective
- Difficult to reuse

- Work in some contexts.
- Work if a lot of central control is imposed.
- Take a lot of manpower.
Pre-LOD Semantic Web

- Foundational ontologies
- Networked ontologies
- Sophisticated ontology languages

Scientific Hypothesis:
These will solve your data and information management problems

Remember that scientific progress is fundamentally about falsification, not verification 😊
Linked Data?
The linked data counter-hype

- “Ontologies don’t work, let’s just link data”

- “Okay, with a little bit of ontologies on top.”

- “The Linked Data Web is the true Semantic Web.”
Linked Data 2011
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia:author</td>
<td>dbpedia:J._R._R._Tolkien</td>
</tr>
</tbody>
</table>
| dbprop:books             | dbpedia:The_Two_Towers, dbpedia:The_Return_of_the_King,  
                          | dbpedia:The_Fellowship_of_the_Ring,  
                          | "Volumes:"                                                               |
| dbprop:country           | England                                                              |
| dbprop:expiry            | 20 (xsd:integer)                                                    |
| dbprop:genre             | dbpedia:Adventure_novel                                             |
|                          | dbpedia:High_fantasy                                               |
| dbprop:hasPhotoCollection| http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/The_Lord_of_the_Rings |
| dbprop:imageCaption      | Tolkien's own cover designs for the three volumes                   |
| dbprop:language          | English                                                              |
| dbprop:mediaType         | Print                                                                |
| dbprop:name              | The Lord of the Rings                                               |
| dbprop:pages             | 1216 (xsd:integer)                                                  |
| dbprop:precededBy        | dbpedia:The_Hobbit                                                  |
| dbprop:pubDate           | 21 (xsd:integer)                                                    |
| dbprop:publisher         | dbpedia:Allen & Unwin                                                |
| dbprop:small             | yes                                                                  |
                          | dbpedia:Template:Pp-vancalism                                        |
| dcterms:subject          | category:Monomyths,  
                          | category:High_fantasy_books,  
                          | category:Middle-earth_books,  
                          | category:British_fantasy_novels,  
                          | category:Fantasy_books_by_series,  
                          | category:1950s_fantasy_novels,  
                          | category:Sequel_novels,  
                          | category:The_Lord_of_the_Rings,  
                          | category:English_novels               |
Information as RDF graph

LOTR hasAuthor Tolkien .
Hobbit hasAuthor Tolkien .
LOTR hasCharacter Bilbo .
Hobbit hasCharacter Bilbo .
## Linked Data: Volume

### Number of Datasets

<table>
<thead>
<tr>
<th>Date</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-09-19</td>
<td>295</td>
</tr>
<tr>
<td>2010-09-22</td>
<td>203</td>
</tr>
<tr>
<td>2009-07-14</td>
<td>95</td>
</tr>
<tr>
<td>2008-09-18</td>
<td>45</td>
</tr>
<tr>
<td>2007-10-08</td>
<td>25</td>
</tr>
<tr>
<td>2007-05-01</td>
<td>12</td>
</tr>
</tbody>
</table>

### Number of triples (Sept 2011)

31,634,213,770 with 503,998,829 out-links

From [http://www4.wiwiss.fu-berlin.de/lodcloud/state/](http://www4.wiwiss.fu-berlin.de/lodcloud/state/)
Geoindexed Linked Data – courtesy of Krzysztof Janowicz
http://stko.geog.ucsb.edu/location_linked_data
October 2013:

Ca. 25,000,000,000 schema.org references on the web.

15% of all pages now have schema.org markup.

That’s just schema.org references …
“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-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)
Linked Data: Variety

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

Bills:h3962

Vote: 2009-887

Votes:2009-887/+ vote:votedBy people/P000197

dc:title

name

rdfs:label

Aye

vote:hasOption

dc:title

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

vote:hasAction

On Passage: H R 3962 Affordable Health Care for America Act
Linked Data federated querying

Query → Upper level ontology

LOD IMDB Dataset

LOD Wikipedia Dataset (DBPedia)

Answer

Joshi, Jain, Hitzler et al. ODBASE 2012
## Bootstrapping-based alignment

### 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>Alignment</th>
<th>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</td>
<td>1</td>
<td>0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>M,D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>F,D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>G,D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>err</td>
<td>err</td>
</tr>
<tr>
<td>S,F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.2</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</td>
<td>0</td>
<td>0</td>
<td>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</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Jain, Hitzler et al, ISWC2010
Linked Data federated querying

Upper level ontology

Query

LOD IMDB Dataset

LOD Wikipedia Dataset (DBPedia)

Answer

Joshi, Jain, Hitzler et al. ODBASE 2012
“Identify films, the nations where they were shot and the population of these countries”

```
SELECT ?film ?nation ?pop
WHERE {
}
```
Querying approach

Works very well, but only in some very limited cases.

Cannot deal with graph representations of even very minimal complexity.
Automated federation?

R2R:
- foaf:Person
  - type
  - name: “Smith, John”

BCO-DMO:
- foaf:Person
  - type: Person_752
  - familyName: “Smith”
  - givenName: “John”
  - name: “John Smith”
Automated federation?
Automated federation?

Copernicus lunar crater located on earth – courtesy of Krzysztof Janowicz http://stko.geog.ucsb.edu/location_linked_data (missing reference coordinate system)

Copernicus is a lunar impact crater named after the astronomer Nicolaus Copernicus, located in eastern Oceanus Procellarum. It is estimated to be about 800 million years old, and typifies craters that formed during the Copernican period in that it has a prominent ray system.

Characteristics

Copernicus is visible using binoculars, and is located slightly northwest of the center of the Moon's Earth-facing hemisphere. South of the crater is the Mare Insularum, and to the south-southwest is the crater Reinhold. North of Copernicus are the Montes Carpatus, which lie at the south edge of Mare Imbrium. West of Copernicus is a group of dispersed lunar hills. Due to its relative youth, the crater has remained in a relatively pristine shape since it formed.

The circular rim has a discernible hexagonal form, with a terraced inner wall and a 30 km wide, sloping rampart that descends nearly a kilometer to the surrounding mare. There are three distinct terraces visible, and arc-shaped landslides due to slumping of the inner wall as the crater debris subsided. Most likely due to its recent formation, the crater floor has not been flooded.
The linked data counter-hype

• “Ontologies don’t work, let’s just link data”

• “Okay, with a little bit of ontologies on top.”

• But then we don’t even know how to effectively query over multiple linked datasets (without using a lot of manpower to manually integrate them).

• It seems rather obvious that we need to get ontologies into the picture, but how to do it while avoiding the drawbacks of strong ontological commitments?
So What Now?
How to establish a flexible conceptual architecture using data and ontological modeling?
“An ontology design pattern is a reusable successful solution to a recurrent modeling problem.”

So-called *content patterns* usually encode specific abstract notions, such as process, event, agent, etc.
Ontology Design Patterns

• Bottom-up homogenization of data representation.

• Avoidance of strong ontological commitments.

• Avoidance of standardization of specific modeling details.

• Well thought-out patterns can be very strong and versatile, thus serve many needs.

We are currently establishing many geo-patterns in a series of hands-on workshops, the GeoVoCamps, see http://vocamp.org/
Ontology Design Patterns

“Horizontal” alignment via patterns
Semantic Trajectories

[Hu, Janowicz, Carral, Scheider, Kuhn, Berg-Cross, Hitzler, Dean, COSIT2013]
Semantics in OWL

\[
\text{Fix} \subseteq \exists \text{atTime}.\text{OWL-Time:Temporal Thing} \sqcap \exists \text{hasLocation}.\text{Position} \\
\quad \sqcap \exists \text{hasFix}^- \cdot \text{SemanticTrajectory}
\]

\[
\text{Segment} \subseteq \exists \text{startsFrom}.\text{Fix} \sqcap \exists \text{endsAt}.\text{Fix}
\]

\[
\top \sqsubseteq 1 \text{startsFrom}.\top
\]

\[
\top \sqsubseteq 1 \text{endsAt}.\top
\]

\[
\text{Segment} \subseteq \exists \text{hasSegment}^- \cdot \text{SemanticTrajectory}
\]

\[
\text{startsFrom}^- \circ \text{endsAt} \sqsubseteq \text{hasNext}
\]

\[
\text{hasNext} \sqsubseteq \text{hasSuccessor}
\]

\[
\text{hasSuccessor} \circ \text{hasSuccessor} \sqsubseteq \text{hasSuccessor}
\]

\[
\text{hasNext}^- \sqsubseteq \text{hasPrevious}
\]

\[
\text{hasSuccessor}^- \sqsubseteq \text{hasPredecessor}
\]
Semantics in OWL

\[\text{Fix} \sqcap \neg \exists \text{endsAt}. \text{Segment} \sqsubseteq \text{StartingFix} \]  
\[\text{Fix} \sqcap \neg \exists \text{startsFrom}. \text{Segment} \sqsubseteq \text{EndingFix} \]  
\[\text{Segment} \sqcap \exists \text{startsFrom}. \text{StartingFix} \sqsubseteq \text{StartingSegment} \]  
\[\text{Segment} \sqcap \exists \text{endsAt}. \text{EndingFix} \sqsubseteq \text{EndingSegment} \]  

\[\exists \text{hasSegment}. \text{Segment} \sqsubseteq \text{SemanticTrajectory} \]  
\[\exists \text{hasFix}. \text{Segment} \sqsubseteq \text{SemanticTrajectory} \]  
\[\exists \text{hasSegment}\. \text{SemanticTrajectory} \sqsubseteq \text{Segment} \]  
\[\exists \text{hasFix}\. \text{SemanticTrajectory} \sqsubseteq \text{Fix} \]
Helpfulness of patterns

Even minimalistic reuse is helpful:

R2R:
- **foaf:Person**
  - type: person/101396
  - name: “Smith, John”

BCO-DMO:
- **foaf:Person**
  - type: Person_752
  - name: “John Smith”
  - familyName: “Smith”
  - givenName: “John”
Patterns

- Help to focus when modeling (one key notion at a time).
- Good ontology modeling implicitly employs the patterns idea anyway. It’s just that you expose the patterns.
- An ontology composed of patterns exposes its internal conceptual structure (as a composition of formal vocabulary pieces).
- Well-designed patterns are widely reusable and adaptable.
- You don’t have to buy a whole ontology when you adopt a few patterns from it.
- You can easily modify a pattern without giving up on a lot of similarity to the original pattern (which can be leveraged for data integration).
- You can separate the patterns from specific (application-driven) modifications.
- You can separate the patterns from specific axiomatically defined “views”.
Patterns Example

NSF EarthCube project “OceanLink”:

- Integration of existing ocean science data repositories.
- For faceted browsing and semantic search.
- To be done in a flexible, extendable, modular way.
- With minimal effort for additional data providers to integrate their content.

National Science Foundation award 1354778 "EAGER: Collaborative Research: EarthCube Building Blocks, Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery."
OceanLink and EarthCube

EarthCube:
Developing a Community-Driven Data and Knowledge Environment for the Geosciences

“concepts and approaches to create integrated data management infrastructures across the Geosciences.”

“EarthCube aims to create a well-connected and facile environment to share data and knowledge in an open, transparent, and inclusive manner, thus accelerating our ability to understand and predict the Earth system.”
OceanLink setup

User Interface

UI Views

OceanLink Patterns

mappings

R2R

BCO-DMO

WHOI Library

AGU

NSF
OceanLink patterns

Some central patterns:

• Cruise
• Trajectory
• Person
• Organization
• Roles of Agents
• Repository Object
• Data Set
• Document

We’re not starting from zero of course.
Ocean Science Cruise (draft)
Cruise trajectory (draft)
Cruise\( (x) \wedge \text{hasTrajectory}(x, y) \wedge \text{hasSegment}(y, z) \wedge \text{isTraversedBy}(z, v) \rightarrow \text{participatesIn}(v, z)\)
Cruise trajectory

\[\text{Cruise}(x) \land \text{hasTrajectory}(x, y) \land \text{hasSegment}(y, z) \land \text{isTraversedBy}(z, v) \rightarrow \text{participatesIn}(v, z)\]

\[
\text{Cruise} \equiv \exists \text{cruise.Self} \subseteq \text{hasParticipant}
\]

\[
\text{hasParticipant} \equiv \text{participatesIn}^{-}
\]
Cruise trajectory

\[
\text{Fix}(x) \land \text{hasAttribute}(x, \#\text{portStopArrival})
\land \text{atPort}(x, y) \land \text{hasSpatialFootprint}(y, z)
\land \text{hasLocation}(x, w) \rightarrow \text{locatedIn}(w, z)
\]
Cruise trajectory

\[
\text{Fix}(x) \land \text{hasAttribute}(x, \#\text{portStopArrival}) \\
\land \text{atPort}(x, y) \land \text{hasSpatialFootprint}(y, z) \\
\land \text{hasLocation}(x, w) \rightarrow \text{locatedIn}(w, z)
\]

\[
\text{Fix} \land \exists \text{hasTrajectory.} \{\#\text{portStopArrival}\} \equiv \exists \text{fixps.Self} \\
\text{hasLocation}^{-} \circ \text{fixps} \circ \text{atPort} \circ \text{hasSpatialFootprint} \\
\subseteq \text{locatedIn}
\]
Ways forward

• Establish a flexible conceptual architecture using data and ontological modeling.
• A principled use of patterns, including
  – the development of a theory of patterns and
  – the provision of a critical amount of central patterns
    may provide a primary path forward.
Thanks!
References


References

References


References


References

