Semantic Web Applications and the Semantic Web in 10 Years

Based on work of Grigoris Antoniou, Frank van Harmelen
Semantic Web Search Engines
SemanticWebDocument, RDFXML, 2006-07-06, 68K, ontoRatio(0.16), metadata, cached

http://www.aifb.uni-karlsruhe.de/Publikationen/viewExternerAutorOWL/id435instance
[DESC] name - Valentina Tamma
SemanticWebDocument, RDFXML, 2006-02-07, 1K, ontoRatio(0.43), metadata, cached

http://www.aifb.uni-karlsruhe.de/Publikationen/viewExternerAutorOWL/id435.owl
[DESC] comment - Instance data for external author Valentina Tamma
SemanticWebDocument, RDFXML, 2006-02-07, 1K, ontoRatio(0.43), metadata, cached
## Swoogle's Statistics of the Semantic Web

### Swoogle Today

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>admin_dt</td>
<td>2013-01-24 00:02:11</td>
<td>Datetime Watched</td>
</tr>
<tr>
<td>url_total</td>
<td>12,065,719</td>
<td>Number of URLs being discovered</td>
</tr>
<tr>
<td>url_pinged</td>
<td>6,135,482</td>
<td>Number of URLs being pinged</td>
</tr>
<tr>
<td>total_swd</td>
<td>3,940,766</td>
<td>Number of Semantic Web Documents (regardless of embedded or containing some errors) be confirmed.</td>
</tr>
<tr>
<td>total_swd_strict</td>
<td>1,936,917</td>
<td>Number of error-free pure Semantic Web Documents</td>
</tr>
<tr>
<td>total_swd_embed</td>
<td>1,516,768</td>
<td>Number of documents (except SWDs, PDF, and JPEG) embedding Semantic Web Data</td>
</tr>
<tr>
<td>triple_total</td>
<td>1,147,784,266</td>
<td>Number of triples could be parsed from all Semantic Web Documents.</td>
</tr>
</tbody>
</table>
Limitations of Swoogle

- Very limited quality control mechanisms
  - Many ontologies are duplicated, no quality information provided

- Limited Query/Search mechanisms
  - Only keyword search; no distinction between types of elements
  - Need for more powerful query methods (e.g., ability to pose formal queries; ability to distinguish between classes and instances, etc…)

- Limited range of ontology ranking mechanisms
  - Swoogle only uses a 'popularity-based' one

- No support for ontology modularization

- Outdated?
Searching objects with Falcons

- Query refinement based on class hierarchy
- RDF reasoning
- Diverse keyword search based on comprehensive keyword index
- => http://ws.nju.edu.cn/falcons/
A Snippet from Falcons search results

Objects 1 - 5 of 100 for your search Beijing

Beijing - Thing, Place, capital, CitiesInChina, Area, PopulatedPlace, HostCitiesOfTheSummerOlympicGames, village, CapitalsInAsia
  • type: Thing
  • label: Beijing
  • page: Beijing
  • comment: Pekín o Beijing es la capital de la República Popular China y tiene más de 17 millones de habitantes. Pekín es una de las cuatro municipalidades de Ch...
  • latitude: 39.90000152587891
  • longitude: 116.4000015258789
  • depiction: Beijing_montage.png
  • hasPhotoCollection: Photos for DBpedia.org resource Beijing
  • subject: Category:Independent_cities
  • abstract: Peking Kíná fővárosa, négy tartományi jogú városának egyike, Kíná közigazgatási szerkezetében megegyezik a tartományokkal. Pekinget északról, nyugatról...

http://dbpedia.org/resource/Beijing
Geonames/DBPedia
**Milton Keynes** ca. 69 m
United Kingdom → Milton Keynes
seat of a first-order administrative division
population: 31967
N 52° 1' 59'' W 0° 42' 0''
52.033333 / -0.7

<table>
<thead>
<tr>
<th>Name</th>
<th>country</th>
<th>feature</th>
<th>km to center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broughton</td>
<td>United Kingdom</td>
<td>populated place</td>
<td>0.92 km</td>
</tr>
<tr>
<td>Willen</td>
<td>United Kingdom</td>
<td>populated place</td>
<td>1.38 km</td>
</tr>
<tr>
<td>Milton Keynes</td>
<td>United</td>
<td>seat of a first-order administrative division</td>
<td>1.7 km</td>
</tr>
<tr>
<td>Moulsoe</td>
<td>United Kingdom</td>
<td>populated place</td>
<td>2.05 km</td>
</tr>
<tr>
<td>Woolstone</td>
<td>United Kingdom</td>
<td>populated place</td>
<td>2.17 km</td>
</tr>
<tr>
<td>River Ouzel</td>
<td>United Kingdom</td>
<td>stream</td>
<td>2.44 km</td>
</tr>
</tbody>
</table>
Supporting Clinical Decision Making

- Use of Semantic Web technologies for clinical decision support in Athens Heart Center
Operational Since January 2006
Active Semantic Electronic Medical Records (ASEMR)

Goals:
• Increase efficiency with decision support
  • formulary, billing, reimbursement
  • real time chart completion
  • automated linking with billing

Technologies:
• Ontologies, semantic annotations & rules
• Service Oriented Architecture
ASEMR - Demonstration

Click to Launch
FilmTrust

- Users can add other persons as friends, but can be one-way action.
- Users rank their friends as “how likely you would see the film which is picked by this friend”
- “Recommended Rating” uses the inferred trust values to compute with TidalTrust, to calculate a weighted average rating for each movie.
- It is an example of collaborative filtering system
Many others

- http://www.reeegle.info
- http://activehiring.labs.exalead.com
- http://poolparty.biz/demozone/general
- http://www.jinni.com

- Most of them are actually LOD applications with „little“ semantics/intelligence 😞
FUTURE

The semantic web in 10 years
The Semantic Web in 10 years

- A workshop at ISWC 2012
- 14 papers by different researchers
- Discussing what the Semantic Web (research) will/should look like at 2022
Interleaving Human-Machine Knowledge and Computation

- Large success of Galaxy Zoo and ReCaptcha show potential for Human Computation with SW technologies

- Major challenges:
  - motivating people
  - Generic interface for human brains
  - Error diversity: how to handle inconsistencies; how many people need to contribute until the result is good (enough)
A truly Read-Write Web for machines as the next generation Web?

- Datasets will be driven by machine updates
- Human-created schemas will become obsolete
- Web of read data $\rightarrow$ Web of read-write data
- Monotonicity of the Semantic Web will remain
- Permission management for reading and writing of data

- Current SW: High precision, low recall
- Rules of thumbs: dangerous lions
- Main challenge: discover imprecise axioms
  - Use inductive learning
  - Crowd-sourcing?
- Use reification for representation of uncertainty
Security Impacts on Semantic Technologies in the Coming Decade.

- Today: Proposals for Security in each layer; only works up to XML
- New job: System Security Engineer
- Future focus:
  - Knowledge base hardening
  - Situational awareness
  - Knowledge base auditing
  - Web service penetration tests
Nowadays: Babylonian Confusion between languages used in ontology engineering

Development of a Distributed Ontology Language
- Meta-language for integration and interoperability
- Combining different logics into one system
- To be ISO-standardized in 2015

Ontohub ontology repository
- Keeping track of ontologies no matter of the language used
Linked Open Data: Are we Drowning in Information and Starving for Know-How?

● Today: Linked Open Know-That
  - Facts, If->Then

● Vision: Linked Open Know-How
  - No need to publish/send sensitive data, e.g. about cancer, but store how to do (a cancer diagnosis)
  - Instead of sending data, you download programs for local computation

● Example query:
  - “how can I use the Facebook API in Python for doing task X?”
How I Would Like Semantic Web To Be, For My Children.

- Develop “real” agents communicating and negotiating with each other
- Dynamic privacy settings, according to the context
- Modern Citizens work together to fight pollution, global warming, etc.
- Automatic traffic coordination by Semantic Web technologies
Semantic Web in 10 years: Semantics with a purpose.

- Get overlap with other research communities as
  - Computational Linguistics (text mining-> understanding)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining of Massive Datasets [2]</td>
<td>zero appearances</td>
</tr>
<tr>
<td>(340 pages)</td>
<td></td>
</tr>
<tr>
<td>Understanding Big Data [3]</td>
<td>three appearances</td>
</tr>
<tr>
<td>(166 pages)</td>
<td></td>
</tr>
<tr>
<td>(156 pages)</td>
<td></td>
</tr>
<tr>
<td>Big Data Glossary [5]</td>
<td>one appearance</td>
</tr>
<tr>
<td>(62 pages)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: number of appearances of the words *semantics* in four leading books on “Big Data”
Semantic Web in Cultural Heritage After 2020.

- ToDos until 2022:
  - Handle uncertainties
  - Personalized semantics (aka opinions)
  - Multilingual knowledge representation and access
  - Automatic and easy-to-use tools
  - Solving copyright and trust issues
The Rise of the Verb

- Not only represent knowledge, but actions
- Example with Siri
  - Today: “move my meeting from 3 to 4”
  - Future: “Mow my lawn”
- In 10 years, every verb will be connected to a set of well-understood actions
- Understanding JavaScript programs
The Terminator's origins or how the Semantic Web could endanger Humanity.

- Possible misuses in the future:
  - The Spam takeover (convincingly personalized); not only messages, but data facts as well!
  - Link spoofing; Malicious rule injection, complex interaction of same-as
  - DDOS by reasoning complexity
  - Identity theft (already today in social networks)

- If everything is handled by agents automatically, “Judgement Day” might come
Enabling the Cognitive Semantic Web.

- Complains about missing semantics
- In 2022, the Semantic Web
  - will be an emergent social network human and artificial cognitive agents interacting with each other in a hybrid environment.
  - Will put you into trouble distinguishing human from artificial agents
Semantics empowered Physical-Cyber-Social systems.

- Natural interactions with humans and agents
- Combining neuroscience with SW
Will it read my mind and cook me breakfast by then?

- State-of-the-art SW is made (and controlled) by computer science researchers
- Future
  - Downloading big whole datasets from the web will be obsolete, if not illegal
  - Services for checking semantic validity
  - Native support for SW by browsers and search engines; intuitive interfaces
  - Mathematical elegance will be balanced by usefulness and socio-technical beauty
  - SW will know everything and might even read minds
Finally THE Scenario: Sensor Data Fusion and Analysis

How do we determine if the three images depict ...
- the same time and same place?
- the same entity?
- a serious threat?
Questions or comments?