Characterizing Semantic Web Applications

Prof. Enrico Motta
Director, Knowledge Media Institute
The Open University
Milton Keynes, UK
Understanding the SW

• Issues
  – What is new/different about the semantic web?
  – What are the key aspects that characterize semantic web applications?
  – What are the key differences between semantic web applications and ‘traditional’ knowledge based systems?

• Results
  – A framework providing a characterization of semantic web applications
  – A classification of a representative sample of SW applications according to our framework
  – A blueprint (set of reqs) for designing SW applications
Semantics on the web
(The Semantic Web)
Professor of Knowledge Technologies
Director, Knowledge Media Institute, The Open University

How to Contact:

Knowledge Media Institute [KMi]
The Open University,
Milton Keynes,
MK7 6AA,
United Kingdom.

Email: e.motta@open.ac.uk
Phone: +44 (0) 1908 653506
Fax: +44 (0) 1908 653169
<akt:Person rdf:about="akt:EnricoMotta">
  <rdfs:label>Enrico Motta</rdfs:label>
  <akt:hasAffiliation rdf:resource="akt:TheOpenUniversity"/>
  <akt:hasJobTitle>kmi director</akt:hasJobTitle>
  <akt:worksInOrgUnit rdf:resource="akt:KnowledgeMediaInstitute"/>
  <akt:hasGivenName>enrico</akt:hasGivenName>
  <akt:hasFamilyName>motta</akt:hasFamilyName>
  <akt:worksInProject rdf:resource="akt:Neon"/>
  <akt:worksInProject rdf:resource="akt:X-Media"/>
  <akt:hasPrettyName>Enrico Motta</akt:hasPrettyName>
  <akt:hasPostalAddress rdf:resource="akt:KmiPostalAddress"/>
  <akt:hasEmailAddress>e.motta@open.ac.uk</akt:hasEmailAddress>
  <akt:hasHomePage rdf:resource="http://kmi.open.ac.uk/people/motta/"/>
</akt:Person>
<akt:Person rdf:about="akt:EnricoMotta">
  <rdfs:label>Enrico Motta</rdfs:label>
  <akt:hasAffiliation rdf:resource="akt:TheOpenUniversity"/>
  <akt:hasJobTitle>kmi director</akt:hasJobTitle>
  <akt:worksInOrgUnit rdf:resource="akt:KnowledgeMediaInstit"/>
  <akt:hasGivenName>enrico</akt:hasGivenName>
  <akt:hasFamilyName>motta</akt:hasFamilyName>
  <akt:worksInProject rdf:resource="akt:Neon"/>
  <akt:worksInProject rdf:resource="akt:X-Media"/>
  <akt:hasPrettyName>Enrico Motta</akt:hasPrettyName>
  <akt:hasPostalAddress rdf:resource="akt:KmiPostalAddress"/>
  <akt:hasEmailAddress>e.motta@open.ac.uk</akt:hasEmailAddress>
  <akt:hasHomePage rdf:resource="http://kmi.open.ac.uk/people/motta/"/>
</akt:Person>
Please get me an appointment with a dealer within 50 miles of my home to arrange a test drive of a Ferrari F430 Spider for Saturday morning.

Enrico’s Semantic Agent
Conceptual Interoperability

Car-Dealership
hasAddress
hasWebAddress...
.....
Schedule.....

Car-Dealership
hasAddress
hasWebAddress...
.....
Schedule.....
Key Aspect of SW #1: Hugeness
Growth of the SW

![Bar chart showing the growth of SW pages from 2003 to 2004](chart.png)
### Swoogle Today

<table>
<thead>
<tr>
<th>admin_dt</th>
<th>2006-06-27 23:58:30</th>
<th>Datetime Watched</th>
</tr>
</thead>
<tbody>
<tr>
<td>url_total</td>
<td>3,946,561</td>
<td>Number of URLs being discovered</td>
</tr>
<tr>
<td>url_pinged</td>
<td>2,448,132</td>
<td>Number of URLs being pinged</td>
</tr>
<tr>
<td>total_swd</td>
<td>1,570,779</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>1,003,095</td>
<td>Number of error-free pure Semantic Web Documents</td>
</tr>
<tr>
<td>total_swd_embed</td>
<td>407,408</td>
<td>Number of documents (except SWDs, PDF, and JPEG) embedding Semantic Web Data</td>
</tr>
<tr>
<td>triple_total</td>
<td>299,824,429</td>
<td>Number of triples could be parsed from all Semantic Web Documents.</td>
</tr>
</tbody>
</table>
Key Aspect of SW #2: Heterogeneity
Key Aspect of SW #2: Heterogeneity
Swoogle 2006

Searching over 10,000 ontologies
Other key aspects of the SW

- **Hugeness**
  - Sem. markup of the same order of magnitude as the web

- **Conceptual Heterogeneity**
  - Sem. markup based on many different ontologies

- **Very high rate of change**
  - Semantic data generated all the time from web resources

- **Heterogeneous Provenance**
  - Markup generated from a huge variety of different sources, by human and artificial agents

- **Various and subjective degrees of trust**
  - Al-Jazeera vs CNN....

- **Various degrees of data quality**
  - No guarantee of correctness

- **Intelligence a by-product of size and heterogeneity**
  - rather than a by-product of sophisticated problem solving
Compare with traditional KBS

- **Hugeness**
  - KBS normally small to medium size

- **Conceptual Heterogeneity**
  - KBS normally based on a single conceptual model

- **Very high rate of change**
  - Change rate under developers' control (hence, low)

- **Heterogeneous Provenance**
  - KBS are normally created ad hoc for an application by a centralised team of developers

- **Various and subjective degrees of trust**
  - Centralisation of process implies no significant trust issues

- **Various degrees of data quality**
  - Centralisation guarantees data quality across the board

- **Intelligence a by-product of size and heterogeneity**
  - In KBS a by-product of complex, task-centric reasoning
Analysis of SW Applications
Requirements for SW Applications

- **Hugeness**
  - SW applications should operate at scale

- **Heterogeneity**
  - SW applications should be able to handle multiple ontologies

- **Very high rate of change**
  - SW applications need to be open with respect to semantic resources

- **Heterogeneous provenance**
  - SW applications need to be open with respect to web resources
Additional Requirements

• SW is an extension of the web, so it makes sense to require that SW applications be compliant with key current web trends
  – Web 2.0 - i.e., providing interactive feature for harnessing collective intelligence (O'Reilly)
  – Web Services
    • Obviously it is also desirable that SW applications are also open with respect to web functionalities
Framework for characterizing SW applications

- Does app operate at scale?
- Can it handle multiple ontologies?
- Is it open to semantic resources?
- Is it open to web resources?
- Is it open to web services?
- Does it include Web 2.0 like features?
Applying the framework to six SW applications

CS AKTive Space, FLINK, Magpie, PiggyBank, AquaLog, PowerAqua
<table>
<thead>
<tr>
<th>Type</th>
<th>Aggregation and visualization of data from multiple sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operates at scale?</td>
<td>Yes, large numbers of data crawled from hundreds of different UK CS sites</td>
</tr>
<tr>
<td>Multi-ontology?</td>
<td>All data extracted and integrated into the AKT reference ontology</td>
</tr>
<tr>
<td>Open to semantic resources?</td>
<td>No, RDF data are generated by the system, rather than reused from existing repositories</td>
</tr>
<tr>
<td>Open to web resources?</td>
<td>No (it is not possible to indicate more sites to the system and expect it to add more data)</td>
</tr>
<tr>
<td>Open to web services?</td>
<td>No (there is no open architecture to add crawlers)</td>
</tr>
<tr>
<td>Web 2.0 like?</td>
<td>No (no tagging or interactive features)</td>
</tr>
<tr>
<td>Type</td>
<td>Semantic Web Browser</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Operates at scale?</td>
<td>Yes, large numbers of data crawled from publication archives, google, FOAF, etc..</td>
</tr>
<tr>
<td>Multi-ontology?</td>
<td>Partially. Can switch from one ontology to another, but only one ontology can be used at the time.</td>
</tr>
<tr>
<td>Open to semantic resources?</td>
<td>Yes</td>
</tr>
<tr>
<td>Open to web resources?</td>
<td>Yes (but quality can degrade as you move away from resources relevant to the current ontology)</td>
</tr>
<tr>
<td>Open to web services?</td>
<td>Yes</td>
</tr>
<tr>
<td>Web 2.0 like?</td>
<td>No (no tagging or interactive features)</td>
</tr>
</tbody>
</table>
Social Network of Enrico Motta

Statistics

- Indegree: 67.0 (41.0) ↑
- Closeness: 0.5 (0.43) ↑
- Betweenness: 8212.68 (4982.34) ↑
- Top Publications: 53.0 (40.0) ↑
- Impact: 11.48 (9.21) ↑

What do the numbers mean?

Rankings

- Indegree: 11 (10) ↓
- Closeness: 155 (7) ↓
- Betweenness: 10 (10) -
- Top Publications: 13 (15) ↑
- Impact: 193 (177) ↓

Links

- Search Google!
- Search Foafnaut!
- Search Citeseer!
- Search A9 (Amazon)!

Tip: You can drag nodes with your mouse. You can also use the scrollbars to pan the image. You can zoom in and out using the +/− buttons, but also with the roller.
<table>
<thead>
<tr>
<th>Type</th>
<th>Aggregation and visualization of data from multiple sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operates at scale?</td>
<td>Yes, large numbers of data crawled from publication archives, google, FOAF, etc..</td>
</tr>
<tr>
<td>Multi-ontology?</td>
<td>No. All data extracted and integrated into a single ontology</td>
</tr>
<tr>
<td>Open to semantic resources?</td>
<td>No, RDF data are generated by the system, rather than reused from existing repositories</td>
</tr>
<tr>
<td>Open to web resources?</td>
<td>No (it is not possible to indicate more sites to the system and expect it to add more data)</td>
</tr>
<tr>
<td>Open to web services?</td>
<td>No</td>
</tr>
<tr>
<td>Web 2.0 like?</td>
<td>No (no tagging or interactive features)</td>
</tr>
</tbody>
</table>
PiggyBank

restaurants + movies
Combined Information

1 filter criterion
• type: Show (remove) Restaurant (remove)

Order Commands

View items at list

1 item(s)
• Abacus Cuisine Of China

date
By hour
Type here to filter
10:00 PM (138)
11:00 PM (20)

type
Type here to filter
Restaurant (26)
Show (158)

address
Type here to filter

Map data ©2005 NAVTEQ™, Tela Atlas

Transferring data from nst.google.com...
<table>
<thead>
<tr>
<th>Type</th>
<th>Semantic Web Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operates at scale?</strong></td>
<td>Yes, data can be collected from semantic and non-semantic sources</td>
</tr>
<tr>
<td><strong>Multi-ontology?</strong></td>
<td>Data can be brought in from different ontologies, unclear whether intg. support is provided</td>
</tr>
<tr>
<td><strong>Open to semantic resources?</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Open to web resources?</strong></td>
<td>Yes (open to screen scraping mechanisms)</td>
</tr>
<tr>
<td><strong>Open to web services?</strong></td>
<td>Yes (open to screen scraping mechanisms)</td>
</tr>
<tr>
<td><strong>Web 2.0 like?</strong></td>
<td>Yes, supports tagging and sharing of bookmarks</td>
</tr>
</tbody>
</table>
The answer to the question:

`peter-scott has-research-interest semantic-web-area`

The value of `has-web-address` for `peter-scott` is/are:

```
"news.kmi.open.ac.uk/peterblog"
```
<table>
<thead>
<tr>
<th><strong>Type</strong></th>
<th>Question Answering System</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operates at scale?</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Multi-ontology?</strong></td>
<td>Partially. Can switch from one ontology to another with zero configuration effort, but only one ontology can be used at the time.</td>
</tr>
<tr>
<td><strong>Open to semantic resources?</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Open to web resources?</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Open to web services?</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Web 2.0 like?</strong></td>
<td>Yes. No tagging, but learning mechanism supports mapping user terminologies to ontologies</td>
</tr>
<tr>
<td>Type</td>
<td>Question Answering System</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Operates at scale?</td>
<td>Yes</td>
</tr>
<tr>
<td>Multi-ontology?</td>
<td>Yes</td>
</tr>
<tr>
<td>Open to semantic resources?</td>
<td>Yes</td>
</tr>
<tr>
<td>Open to web resources?</td>
<td>No</td>
</tr>
<tr>
<td>Open to web services?</td>
<td>Yes</td>
</tr>
<tr>
<td>Web 2.0 like?</td>
<td>Yes. No tagging, but learning mechanism supports mapping user terminologies to ontologies</td>
</tr>
<tr>
<td>Feature</td>
<td>Response</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Operates at scale?</td>
<td>All</td>
</tr>
<tr>
<td>Multi-ontology?</td>
<td>PowerAqua, Magpie and AquaLog (partially), PiggyBank (unclear)</td>
</tr>
<tr>
<td>Open to semantic resources?</td>
<td>PowerAqua, Magpie, AquaLog, PiggyBank</td>
</tr>
<tr>
<td>Open to web resources?</td>
<td>PiggyBank, Magpie</td>
</tr>
<tr>
<td>Open to web services?</td>
<td>PiggyBank, Magpie, PowerAqua</td>
</tr>
<tr>
<td>Web 2.0 like?</td>
<td>PiggyBank, AquaLog, PowerAqua</td>
</tr>
</tbody>
</table>
Conclusions

- Even the earliest SW applications recognised scale as a key requirement to address
- Semantic portals more similar to large scale KBs, than to our blueprint for SW applications
- The heterogeneous nature of the SW more and more taken into account by SW applications
- Overall trend is positive
  - Latest tools more closely address our requirements
- Automatic data acquisition remains the feature most often missing from SW applications
  - However, it may matter less and less....