PROPEL: Topic and trend analysis

Javier D. Fernández, Sabrina Kirrane, Axel Polleres
Interviews & Workshop

23 interviews:

- **Domains**
  - Consulting, Engineering, Environment, Finance and Insurance, Government, Healthcare, ICT, IT, Media, Pharmaceutical, Professional Services, Real Estate, Research, Startup, Tourism, Transports & Logistics

- **Roles**
  - Business Intelligence, CEO, Chief Engineer, Data and Systems Architect, Data Scientist, Director Information Management, Enterprise Architect, Founder, General Secretary, Governance, Risk & Compliance Manager, Head of Communications and Media, Head of Development, Head of HR, Head of R&D, Innovation Manager, Information Architect, IT Project Manager, Management, Managing director, Marketing Analyst, Principle System Analyst, Project Coordinator, Researcher, Technical Specialist

- Took place on the 10th of May 2016
  - private sector (solution providers and users)
  - research sector (technicians and strategic people)

- Business barriers and drivers
- Technological challenges and opportunities
User Stories

4 out of 60 user stories we collected in the interviews:

- **Horizontals**
  - Business processes (e.g. product logistics and supply chain management)
  - Human resources (e.g. expert and resource management)

- **Verticals**
  - Media & Publishing
  - Healthcare & Pharma
Technologies in need...

Analytics
Computational linguistics & NLP
Concept tagging & annotation
Data integration

Data management
Dynamic data / streaming
Extraction, data mining, text mining, entity extraction
Logic, formal languages & reasoning

Human-Computer Interaction & visualization
Knowledge representation
Machine learning
Ontology/thesaurus /taxonomy management

Quality & Provenance
Recommendations
Robustness, scalability, optimization and performance
Searching, browsing & exploration

Security and privacy
System engineering

We pretty much ended up in all areas that SW touches upon!
“**I would like to be able to exchange information and coordinate production and logistics with suppliers and customers…**”

“**…so that I can improve efficiency, effectiveness and flexibility of my inventory management and operations**”
Human Resources

“I would like identify expertise within our large organisation and be able to pinpoint the relevant experts…”

“…so that I can identify top trends within the organisation and expertise for the organisation as a whole”
I would like to display personalized content as precise as possible

So that my readers stay as long as possible on my website.
I would like to Integrate disparate systems that are:
- Hard to integrate
- Widespread
- Contain the same data that contradicts each other

So that I can gain insights from other clinical trials
Let’s take a step back...

- What can we offer as a community?

Taking an introspective view
COMMUNITY ANALYSIS
### Foundations in 2006-2015

<table>
<thead>
<tr>
<th>Topic</th>
<th>% over total occurrences in proceedings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge representation &amp; data creation/publishing/sharing</td>
<td>22.7%</td>
</tr>
<tr>
<td>Data management</td>
<td>15.7%</td>
</tr>
<tr>
<td>System engineering</td>
<td>11.1%</td>
</tr>
<tr>
<td>Searching / Browsing / Exploration</td>
<td>9.0%</td>
</tr>
<tr>
<td>Data integration</td>
<td>7.9%</td>
</tr>
<tr>
<td>Ontology/Thesaurus/Taxonomy management</td>
<td>7.5%</td>
</tr>
<tr>
<td>Formal logic / Formal languages / Description logics...</td>
<td>4.4%</td>
</tr>
<tr>
<td>Human-Computer Interaction &amp; Visualization</td>
<td>3.2%</td>
</tr>
<tr>
<td>Extraction, data mining, text mining, entity extraction</td>
<td>3.1%</td>
</tr>
<tr>
<td>Concept tagging, annotation</td>
<td>2.6%</td>
</tr>
<tr>
<td>Quality</td>
<td>2.3%</td>
</tr>
<tr>
<td>Dynamic data / Streaming</td>
<td>2.1%</td>
</tr>
<tr>
<td>Robustness, scalability, optimization and performance</td>
<td>1.8%</td>
</tr>
<tr>
<td>Analytics</td>
<td>1.7%</td>
</tr>
<tr>
<td>Computational linguistics &amp; NLP systems</td>
<td>1.7%</td>
</tr>
<tr>
<td>Machine learning</td>
<td>1.4%</td>
</tr>
<tr>
<td>Security and privacy</td>
<td>1.3%</td>
</tr>
<tr>
<td>Recommendations</td>
<td>0.4%</td>
</tr>
</tbody>
</table>
Community Analysis

- Monitoring SW community major venues:
  - ISWC (since 2006), ESWC (since 2006), SEMANTiCS (since 2007), JWS (since 2006), SWJ (since 2010)

- 3 seminal papers:
Semantic Web/Linked Data over time...

Subtopics:
- Expressing Meaning
- Knowledge Representation
- Ontologies
- Agents
- Evolution of Knowledge
Knowledge Representation & Reasoning

Trends in topics 2006-2015

- ontology
- web service
- agents
- artificial intelligence
- automated reasoning
- machine-readable
- intelligent user interfaces
The Semantic Web in Action
Corporate applications are well under way, and consumer uses are emerging

By Lee Feigenbaum, Ivan Herman, Tonya Hongsermeier, Eric Neumann and Susie Stephens

Early adopters:
MITRE
Chevron
British Telecom
Boeing
Ordnance Survey
Eli Lily
Pfizer
Agfa
Food and Drug Administration
National Institutes of Health

Software adopters/products:
Oracle
Adobe
Altova
OpenLink
TopQuadrant
Software AG
Aduna Software
Protége
SAPHIRE
LD Adopters - Companies

Early adopters (as of seminal papers) in papers 2006-2015

Year
Quantities
0 1 2 3 4 5 6 7

MITRE
British Telecom
Boeing
Food and Drug Administration
Pfizer
Chevron
National Institutes of Health

Software adopters (as of seminal papers) in papers 2006-2015

Year
Occurrences
0 20 40 60 80 100 120 140 160 180 200

Oracle
Altova
OpenLink
Software AG
TopQuadrant
Adobe
Protégé
SAP HPI
Adobe Software
LD Adopters - Companies

Conference Sponsors that appear in papers 2006-2015

Companies

- Google
- Oracle
- Yahoo
- SAP
- IEEE Intelligent Systems
- Franz
- Bing
- Expert System
- IBM Research
- Poolparty

Occurrences

0  200  400  600  800  1000  1200  1400  1600

Google
Oracle
Yahoo
SAP
IEEE Intelligent Systems
Franz
Bing
Expert System
IBM Research
Poolparty
LD Adopters - Domains

Topics grouped by domain 2006-2015

E.g.:
- dblp
  - computer science bibliography
  - Journal of Biomedical Semantics
  - Home > Journals
  - Journal of Biomedical Semantics @ BMC
    - Volume 7: 2016
    - Volume 6: 2015
    - Volume 5: 2014 - Supplement
    - Volume 4: 2013 - Supplement
    - Volume 3: 2012 - Supplement
    - Volume 2: 2011 - Supplements
    - Volume 1: 2010 - Supplement

Well, they publish in other venues…
The authors claim that "early research has transitioned into these larger, more applied systems, today’s Semantic Web research is changing: It builds on the earlier foundations but it has generated a more diverse set of pursuits".

Viewpoint
A New Look at the Semantic Web

Seeking to make Web data “smarter” by utilizing a new kind of semantics.
Looking to the future

**Trends in topics 2006-2015**

- Large/big data
- Event
- Stream
- Temporal
- Large scale
- Volume
- Variety
- Velocity

**Trends in topics 2006-2015**

- Quality
- Provenance
- Trust
Roadmap for Enterprise LD?

1. Linked Data security and privacy requirements
2. Determining and resolving data quality issues
3. Managing large amounts of data and associated metadata such as provenance and temporal data
4. Visualisation requirements
5. Understand the different application areas and their maturity in terms of real world deployment
6. Showcases in terms of startups, projects, applications and systems based on Linked Data technologies
7. Linked Data and recommender system models
8. Dynamic linked datasets

https://www.linked-data.at/