

Sistemi di Elaborazione dell'informazione II

Corso di Laurea Specialistica in Ingegneria Telematica

II anno – 4 CFU

Università Kore – Enna – A.A. 2008-2009

Alessandro Longheu

<http://www.dit.unict.it/users/alongheu>

alessandro.longheu@dit.unict.it

Presente e futuro del Web Semantico

A. Longheu – Sistemi di Elaborazione delle Informazioni II

Presente e futuro del SW

Currently...

- We have a solid specification since 2004
- Lots of tools are available
- There are lots of tutorials
- Active developers' communities
- Some mesasures claim that there are over 10⁷ Semantic Web documents... (ready to be integrated...)
- Large ontologies are being developed (converted from other formats or defined in OWL)
 - eClassOwl: eBusiness ontology for products and services, 75,000 classes and 5,500 properties
 - the Gene Ontology: to describe gene and gene product attributes in any organism
 - BiOPAX, for biological pathway data
 - UniProt: protein sequence and annotation terminology and data
- One should never forget: ontologies/vocabularies must be shared and reused!
- Querying RDF graphs becomes essential ; SPARQL is almost here

Presente e futuro del SW

- There are a **number of problems** however:
 - how to get RDF data
 - missing functionalities: rules, “light” ontologies, fuzzy reasoning, necessity to review RDF and OWL,...
 - misconceptions, messaging problems
 - need for more applications, deployment, acceptance
- A huge amount of data in Relational Databases Although tools exist, it is not feasible to convert that data into RDF Instead: SQL \Leftrightarrow RDF “bridges” are being developed

3

Presente e futuro del SW

- **Some of the messaging on Semantic Web** has gone terribly wrong :
 - “the Semantic Web is a reincarnation of Artificial Intelligence on the Web”
 - “it relies on giant, centrally controlled ontologies for “meaning” (as opposed to a democratic, bottom–up control of terms)”
 - “one has to add metadata to all Web pages, convert all relational databases, and XML data to use the Semantic Web”
 - “it is just an ugly application of XML”
 - “one has to learn formal logic, knowledge representation techniques, description logic, etc, to use it”
 - “it is, essentially, an academic project, of no interest for industry”

4

Presente e futuro del SW

- SW has indeed a strong foundation in research results, But remember:
 - (1) the Web was born at CERN...
 - (2)...was first picked up by high energy physicists...
 - (3)...then by academia at large...
 - (4)...then by small businesses and start-ups...
 - (5) “big business” came only later!
- network effect kicked in early...
- Semantic Web is now at #4, and moving to #5!

5

Presente e futuro del SW

- **Data integration** comes as one of the SW application areas, very important for large application areas (life sciences, energy sector, eGovernment, financial institutions), as well as everyday applications (eg, reconciliation of calendar data)
- **Life sciences** example:
 - data in different labs...
 - data aimed at scientists, managers, clinical trial participants...
 - large scale public ontologies (genes, proteins, antibodies, ...)
 - different formats (databases, spreadsheets, XML data, XHTML pages)

6

Presente e futuro del SW

- the need has increased for **shared semantics** and a web of data and information derived from it. One major driver has been e-science, i.e. life sciences research demands the integration of diverse and heterogeneous data sets (ontologies) that originate from distinct communities of scientists in separate subfields. Experience suggests that **an incubator community with a pressing technology need is an essential prerequisite for success**. In the original Web, this community was high energy physicists who needed to share large document sets.
- RDF, OWL but also there is the need for rules, in particular the **Rule Interchange Format (RIF)**, an attempt to support and interoperate across a variety of rule-based formats.

7

Presente e futuro del SW

- **Advances in AI** exists: AI researchers have extended logics and modified them to capture **causal, temporal, and probabilistic knowledge**.
- In most cases, we aren't able to look up a URI and have the data returned: the **data exposure revolution has not yet happened**.
- We need to regard such ontologies as **living structures**. Communities and practice will change norms, conceptualizations, and terminologies in complex and sociologically subtle ways, hence the cost of ontology development and maintenance is critical.
 - Some numbers: If we assume that ontology building costs are spread across user communities, the number of ontology engineers required increases as the log of the user community's size. The amount of building time increases as the square of the number of engineers. The consequence is that the effort involved per user in building ontologies for large communities gets very small very quickly.

8

Presente e futuro del SW

- **Not only ontologies, but also folksonomies.** Folksonomies arise when a large number of people are interested in particular information and are encouraged to describe it—or tag it (they may tag selfishly to organize their own content retrieval or altruistically to help others). Rather than a centralized form of classification, users can assign keywords to documents or other information sources. These applications, driven by decentralized communities from the bottom up, are sometimes called **Web 2.0** or **social software**.
- **Folksonomies serve very different purposes from ontologies.** Ontologies are defined through a careful, explicit process that attempts to remove ambiguity.
- Some people perceive ontologies as topdown, somewhat authoritarian constructs unrelated, or only tenuously related, to people's actual practice, then it's understandable that emergent structures like folksonomies begin to seem more attractive

9

Presente e futuro del SW

Substantial research challenges:

- How do we effectively query huge numbers of decentralized information repositories of varying scales?
- How do we align and map between ontologies?
- How do we construct a Semantic Web browser that effectively visualizes and navigates the huge connected RDF graph?
- How do we establish trust and provenance of the content?
- We must not lose sight of the fact that the Web, and indeed many of our most important digital environments, depends fundamentally on certain general assumptions about social behavior
- The critical factors that led to the Web's success will also be important to the success of our Semantic Web enterprise. As we've seen, some of these factors are social; others have their origin in elementary and fundamental design decisions about the Web's architectural principles.
- *Will the Semantic Web be the future or it will be another dream?*

10