What does it mean to be semantic?
(On the role of semantics in the Semantic Web)

Enrico Motta
Knowledge Media Institute
The Open University
United Kingdom

ESWC 2013, Montpellier
Main objective of the talk

- To reflect on the nature of Semantic Web applications and discuss questions such as:
  - What does it mean to do semantics on the web? – i.e., what are the innovative elements that distinguish Semantic Web applications from other systems, in particular conventional DBs and KBS?
  - What makes Semantic Web applications interesting and powerful – i.e., in what ways the use of web semantics provides us with a competitive advantage over other approaches?
Motivation

Why should we reflect on these issues?
The SW as a teenager: time to reflect

• For the past 14 years we have been busy nurturing and protecting our child and telling everybody that she was going to change the world.

• The child is now entering her teens, is reasonably healthy, and we are no longer so worried about her survival. Hence, it is time to worry instead about whether she is growing beautiful and intelligent....

• Also because she may look a bit different from what at least some of us had anticipated....
May 17, 2001

The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

By Tim Berners-Lee, James Hendler and Ora Lassila

The entertainment system was belting out the Beatles' **<blank>** when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other local devices that had a volume control. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web.)

In a few minutes the agent presented them with a plan. Pete didn't like it—University Hospital was all the way across town from Mom's place, and he'd be driving back in the middle of rush hour. He set his own agent to redo the search with stricter preferences about location and time. Lucy's agent, having complete trust in Pete's agent in the context of the present task, automatically assisted by supplying access certificates and shortcuts to the data it had already sorted through.
What’s the big deal?

One could wonder what’s the problem. Why is this an issue? We know what SW applications are about...

- They are about using OWL/RDF
- They are about ontologies
- They are about linked data
- Etc..

Let’s look at a couple of concrete examples from the last ISWC...
Making Sense of Scholarly Data

Semantic Web

- Publications: 22,143
- Citations: 120,704
- Plot authors and publications
- Plot average citations vs authors and publications
- Explore authors
- Explore publications

Semantic Relationships
- Reload Semantic Web relationships
- Deactivate semantic layer

Enrico Motta
Publications normalized by topics

zoom [5y | 10y | 15y | 20y | 30y | All]

1999: 648% Expert System
• Provides a variety of ways to explore scholarly data, e.g.:
  – To identify interesting research trends
    • E.g., which areas are expanding, which ones are shrinking, which interesting ‘migration patterns’ exist between areas, etc.
  – To discover interesting connections and similarities between authors beyond simple co-authorship (e.g., authors who share a similar academic trajectory)
  – To support fine-grained academic expert search

• Exploits a novel algorithm, Klink, which integrates semantics and statistics to automatically mine semantic relations between research topics from very large corpora and then uses these to support knowledge-based exploration
EventMedia Live – 2012 SW Challenge Winner
The more seriously funny experience was at the ISWC demo session. **The two demos that most impressed me were systems for (i) browsing upcoming events (concerts etc.) and (ii) browsing academics and their publications.** Both of these systems were characterized by rich data models and nicely designed user interfaces that delivered valuable information and insights from their chosen domains.
The more seriously funny experience was at the ISWC demo session. **The two demos that most impressed me were systems for (i) browsing upcoming events (concerts etc.) and (ii) browsing academics and their publications.** Both of these systems were characterized by rich data models and nicely designed user interfaces that delivered valuable information and insights from their chosen domains.

**The funny part is that neither of these applications should really be called a “Semantic Web application.”** Someone unaware of the Semantic Web ...would see a traditional data management and visualization problem that they would solve using traditional database tools (SQL) and web APIs. The fact that these tools are storing their data in a triple store instead of a SQL database is irrelevant to the user experience. And the fact that at least one of them is exposing a SPARQL endpoint for querying the data they are managing is good citizenship, helpful to the next project, but not important for this one.
There is more to a SW app than triple stores...

Now, funnily enough I do not necessarily agree with David that Rexplore (or for that matter, EventMedia Live) is not a SW application, however I basically agree with the essence of its comment:

The fact that these tools are storing their data in a triple store instead of a SQL database is irrelevant to the user experience. And the fact that at least one of them is exposing a SPARQL endpoint for querying the data they are managing is good citizenship, helpful to the next project, but not important for this one.

One would certainly think there is more to a SW application than simply using a triple store!
Semantic Web Challenge 2012 - Call for Participation

Submissions are now invited for the 10th Semantic Web Challenge, the premier event for demonstrating practical progress towards achieving the vision of the Semantic Web. The 10th Semantic Web Challenge which will take place at the 11th International Semantic Web Conference in Boston, USA.

As in previous years, the Semantic Web Challenge will consist of two tracks: the Open Track and the Billion Triples Track. The key difference between the two tracks is that the Billion Triples Track requires the participants to make use of the data set that has been crawled from the Web and is provided by the organizers. The Open Track has no such restrictions. The Challenge is open to everyone from industry and academia. The authors of the best applications will be awarded prizes and featured prominently at special sessions during the conference.

Important Dates

- **Friday October 12, 2012, 23:59 CET**: Submissions due
- November 13-15: Semantic Web Challenge takes place at ISWC 2012

Challenge Criteria

The Challenge is defined in terms of minimum requirements and additional desirable features that submissions should exhibit. The criteria for the Semantic Web Challenge 2012 are described on the Criteria page.
Discussing Rexplore with Jim Hendler
Discussing Rexplore with Jim Hendler

Enrico, how come you stopped doing SW research?
Discussing Rexplore with Jim Hendler

Jim, are you kidding? Rexplore follows exactly your philosophy: A little semantics goes a long way
Discussing Rexplore with Jim Hendler

Yes, and no semantics goes even further
There certainly seems to be a clash of intuitions about what SW applications are about and, given the relatively advanced level of maturity of the field, it may be a good idea to try and clarify:

- The nature of SW applications
- The value proposition offered by the use of semantics on the web
Disclaimer/Caveat #1

• I am not a philosopher, so I am not interested in discussing “What does it mean to be semantic” in an abstract way.

• My aim is to give pragmatic, engineering-oriented answers to these questions.
Disclaimer/Caveat #2

• I am not going to attempt to ‘define’ the discipline!

“Disciplines are recognized, not defined”
Herb Simon

• So, it is less a matter of defining SW research than highlighting the key principles of what we are doing to ensure that we capitalize on our strengths and build systems that push the envelope
Disclaimer/Caveat #3

- I will present one particular viewpoint drawing on my background in knowledge-based systems
  - This is just one perspective, others are also possible

- In particular, my **Knowledge-level** (Newell, 1982) perspective focuses on the role of semantics as an enabler of intelligent functionalities
  - Abstracts from specific symbol-level architectures and data structures
  - Takes for granted that there is lots of excellent work on specific infrastructure issues, e.g., ontologies, ontology and data integration, Web KR, query languages, large-scale reasoning, etc....
Frank’s talk at ISWC 2011

- Aims to identify “laws about the information universe”

- Several principles emerge
  - Factual knowledge is a graph
  - Terminology is hierarchical and exhibits low complexity
  - Small Hierarchy/Big Graph
  - Heterogeneity is unavoidable but solvable

- Focus primarily on the data side (SW as massive heterogeneous information space) in contrast with our focus on applications and the role of semantics
Knowledge-Based System Perspective

• My background is knowledge engineering and indeed one of the really cool aspects that attracted me to the SW was that it seemed to me it provided an opportunity to address fundamental AI issues in an exciting new, open, and distributed context, thus moving away from classic, ‘closed’ KBS.
Knowledge Representation Hypothesis in AI

Emphasis on causal and essential relationship between an explicit body of knowledge and the resulting intelligent behaviour

*Brian Smith, 1982*
Task-centric view of Symbolic Intelligent Systems

Body of Knowledge

Intelligent Problem Solving

Classification
Diagnosis
Planning
Design
Scheduling
Monitoring
Etc...
Criteria for recognising intelligent knowledge-based systems

- Explicit representation of knowledge
- Causal link between body of knowledge and problem solving behaviour
- Task-specific criterion
  - Systems tackling tasks complex enough to require knowledge-based problem solving
Intelligent Semantic Web Systems (?)

Intelligent Problem Solving

Body of Web KR
Intelligent Semantic Web Systems (?)

- Several problems with this simple view
  - Semantics on the web is not the same as web semantics
    - Putting a classic, closed KBS on the web does not add anything new
  - Also, in contrast with the AI scenario, it is difficult to come up with a catalogue of tasks defining the range of intelligent activities that SW systems can support in the same way as it was done in the KBS community.
  - SW tasks appear to be more adhoc in nature
    - Analysis by Van Harmelen et al (KCAP 2009) identified the following tasks:
      - Search
      - Browsing/Querying
      - Data Integration
      - Personalization
      - Service Selection
      - Service Composition
      - Semantic Enrichment
Key postulates

Just having a body of Web-KR up on the web to drive some functionality in a manner analogous to traditional, closed KBS is not enough to talk about a ‘real’ SW application

The intuition is that SW applications ought to have at least elements to do with openness to other data sources and/or scale (+ all the other elements that derive from these two)
Toward a New Generation of Semantic Web Applications

Mathieu d'Aquin, Enrico Motta, Marta Sabou, Sofia Angeletou, Laurian Gridinoc, Vanessa Lopez, and Davide Guidi, Open University

Although research on integrating semantics with the Web started almost as soon as the Web was in place, a concrete Semantic Web—that is, a large-scale collection of distributed semantic metadata—emerged only over the past four to five years. The Semantic Web's embryonic nature is reflected in its existing applications. Most of these applications tend to produce and consume their own data, much like traditional knowledge-based applications, rather than actually exploiting the Semantic Web as a large-scale information source.1

These first-generation Semantic Web applications1 typically use a single ontology that supports integration of resources selected at design time. An early influential example from the academic world is CS Aktive Space (http://cs.aktivespace.org). This application combines data about UK computer science research from multiple, heterogeneous sources (such as databases, Web pages, and RDF data) and lets users explore the data through an interactive interface. Corporate Semantic Web application areas include the car industry (such as Renault's system for managing project history), the aeronautical industry (such as Boeing's use of semantic technologies to gather corporate information), and the telecommunication industry (such as British Telecom's system for enhancing digital libraries).

Although corporate Semantic Web applications are typically not considered part of the Web, this may change. Some organizations are already attempting to build and manage their own Semantic Web systems. For example, the Royal Netherlands Meteorological Institute is building a Semantic Web application to extract historical climate data from a number of archives.

It is clear that the Semantic Web's future will depend on how well its current and future challenges—such as scalability, interoperability, and the need for a seamless integration of human and machine processing—are addressed.


A new generation of applications offers insight into the Semantic Web's current and future challenges—as well as the opportunities it might provide for users and developers alike.
Knowledge Engineering in the age of the SW

Systems which exploit large-scale, distributed knowledge to do interesting stuff

Intelligent Behaviour
# Features of NGSW Systems (vs Classic KBS)

<table>
<thead>
<tr>
<th></th>
<th>Classic KBS</th>
<th>NGSW Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontology/Data Selection</strong></td>
<td>Static/Design-time/Closed</td>
<td>Dynamic/Run-time/Open</td>
</tr>
<tr>
<td><strong>Provenance</strong></td>
<td>Centralized</td>
<td>Distributed</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>Small/Medium</td>
<td>Very Large</td>
</tr>
<tr>
<td><strong>World View</strong></td>
<td>Homogeneous</td>
<td>Heterogeneous</td>
</tr>
<tr>
<td><strong>Repr. Schema</strong></td>
<td>Homogeneous</td>
<td>Heterogeneous</td>
</tr>
<tr>
<td><strong>Quality of data</strong></td>
<td>High</td>
<td>Very Variable</td>
</tr>
<tr>
<td><strong>Degree of trust</strong></td>
<td>High</td>
<td>Very Variable</td>
</tr>
</tbody>
</table>
Major shift in the type of reasoning

<table>
<thead>
<tr>
<th>Intelligence</th>
<th>Classic KBS</th>
<th>NGSW Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A function of applying sophisticated logical reasoning to tackle complex real-world tasks</td>
<td>A side-effect of the ability to reason with large scale data and heterogeneous quality and representation, typically integrating (a bit of) logical reasoning with other technologies—e.g., statistics and NLP</td>
</tr>
</tbody>
</table>
NGSW Applications:
Realising the research programme
Watson: A Gateway to the Semantic Web

- Semantic Web Search Engine
- Development Platform for NGSW Applications
- Research Platform for studying the Semantic Web

NG SW Application
Watson-based Applications
Folksonomy Tagspace Enrichment
Using the SW as a source of background knowledge for Ontology Alignment
Discovering alignments across multiple ontologies on the SW

Semantic Web

Strategy

Cholesterol

Steroid

Lipid

Organic Chemical

Cholesterol ⊆ Organic Chemical

Scarlet
Ontology Evolution using Background Knowledge
Relation Discovery Examples

New Concepts

- Contact
- Representative
- Communicator

WordNet

Background Knowledge

- Negotiator

Ontology Concepts

- Person

Scarlet

- Place
- Event

occuredIn
NGSW Applications

- Systems able to exploit distributed semantic information at very large scale, (typically) by combining semantic and statistical methods and tackling the challenges associated with operating in a massively distributed and heterogeneous environment

- Focus on openness and scale with respect to using large amounts of massively distributed SW sources

- Essentially reflect the view of the SW as “Knowledge Medium”

---

**The Next Knowledge Medium**  
Mark Stefik  

*Intelligent Systems Laboratory, Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304*

“An information network with semi-automated services for the generation, distribution, and consumption of knowledge”

---

*AI Magazine 7(1), 1986*
Making Sense of Scholarly Data
Motivations for Rexplore

• Some limitations of current tools
  – Poor support for ‘horizon’ tasks, concerned with making sense of what goes on in research, as opposed to searching for specific authors and publications
  – Poor support for fine-grained academic search – e.g., parametrized with respect to fine-grained expertise, degree of experience, coverage of multiple research areas, geography, academic ranking criteria, etc
  – Poor support for discovering relations and similarities between researchers, beyond obvious co-authorship relations
  – Poor HCI, loss of context during navigation, etc.
  – Etc...
Limitations of existing tools (lack of semantics!)

• In order to find experts in an area, e.g., Semantic Web, I need systems which have some level of understanding that the Semantic Web is a research area

• But understanding what is a research area also means understanding what is not a research area
  – E.g., “case study” is often used as a tag for papers, but it is not actually a research area

• In addition, research areas have a structure.
  – For instance, knowing which are the sub-areas of the SW would allow a system to provide more precise answers about the competence of the relevant researchers and the important trends within the SW
  – It is also the case that multiple labels can be used for the same research area
    • E.g., “Ontology Alignment” and “Ontology Matching” are labels for the same research area
Very high level research fields

Only co-authorship is provided

Old name for IJHCS (changed long ago!)

This journal has nothing to do with my research areas

Case Study is not a research area

KB and KBS are the same research area
Semantic Analysis

• No explicit notion of Research Area, aside from very abstract topics, such as Computer Science, WWW, etc.
• No semantic relations between research areas
• Homogenous treatment of keywords, regardless of whether they denote research areas or other entities (e.g., project, type of study, application area, etc.)
• Only a limited set of static relations between authors
• Limited degree of disambiguation between homonymous authors
• Many co-reference errors concerning authors
  – i.e., situations where distinct authors are actually the same person
• No disambiguation between homonymous areas
  – E.g., linked data!
Identifying and linking research topics with **Klink**

- The Klink algorithm automatically identifies research topics in the literature, mines semantic relations between research topics and uses the resulting data to support knowledge-based exploration, pattern extraction and author clustering in Rexplore.

- Three types of semantic relations are currently mined:
  - \textit{Skos:broaderGeneric} \((A, B)\) – \(A\) is a sub-area of \(B\).
    - E.g., “Semantic Web Services” is a sub-area of “Web Services”
  - \textit{relatedEquivalent} \((A, B)\) – \(A\) and \(B\) are normally used to denote the same research area. This is defined as a sub-property of \textit{skos:related}.
    - E.g., “Ontology Matching” and “Ontology Mapping” denote the same area
  - \textit{contributesTo} \((A, B)\) – The outputs from area \(A\) are relevant to research in area \(B\). In other words, if I am interested in area \(B\), I may also examine papers/authors in area \(A\). This is also defined as a sub-property of \textit{skos:related}.
    - E.g., Research in “Ontology Engineering” contributes to research in “Semantic Web”
About 1500 topics linked by almost 3000 semantic relationships
ACM and other similar classifications

- These exhibit several major limitations:
  - The relations between entries are unclear
    - They are meant to be sub-areas, but for many of them it can be argued that they are not really sub-areas
  - The different types of relationships are not distinguished
  - Rather shallow
    - Most areas we know about are not listed — e.g., only 4 topics are classified under Semantic Web
  - Static, manually defined, hence they get obsolete very quickly
Visualizing research trajectories

Enrico Motta
Publications normalized by topics

- total publications
- Semantic Web
- Knowledge Base
- Artificial Intelligence
- Information Retrieval
- Expert System

Zoom 5y 10y 15y 20y 30y All

% Difference against average topic pub.


Shared Research Trajectories

The authors who are most similar to a particular author with respect to the evolution of their research interests over time.
Comparing publication trends

Semantic Web
broaderGeneric Topics – publication trend

Linked Data  Web Ontology Language

Zoom 5y 10y 15y 20y 30y All

Publications

200
100
0

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
Semantics as analytical engine

• Systems which like Rexplore are able to handle large amounts of distributed web data (typically) by combining semantic, statistical, and computational linguistics methods.

• These systems use semantics as an analytical tool to identify distinctions and give meaning to regularities in the data to support problem-solving, sense-making or decision-making.

• Focus on openness and scale with respect to their ability to impose semantics (thus providing model-theoretical interpretations) on large amounts of massively distributed web data, which are not already formalised using SW standards.
A different class of SW applications (semantics as analytical engine)

Semantics as a medium to impose formal distinctions on large amounts of unstructured information to make explicit the underlying semantic patterns and enable intelligent functionalities

Intelligent Functionalities
Implicit vs explicit semantic patterns

Analytical

Non-Analytical
Conclusions

• Many of us share the intuition that SW applications ought to have elements to do with openness to data sources and scale—plus, optionally, others

• Our “NGSW Applications” research programme investigated the feasibility and the value derived from building applications exploiting large-scale distributed semantic structures

• However, semantics is also an effective tool to help users to make sense of large web data, acting as a device to introduce distinctions and give meaning to structures that may have been generated using other technologies (e.g., statistical methods)
SeaFood disjointWith Meat

SeaFood subClassOf Meat

Vegan subClass Vegetarian