The semantic web(s)

"A web talking to machines"

The goal of the semantic web is to be "A web talking to machines", i.e. in which machines can provide a better help to people because they can take advantage of the content of the Web. Computers should thus be able to exploit the information expressed on the web in a meaningful way. This definition is easily related to what already exists on the web: wrappers for extracting data from regularly structured pages, natural language analysis for extracting web page contents, indexing schemes, syndication facilities for broadcasting identified web resources. Much of this is painful and fragile: the semantic web should make it easier and robust.
The semantic web can also be thought of as an infrastructure for supplying the web with formalised knowledge in addition to its actual informal content. This is obviously contributing to the proposed goal. No consensus exists on how far the formalisation should go: it ranges from precise metadata schemes (like the Dublin core metadata markers) to full-fledged logical representation languages. This exists only for particular applications (SHOE, Ontobroker) and is currently limited to a small subpart of the web.
One of the challenges of the current semantic web developments is the design of a framework in which all these understanding can collaborate.

Applications

Like the web, the semantic web is not an application, it is an infrastructure on which many different applications (like eCommerce) will develop. Characterising the "killer application" of the semantic web will be as hazardous as predicting that of the web ten years ago. The usage does not precede the technology, but it explodes when the technology is available.
There seems to be two approaches to the application of the semantic web:

- Semantic web applications for the organisations such as the development of ontology-based marketplaces for business to business electronic commerce, or the bioinformatic knowledge grid in which biological data and knowledge bases are seamlessly interconnected and computing resources are available.
- Semantic web applications for the masses such as intelligent personal assistant gathering and filtering relevant information and composing it in a coherent picture with regard to the user preferences (the travel assistant scenario).

More generally, knowledge management, personal or corporate, can take advantage of the semantic web. The semantic web will provide value to any semantically annotated resource by facilitating its retrieving when appropriate.
We present below four scenarii, that could be seeding and testbench applications for the semantic web, and consider the technical challenge for achieving major breakthroughs involved in building one of these semantic webs.

Requirements

Identifying the semantic web with a particular technology (search engine, knowledge representation, natural language processing, etc.) or language (XML, RDF, DAML+OIL, etc.) is certainly not accurate.
The realisation of a semantic web will require several layers of developments presented in figure 1. An infrastructure will allow identifying, locating, and transforming resources in a robust and safe way. Languages are necessary for expressing the content of the semantic web;

the semantics of these languages are sanctioned by inferences engines. Resources such as ontologies, transformations, and database must feed these two base layers. The resources of the semantic web are exploited by applications that run on devices.

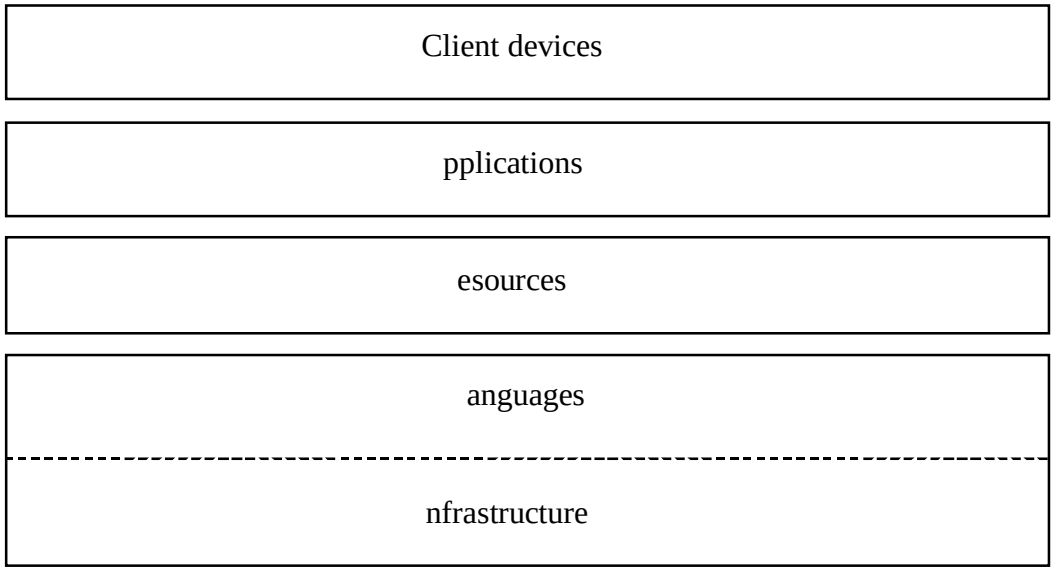

Figure 1. High-level layered view of the semantic web.
This schema roughly corresponds to the working groups that have been set up during the workshop.

Who is involved?

The development of the semantic web will involve efforts in many areas of computer science: web technology, knowledge representation, database, automated theorem proving, knowledge extraction, natural language processing, etc.
Moreover, the semantic web must not be separated of many aspects: personalisation (and thus privacy issues), mobility (and thus reliability issues), publication (and thus security issue). These issues and topics are quite traditional but the semantic web drive them to their extreme because the use of semantically grounded languages make the use of the semantic web more powerful and the threats more acute.
Nevertheless, for the semantic web to happen, it is not just a matter of technology. It involves technology, economics, and social matters. The point is to find the compromise between these domains that could lead to a value adding, appealing, and easy to use semantic web.
All these aspects must be articulated in a delicate alchemy for the semantic web to take off and provide real value to its users. This is why many applications should be encouraged on top of a solid infrastructure so that the fittest raises the utility of the semantic web.

Potential

Expectations are very high for the semantic web because information overload currently reduces the usability of the web and the lack of interoperability between web services is an obstacle to realizing its best promises. The improvements resulting from the development and use of the semantic web in some areas might provide values for both users and providers and result in an increased activities (economic, social, cultural, etc.) like the web did in the 90s.

The semantic web is going to happen. It is going to happen because it is necessary for improving the information infrastructure. It is necessary for commercial companies which wants to better interoperate (either in business to business electronic commerce or in the worldwide enlarged enterprise). It is necessary for the citizen who wants a better service from the “suppliers” and her administration and a more performant protection of her privacy.

When? Now, in ten years or in a century. This mainly depends on what is expected from a semantic web. The semantic web is happening now, the semantic webs are happening now, if one thinks of the many initiatives for marking up resources (syndication, open catalogues, and annotations). Several years ahead if one thinks of agents realising one of the scenarii presented below.

The semantic web will have fully succeeded when no one talks anymore of the semantic web, but simply call it “The web”.

Summary of the recommendations

General recommendations

- For the research, **support world-wide collaboration** between researchers. Obviously because it allows to reach consensus on a global level required for the web (and not at the continental one, see the GSM/CDMA incompatibilities). There is also a **need for non-project focussed funding**. In computer science, the research is too often directed towards prototype building though some funding for producing reports, surveys and studies is necessary.
- For the tools, **encourage open source development** of high quality components **and** non-profit **shelter organisations** for software development (aka Apache). For the web CERN and NSCA have played this role at the beginning. It is possible that this model could be applied to ontologies as well.
- For the applications, support efforts for **building seeding applications** of the semantic web. Several scenarii have been provided in the present report. It is important that these applications are not developed in isolation. We first need a set of existing applications for improving on them.
- For the education, provide support for evangelisation (e.g. teaching material, company "educating").

Languages

- Organizing in a coherent manner the multiplicity of languages of different expressivity and purpose;
- Reconciling the various modelling styles existing (including those from software engineering);
- Developping and articulating the different possible reasoning services;
- Developping tolerant inference and levels of tolerance;
- Investigating the notion of identity in relation with the semantics of languages.

Infrastructure

- Localising and matching resources: identity model;
- Checking proofs and policies: representation of policies, proofs and properties;
- Propagating trust, proofs and rewards: trust model and knowledge level right management;
- Experimenting transformation infrastructure (proof-carrying transformations, processing and composing).

Ontology management

- Acquiring ontologies from primary sources (texts, multimedia, images);
- Supporting reuse and evolution: comparison, merging, versions, and conceptual refinement;
- Theoretical issues about identity;
- Theories and metrics for comparing ontologies;
- Development of strategic ontologies.

Human factors

- Sustained use of technologies in authentic work contexts;
- Growth models for the Semantic Web;
- Next generation application development environments;
- 'Incidental' knowledge capture (→ automatic metadata generation and knowledge base population);
- Understanding the *use* of metadata;
- Coping with 'messy metadata' (from people and machines);
- Supporting collaborative, emergent consensus?
- Maintaining and analyzing history?
- Controlling information overload;
- Imagining 'Lightweight' semantic collaboration.

// here I could imagine gathering the recommendations by themes that arise in several domains (this could also go into the executive summary):

- **Identification and localisation** is an important topic for the semantic web reasoning and computing. This topic involves works on language, infrastructure and ontological premises.
  - Investigating the notion of identity in relation with the semantics of languages.
  - Localising and matching resources: identity model;
  - Theoretical issues about identity;
- **Relationships between semantic web annotations** across languages, styles and modelling. This involves the study of language and language style relationships, the building of articulation theories and the development of safe transformation methods.
  - Organizing in a coherent manner the multiplicity of languages of different expressivity and purpose;
  - Reconciling the various modelling styles existing (including those from software engineering);
  - Developping and articulating the different possible reasoning services;
  - Experimenting transformation infrastructure (proof-carrying transformations, processing and composing).
  - Supporting reuse and evolution: comparison, merging, versions, and conceptual refinement;
  - Theories and metrics for comparing ontologies;
- **Tolerant and safe reasoning** adapted to the web and the evaluation of the accuracy of the result. This involves coping with messy metadata and the open character of the web.
  - Developping tolerant inference and levels of tolerance;
  - Coping with 'messy metadata' (from people and machines);
  - Checking proofs and policies: representation of policies, proofs and properties;
  - Propagating trust, proofs and rewards: trust model and knowledge level right management;
- **Facilitation of the semantic web bootstrap**, by ontology and data acquisition from primary sources, incidental knowledge capture, new application development tools.
  - Acquiring ontologies from primary sources (texts, multimedia, images);
  - Development of strategic ontologies.
  - Sustained use of technologies in authentic work contexts;

- Growth models for the Semantic Web;
- Next generation application development environments;
- 'Incidental' knowledge capture (→ automatic metadata generation and knowledge base population);
- Understanding the *use* of metadata;
- Supporting collaborative, emergent consensus?
- Maintaining and analyzing history?
- Controlling information overload;
- Imagining 'Lightweight' semantic collaboration.

//

// The even shorter version

— **Identification and localisation** is an important topic for the semantic web reasoning and computing. This topic involves works on language, infrastructure and ontological premises.

— **Relationships between semantic web annotations** across languages, styles and modelling. This involves the study of language and language style relationships, the building of articulation theories and the development of safe transformation methods.

— **Tolerant and safe reasoning** adapted to the web and the evaluation of the accuracy of the result. This involves coping with messy metadata and the open character of the web.

— **Facilitation of the semantic web bootstrap**, by ontology and data acquisition from primary sources, incidental knowledge capture, new application development tools.

- //