



IST: Future & Emerging technologies

Report of the EU-NSF strategic workshop on the

Semantic web

Held at Sophia-Antipolis, France, October 3rd-5th, 2001

A proposal made by a workshop of scientists

under the auspices of the Future and Emergent Technologies Unit of the Information Society
DG of the European Commission and the National Science Foundation of the United States of
America

Scientific coordinator and editor: Jérôme Euzenat

Organizational coordinators: Jean-Eric Pin and Remi Ronchaud

Table of Contents

Foreword	5
Executive summary	7
The semantic web(s)	9
<i>“A web talking to machines”</i>	9
<i>Applications</i>	9
<i>Requirements</i>	9
<i>Who is involved?</i>	10
<i>Potential</i>	10
Application scenarii	12
<i>ECommerce and the semantic web</i>	12
<i>Towards a bioinformatics knowledge grid</i>	12
<i>A semantic web of personal universes</i>	15
<i>Knowledge Management</i>	16
Languages	19
<i>Scope of inquiry</i>	19
<i>Major research questions</i>	19
<i>Recommendations for future research</i>	20
Infrastructure	21
<i>Scope of inquiry</i>	21
<i>Major research questions</i>	21
<i>Recommendations for future research</i>	25
Ontology management	26
<i>Scope of inquiry</i>	26
<i>Major research questions</i>	26
<i>Recommendations for future research</i>	28
Human factors	29

<i>Scope of inquiry</i>	29
<i>Major research questions</i>	29
<i>Recommendations for future research</i>	33
Relations with other initiatives	35
<i>Scientific communities</i>	35
<i>Institutional initiatives</i>	35
<i>Other initiatives</i>	36
Summary of the recommendations	37
<i>General recommendations</i>	37
<i>Languages</i>	37
<i>Infrastructure</i>	37
<i>Ontology management</i>	37
<i>Human factors</i>	38
Appendix: resources	40
<i>Web sites</i>	40
<i>Journals and special issues</i>	40
<i>Books</i>	40
<i>Articles</i>	40
<i>Conference and workshops</i>	40
<i>Reports</i>	41
<i>Cited material</i>	41
Appendix: acronyms	42

Foreword

This report is the result of a joint European Union Future Emergent Technology program (EU-FET) and National Science Foundation (NSF) strategic workshop organised by the European Consortium in Informatics and Mathematics (ERCIM). The workshop has been held at Sophia-Antipolis (France), October 3rd-5th, 2001. Jérôme Euzenat, helped by Dieter Fensel and Eric Miller, welcomed 20 European and US researchers from the field of knowledge acquisition and representation, database, worldwide web and man-machine communication.

The semantic web is a web whose content can be processed by computers. It can be thought of as an infrastructure for supplying the web with formalised knowledge in addition to its actual informal content. It aims at complementing the web in areas where formal knowledge can be useful. The workshop was aimed envisioning the future of the “semantic web” emergent research area in order to isolate required breakthrough and to put forward recommendations to the funding bodies.

The workshop was composed of two days of participant presentations on a negotiated topic. These presentations were grouped into four sessions (Languages, Resources and infrastructure, Clients and human interface, The semantic web in application areas). After each session, a general discussion was held in order to isolate topics to be further discussed. On the third day, participant were split into four working groups (Language, Infrastructure, Human-related issues, Ontologies) and elaborated research perspective and agendas for the years to come.

Instead of presenting a summary of each presentation session of the workshop, we have preferred to gather the presentation summary in appendices and focus on core topics that have been discussed by dedicated working groups the last day of the workshop. They are concerned with language, infrastructure, ontologies and human factors.

There are a few application scenarii that have retained the attention of the audience: semantic web for electronic commerce, knowledge management and bioinformatics. Some of these applications could be seeding applications (both test benches and early adopters for semantic web techniques: the bioinformatics community could be for the semantic web what the physics community has been for the web). The report has thus devoted a section to these application scenarii.

For each working group, group reporter has provided the redaction of the report: Jérôme Euzenat for “infrastructure”, Nicola Guarino and Rudi Studer for “ontology management” and Simon Buckingham-Shum for “human factors”. In addition, individuals contributed to the application parts: Carole Goble for “Bioinformatics and the semantic web”, Jérôme Euzenat for “A semantic web of personal universes” and Rudi Studer for “Knowledge Management”. Jérôme Euzenat has coordinated the redaction of the report. The workshop participants were: Michel Biezunski, Simon Buckingham-Shum, Vassilis Christophides, Stefan Decker, Jérôme Euzenat, Dieter Fensel, Carole Goble, Nicola Guarino, Ian Horrocks, Henri Lieberman, Brian McBride, Deborah McGuinness, Eric Miller, Enrico Motta, David Pearce, Hans-Georg Stork, Rudi Studer, Bhavani Thuraisingham, and Frank van Harmelen.

This report has circulated among the workshop participant and been amended. It can be considered as a view of the workshop as a whole.

Thanks (Gio, Heiner, people from ERCIM).

Extensive information about the workshop, including the program, full overhead presentation used at the workshop, the short vita and position statements of participants, can be found at the URL:

<http://www.ercim.org/EU-NSF/semweb.html>

Executive summary

The **semantic web** is a web whose content can be processed by computers. It can be thought of as an infrastructure for supplying the web with formalised knowledge in addition to its actual informal content. It aims at complementing the web in areas where formal knowledge can be useful. When computers will exploit the information expressed on the web in a meaningful way, they will provide a better help to users. Much of this is currently painful and fragile: the semantic web should make it easier and robust. It will be beneficial for the society as a whole: for the economy because it will allow companies to better interoperate and to quickly find the best opportunities. It will benefit citizens because it will support them in their day-to-day work, leisure and interaction with organisation and because it will help them to enforce the degree of control they want (over their personal data, preferences, etc.).

This report is the synthesis from a strategic workshop on the semantic web which has been organised by the European Consortium in Informatics and Mathematics (ERCIM) for the European Union Future Emergent Technology program (EU-FET) and the US National Science Foundation (NSF). The workshop which was held at Sophia-Antipolis (France), October 3rd-5th, 2001, gathered 20 European and US researchers from the field of knowledge acquisition and representation, database, worldwide web and man-machine communication. The participants considered the various aspects of **languages, infrastructure, human-related issues, ontologies** as well as applications and proposed what follows.

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. Instead there can be several potential seeding and test applications for the semantic web: **business to business electronic commerce, bioinformatic knowledge grid, personal semantic assistants**, or more generally **knowledge management**. The development of a few pilot applications could boost the taking off of the semantic web.

The most important topics to be investigated for supporting the semantic web development have been clustered in four large categories:

- **Identification and localisation** is an important topic for semantic web reasoning and computing. This topic involves works on language, infrastructure and ontological premises: identity must be taken into account in the semantics of representation languages and its properties must be verified in ontologies and the infrastructure. The infrastructure must support the localisation of and access to identified resources.
- **Relationships between semantic models** across languages, styles and modelling. This line of work has to think about heterogeneity from the beginning of the semantic web. Because, semantic description of information and knowledge is available, heterogeneity can be dealt with. This involves developing layered and modular representation languages for the semantic web, studying the impact of modelling style on interoperability, providing a transformation infrastructure, articulating and composing services and transformations, supporting reuse and evolution by metrics for comparing models and distributed version models.
- **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

with tolerant computing techniques, providing an infrastructure for implementing safe computing with proven properties, and developing new computational models for trust, proofs and rewards on the web.

- **Facilitation of the semantic web bootstrap**, which is a critical point for the semantic web. This can be achieved by the development of strategic ontologies or libraries of well-crafted ontologies, text mining and ontology learning, studies of the growth model and acceptance factors of the semantic web, incidental knowledge capture, supporting consensus building tools and lightweight collaboration.

Concerning the modalities of research funding, the following recommendations have been made:

- **Support worldwide collaboration** between researchers, because it allows to reach consensus on a global level required for the web (and not at the continental one). **Funding non project-focussed** work is necessary for producing reports, surveys and studies.
- **Encourage open source development** of high quality components **and non-profit shelter organisations** for software development (as Apache).
- Support efforts for **building seeding applications** of the semantic web. We first need a set of existing applications for improving on them.
- Provide support for evangelisation (e.g. teaching material, company “educating”).