Semantic Web

European Commission -
US National Science Foundation

Strategic Research Workshop

Research Challenges and Perspectives of the Semantic Web

Sophia Antipolis, France,
3-5 October 2001

Workshop Report and Recommendations

Research challenges and perspectives of the

Semantic Web

Report of the EU-NSF strategic workshop

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

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

Final report

Scientific coordinator and editor: Jérôme Euzenat

Organisational coordinators: Jean-Eric Pin and Remi Ronchaud

Modified: Friday 11 January 2002

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 Research Consortium in Informatics and Mathematics (ERCIM). The workshop was held in 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 representation and engineering, 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. The workshop was aimed at envisioning the future of the "semantic web" emergent research area in order to identify required breakthroughs and to put forward recommendations to the funding bodies.

The workshop was composed of two days of participant presentations on an agreed 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 identify topics to be further discussed. On the third day, participants were split into four working groups and elaborated research perspectives 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 summaries in appendices and focus on core topics that were discussed by the dedicated working groups on the last day of the workshop. These groups were concerned with Languages and inferences, Infrastructure, Human-related issues, and Ontology.

A few application scenarii have caught 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, a reporter has provided a summary report: Frank van Harmelen for "languages and inferences", Jérôme Euzenat for "infrastructure", Nicola Guarino and Rudi Studer for "Ontologies" and Simon Buckingham Shum for "human factors". In addition, individuals contributed to the application parts: Dieter Fensel for "e-commerce and the semantic web", Carole Goble for "Bioinformatics and the semantic web", Jérôme Euzenat for "A semantic web of personal universes" and Rudi Studer for "Knowledge Management". 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. Jérôme Euzenat has coordinated the production of this report.
This report has circulated among the workshop participant and been amended and can be considered as a view of the workshop as a whole

The editor also thanks Gio Wiederhold and Heiner Stuckenschmidt who provided input to the workshop by answering a questionnaire and the ERCIM team (Jean-Eric Pin, Bruno Le Dantec, Peter Kunz, Rémi Ronchaud) for making this workshop possible.

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 web today enables people to access documents and services on the Internet. Today's methods require human intelligence. The interface to services is represented in web pages written in natural language much must be understood and acted upon by a human. The **semantic web** augments the current web with formalised knowledge and data that can be processed by computers. Some services will mix human readable and structured data so that they can be used by both humans and computers. Others will support only formalised knowledge and will only be used by machines. This will enable:

— computers to assist human users in tasks; the computers can "understand" the data in ways they cannot today,

— the creation of a more open market in information processing and computer services enabling the creation of new applications and services from combinations of existing services.

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 in Sophia-Antipolis (France), October 3rd-5th, 2001, gathered 20 European and US researchers from the field of knowledge representation and engineering, database, worldwide web and man-machine communication. The participants considered the various aspects of **languages and inferences**, **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 electronic commerce) 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 takeoff of the semantic web.

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

— **Identification and localisation** is an important topic for semantic web reasoning, annotating and computing. It amounts to agreeing on how some resources can be identified, how two identifiers can be compared or equated and how web resources can be localised for processing. This involves works in language, infrastructure and ontological areas: identity must be taken into account in the semantics of representation languages, and the assumptions related to object identity must be made explicit in ontologies and manageable by the infrastructure. The infrastructure must support the localisation of and access to identified resources.

— **Relationships between semantic models** across languages and modelling styles must be overcome. Heterogeneity must be considered as an intrinsic feature of the semantic web: no language will be suitable for all purposes, no model will be applicable to all cases and no ontology will cover the infinity of potential applications. Because a 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 assumptions 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** is needed for dealing with the web and the accuracy of the results must be assessed. A variety of reasoning methods will be necessary for different applications (from fetching to theorem proving) and the quality of their required results will vary (from anything-will-do to 100% certified). 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.
— **Facilitating semantic web adoption** is a critical point for the semantic web. It will first depend on the availability of resources and then on the ease of use of the semantic web. This can be supported by the development of foundational ontologies and well-crafted ontology libraries, text mining and ontology and metadata 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 reaching consensus on the world 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 (like Apache).
— Support efforts for **building seeding applications** of the semantic web. We first need a set of existing applications for improving on them.
— Provide **educational support** (e.g. teaching material, company “educating”, starter kits).

Table of Contents

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. The information on the web should thus be expressed in a meaningful way accessible to computers. 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 smart 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. No consensus exists on how far the formalisation should go: it ranges from metadata schemes (like the Dublin core metadata markers) to full-fledged logical representation languages. This exists only for particular applications (e.g. 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, because the full benefit of the semantic web can only be attained when computers relate resources from various sources.

Applications

Like the web, the semantic web is not an application; it is an infrastructure on which many different applications (like electronic commerce) 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 seem 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 test applications for the semantic web, and consider the technical challenge for achieving major breakthroughs involved in building one of these semantic webs.

Requirements

The key requirement for the semantic web is interoperability. If machines have to take advantage of web resources, they must be able to access them and use them. As a consequence, the resources must be open and understandable. The resources should not be hidden in a

proprietary system which only publishes information in a human or tool-oriented format: they must be invokable and published in an open, structured and rich format that let the machines make the best out of them.
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 is sanctioned by inferences engines. Resources such as ontologies, transformations, metadata, and databases must feed these two base layers. The resources of the semantic web are exploited by applications that run on devices.

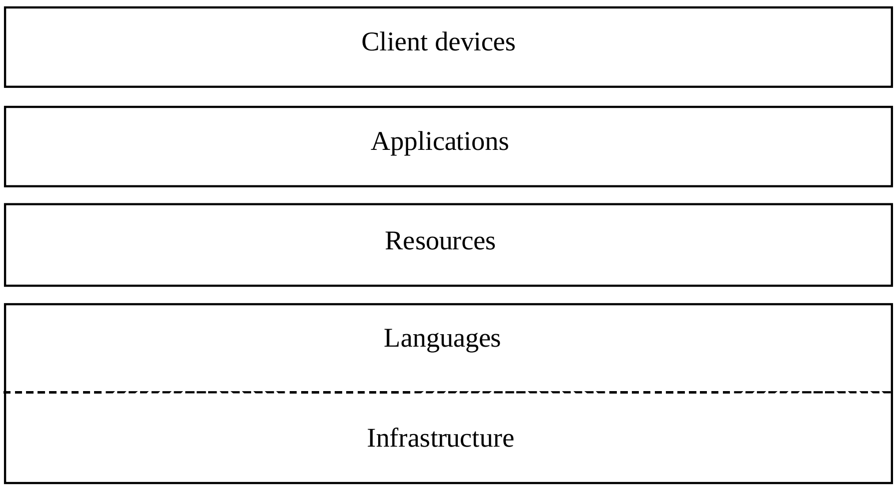

Figure 1. High-level layered view of the semantic web.
This schema roughly corresponds to the working groups that have been set up during the workshop.

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 engineering, computer-supported cooperative work, human-computer interaction, natural language processing, etc.
Moreover, the semantic web must not be separated from 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 computer actions 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. A trade-off must be found 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 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 want to better, interoperate (either in business to business electronic commerce or in the worldwide-enlarged enterprise). It is necessary for the citizens who want a better service from the “suppliers” and their administrations and a more efficient protection of their 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). It is 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 calls it “The web”.

Application scenarii

One can find a number of application scenarii for the semantic web. We present below several application areas in which the semantic web shall be beneficial.
Theses applications could be both test benches and early adopters for semantic web techniques.

E-Commerce and the semantic web

The web has drastically changed the online availability of data and the amount of electronically exchanged information. It has revolutionised access to personal information and knowledge management in large organisations. Also, it has started to change the commercial relationships between suppliers and customers. Around 1% of the overall sales figures in B2C in the US are done on-line. This is still a small fraction but its fast growth looks very likely given that the number of Internet users grew from 100 to 300 million between 1997 and 2000. Forecasts for the dollar value of B2B in the US range between $0,5 and $3 trillion for 2003. Currently, a large fraction of the B2B transactions is still realised by traditional non-Internet networks, such as those conducted over EDI systems. In this traditional paradigm, direct 1-1 connections and mappings are programmed. However, this traditional paradigm nowhere near employs the full power of electronic commerce and it is quite likely that it will soon to be out-ranged by more timely, Internet and web-based types for transactions. Internet-based electronic commerce provides a much higher level of *flexibility* and *openness* that will help to optimise business relationships. Instead of implementing one link to each supplier, a supplier is linked to a large number of potential customers. Therefore, a supplier or customer can choose between a large number of potential customers and can optimise his business relationships.

Bringing electronic commerce to its full potential requires a *Peer-to-Peer* (P2P) approach. Anybody must be able to trade and negotiate with anybody else. However, such an open and flexible electronic commerce has to deal with many obstacles before it becomes reality.

- Mechanised support is needed in finding and comparing vendors and their offers. Currently, nearly all of this work is done manually which seriously hampers the scalability of electronic commerce. Semantic Web Technology can make it a different story: Machine processable semantics of information allows mechanising these tasks.
- Mechanised support in dealing with numerous and heterogeneous data formats. Various “standards” exists on how to describe products and services, product catalogues, and business documents. Ontology technology is required to define such standards better and to map between them. Efficient bridges between different terminologies are essential for openness and scalability.
- Mechanised support in dealing with numerous and heterogeneous business logics. Again, various “standards” exist that define the business logic of a trading partner. A simple example: A trading agents using RosettaNet expects an acknowledgement after sending a purchase order, however, an agent using EDI will never send such an acknowledgement. Mediation is needed to compensate these differences, allowing partners to cooperate properly.

Therefore, applying semantic web technology to bring electronic commerce to its full potential is such a promising activity. Semantic web technology has the potential to solve some of the key obstacles in effective and efficient electronic commerce and electronic commerce is an area with a large economic potential. It is also the natural way to link semantic web technology and web services. Transferring the latter from hype to a working technology requires the solutions of the problems we mentioned above.

Towards e-science knowledge grids

E-Commerce reflects the globalisation of business and the way that commerce is changing. Similarly the way that science is done in biology is changing. ***E-Science*** is the use of electronic resources—instruments, sensors, databases, computational methods, computers—by scientists working collaboratively in large distributed project teams in order to solve scientific problems. Large-scale science, as illustrated by the Human Genome Project, will increasingly be carried out through distributed global collaborations enabled by the Internet that will require access to very large data collections, very large scale computing resources and high performance visualisation. In practice biology has already moved to large interdisciplinary teams distributed throughout the world working together on specific problems. Post-genomics and high throughput experimentation is promising to overwhelm the community with an avalanche of data that needs to be organised and harnessed. The data is often complex, generated through different media, variable in quality, stored in many places, difficult to analyse, often changing and mostly comprised of incomplete data sets. Analysis methods to handle the different types of data are constantly and rapidly evolving. The questions we ask of the data, and the computational analyses to ask them, are more complicated: multiple species rather than single species; whole genome rather than single gene; whole metabolic lifecycle rather than single biological process. The computational power needed to model metabolic pathways or cells will be huge. Consequently, the traditional scientific experimental methods are supplemented with "in silico experiments", for example, the prediction of genes and the metabolic pathways they encode from the genomic DNA of an organism.

In the early 1990s web technology was rapidly taken on board by the biological community as a way of disseminating data and analysis methods that were readily accessible to the wider biology community. The Web enabled individual scientists to answer simple "low volume" questions over large but relatively simple data sets without needing a profound knowledge of computer science. The sharing of data repositories and tool libraries became straightforward. Widespread collaboration was possible even if it was just by publishing a simple web page. However, standard web technology is now straining to meet the needs of biologists. The next step is a much more powerful infrastructure to generally support further growth of e-Science: the Grid. The Grid should enable collaborative groups of scientists to ask complex questions over complex data sets without a profound knowledge of computer science.

What is the Grid?

"The Grid" is a vision of "…flexible, secure, coordinated resource-sharing among dynamic collections of individuals, institutions, and resources—what we refer to as *virtual organisations*". Resource in this context includes computational systems and data storage and specialised experimental facilities. Now the Grid is seen as more as a platform to support coordinated resource sharing and problem solving on a global scale for data-intensive and compute-intensive applications.

The major differences between the Grid and the Web are in the increased computing power available; the increased volume of data that can be handled and the speed with which data can be transferred between nodes on the Grid. The Grid will also provide vast capacity to store and retrieve data from a variety of sources and will allow the presentation of data obtained in the same format, regardless of its source. The main thing is that for the Grid to work it must work seamlessly and transparently: the scientist won't care where calculation is done or where data is actually held, it will just happen. The success of the Grid will be when a bioinformatician, like a database curator, finds it easier to use than not, and a biologist only knows it's there when it breaks.

A Grid-enabled scenario

Let's use a scenario to present the potential of a system that uses the Grid (another one can be found in the EPSRC report, see the "resources" section). Robert is a biologist in a team examining yeast gene expression. Before conducting a microarray experiment he has checked whether any other similar experiment has taken place and if the data was already available. The system recommends a set of parameters for the machine. A sample is logged into a database and labelled. The microarray machine, recognising Robert from the log, sets parameters to those he has used on previous runs. The parameters are recorded with the output results, which are stored in his personal database alongside the image results.

The results are immediately accessible by Robert from his office where he analyses them with a number of specialist statistical computations and a complex interactive time-series visualisation both of which dynamically exploit a number of available computational resources to get better performance. The visualisation is examined collaboratively with a colleague on a remote site. Both scientists attach online personal notes to the results. Several products with up regulated expression look interesting. A search using the SRS database portal identified this gene as encoding a transcription factor. Papers, in free text, quoted to the database entries and extracted online from the Medline digital library reveal that, in certain circumstances, it could control genes related to the yeast gene of interest. The system recommends other scientists who have published work or experiments that are related.

The system inspects Robert's labs various transcriptome databases, and discovers genes that were co-regulated with the original gene also share a target site. This information is added to a yeast database with a link to the workflow of database interrogations and analysis tools that lead to the discovery, including versions of databases, parameter settings, versions of the algorithms and the lab that made the discovery.

Other scientists with appropriate access rights to this database who have run an analysis that included the gene in the last month are automatically notified with this new information. Another scientist incorporates the results into a simulation of a metabolic pathway they are running, using a problem-solving environment. The simulation is monitored by various colleagues around the world, who record both private and public observations. The simulation and its results are added to a public database, and trigger new simulations automatically.

Requirements

This scenario illustrates six major characteristics, and challenges, of the proposed Grid:

An open platform to facilitate interoperability: the Grid plans to be a universal platform bridging heterogeneous systems. The Grid connects all the players in a scientific endeavour: the instruments and sensors; the databases and documents; the machines and networks and the people (e.g. via video). This platform must be scalable, be able to evolve to be future proof and be fault-tolerant, robust, persistent and reliable. Metadata (data about the data) describes the environment, the services available and the ways they can be combined and exploited. Resources are advertised, brokered, monitored and removed.

Large scale distributed information management: the Grid should store and process the huge volumes and diversity of content efficiently. Content can be combined from multiple sources in unpredictable ways depending on the users' needs, and users should be able to discover, transparently access and process relevant content wherever it is located on the Grid. New methods are needed for archiving, mining, manipulating and sharing information derived from multiple sources.

The explicit management of experimental process or "workflow": The "workflows"—how database searches and analysis tools flow together to generate a result—are as important and exchangeable as the results they generate. Recording, and sharing, workflows helps: improve experimental practice by avoiding unnecessary replication of *in silico* experiments (or in vitro experiments for that matter); assist in setting up of equipment or computational processes in appropriate ways; and ensure that conclusions are not drawn that are not fully justified by the techniques used.

Coordinated distributed resource sharing: computationally intensive data analysis and predictive modelling can take advantage of spare resources available on machines connected to the Grid. Resources are discovered, allocated and disbanded dynamically and transparently to the user.

Collaborative science: users will form, maintain and disband communities of resources, use video conferencing and shared collaborative environments to jointly solve problems.

Governance services: a distributed environment on the scale of the Grid requires a number of core services built into its fabric to govern the whole scientific environment: ownership and watermarking; provenance, quality, audit, versioning; authentication, security and confidentiality; change management and propagation, personalisation and configuration and so on.

The Web was originally developed at CERN as a scheme for enabling Physicists to exchange ideas. It worked because the Physics community had a real problem and the computer scientists worked with them to solve their problem, not some other problem.

A semantic web of personal universes

Beside these organisational semantic webs where organisations invest a lot of time for providing the semantic web with data and knowledge (because they expect some income from the semantic web), there is room for personal involvement into the semantic web (because it will provide some income to individuals). Personal involvement is more difficult to obtain from users, but once initiated, it reveals more fruitful.

We present below a scenario involving personally handled data. A different one can be found in the 'scientific american' article (see the "resources" section)

Scenario

During her stay at Honolulu, Clara run into several interesting people with whom she exchanged vCards. When time to rest came in the evening, she had a look at her digital assistant summarising the events of the day and recalling the events to come (and especially her keynote talk of the next day). The assistant popped up a note with a link to a vCard that reads: “This guy’s profile seems to match the position advertisement that Bill put on our intranet. Can I notify Bill’s assistant?” Clara hit the “explain!” button. “I used his company directory for finding his DAML enhanced vita: he’s got the required skills as a statistician who led the data mining group of the database department at Montana U. for the requirement of a researcher who worked on machine learning.”, Clara hit then the “evidence!” button. The assistant started displaying “I checked his affiliation with university of Montana, he is cited several times in their web pages: reasonably trusted; I checked his publication records from publishers’ DAML sources and asked bill assistant a rating of the journals: highly trusted. More details?”. Clara had enough and left her assistant inform Bill’s.

Bill’s and Peter’s assistants arranged a meeting in Paris, just before ISWC in Sardinia. Thanks to Peter assistant knowing he was vegetarian, they avoided a faux pas. Bill was surprised that Peter was able to cope with French (his assistant was not authorised to unveils that he married a woman from Québec). Bill and Peter had a fruitful meeting and Bill will certainly be able to send Peter an offer before he came back to the US.

Before dinner, Peter investigated a point that bothered him: Bill used the term “Service” in an unusual way. He wrote: “Acme computing will run the trust rating service for semanticweb.org” (a sentence from Bill). His assistant found no problem so he hit: “service”, the assistant displayed “service in {database} equivalentTo: infrastructure.”. Peter asked for “metainfo”, which raised “Updated today by negotiating with Bill’s assistant.” Peter again asked for “Arguments!”: “Service in {database} conflicts with service in {web}”. “Explain!” “In operating system and database, the term services covers features like fault-tolerance, cache, security, that we are used to put in the infrastructure. More evidence?”.

Peter was glad he had not to search the whole web for an explanation of this. The two assistants detected the issue and negotiated silently a solution to this problem. He had some time left before getting to the théatre de la ville. His assistant made the miracle to book him a place for a rare show of Anne-Theresa De Keermaeker’s troupe in Paris. It had to resort to a particular web service that it found through a dance-related common interest pool of assistants. The service workflow had been acquired by one of the fanatic’s assistant during his use of the service. He had accepted to share his know-how with his fellow amateurs.

Peter put his assistant to rest (or rather, he notified his assistant that he wanted to rest) and went with a free mind to look for a nice place to eat in the warm spring evening. In fact, he never leaves his assistant choosing the wine for him.

Requirements

The scenario involves a lot of technology: powerful personal assistants, wireless connectivity, knowledge exchange and negotiation protocols. This technology does not call for outstanding research. The natural development of technology will provide them soon and some programs (such as the idea of ambient intelligence) have them as goal.

It also requires a lot of data and a lot of knowledge accessible to computers. This is the semantic web idea. Having this data and knowledge available is certainly a great challenge for the semantic web development.

Aside from these sources, this scenario requires some mechanisms that will help people trust and use these systems. These mechanisms are mainly:

— a model for protecting data from unauthorised use (e.g., “the woman from Québec”, disclosing the service workflow);

— a model for trusting the data (e.g., “trust the university site more than the personal vCard, trust publishers’ journal tables more than personal list of publications, ask for an evaluation of the merit of this information by the person to who it is meant”);
— an ability to filter information from trust and relevance model and to explain the results;
— the ability to negotiate with other agents the ontology content (e.g., the two assistants have exhibited a peer-to-peer ontology synchronisation on the terms at hand).

Knowledge management

Nowadays, knowledge is one of the most crucial success factors for enterprises. Therefore, Knowledge Management (KM) has been identified as a strategically important means for enterprises. Clearly, KM is an interdisciplinary task, including human resource management, enterprise organisation and culture as well as IT technology. However, there is a widespread consensus that IT technology plays an important role as an enabler for many aspects of a KM solution.

In the past, IT technology for knowledge management has focused on the management of knowledge containers using text documents as the main repository and source of knowledge. In the future, semantic web technology, especially ontologies and machine-processable relational metadata, pave the way to KM solutions that are based on semantically related knowledge pieces of different granularity: Ontologies define a shared conceptualisation of the application domain at hand and provide the basis for defining metadata that have a precisely defined semantics and are therefore machine-processable. Although first KM approaches and solutions have shown the benefits of ontologies and related methods there still exists a large number of open research issues that have to be addressed in order to make semantic web technologies a complete success for KM solutions:

Industrial KM applications have to avoid any kind of overhead as far as possible. Therefore, a **seamless integration** of knowledge creation, e.g. content and metadata specification, and knowledge access, e.g. querying or browsing, into the working environment is required. Strategies and methods are needed that support the creation of knowledge as side effects of activities that are carried out anyway. This requires means for **emergent semantics**, e.g. through ontology learning, which reduces the overhead of building-up and maintaining ontologies.

Access to, as well as presentation of, knowledge has to be **context-dependent**. Since the context is set-up by the current business task and thus by the business process being handled, a tight integration of business process management and knowledge management is required. KM approaches being able to manage knowledge pieces provide a promising starting point for smart push services that will proactively deliver relevant knowledge for carrying out the task at hand.

Contextualisation has to be supplemented by **personalization**. Taking into account the experience of the user and his or her personal needs is a prerequisite, on the one hand, for avoiding information overload and, on the other hand, for delivering knowledge on the right level of granularity.

The development of knowledge portals serving the needs of companies or communities is still a more or less manual process. Ontologies and related metadata provide a promising conceptual basis for generating parts of such knowledge portals. Obviously, conceptual models of the domain, the users and the tasks are needed among others. **Generation of knowledge portals** has to be supplemented with the (semi-)automatic evolution of portals. Since business environments and strategies change rather rapidly, KM portals have to be kept up-to-date in this fast changing environment. Evolution of portals also includes the aspect of "forgetting" outdated knowledge.

KM solutions will be based on a combination of intranet-based and mobile functions in the very near future. Semantic web technologies are a promising approach to meet the needs of the mobile environments, like e.g. location-aware personalization and adaptation of the presentation to the specific needs of mobile devices, i.e. the presentation of the required information at an appropriate level of granularity. In essence, employees should have access to the KM application **anywhere** and **anytime**.

Peer-to-Peer computing combined with semantic web technology will be an interesting path to get rid of the more centralised KM solutions that are currently implied by ontology-based solutions. P2P scenarii open up the way to derive consensual conceptualisations among employees within an enterprise in a bottom-up manner.

Virtual organisations become more and more important in business scenarios that are characterised by decentralisation and globalisation. Obviously, semantic interoperability between different knowledge sources as well as trust is a must in inter-organisational KM applications.

The integration of KM applications, e.g. skill management, with **e-learning** is an important field enabling a lot of synergy between these two areas. KM solutions and e-learning have to be integrated from both an organisational and an IT point of view. Clearly, interoperability and/or integration of (metadata) standards are needed to realise such integration.

Knowledge management is obviously a very promising area for exploiting semantic web technology. Document-based KM solutions have already reached their limits, whereas semantic technologies open the way to meet the KM requirements of the future.

Languages and inferences

Scope of inquiry

As stated in the introduction, "languages are necessary for expressing the content of the semantic web", and the semantics of these languages are exploited by inference engines.
In the young history of the semantic web, research and development on languages to support the semantic web has been perhaps one of the most active areas so far. Results in this area have been leading the other areas required for the semantic web, and this is indeed appropriate: the presence of well-defined languages is a necessary prerequisite for the interoperability that is the key issue of the semantic web.
This means that a first proposal for a set of semantic web languages is in place (XML, RDF, RDF Schema, DAML+OIL, the latter being subject to standardisation by a W3C working group). However, this is only a very first step.
The scope of the work on "languages for the semantic web" is much more than deciding on a standardised syntax alone. Besides a standardised syntax, a precise semantics is also required. Because the languages are intended for interoperability between machines, we cannot rely on an implicit shared understanding of terms between humans with a shared linguistic and cultural background. Instead, the semantics of these languages must be formally defined. These semantics can be defined in multiple forms: current proposals for semantic web languages have used a model-theoretic and axiomatic form for specifying the semantics. Other forms are possible, such as operational semantics.

Major research questions

Here we identify a number of major research questions which have not been solved yet by the current efforts in this area, and which need urgent attention.

Multiplicity of languages

Different languages apply to different situations. Some applications require expressive languages with expensive computational costs, while others need simple languages. Figure 2 displays various languages that could be used in the semantic web. It is important that application designers are guided in the choice of language that is most appropriate to their task. Research should be carried out on the following lines:

— Designing languages that stack easily or at least that can be combined and compared easily. The design of DAML+OIL has shown this path. This stacking needs to apply not only to syntax, but also to semantics, inference procedures, partial interpretability, etc.
— Make explicit the relations between the languages and which needs they can fulfil, so that application developers can choose the appropriate language.

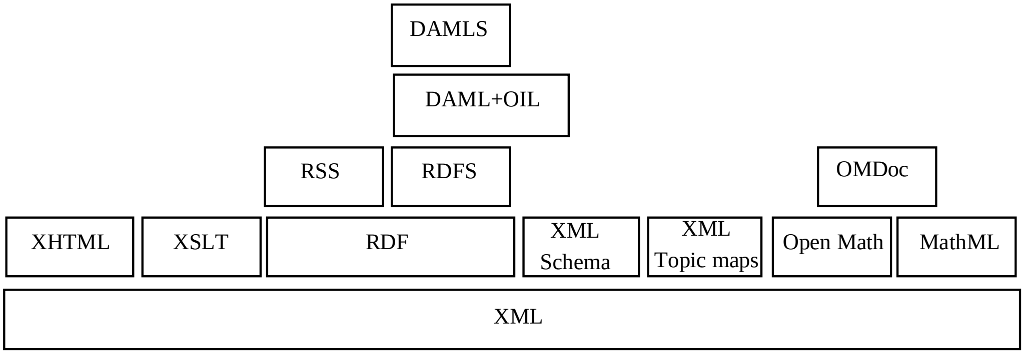

Figure 2. A set of languages that are available for implementing the semantic web.

The language-stack (such as the one depicted in figure 2) should enable both a low step-in cost (possible to participate at a low cost with limited functionality) and a high ceiling (possible to expand into powerful functionality at increased costs).
Somewhat in contrast to the above, the current situation suffers from rather an explosion of languages. The XML language alone is really a family of standards (XML, XSL, XSLT, XPath, XPointer, XLink, XQuery, etc), and a similar explosion is now threatening to happen at every subsequent layer (RDF, DAML+OIL). At some point in the future, a shakeout and reconciliation of many of these separate languages will be required.

Reconciling different modelling styles

Different communities in different situations adopt different knowledge modelling styles: axioms (from logic), objects (from software engineering), constraints (from artificial intelligence), view-queries (from databases). It is important to know how to combine these modelling styles and how to implement one style into another. This is partly a technical process (requiring translations between the different modelling paradigms), but it is also large a social process, where different communities must be educated about the benefits and limitations of their own modelling styles, and those from others.

Different reasoning services

As with languages, different reasoning services are required by different applications. These can vary in substance or quality requirement. Examples of different reasoning services are:

- querying consequences of a domain description,
- checking for consistency,
- matching between two separate descriptions,
- determining the degree of similarity between descriptions,
- detecting cycles and other possible anomalies,
- classifying instances in a given hierarchy,
- etc.

All these different reasoning service can be executed with different requirements on their quality and cost:

- exact and complete reasoning (only feasible for limited applications, or safety critical ones),

- approximate inferences which give incomplete answers in a much shorter time than it would take to compute complete answers (these are important for large scale applications, or applications under time constraints),
- anytime inference algorithms which can be interrupted at "any time", giving the best available answer at that time (in this case users can make their own trade-off between the time they want to wait and the quality of the answer they want to have),
- resource-bounded reasoning, where the quality of the answer is determined by the amount of resources that are allocated to the computation in advance.

Traces of much relevant work on all these issues can be found in artificial intelligence and software engineering but not in a form that they are already practically applicable to semantic web applications.

Identity

When addressing and reasoning about objects and individuals that occur on the web and in the world, there is of course a need to identify them. This requires first of giving all individuals non-ambiguous names. The current web has made a large step by introducing a globally unique naming scheme through uniform resource identifiers (URI) whose more prominent instantiation are uniform resource locators (URL) used on the web. However, besides giving unique names, it must also be possible to identify when two separate names actually refer to the same object in the world. We must be able to state such identities and use them in further processing. Major questions that play a role here are:

- Formal semantics of URI identity (how a URI must be interpreted);
- Scope of the equality statements (i.e., anything more subtle than simply world wide);
- Rigid designators and temporal identity (i.e., names that refer to the same role although fulfilled by different entities, such as "the president of the United States").

Modularity

The semantic web can only get off the ground through large-scale re-use of resources. Re-use of ontologies and service descriptions must provide leverage for the growth of the semantic web. However, current re-use is only possible through a complete import of somebody else's ontology. Even if one would want to reuse only a tiny fraction of some large ontology, currently one is obliged to import the entire ontology, with obvious drawback for maintenance and computational efficiency. Much more sophisticated modularity concepts have been investigated in software engineering, both for data-structures and for program-constructs. Similar mechanism must be developed for the semantic web as a matter of urgency.

Recommendations for future research

By its very nature, the work on standardising languages for the semantic web is not only a technical but also a social process.

It is important to emphasise cross-Atlantic collaboration, and even expanding this into collaboration with Asia and Australia. The current funding models make it difficult in practice to get funding for projects with partners disseminated across America, Europe, Japan or Australia. This is a severe limiting factor on the work on establishing world-wide standards.

Explicit funding must also be available for the required follow-up processes after a standard has been defined. These include the development of teaching material, PR activities, training at both academic, technical and management levels, dissemination workshops, etc.

Many of these activities do not lend themselves very well for to a project-oriented funding scheme, and other models should be considered.

Infrastructure

Scope of inquiry

The web is a combination of an infrastructure (provided by the IP and HTTP protocols) and a language (HTML). A currently open question concerns the need for a specific infrastructure for the semantic web that could be provided by modifying the actual web protocols or by other specific protocols (or conservative extensions of the actual protocol).
The infrastructure does not have to provide ontologies, transformations and inference engines repositories. But it would be better to have a standard way to approach these repositories on the semantic web. The design of an infrastructure has to answer to a set of questions that are considered below:

- How should semantic web resources be identified and located?
- What computation is carried out to support applications?
- Can interoperability be supported by a transformation infrastructure?
- What protection of knowledge must be offered?
- How to guarantee safe computation?
- How computers can choose which knowledge source to trust?

Our initial scope of investigation was named “infrastructure and services”. We found that the term “service” was used, in the database and system field, for what several participants call “infrastructure” and that the word “service” was used in the web field as a synonym for a “server” in the client-server philosophy. In the semantic web, the word “service” has been extended to any application on the web (just like the common life term “service”).
So the focus of the discussion below in called “infrastructure” but it covers what is usually called service in other fields (e.g., transactions, security, authentication). We will talk below of *the* semantic web infrastructure, even if it can be disconnected.

It can be remarked that the questions considered here are questions that can be raised about the current web. So what new does the semantic web brings to these questions? The answer is inference. As soon as the web has gathered some kind of semantics that is accounted for by the computers, they must draw the inferences that are sanctioned by the semantics.
Built-in semantics considerations shed new light on the considered questions, because a syntactic check for confidentiality, for instance, is not enough. The computer should not just check that a classified piece of information is not communicated, but that it is not deducible! What was before devoted to the classifying authority can be left to the computer because it can access the semantics.
The questions now become very complex ones that will require strong involvement of several research communities in order to progress significantly. However, such progress should be invaluable, because this task once done by computers should be better handled.

Major research questions

Localisation and identification

A point strongly debated in the semantic web arena is that of identification of the concepts or objects considered. In the current web, URIs identifies resources. While lack of identity reference is not problematic on the current web, it becomes problematic in a formal web (where one assert that there is only one planet at one position and has to reconcile the morning and the evening stars).

The problems here can sound philosophical ones, but they have a concrete impact on the building of the semantic web. The debated issues are:

— How to establish object identity: the fact that the URI structure is strongly decentralised (anyone can create an URI) means that there can be several URIs for the same resource. This is not really a problem when locating a resource, but becomes a problem when reasoning about the objects (especially with cardinality).
— URIs are often used as proxies. They are locators for a web site (that of "W3C"), designators for a real world entity (the World Wide Web consortium), namespace for programs (developed at W3C)… It is not easy to determine from context what the URI identifies and this is really a problem both for locating and reasoning. In practice there is negotiation when accessing but this negotiation does not occur, so far, when reasoning.
— The URIs are used in some recent works for identifying "real-world" objects (on the web). Since no one can "access" these real-world objects, all the features provided by accessibility (negotiation) are lost. Due to the decentralised nature of URI definition, there is no rule for identification (not one URI for identifying "W3C"). It will be very useful to have a normalised way to address that object without having to resort on a centralised resource. If a solution is found for that problem, there is no limitation to its application to abstract notions (e.g., to the concept of limit).
— The web infrastructure is based on a very rustic naming model. However, names are not tied to concepts and there can be several parallel models. Once the semantic web will not be restricted to web resources but must encompass real-world ones, it would be useful to disconnect concept identifiers from concept names. People could then call identifiable concepts with different names depending on context (company, language, etc.). The naming model of Topic Maps allows this. There is a need to adapt it to the semantic web languages.
— Knowing that web resources exist is not enough for processing: they must be located and available. There is a need for getting web resources: "where can I find a knowledge base about France", "what can book me a flight to Sophia-Antipolis" are meta-questions that the agents will have to ask. Getting the answer necessitates kinds of registry. There can be peer-to-peer connections, centralised registry (UDDI), and multiple registries, search services or combination of these. The infrastructure must support this (and the agent must at least know a search service that can answer its question and will scale).

The infrastructure of the semantic web will have to answer to these questions in order to define a well-grounded model.

Processing model

The semantic web could easily consume a large amount of resources for querying the web, computing consequences and storing them. It is thus necessary to take into account from the architectural point of view several requirements that will help to support this:

— Storage, caching, optimisation, query…
— Distribution of the computation on many computers if necessary.
— Transactions: web services require that several actions are performed in a unique transaction.

These actions traditionally resort to system and are largely considered by work on grid computing. However, the semantic web infrastructure must perform some inference for supporting applications in a more efficient way. For instance, caching could depend on the content; distribution can depend on the use of the inferences that are performed. The question is: How much must be left to applications and how much must be provided by the infrastructure.

This is the same kind of dilemma as the one that arose in multi-agent systems with regard to agent communication languages: how much semantics must be put in the content and how much must be put in the performatives. The agent community has found a fragile equilibrium on that topic. The same should be achieved in the context of the semantic web.

Semantic transformations

The knowledge will be available on the web under many different forms (languages) and under different modelling (axiomatisation). Moreover, some applications will need to adapt the knowledge found on the web. Adaptation is already at work for the current web. The current web solves that problem by writing wrappers or guesses, which are always written in an adhoc manner. It is expected that the semantic web will help building these tools because it will provide the semantic-based answer to a query instead of a structure-based answer.

It is necessary, for taking advantage of the available knowledge to be able to transform it and import it under particular processing environments (which require different granularities or extent). This task is better taken into account within the infrastructure than individually in each context. There are several tools that the infrastructure can provide.

First, the crucial need is to be able to merge ontologies for integrating data. This can be achieved through transformations and articulation theories (which express a theory in terms of another). The infrastructure could provide a way to find these resources (i.e. be able to identify them from a description of the problem) and to assemble them. This could be achieved through the development of transformation services (mediators) and libraries of transformations.

However not all transformations are suited to every task. The transformations can be characterised by the properties they satisfy (like preserving meaning, weakening structure, or filtering information). When a transformation is needed for a target, it is required to satisfy such particular properties. This model is general enough for being applicable to simple extraction from databases as well as complex ontology merging with change of granularity and languages. A first research effort should provide characterisation of transformations in terms of useful properties and providing general schemes for expressing these transformations and properties in languages that the machine can understand.

These properties cannot be taken for granted as soon as they are assigned to a transformation. A suspicious program might want to have more grounded reasons to use a transformation. For that purpose, it will need a proof that the transformation satisfies the advertised property. It will then be able to check that proof and process the transformation.

Let’s take the example of ontology versioning. It is very dangerous to use a knowledge base with a new version of the ontology. However if the ontology maintainers provide a transformation from old to new versions, it becomes possible to use an old knowledge base once transformed with the new ontology (and other new resources).

So, the infrastructure could take in charge the localisation of transformations, their properties and their proofs, the checking of the proofs and the composition of new transformations. Research is of course needed on transformation composition, proof-carrying transformations, proof languages and proof-checkers.

Knowledge protection

The semantic web will use deduction capabilities to derive far fetching conclusions about information not explicitly provided by people. This will help providing better services to users. This strength of the semantic web can be a threat for privacy. People will not use the semantic web if they think that it can break their privacy and providers will not develop for the semantic web if people do not use it!
This problem can be solved if service providers express their policy (what do they do with data and knowledge that are communicated to them) and users their preferences (what do they want to communicate at what condition). The W3C has supported the development of the P3P and CC/PP languages just for that purpose. Yet these kinds of languages are based on specific expressions related to the current use of web sites. The expressive power brought by the semantic web should enable to express very precisely policies and preferences and its deductive power should ease the deduction of consequences of particular agreements.
A protocol for matching policies and preferences is needed for guaranteeing that users can communicate data with the best possible insurance. It is needed for service providers to reach the customer base they deserve.

Spreading knowledge all over the web is useful because it can help solve problems and provide services. However, some actors might want to protect knowledge for various reasons (mainly because it provides them a competitive advantage). Some others might want to distribute it on a commercial basis or want their contribution to be acknowledged. There is a need for access control and right management. There exist general techniques for achieving this, but, in the context of a semantic web, knowledge can be more difficult to protect because what is not explicitly stated can be deduced. Research is needed in efficiently protecting knowledge: disclosing some knowledge and being certain that some other knowledge cannot be deduced from it.

Robustness, safety, and trust

In the human-browsable web, clicking on a dangling link can only trigger a 404 error and the browsing human, while upset is able to recover and to look somewhere else (let aside correcting the ill-formed URLs). In a machine-readable web, many characteristics of the searched resource can alter the performance of the machine: version, trust level, language, mathematical properties, cost of use, authentication. Although an altered human-readable web page is still readable by human, any of the reason to not meet the expectation of a program is a cause for failure. Bus error is the less hurting solution because we will then know that something went wrong. More malicious is the absence of clash when using bogus knowledge…
The infrastructure requires being able to provide quality of service. While this is a well-known term, it will now depend on other parameters than those usually measured. The methods for achieving this quality of service are similar to those actually used (certified methods based on proofs and authority, robust methods based on statistic and rough computation). They will have to be enhanced with knowledge of the parameters that can affect the task (for instance, if I spend all the money available for this task on search, I will not have anymore money on deducing).

The semantic web being processed mainly by machines cannot depend only on the user for trusting or not the site content. It will have to decide if a particular resource found on the web

is trustworthy before using it. Trustworthiness concerns many aspects of the resource: the fiability of a program, the correctness of the content, etc.
Trust can be established in many ways including proving that the resource satisfies the required property, applying trust propagation mechanisms based either on indices (how many sites refer to this one, how many papers cite this one), trust assertions from other trusted parties (non formal arguments à la epinions.com or formal argumentation systems), semantic arguments (involving reasoning about the issue).
As an example, let's come back to the Ontology version problem. If the maintainers publish a checkable proof of the transformation between two versions of their ontology, it becomes possible to trust the transformation. Moreover, if we have a certificate from a trusted party that the initial ontology satisfies properties that we expect, then the program will be able to derive that it can trust the transformed knowledge and use it for its own purposes.
However, trust is not an all-or-nothing property: one can be more or less confident in an assertion. So the trust representations and models have to be context dependent. For achieving this, the semantic web will have to rely on a trust propagation model that helps evaluate the reasons for trusting and the way these reasons can be combined.

Recommendations for future research

From an infrastructure viewpoint, the key point is to design an infrastructure that:

- scales,
- does not commit to a specific model (of trust, rights, properties, etc.);
- proves to be usable.

To that respects we identified the future research directions at three levels.
Work on models:

- trust,
- identity,
- properties.

Work on tools:

- for localising and matching resources,
- for checking proofs and policies,
- for propagating trust, proofs and rewards.

Experimentation:

- proof-carrying transformations (processing and composing),
- trust propagation,
- right management,
- robustness (to openness, fuzziness, granularity mismatches, etc.).

An additional specific recommendation of this working group is the support of open source development of high quality components and of a non-profit shelter organisation for this software.

Ontologies

Scope of inquiry

An ontology is a formal conceptualisation of a domain that is usable by a computer. This involves an explicit description of the assumptions regarding both the domain structure and the terms used to describe the domain. Ontologies are central to the semantic web because they allow applications to agree on the terms that they use when communicating. They are a key factor for enabling interoperability in the semantic web.
Ontologies will have to grow and develop with the semantic web and this needs support. Several open questions related to the development of ontologies are considered here.

Major research questions

The research questions have been divided into ontology development methodology, major theoretical issues, needs for strategic ontologies, and requirements for better tools.

Ontology lifecycle

Ontologies aim at modelling and structuring domain knowledge that provides a commonly agreed understanding of a domain, which may be reused and shared across applications and groups of people. As for any professional software artefact, their employment in real-life applications requires comprehensive engineering methodologies that need to cover all the aspects of an ontology's *lifecycle*. Thus, the lifecycle contains several phases, ranging from requirements analysis and initial ontology design to conceptual refinement, evaluation and, last but not least, evolution (see Figure 3).

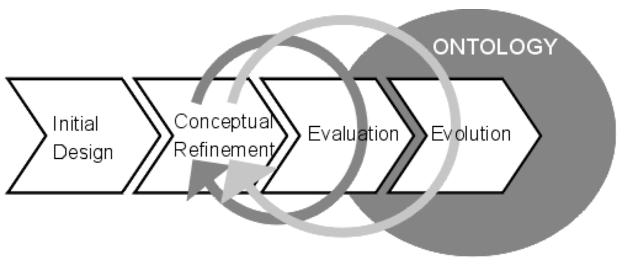

Figure 3: Ontology lifecycle

The engineering of ontologies starts with a **requirements analysis**. This is a crucial step since *the* ontology for a domain does not exist, rather various ontologies might be developed depending on the specific requirements of the applications and users. Obviously, methods related to requirements engineering, e.g. competency questions, may be applied for that purpose as well, however specific aspects that are to be addressed are requirements like richness, i.e. the question of how detailed the domain should be modelled, coverage, i.e. the breadth of the ontology, cognitive adequacy, i.e. what are structures and concepts that fit to the needs of prospective users. A key issue here is the balance between specificity and reusability: this seems to be especially relevant in the semantic web perspective, since the users of an ontology may not be known in advance.
Based on the requirements specification the **initial ontology design** may be approached. An important topic here is the analysis of various kinds of knowledge sources, i.e. free texts, semi-

structured sources and structured sources like databases. Here ontology learning, i.e. the combination of linguistic analysis, information extraction, statistical techniques and machine learning is a very challenging, yet promising area of research. When exploiting multimedia sources, text analysis techniques have to be supplemented with picture and video analysis techniques. During initial design, too, reusability is a specifically tough aspect that has to be addressed.

The output of the initial design step, i.e. some preliminary conceptual structures, may directly feed into the next step, viz. the **conceptual refinement**. To support efficient and effective refinement, semi-automatic means for restructuring and enriching data are needed that exploit already existing conceptual sources, like thesauri, database schemata, upper-level ontologies or further domain specific ontologies; this means that ontology learning is an important method for conceptual refinement as well. Formal ontological principles guiding the definition, modification and enrichment of conceptual structures have to be specified in a way that they are applicable in practical settings. Reuse of parts of ontologies asks for appropriate structuring principles, i.e. modularization concepts that support composition links between ontology modules that are more sophisticated than just “import”.

In many cases, this refinement process needs to be supported by tools and methods for consensus building and for comparing and analysing the implications of different conceptual views. Unfortunately, there is a lack of tools and techniques in this area. Furthermore, in a lot of application scenarios the alignment of different ontologies might be much more appropriate than the integration of different ontologies into a single target ontology. Again, semi-automatic means that handle syntactic and semantic aspects are of high importance in order to lower the overhead of that alignment process. In dynamic applications scenarios, like e.g. mobile applications, “alignment on the fly” is a promising, yet not well understood approach.

During refinement and evolution (see below) of ontologies aspects of sloppiness have to be taken into account as well. Whereas existing methods aim at delivering precisely defined ontologies, application scenarios, e.g. peer-to-peer based applications, might be confident with ontologies and alignment of ontologies that are to some extent sloppy. However, there do not exist methods for addressing these sloppiness aspects.

The **evaluation** phase needs ways of formally evaluating an ontology with respect to its requirements. Many different dimensions can be assessed by evaluation: psychological adequacy, usage evaluation, deepness, agreement between ontologies, etc. Moreover, the evaluation can be performed in situation (i.e. in a particular application) or in abstracto (in an ideally neutral way). One may distinguish between different kinds of evaluation scenarii, e.g. the evaluation of the usage of the ontology in different applications and by different users, or, experiments for checking the psychological adequacy of certain ontological choices. Theoretical means, i.e. metrics, are needed for comparing different ontologies according to their expressivity, accuracy, domain richness and cognitive adequacy, to name but a few aspects.

In the real world, things are changing, and so do ontologies. To handle the **evolution** and maintenance of ontologies, one needs to explore and to formalise the kinds of relationships that may arise between different ontology versions. A productive versioning management includes also the recognition of change needs by checking the usage patterns of ontologies and by semi-automatically exploring newly available sources with respect to not yet modelled concepts and relations. Usage patterns may also indicate that some ontology parts have become outdated. Last but not least, these changes need somehow to be propagated to the metadata produced according to each ontology, i.e. to the annotations, in order to guarantee consistent and up-to-date metadata annotations.

Theoretical issues

Several theoretical issues concerning ontology development are still open, and require to be addressed for an effective use of ontologies in the semantic web. A first need concerns the definition of **formal ontological principles**, general criteria to be used to impose a discipline on the ontology development process and make possible to compare different ontological criteria on a rigorous and neutral basis. Related to this, there is the need for developing libraries of **foundational ontologies**, describing (and relating together) basic modelling choices related to general issues like the ontology of time, space, events, and objects. What we envision, in this respect, is not so much a monolithic top-level ontology to be used by all applications (an unrealistic and naïve perspective), but rather the possibility of making explicit the nature and the content of ontological choices simply by committing to specific foundational modules, whose characteristics and implications are already well understood (although not necessarily agreed upon) by a large community. In this way, the ontological commitment of a certain theory could be "certified" in a way similar to what happens for other engineering artefacts.
More generally, the challenge for the next years is the development of a **unified conceptual analysis methodology** which considers all the methodological efforts that have been developed in computer science (for database systems, object-oriented systems, knowledge-based systems, etc.) in the light of common ontological principles.
Finally, specific issues especially relevant for the semantic web are related to the notion of *identity* (already mentioned above), and to the notions of *relevance* and *granularity*, necessary to cope with multiple conceptualisations.

Strategic domains for ontology development

Besides the general foundational ontologies discussed above, more specific ontologies need to be developed to support the semantic web activities. The main ontologies whose development must be supported are:

— Ontology of information and information processing: this involves the semantic web infrastructure, as well as the relationships between such infrastructure and the real world (semiotic relations, among other things),
— Ontology of social entities: this mainly involves the ontology of people, organisations, and laws,
— Ontology of social co-operation and interaction: this involves the various ways of interactions among human and artificial agents and organisations.

Requirements for tools and methods

Building and especially maintaining ontologies will require tools and methodologies. Some requirements on these are listed below:

— Visualisation of complex ontologies is an important issue because it is easy to be overwhelmed by the complexity of ontologies. Moreover, tools should be able to provide comparative displays of different ontologies (or ontology versions). Specific tools based on the objective semantics and the perceived meaning should be developed.

- Cooperative development environments should support consensus building and the comparison of different conceptual views or degree of consensus.
- Managing argumentation and design rationale can be thought of as part of the above. Design rationales should be attached to ontologies so that the application developers (and, maybe, the applications) can use them to choose the adequate ontology.
- Support for modularity and transformation of ontologies by recording the links between modules and proposing transformations to export part of ontologies in a context with specific requirements. This can be based on the results of the "inter-ontology linking" and "evolution" parts above.

Recommendations for future research

- Ontology acquisition from multiple primary sources (texts, multimedia, images) by means of learning techniques,
- Ontology comparison, merging, versioning, conceptual refinement, and evaluation,
- Formal ontological principles as a basis for a unified conceptual modelling methodology
- Theoretical issues about identity,
- Libraries of foundational ontologies simplifying the documentation of basic ontological choices,
- Development of specific ontologies in strategic domains: information and information processing, social entities, social co-operation and interaction.

Human factors

Scope of inquiry

We use the general term **Human Interfaces** to cover:

- **User Interfaces** to single- and multi-user applications: the software and hardware with which people interact to communicate with a semantic web technology.
- **Community/Organisational Interfaces—**the broader interaction that organisational and other groups have with semantic web applications: the factors that make sustained adoption of such technologies more likely, and the impact that it has on their work practices, and their perceptions of the technology.

Major research questions

Sustained use of technologies in authentic work contexts

The semantic web, like so many other technology research areas, is strong on innovation (e.g. small scale, short-term prototypes), and weak on longer-term evaluation. Whilst this is a function of the immaturity of many systems, as the infrastructure stabilises a sign of its maturity will be cases demonstrating *sustained use* of technologies in authentic work contexts, in order to understand how the *tools and user communities co-evolve* into an effective work system.

This should be a research priority, and implies the need to draw in researchers from communities such as human-computer interaction, computer-supported collaborative work, and social informatics.

Growth models for the semantic web

The web exploded so rapidly in size because the cost-benefit trade-off was right. Untrained people could see the benefits of publishing, and were able to learn by example well enough to get going rapidly with HTML. However, the cost-benefit trade-off may be very different for the semantic web.

A number of research questions and possibilities emerge on this matter:

- **Mimic the Web.** Can the semantic web exploit the Web's explosive "copy-edit-paste" model? Can some of the benefits of semantic web technologies be spread rapidly to the "wild web" at large? What would it mean to copy-edit-paste RDF from one person's page to your own? What services would be needed to support the creation of coherent RDF when it is transplanted? What mark-up tools can assist people in moving existing web sites into the "semantic" world?
- **Don't try to mimic the Web.** Is it appropriate to even try to mimic this? Will the reality be multiple, but relatively small semantic webs, within trusted communities who have consensus on worldview and purpose, with "professional" level codification. In this view, the semantic web is very different to the web that we currently know.
- **Hybrid model.** The "copy-edit-paste" model will work within communities who subscribe to a particular set of semantic web services. This is how semantic web literacy will spread.
- **Supporting the spectrum of developers and users.** It is well established from studies of end-user communities that a spectrum of expertise emerges, from expert developers, via "tinkerers" and "gardeners" who can make minor changes, to passive consumers

who do not modify their environment in any way. What tools are needed to help foster such communities of practice around semantic web applications?

Next generation application development environments

Developers of semantic web applications are a crucial class of "end user" who need to be considered. Knowledge modelling and the design of reasoning services are conducted in impoverished environments at present. Since there is already a large "installed base" of web developers who are used to integrated development environments, one argument is that they can be helped in making the transition to building semantic web applications if intelligible languages and effective design tools are provided.

This research effort needs to draw on the significant body of knowledge already built on the psychology of programming (e.g. cognitive processes; software visualisation; debugging tools). Examples:

— **The HTML model.** We have seen how HTML tools have evolved from editors to sophisticated layout and site management environments. What are the requirements of semantic web developers?

— **Implications for languages?** The semantic web is often conveyed as a set of layers, each with its own formalisms. Whilst this "layered cake" model may not be the best way to communicate the vision, it is undeniable that the infrastructure is complex, and this is an adoption barrier. Developers need support in managing the baroque range of languages. Many now work with XML, and there is some evidence that they want to "stick with what they know" rather than learn, for instance, RDF. A research challenge, therefore, is to better understand how "non semantic" web developers perceive the semantic web, and explore strategies for delivering its functionality in "palatable" ways. What can we learn from the successes and failures in efforts to establish other common languages/standards?

"Incidental" knowledge capture (i.e. automatic metadata generation and knowledge base population)

The semantic web depends on large amounts of well-structured data (i.e. at an appropriate granularity, appropriately classified and linked), whether this is basic metadata or ontology-based in a more systematic, formal sense. Knowledge capture is therefore a challenge of the first order. Whilst knowledge engineers, librarians and information scientists can do this work, they are in short supply. If we draw inspiration from the Web's explosive growth, a strategy is to make non-specialists part of the codification effort. However, it is an established fact that "normal" people neither like, nor have the time or skills, to create metadata in their busy lives (how many of us bother for own our documents or web sites?). This is even more the case for designing and populating coherent ontologies. Formalised knowledge capture is an "unnatural act" for most people.

The term *"incidental"* refers to the goal of generating metadata, and semantic annotation, as a *by-product* of people's normal work activities. Ideally, they will simply do their work, and smart technologies will "capture" the product and process in useful ways. The challenge is to develop strategies that match different domains, user groups and tasks that maximise the quality of knowledge capture with minimum effort on the part of users.

Although "pure" incidental knowledge capture is the goal, in reality this is often impossible. Assuming some changes are necessary in individual or group behaviour, the cost-benefit trade-off must be negotiated successfully to demonstrate uptake. If training is required because the tools are so novel, then again, the cost-benefit trade-off must be negotiated: the training has to be shown to be valued by the trainees, and effective.

Example strategies:

- When aspects of the work domain can be usefully modelled (e.g. tasks, roles, artefacts, states, workflow, user goals), agents may be able to monitor the online activities of single users or groups, and automatically or semi-automatically generate metadata based on context. Specific challenges include evaluations of the integrity of such modelling techniques, and the design of agent user interfaces that do not intrude, can be integrated into everyday environments, and help users to review and refine what is being captured so that they *trust* the system.
- Much knowledge-intensive activity is not conducted via computers, e.g. individuals talking on the phone, or in meetings. Assuming there is information worth capturing, approaches to explore in these contexts include technologies for tracking document use and the co-location of individuals, and capturing significant events in planned and unplanned conversations/meetings. Again, the emphasis is on enabling people to do their work with minimal disruption, and to prove that knowledge capture augmentation is possible.

Understanding the **use** *of metadata*

We need to better understand how metadata is created and used in practice.

Examples:

- "Articulation work" is a term from the study of work practices that refers to the processes involved in actually using an artefact, such as a formal classification scheme, as part of work activities. Such analyses often demonstrate that groups negotiate amongst themselves conventions for the application of such schemes. This knowledge needs to be brought to bear on semantic web applications as they emerge.
- Mechanisms need to be found to support the *layering of metadata* on resources. There may be many types of metadata from multiple perspectives, authors and agents, with varying access rights.
- "Dressing and undressing" objects with metadata/roles. *There are fashions in metadata.* Since metadata is designed to support resource discovery, maintaining a digital presence (e.g. to funding bodies or management) implies that as vocabulary shifts, metadata must adapt. In the economy of attention, if one's own resources don't show up "near the top" of a search, they may remain invisible.
- Binding metadata to objects. In some domains (such as bio-informatics) objects are passed around many people, who may annotate them. It is crucial to devise trusted methods to *bind* metadata to objects to ensure that history, dependencies, and ownership is preserved (with obvious implications for issues such as patent rights).

Coping with "messy metadata" (from people and machines)

When untrained people and imperfect machines are generating metadata or populating ontologies, the results will be "messy". Depending on the context (e.g. the criticality of maintaining high quality capture), three non-exclusive strategies are:

- **Avoid it.** The cost of incorrect metadata is too high (e.g. medicine or safety-critical applications).
- **Tolerate it.** The cost-benefit trade-off is good enough that imperfections do not disrupt system utility.

- **Clean it.** This may be a cooperative process between humans and intelligent agents, or a purely discursive process, e.g. expert conferences to decide on a new taxonomy, or how to codify new discoveries. The latter (found in some areas of bio-informatics) is an interesting example of how new organisational infrastructure will evolve around the semantic web as it takes root in a community.

How to support collaborative, emergent consensus?

Ontologies are generally designed by a small group of experts, in order to ensure consistency, coverage, etc. However, there is another niche of semantic web applications for those domains in which there is insufficient consensus to declare an ontology. In such situations, particularly in networked communities, the semantic web may be able to play a role in supporting *emergent consensus* by providing tools for collective sensemaking that tolerate inconsistency, ambiguity, incompleteness, and other such characteristics normally eliminated from ontologies and metadata.

Examples:

- **The domain is dynamic.** This is typical of many real world problems, which are subject to constant change due to unpredictable environmental factors. This can raise the cost of maintaining the ontology so high that it is not worth the effort. Representational schemes, methods and tools for ontology evolution and propagation of changes have contributions to make in this context.
- **The domain is contested.** A domain may have been studied for a long time, but be still partially understood, and from multiple perspectives. The classic example is the scientific/scholarly research field in which resolving disagreement on domain boundaries, concept distinctions and status is the defining activity of the stakeholders. Representational schemes, methods and tools to support such discourse/argumentation have contributions to make in this context.

How to maintain and analyse history?

Understanding the history behind a resource is often critical to understanding how to use and interpret it appropriately, whether that resource is a metadata scheme, an ontology, a recommendation from a knowledge based system, a design decision, the state of an application, or the state of current knowledge in a field.

Examples:

- **“Rollback” in scientific analysis.** This refers to services that could help answer the question, “what did we know then?” Semantic web applications to support ‘e-science’ should provide researchers with advanced tools to analyse the history behind and context surrounding research results.
- **Giving explanations.** A well-established field in artificial intelligence, the knowledge in this community about explanation generation needs to be brought to bear in semantic web applications.
- **Debugging semantic web applications.** Developers of semantic web applications need much more sophisticated environments to support their work (see below). A specific aspect of this—well proven in research into conventional code development tools—is the provision of historical traces of program execution in order to understand the state of the code at different points in the application’s state.
- **Design rationale.** Also known as design history, this is another established research field, concerned with the effective capture of reasoning behind design decisions, to

support subsequent analysis and maintenance. What kinds of design rationale are of particular importance to semantic web applications? What support can be given to its capture and recovery by different communities?

Controlling information overload

One of the most pressing problems today is information overload. If semantic web technologies can make a substantial contribution to tackling this problem it would represent an example of the “killer app” that many would appreciate.
Examples:

— **Personalisation agents/profiles.** This is a generic issue, independent of specific application. There is a large and growing research literature on agent user interfaces that needs to be brought to bear on semantic web agents. Issues that need to be better understood include interaction techniques for defining agents (e.g. programming by example), debugging agents (conveying programs to novice programmers), and representing both declarative knowledge (well understood) and procedural knowledge (poorly understood).

— **Email management.** Electronic mail is arguably the ubiquitous technology that everyone online has. For some, it has evolved from merely a communication tool to personalised information and diary management environment. How would a semantically enriched email environment assist in managing thousands of messages, and related activities?

— **Collaborative filtering.** How can colleagues and communities usefully share agents and other filtering devices, without this resulting in confusing agent behaviours, or aggravating the information overload problem?

“Lightweight” semantic collaboration

Analogous to early computers, semantic web applications tend to be “heavyweight” at present, as we struggle to understand their properties and get real examples to work. They take a lot of work to get running. Analogous to computing’s evolution to portable, ubiquitous devices with good user interfaces, semantic web applications may eventually become more “lightweight” (from the end-user’s perspective), lowering the adoption threshold.
Example:
“Good-enough” semantic interoperability. Groups who need to come together to collaborate for a short period may want the benefits of semantic web applications but have limited resources that they are willing to invest in achieving interoperability with each other’s tools/resources. The challenge is to devise infrastructures that can achieve ‘good enough’ interoperability to give value. This is likely to implicate underlying representational schemes (e.g. upper ontologies) plus quality user interfaces that communicate effectively the nature of the interoperability.

Recommendations for future research

The following areas, which have been detailed, are worth investigating in depth:

— Sustaining use of technologies in authentic work contexts,

— Developing growth models for the semantic web,

— Designing next generation application development environments,

— “Incidental” knowledge capture (i.e. automatic metadata generation and knowledge base population),

— Understanding the *use* of metadata,
— Coping with “messy metadata” (from people and machines),
— Supporting collaborative, emergent consensus and “lightweight” semantic collaboration
— Maintaining and analyzing history,
— Controlling information overload.

Relations with other initiatives

The semantic web is a transversal topic. It does not rely on only one domain like computer networks or artificial intelligence. Rather, it calls for the skills developed within many different fields that can contribute to the advance of the semantic web idea. Even if a relatively important community has already adhered to the semantic web challenge, the future semantic web programs must be addressed to all the fields that can contribute the semantic web.

Scientific communities

The workshop participant identified many such fields listed below:

- Artificial intelligence: reasoning mechanisms, knowledge representation languages, approximate and rough computing, learning and resource discovery, etc.
- Databases: storage, fault tolerance, security, transactions, query languages, etc.
- Web: profiling, identification, XML-based languages and technologies, etc.
- Agents: distributed computing, communication languages, interaction and cooperation protocols, etc.
- Theoretical computer science and (computational) logic: languages, theorem proving, semantics, etc.
- System: reliability, mobility, security (web security), etc.
- Computational linguistics and pattern recognition: knowledge acquisition from primary sources, using lexical resources for ontology development, pragmatics, question and answering, etc.
- Document engineering and digital libraries: transformation, mark-up, indexing, etc.
- Human-computer interface: computer-supported collaborative work, work factor evaluation, communication studies, etc.
- Social and human sciences: ontology validation experiments, social informatics, etc.

Institutional initiatives

Many funding and institutional bodies have already launched initiatives related to the semantic web:

- US DARPA “Agent Mark-up Language” (DAML) program for building languages and tools for facilitating the concept of the semantic web.
- EU IST “OntoWeb” network linking research and industrial teams involved in the semantic web development.
- W3C “Semantic web” activity to serve in the design of specifications and the open and collaborative development of technology.
- Semantic Web Agreement Group (SWAG), an independent development effort for building the semantic web infrastructure.
- EU IST key action line on “Multimedia content and tools” and its “Semantic web technology” program (2001) and “Knowledge technologies” program (2002).
- Japanese Interoperability Technology Association for Information Processing (INTAP) “Semantic web” task force for proposing semantic web activities to the corporate and institutional bodies.
- French CNRS “specific action” on the “semantic web” for proposing future research lines.

Other initiatives

Beside these purely scientific categories there are related initiative targeting application sectors or related technologies:

— NSF "digital library" initiative and UE-ERCIM DELOS network of excellence in "digital libraries". The PITAC report on "digital libraries: universal access to human knowledge" lists as its first recommendation "support expanded digital library research in metadata and metadata use, scalability, interoperability, archival storage and preservation, intellectual property rights, privacy and security and human use.
— EU IST C-Web project on "community webs" which produces a platform for supporting communities on the web (with formal ontologies and RDF annotations).
— MIT Oxygen and EU Ozone projects on ambient intelligence and ubiquitous computing.
— EU IST key action line on "new methods of work and electronic commerce" including "organisational knowledge management" and "electronic government".
— Industrial initiatives in the domain of e-commerce, such as UN and OASIS ebXML, or web services, such as UDDI.
— UK research council "e-Science" programme, EU IST "grid technology and their applications" cross-programme theme and GEANT project.
— The EU IST Advisory Group (ISTAG) has proposed integration projects as new instruments for the 6^{th} framework program. Among the possible integrated projects are "dependability, trust, security and privacy" and "community memory/sharing of knowledge".

Summary of the recommendations

The workshop recommendations have been grouped here with regard to four lines of research that look very important and that involve several working groups at once.

Identification and localisation

Identification is an important topic for semantic web reasoning, annotating and computing. It amounts to agreeing on how some resources can be identified, how two identifiers can be compared or equated and how web resources can be localised for processing. This involves works in language, infrastructure and ontological areas:

- Investigating the notion of identity in relation with the semantics of languages,
- Localising, accessing and matching resources,
- Explicit assumptions about identity in ontologies.

Relationships between semantic models

Heterogeneity must be considered as an intrinsic feature of the semantic web: no language will be suitable for all purposes, no model will be applicable to all cases and no ontology will cover the infinity of potential applications. Because, semantic description of information and knowledge is available, heterogeneity can be dealt with. This involves:

- Organising in a coherent manner the multiplicity of languages of different expressivity and purpose,
- Reconciling the various modelling styles existing (including those from software engineering),
- Investigating the engineering of articulation theories,
- Developing safe transformation methods,
- Developing and articulating the different possible reasoning services (e.g., querying, updating, deducing, inducing),
- Experimenting transformation infrastructure (proof-carrying transformations, processing and composing),
- Supporting reuse and evolution: comparison, merging, versions, and conceptual refinement of ontologies and metadata,
- Designing theories and metrics for comparing ontologies.

Tolerant and safe reasoning

A variety of reasoning methods will be necessary for different applications (from fetching to theorem proving) and the quality of their required results will vary (from anything-will-do to 100% certified). Tolerant and safe reasoning methods adapted to the web must be developed and the accuracy of their result must be characterised. This involves:

- Coping with “messy metadata” (from people and machines) and the open character of the web,
- Developing tolerant inference and levels of tolerance,
- Checking proofs and policies: representation of policies, proofs and properties,
- Propagating trust, proofs and rewards: trust model and knowledge level right management.

Facilitating semantic web adoption

Semantic web adoption is a critical point. It will first depend on the availability of resources and then on the ease of use of the semantic web. This can be achieved by:

- Acquiring ontologies and metadata from primary sources (texts, multimedia, images),
- Developing foundational ontologies and well-crafted ontology libraries,
- Sustaining use of technologies in authentic work contexts and developing growth models for the semantic web,
- Designing next generation application development environments,
- "Incidental" knowledge capture (i.e. automatic metadata generation and knowledge base population),
- Understanding the *use* of metadata and controlling information overload,
- Supporting collaborative, emergent consensus and imagining "lightweight" semantic collaboration,
- Maintaining and analyzing history.

General recommendations

- For research, **support worldwide collaboration** between researchers. Obviously because it allows to reach consensus at the 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 tools, **encourage open source development** of high quality components **and** non-profit **shelter organisations** for software development (like 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 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 education, provide **educational support** (e.g. teaching material, company "educating", starter kits).

Appendix I: resources

Web sites

http://www.semanticweb.org	a central resource about the semantic web.
http://www.ontoweb.org	EU IST OntoWeb initiative
http://www.daml.org	DARPA DAML program
http://www.w3.org/2001/sw/	W3C semantic web activity
http://swag.webns.net	SWAG site

Journals and special issues

European transactions on artificial intelligence: semantic web area, http://www.ida.liu.se/ext/etai/seweb

Dieter Fensel, Mark Musen (eds.), special issue on the semantic web, *IEEE Intelligent systems* 16(2):24-79, 2001 http://www.computer.org/intelligent/ex2001/x2toc.htm

Dan E. O'Leary, Rudi Studer (eds.), special issue on knowledge management, *IEEE Intelligent systems* 16 (1), 2001 http://www.computer.org/intelligent/ex2001/x1toc.htm

Books

Johan Hjelm, Creating the semantic web with RDF, John Wiley (UK), 2001

Dieter Fensel, Ontologies: silver bullet for knowledge management and electronic commerce, Springer-Verlag, Berlin (DE), 2001

Dieter Fensel, James Hendler, Henri Lieberman, Wolfgang Wahlster, Semantic web technology, The MIT press, Cambridge (MA US), 2002

Articles

Tim Berners-Lee, James Hendler, Ora Lassila, The semantic web, *Scientific american* 284(5):35-43, 2001, http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html

Stephan Decker, Sergey Melnik, Frank van Harmelen, Dieter Fensel, Michel Klein, Jeen Broekstra, Michael Erdmann, Ian Horrocks, The semantic web: the roles of XML and RDF, *IEEE Internet computing* 4(5):63-74, 2000 http://www.computer.org/internet/ic2000/w5063abs.htm

Sean Palmer, The semantic web: an introduction, http://infomesh.net/2001/swintro/

Ian Foster, Carl Kesselman, and Steven Tuecke, The anatomy of the Grid: Enabling scalable virtual organizations. *International journal of supercomputer applications*, 2001. http://www.globus.org/research/papers/anatomy.pdf

Conference and workshops

There had been a few workshops on the semantic web so far:

— First international workshop on the semantic web (SemWeb) in conjunction with ECDL 2000, Lisboa, PT, September 2000 (http://www.ics.forth.gr/proj/isst/SemWeb). The second workshop in this series has been organised at Hong Kong, CN, in conjunction with the WWW2001 Conference (http://semanticweb2001.aifb.uni-karlsruhe.de/). A report on these Workshops has been published in *SIGMOD Record* 30(3): 96-100 (http://www.acm.org/sigmod/record/issues/0109/r2-semantic_web.pdf). A 3rd issue will happen in relation to the WWW2002 conference.

— Semantic Web Technologies Workshop, European Commission, Luxembourg, November 2000 (http://www.cordis.lu/ist/ka3/iaf/swt_presentations/swt_presentations.htm)

— Semantic web working symposium, Stanford, CA US, July 2001 http://www.semanticweb.org/SWWS/

— Reviewed track on the semantic web at WWW 2002, Hawaï, May 2002 http://www11.org

A new international conference devoted to the semantic web (ISWC) has been set up after SWWS and will take place in June 2002 in Sardinia (IT). http://iswc.semanticweb.org
Many other workshops are flourishing now. The semanticweb.org web site is a good place to find them.

Reports

EU ISTAG (K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, J.-C. Burgelman eds.), Scenario for ambient intelligence in 2010, 2001 http://www.cordis.lu/ist/istag.htm

US PITAC, Digital libraries: universal access to human knowledge, 2001 http://www.ccic.gov/pubs/pitac/

US-UK US-UK, grid workshop, 2001 http://www.isi.edu/us-uk.gridworkshop/

David De Roure, Nicholas Jennings, Nigel Shadbolt, Research agenda for the semantic grid: a future e-science infrastructure, Report commissioned for EPSRC/DTI e-science core programme, 2001 http://www.ecs.soton.ac.uk/~dder/semgrid.pdf

Program consultation meeting report, Knowledge technologies, European commission: DG information society, 2001 ftp://ftp.cordis.lu/pub/ist/docs/pcm-9finalreport.pdf

Appendix II: acronyms

AI	Artificial Intelligence
B2B	Business to business
B2C	Business to customers
CC/PP	(W3C) Composite Capabilities/Preference Profiles
CDMA	Code Division Multiple Access (mobile phone standard)
CERN	(EU) European Organisation for Nuclear Research (Centre d'Études et de Recherche Nucléaire)
CNRS	(FR) Centre National de la Recherche Scientifique
DAML	DARPA Agent Markup Language
DAML/S	DAML Service description language
DAML+OIL	DAML+OIL
DARPA	(US) Defense Advanced Research Projects Agency
DELOS	(ERCIM/EU) Network of Excellence in Digital Libraries
DG	(EU) Direction Générale
DNA	Desoxyribo-Nucleic Acid
e-	electronic
ebXML	e-business XML
ECDL	European Conference on Digital Libraries
EDI	Electronic Data Interchange
EPSRC/DTI	(UK) Engineering and Physical Sciences Research Council
ERCIM	European Research Consortium in Informatics and Mathematics
EU	European Union
FET	(EU) Future emergent technologies
GEANT	(EU) pan-European Gigabit Research and Education Network
GSM	Global System for Mobile communication (mobile phone standard)
HTML	(W3C) HyperText Markup Language
HTTP	(W3C/IETF) HyperText Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
INTAP	(JP) Interoperability Technology Association for Information Processing
IP	(IETF) Internet Protocol
ISO	International Standards Organization
IST	(EU) Information Society Technology
ISTAG	(EU) IST Advisory Group
ISWC	International Semantic Web Conference
IT	Information Technologies (alt. Italy)
KM	Knowledge Management
MIT	Massachussets Institute of Technology
NSCA	(US) National Super Computing Applications
NSF	(US) National Science Foundation
OASIS	Organization for Structured Information Standards
OIL	Ontology Inference Layer
OMDoc	Open Math Document format

P2P	Peer to Peer
P3P	(W3C) Platform for Privacy Preferences
PDA	Personal Digital Assistant
PITAC	(US) President's IT Advisory Committee
Q&A	Question and Answering
RDF	(W3C) Resource Description Framework
RDFS	(W3C) RDF Schema
RSS	RDF Site Summary
SETI	Search for ExtraTerrestrial Intelligence
SHOE	Simple HTML ontology extension
SUO	Standard Upper Ontology
SW	Semantic Web
SWAG	Semantic Web Agreement Group
SWWS	Semantic Web Working Symposium
TM	(ISO) Topic Maps
UDDI	Universal Description, Discovery and Integration of Business for the Web
UK	United Kingdom (of Great Britain and Northern Ireland)
UN	United Nations organization
URI	(IETF/W3C) Uniform Resource Identifier
URL	(IETF/W3C) Uniform Resource Locators
US	United States of America
W3C	WWW Consortium
WWW	World Wide Web
XHTML	(W3C) XML HTML
XML	(W3C) eXtensible Markup Language
XMLS	(W3C) XML Schema
XSL	(W3C) XML Stylesheet Language
XSLT	(W3C) XSL Transformations

Appendix III: Program of the workshop

Wednesday, October 3rd

Opening notes

Languages

The semantic web from W3C perspective
    Eric Miller
RDF, RDF-Schema and the possible evolutions
    Brian McBride
Currents developments of DAML+OIL
    Frank van Harmelen
Inference systems
    Ian Horrocks
Topic maps
    Michel Biezunski

Resources and infrastructure

Ontological resources and top-level ontologies
    Nicola Guarino
Perspective in document annotations and metadata
    Eric Miller
Gathering and assembling knowledge structure
    Stefan Decker
Database technology for the SW
    Vassilis Christophides
Interoperability in an open semantic web
    Jérôme Euzenat

Thursday, October 4th

Clients and human interface

Adaptive SW
    Henri Lieberman
Human-Centred aspects of the SW
    Simon Buckingham-Shum
Secure semantic web
    Bhavani Thuraisingham

The semantic web in application areas

Ontology-based retrieval of resources
    Deborah McGuinness
SW for e-science and education
    Enrico Motta
Bioinformatics and science databases

Carole Goble
Semantic web for generalized knowledge management
Rudi Studer
Semantic eCommerce
Dieter Fensel

Friday, October 5th

Synthesis of the discussions/working groups

Language
Deborah McGuinness, Frank van Harmelen, Ian Horrocks, Michel Biezunski, Vassilis Christophides
Infrastructure
Jérôme Euzenat, Bhavani Thuraisingham, Brian McBride, Stefan Decker
Human-related issue
Simon Buckingham-Shum, Henri Lieberman, Enrico Motta, Carole Goble
Ontologies
Nicola Guarino, Rudi Studer,

Closing

We add below a short CV and position abstract of each of the position summary provided by the participants (mainly beforehand).
The full overhead presentation used at the workshop can be found at the URL:

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

Appendix IV: CV and position of participants

Michel Biezunski

Coolheads Consulting, Allen, USA
mb@infoloom.com
http://www.topicmaps.net

Short CV

Ph.D, 1981, Physics Education, Universite Paris VII.
Research in History/Philosophy of Science, focused on the reception of the theory of relativity.
Professional experience in Publishing
Professor at the UTCS (Universite de Technologie de Compiegne, Sevenans), 1989-1992
Founder and director of High Text, an Electronic Document Engineering Company, Paris, 1993-1997.
Founder and president of Infoloom, a US-based company, 2000-
Co-Editor of the ISO/IEC 13250 Topic Maps standard (published in 2000)
Co-founder and co-editor of TopicMaps.org and the XTM Authoring Group (2000)
Numerous presentations and workshops on Topic Maps, mainly at GCA conferences.
Invited speaker (with Steven R. Newcomb) at the Semantic Web Working Symposium, Stanford, July-Aug. 2001.

Current Research Activities: facilitate semantic addressability of notions, develop a processing model for topic maps, provide a common interchange languages for ontologies.

Key words: Topic Maps, Addressability, Knowledge Networks, Semantic Connectivity.

Abstract

I will present why Topic Maps are likely to play an increasing role within the context of the Semantic Web.
Directions : enhance the quality of services for web users, including high-value corporate assets. Make any unit semantically addressable without imposing a unique world-view.
Problems : making various communities agree on a common background platform is not easy. This requires
defining the proper level of abstraction for ensuring efficient interchange. Separating content from processing made on the content is not easy to explain, and is somewhat different from the main directions taken by web services including the Semantic web.
Challenges : establish a link between various research projects done within the academic world and the interests of business users.
Opportunities : The « semantic real estate » (similar in a way to the domain name space) is wide open and those who will the first to occupy and exploit it will make a difference.

Vassilis Christophides

Institute of Computer Science, FORTH, Heraklion, GR
christop@ics.forth.gr

Short CV

Vassilis Christophides studied Electrical Engineering at the National Technical University Athens, Greece (NTUA). He received his DEA in computer science from the University PARIS VI and his Ph.D. from the Conservatoire National des Arts et Metiers (CNAM) of Paris, France. From 1992 until 1996, he did his research work at the French National Institute for Research in Computer Science and Control (INRIA) in the topics of Object Database and Documents Management Systems. Since 1997 he is adjunct professor at the Computer Science Department of Crete University, Heraklion, Greece. His interests include, Data and Knowledge Base Systems, Formal Models and Query Languages for semi-structured data, Wide-area Distributed Systems, Query Optimization, Data and Process Mediation. He co-chaired the 1st International Workshop on the Semantic Web in Lisbon and served on the commitees of subsequent workshops on the Semantic Web.

Key words: Models and Query Languages, Metadata Management, Wide-area Distributed Systems, Data and Process Mediation

Abstract

The primary goal of the Semantic Web activity so far has been the definition of infrastructure, standards and policies facilitating an explicit description of meaning of Web resources that can be processed by both humans and automated tools. Complementary to the demand for universal access to information is the ever increasing need for semantics-based access to services. These efforts towards the next evolution step of the Web have given rise to a large number of research problems that relate to models, architectures, applications and services for the Semantic Web. Research activity has been flourishing around these problems, with numerous impressive results, but still leaving us with an even greater number of challenges and opportunities. Classical data management practices are branching out of traditional frameworks, facing an unseen before demand for open-ness, expressiveness, flexibility and scalability.

Database technology is one of the main driving forces and - in our view - the **core enabling technology for building infrastructure and services on the Semantic Web**. Specifically, we regard the following problems as cornerstone issues for the realization of the Semantic Web:

- *Infrastructure for Data and Process Mediation* : to facilitate automated description, discovery, brokering and composition of e-services
- *Metadata Management* : to enable management and maintenance of superimposed resource descriptions and schemas of heterogeneous data sources
- *Ontology Evolution and Metadata Revision* : to support the dynamics of schemas and vocabularies (expansion, revision)
- *Transactional Aspects* for e-services : revision of ACID properties of transactions
- *Formal Foundations for Web Metadata Standards* : efficiently implementable formal models (expressiveness vs. efficiency tradeoff)
- *Semantics-aware Query Languages versus Inference Services* : to enable sophisticated searching and browsing, declarative content-based access, resource discovery and matchmaking
- *Persistent Storage* : to enable the efficient management of voluminous resource description bases and support effective sophisticated querying (schema-specific storage, index structures)

Stefan Decker

Stanford University, USA
stefan@db.stanford.edu
http://www-db.stanford.edu/~stefan/

Short CV

Stefan Decker did his Ph.D. studies in Computer Science at the University of Karlsruhe, Germany where he worked on ontology based access to distributed information on the Web. He is a postdoctoral fellow at the department of computer science at Stanford University, where he leads (with Prof. Gio Wiederhold) the OntoAgents project in the DARPA DAML program. He works on Semantic Web and mediation technology. His research interests include knowledge representation and database systems for the World Wide Web, information integration and ontology articulation and merging.

Abstract

On the Semantic Web a plethora of information models like RDF,UML, ER, OIL, DAML+OIL, RDF Schema, XML Schema etc. are used for the exchange of structured information, varying greatly in their expressive power. The advent of XML leveraged a promising consensus on the encoding syntax for machine-processable information. However, interoperating between different information models on a syntactic level proved to be a laborious task. We suggest a layered approach to interoperability of information models that borrows from layered software structuring techniques used in today's internetworking. We identify the object layer that fills the gap between the syntax and semantic layers and examine it in detail. We suggest the key features of the object layer like identity and binary relationships, basic typing, reification, ordering, and n-ary relationships. Finally, we examine design issues and implementation alternatives involved in building the object layer.

Jérôme Euzenat

INRIA Rhône-Alpes, Montbonnot, FR
Jerome.Euzenat@inrialpes.fr
http://www.inrialpes.fr/exmo/

Short CV

Jérôme Euzenat is researcher at INRIA Rhône-Alpes (Montbonnot, France). He has worked as a software engineer and consultant for several companies and lectured on knowledge representation at several universities (Joseph Fourier, INPG). He holds a PhD (1990) and habilitation (1999) in computer science, both from Grenoble 1 university. Jérôme Euzenat has worked and published about reason maintenance systems, object-based knowledge representation, temporal granularity and knowledge base co-operative editing. His all time interests are tied to the relations holding between concurrent representations of the same situation.

Jérôme Euzenat currently leads the EXMO action investigating the exchange of formal knowledge mediated by computers. This covers formally annotated documents, knowledge servers, content-based cooperative work and transformation of representations. It is more focussed on the preservation of specified properties (e.g., meaning, order, or confidentiality) the transformation of representations. This aims at contributing to a more accurate and faithful communication of formal representations, a central topic of the semantic web.

Key words: Transformation, semantic interoperability, customisable knowledge representation, knowledge communication.

Abstract: Interoperability in an open semantic web

The semantic web is an infrastructure based on a network of connected formalised knowledge repositories. Because of different expressivity and efficiency requirements, the variety of languages for expressing knowledge seems unavoidable. In order to keep knowledge from the web interoperable, it is necessary to have a semantically sound way to exchange and transform knowledge. There are many solutions to this problem (including standardising to one pivot language).

I think that, in the same way XML is an extensible meta-language, knowledge representation languages for the web must be extensible and cannot have beforehand a fixed semantics. Each application should be able to choose its representation language and to expose its syntax, semantics and other rules of use (e.g., GUI display conventions). If this extensibility is taken into account from the beginning of the semantic web, it will be possible to build an integrated semantic web with various levels of languages.

We can put forth several possible way of dealing with this problem like family of languages acting as a pivot language, knowledge patterns or proof carrying transformations. All these techniques are based on the idea that it is possible to specify (syntactic) transformations from one representation language to another and that it is possible to prove the properties that the transformations satisfy. If a transformation is written in an intelligible language, the publication of the proofs enables proof checking to be performed, before using the transformation with complete trust. Of course, the satisfaction of semantic properties (like consequence preservation) requires the exposition of the language semantics.

This perspective requires work at the theoretical level (about proof of particular properties, assisted proof generation from transformation or composition of properties) and infrastructure level (language-oriented transformation lookup, proof representation languages or fast proof checkers).

Dieter Fensel

Vrije Universiteit Amsterdam, NL
dieter@cs.vu.nl
http://www.google.com/search?q=dieter

Short CV

Dieter Fensel is an associated professor at the Vrije Universiteit Amsterdam in the area of business informatics. After studying mathematics, sociology and computer science in Berlin, he joined in 1989 the Institute AIFB at the University of Karlsruhe. His major subject was knowledge engineering and his PhD thesis in 1993 was about a formal specification language for knowledge-based systems. From 1994 until 1996 he visited the group of Bob Wielinga at the SWI Department in Amsterdam. During this time his main interest were problem-solving methods of knowledge-based systems. In 1996, he come back as a senior researcher at the Institute AIFB finalizing his Habilitation in 1998. Currently, his foccus is on the use of Ontologies to mediate access to heterogeneous knowledge sources and to apply them in knowledge management and electronic commerce. He is involved in several national and internal research projects, for example, in the running IST projects H-Techsight, IBROW, On-to-Knowledge, Ontoweb, and Wonderweb. Dieter Fensel is the author of the books Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2000; Problem-Solving Methods: Understanding, Development, Description, and Reuse, Lecture Notes on Artificial Intelligence (LNAI), no 1791, Springer-Verlag, Berlin, 2000; and The Knowledge Acquisition and Representation Language KARL, Kluwer Academic Publisher, Boston, 1995.

Key words: Semantic Web, Electronic Commerce

Abstract

Electronic marketplaces for B2B electronic commerce bring together lots of suppliers and buyers, and mediate their business transactions. This requires the marketplaces to be able to integrate a large number of documents, specified according to certain document standards used by marketplace participants. Existing solutions for document integration require creation and maintenance of numerous complicated transformation rules, which hampers their scalability required by the B2B area. ***Semantic web technology*** can provide two important solutions for this area. Scalable mapping support for realizing the full potential of electronic commerce based on a P2P approach and formal product and service descriptions that significantly increase the degree of mechanization in electronic commerce. Intelligent electronic commerce may become the killer application for semantic web technology

Carole Goble

University of Manchester, UK
carole@cs.man.ac.uk

Short CV

Carole Goble is a Professor of Computer Science at The University of Manchester and leader of the Information Management Group. Her interests are in the application of metadata, knowledge representation and ontologies to multimedia, conceptual hypermedia, intelligent information integration and retrieval. Her work is applications driven, chiefly in Bioinformatics and the Semantic Web. Her projects include the mediated transparent access to multiple bioinformatics information sources through a single ontology (TAMBIS). She ash lead developments of tools and applications for the Semantic Web ontology languages OIL and DAML+OI, including OilEd and the COHSE hypermedia system. She is a leading member of the UK's e-Science initiative and is the leader of an e-Science testbed for a collaboratory workbench for biologists (MyGrid) that will be a testbed for the GRID, the Semantic Web, DAML+OIL and Web Services. She is a co-investigator of the EU Network of excellence OntoWeb and the EU WonderWeb project. She is co-founder of a bioinformatics company Sagitus Solutions Ltd, and a Semantic Web tools company, Network Inference. She is and active member of the BioOntologies Consortium and is Vice Chair of the Semantic Web track of WW2002.

Key words: eScience, semantic web, bioinformatics, application, Grid

Abstract: Bioinformatics: a community case study for the Semantic Web

The Web was incubated by a scientific community: Physics. The Semantic Web could benefit from being pioneered in the same way, and another scientific community, Biology, seems a promising choice. The biology community is globally distributed and highly fragmented. Most biological knowledge resides in a large number of modestly sized heterogeneous and distributed resources, including: published biological literature (increasingly in electronic form), specialised databases curated by a small number of experts, and embedded within, or generated by, specialist analytical tools. The complex questions and analyses posed by biologists cross the artificial boundaries set by these resources. The pressing need is *interoperation* and *fusion*; however, current data resources have been designed to be meaningful to, and navigated by, people rather than automated processes (a.k.a. agents). To provide automated support for intelligent brokering, searching and filtering, and ultimately the discovery of new knowledge, means adding computationally accessible descriptions of meaning to the data, tools and repositories. This sounds reassuringly familiar to those versed in the Semantic Web rhetoric. Thus the issues to be addressed by the bioinformatics community are a version of issues to be addressed by the whole web community. The scientific community has major concerns regarding digital signatures (for intellectual property), provenance, change propagation, trust. The shift to "in silico" experimentation requires finding and weaving services. The community is embracing web services coupled with workflow to do this. Much of the biological data is self-described marked up text (pre-dating XML), and hence ontologies for disambiguating database entries and annotation is accepted as standard practice. The biologists have a well-established mark-up culture where resources are annotated (often by hand) by curators, and are well aware of the need for automated metadata extraction. A recent

editorial in the Journal of Bioinformatics suggests that the paradigm shift in biology from publishing papers to publishing data be extended to publishing knowledge.

I hope to explore Biology's potential as a receptive and well-organised community with clearly articulated information and knowledge needs that can pioneer the Semantic Web. The GRID, proposed as the next generation Internet, is another movement prevalent in e-Science. Its Knowledge Layer seems to be closely related to the Web Services world, and I will draw attention to this relationship.

Nicola Guarino

LADSEB-CNR, Padova, IT
Nicola.Guarino@ladseb.pd.cnr.it
http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html

Short CV

Nicola Guarino graduated in Electrical Engineering at the University of Padova in 1978. He is active in the ontology field since 1991, and has played a leading role in the AI community in promoting the study of the ontological foundations of knowledge engineering and conceptual modelling under an interdisciplinary approach. His current research activities regard ontology design, knowledge sharing and integration, ontology-driven information retrieval, and ontology-based metadata standardisation. He is general chairman of the International Conference on Formal Ontology in Information Systems (FOIS), and associated editor of the Semantic Web area of the Electronic Transactions on Artificial Intelligence and of the International Journal of Human-Computer Studies. He has published more than 60 papers on scientific journals, books and conferences, and has been guest editor of 3 special issues on scientific journals related to formal ontology and information systems.

Key words: Formal ontology, conceptual modelling, ontology design methodology, knowledge engineering, semantic web

Abstract:

I believe that, in the knowledge society, mutual understanding is more important that mass interoperability. The current notion of the Semantic Web is a bit too much to computer-computer knowledge integration, with little importance given to human understanding. On the other hand, knowledge trustability seems to be a crucial problem nowadays, and cognitive transparency is in my opinion the key to trustability. My main point is that ontologies should aim at cognitive transparency in order to be effectively used for the semantic web. This can be achieved by investing adequate resources in (i) principled ontology development methodologies, aiming at a unified conceptual analysis methodology common to database systems, object oriented systems, and knowledge-based systems ; (ii) libraries of upper level foundational ontologies whose characteristics and implications are already well understood (although not necessarily agreed upon) by a large community; (iii) specific ontologies developed for strategic domains, such as: the ontology of information and information processing, the ontology of social entities (people, organizations, laws), the ontology of social interactions (among human and artificial agents and organizations)

Ian Horrocks

University of Manchester, UK
horrocks@cs.man.ac.uk
http://www.cs.man.ac.uk/~horrocks

Short CV

Ian Horrocks is a Lecturer in Computer Science at the University of Manchester, UK. He graduated in Computer Science from Manchester in 1982, winning the prize for most outstanding graduate. After working in industry he returned to Manchester to complete a PhD in 1997. His FaCT system revolutionised the design of DL systems, redefining the notion of tractability for DLs and establishing a new standard for DL implementations. He is/has been involved in numerous national and international research projects including Camelot, DWQ, OntoWeb and
DAML, and is the coordinator of the IST WonderWeb project. He has published widely in leading journals and conferences, winning the best paper prize at KR'98, and is the author of two chapters in the forthcoming ``Handbook of Description Logics''. He is a member of the programme/editorial committees of several international conferences, workshops and journals, and is the program chair of the 2002 International Semantic Web Conference. His current research interests include knowledge representation, automated reasoning, optimising
reasoning systems and ontological engineering, with particular emphasis on the application of these techniques to the World Wide Web. He is a member of the OIL language steering committee, the Joint EU/US Committee on Agent Markup Languages, the W3C Web Ontology Language working group, and is an editor of the DAML+OIL language specification.

Abstract: The Semantic Web: (Ontology) Languages and Reasoning

Ontologies, and hence ontology languages, are widely recognised as being key components in the development of the Semantic Web. In response to the requirement for a suitable web enabled ontology language, groups of researchers in Europe and the USA developed the OIL and DAML languages respectively. These effort were subsequently merged, with the DAML+OIL ontology language being the result. This language has now been adopted by the W3C web ontology language working group as the basis for a W3C ontology language standard.

DAML+OIL is based on existing web standards, in particular RDF. It extends RDF by defining a semantics for a set of RDF classes and properties that capture common idioms from object oriented (e.g., frame based and description logic) knowledge representation, and by allowing XML Schema datatypes to be used instead of RDF literals (which are simply strings). The semantics can be specified in several ways, in particular via a model theory, via an axiomatisation, or via a mapping to a description logic.

As in any description logic, the expressive power of DAML+OIL is determined by the kinds of class (and property) constructors supported, and by the kinds of axiom (asserted facts) that are allowed. As well as supporting a wide range of class constructors, DAML+OIL also supports transitive properties, axioms asserting sub/super property relationships, and a wide range of class axioms, all of which can be reduced to asserting sub/super class relationships between arbitrarily complex class descriptions. This combination is equivalent to a very expressive description logic: the logic SHIQ extended with nominals (extensionally defined classes) and datatypes.

Consistency/subsumption reasoning in this logic is know to be decidable (it is contained in C2, a decidable fragment of FOL), but many challenges remain for implementors of "practical" reasoning systems, i.e., systems that perform well with the kinds of reasoning problem generated by realistic applications. Although reasoners for SHIQ are already available, dealing with inverse properties (the "I" in SHIQ) is known to be problematical, and it is not yet been demonstrated that these reasoners can deal with realistic problems. Extending SHIQ with datatypes seems to be harmless, but nominals are another matter: the resulting logic is know to have high worst case complexity (NExpTime), it no longer has the tree model property, and there is no known algorithm that appears suited to highly optimised, "practical" implementations. Moreover, even for the SHIQ logic, there may still be problems dealing with large ontologies and, in particular, with very large numbers of individuals.

Current research is beginning to address some of the above problems: new algorithms and optimisation techniques have improved the performance of SHIQ reasoners, work is underway on advanced optimisation techniques to deal with inverse properties, and experiments have shown that it is possible for such systems to deal with very large ontologies (e.g., 100,000 classes) when the expressive power of the language is used parsimoniously. Work is also being done on other inference problems, in particular those associated with query answering and knowledge extraction (e.g, computing useful concepts from example instances). However, much work remains to be done if the full power of DAML+OIL is to be exploited by software agents.

Henry Lieberman

MIT Media Lab, Cambridge, USA
lieber@media.mit.edu

Short CV

Henry Lieberman is a Research Scientist at the MIT Media Lab, where he heads the Software Agents Group, investigating how programs can act in the role of assistants to the user, rather than simply as tools. Agent programs can learn from interacting with the user, and provide proactive assistance. He recently edited the book,
Your Wish is My Command: Programming by Example (Morgan Kaufmann, 2001), which describes how new behavior can be taught to a machine by presenting concrete examples, and having the machine generalize a program. He has published many papers on a wide variety of subjects, including tools for intelligent visual design, information visualization, programming environments, debugging, agents for the Web, digital photography, computer music, and more.

Prior to joining the Media Lab in 1987, he was at the MIT Artificial Intelligence Lab since 1972, where he worked on Actors with Carl Hewitt, an early object-oriented parallel computing formalism for AI. He invented the first real-time garbage collection algorithm and introduced the notion of prototypes for object-oriented programming. He worked with Seymour Papert and Hal Abelson and the group that developed the educational language Logo and educational applications. He wrote the first color and bitmap Logo graphics systems, and the first flood-fill graphics algorithm. He holds an Habilitation degree from the Université Paris VI- Pierre et Marie Curie and was a Visiting Professor there.

Key words: dynamic semantics, declarative semantics, programming by example

Abstract

The static model of the Web as a statically linked network of pages is now breaking down as programs attempt to dynamically generate information, and as human browsing is increasingly assisted by intelligent agent programs. The next phase of the Web, the Semantic Web, lies in encoding properties of and relationships between objects. However, to date, this semantics is primarily declarative. It is expected to change only relatively slowly as Web pages are created, destroyed, or modified, typically by explicit, relatively coarse-grained human action. Less concern has been given to dynamic semantics of the Web, which is equally important. Dynamic semantics have to do with the just-in-time creation of content, actions which may be guided by user-initiated interface operations, time, users' personal profiles, or data on a server. The challenge for the future is in methods for cleanly integrating the static and dynamic aspects of the Web.

Dynamic semantics will enable the new generation of intelligent agent software that will increasingly inhabit the Web. Though there will be fully autonomous agents, the more interesting agents will cooperate interactively with humans, helping them to achieve their goals more efficiently than a user could on their own. Beyond that, we envision that Web end users and Web application developers will not be routinely writing code directly in any textual language, but instead avail themselves of Programming by Example and other interactive agent

interfaces that will hide the details of the formal languages. Transition to these future interfaces will be greatly aided by a foundation that can cleanly integrate static and dynamic semantics.

Brian McBride

Hewlett Packard, Bristol, UK

Short CV

Brian McBride is a researcher at Hewlett Packard Laboratories (Bristol, UK). He has experience of product development, particularly object oriented office software systems and industrial research on distributed object system architecture and development (CORBA), mobile communications (wireless service gateways) and RDF. He holds a Batchelor's degree in Computer Science from the Unversity of Edinburgh.

He is the founder of Hewlett Packard Laboratories' Semantic Web group and original developer of the Jena Semantic Web toolkit. He is co-chair of the W3C RDFCore working group which is revising and completing the RDF specifications.

Key words: Topic Maps, Addressability, Knowledge Networks, Semantic Connectivity.

Abstract

The "semantic web" is a term which attracts a broad constituency of researchers to work towards a broadly cohesive goal. Whilst it is unlikely that they share the same precise vision of what the "semantic web" will be, there is agreement that it is in part about creating a global infrastructure where machines have access to the semantics of the data they are processing.

To accomplish this goal we need:

- a common, globally scalable naming system
- a common language for the representation of facts **-** the A-box
- a common language for the representation of information schemas **-** the T-box
- a common protocol stack which permits independently developed systems to exchange semantically rich information.

In particular, it is desirable to reduce the fragmentation amongst different efforts to achieve this goal.

It is not enough to figure out how to build such an infrastucture, it is also our goal to bring that infrastructure into being. This requires:

- demonstration of the value of such semantically rich information exchange
- enabling its feasibility through the development of powerful, usable, industrial strength, freely available open source toolkits
- seeding its development with the provision of key enabling services

Deborah L. McGuinness

Knowledge Systems Laboratory, Stanford University, Stanford, USA
dlm@ksl.stanford.edu
http://ksl.stanford.edu/people/dlm

Short CV

Dr. Deborah McGuinness is associate director and senior research scientist of the Knowledge Systems Laboratory at Stanford University. She has been working in knowledge representation and reasoning environments for ontology creation and maintenance for 20 years. She has built and deployed numerous ontology environments and ontology applications, including some that have been in continuous use for over a decade at AT&T and Lucent. (Her AT&T web site has more information on this topic.) She is a co-author of the current ontology evolution environment from Stanford University and is also a co-author of one of the more widely used description logic systems (CLASSIC) from Bell Laboratories. She is the co-editor of the web ontology language for the DARPA Agent Markup Language program. The initial release is available here and continuing updates are released from the Joint EU/US Agent Markup Language Committee, of which she is also a member. She is the project leader for the Rapid Knowledge Formation project at Stanford's Knowledge Systems Laboratory. The project's goal is to allow distributed teams of subject matter experts to quickly and easily generate, use, and modify knowledge bases without the aid of knowledge representation and reasoning experts. She is also the project leader for the knowledge systems laboratory's project on Tools for DAML-Based Services, Document Templates, and Query Answering for the Darpa Agent Markup Language project. That project's goal is to create technologies that will enable software agents to dynamically identify and understand information sources, and to provide interoperability between agents in a semantic manner. She is co-designing the DAML-O language (announced on the daml web site) which is based on part on OIL - the ontology inference layer. She is on the steering committee of the OIL consortium. The oil language is available. She is also project manager for the ontology project with Cisco Systems.

Key words: Ontologies, Semantic Web, Markup Languages, Description Logics

Abstract

Deborah's current research directions have been aimed at emerging opportunities on the web. The web must move from being presented for human interpretation to being presented for agent interpretation. This means that information and services need to provide more in the way of telling agents when to use them and how to interact with them. One solution to this is to have expressively powerful markup languages and have environments for creating, maintaining, and reasoning with the markup information. These environments need to work in distributed settings and work in the presence of incomplete and inconsistent information. They also need to work in the presence of multiple vocabularies. Deborah has focused both her academic research and her consulting work around generating, evolving, merging, and analyzing structured information, such as that representable in the DAML+OIL ontology language. Some main areas of growth in these areas will be in the collaboration, diagnostics, truth maintenance, persistence, query languages, rule extensions and other reasoning, and filtering and presentation.

Applications areas that are expected to drive this work include business to business web interactions, supply chain applications (requiring strong interoperability and some configuration requirements), configuration in general,and agent-based web services.

Eric Miller

W3C World Wide Web Consortium, Cambridge, MA USA
em@w3.org
http://www.w3.org/people/em/

Short CV

Eric Miller is the Activity Lead for the W3C World Wide Web Consortium's Semantic Web Initiative.

Eric's responsibilities include the architectural and technical leadership in the design and evolution of Semantic Web infrastructure. Responsibilities additionally include working with W3C Working Group members so that both working groups in the Semantic Web activity, as well as other W3C activities, produce Web standards that support Semantic Web requirements. Additionally, to build support among user and vendor communities for the Semantic Web by illustrating the benefits to those communities and means of participating in th creation of a metadata-ready Web. And finally to establish liaisons with other technical standards bodies involved in Web-related technology to ensure compliance with existing Semantic Web standards and collect requirements for future W3C work.

Before joining the W3C, Eric was a Senior Research Scientist at OCLC Online Computer Library Center, Inc. and the co-founder and Associate Director of the The Dublin Core Metdata Initiative, an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models.

Eric holds a Research Scientist appointment at MIT's Laboratory for Computer Science.

Enrico Motta

Knowledge Media Institute, The Open University, UK
e.motta@open.ac.uk

Short CV

Dr. Enrico Motta is Director of the Knowledge Media Institute at the Open University in England, a 70-people strong R&D centre, which specialises in web, media and knowledge technologies. His first degree was in Computer Science, obtained from the University of Pisa in Italy, and he later received a PhD in Artificial Intelligence from the Open University. Dr Motta has been conducting research in knowledge systems since 1984 and has published extensively on topics such as knowledge acquisition, knowledge management, formal knowledge representation and agent systems. His current research focuses on the integration of web and knowledge technologies, to support knowledge sharing and reuse. Application domains include e-commerce, organizational learning, scheduling systems, electronic news publishing, medical decision making and scholarly discourse. Dr Motta is on the editorial board of the International Journal of Human-Computer Studies, a member of several international programme committees and a reviewer for the top journals in the field of knowledge technologies. He has authored a book on knowledge sharing and reuse, entitled “Reusable Components for Knowledge Modelling”, which is published by IOS Press.

Key words: Knowledge Technologies, Semantic Web Services, Organizational Learning, Ontologies, Knowledge Sharing and Reuse

Abstract

The semantic web idea crucially depends on the adoption and specification of ontologies seen as the glue to link services and resources semantically. Even if we simply focus on these two processes, specifying and populating ontologies, a number of important issues arise: “how do communities agree on ontology specifications?”, “what about those domains where there is little consensus, or where the underlying conceptual framework evolves rapidly?”, “to what extent will it be possible to support automatic merging and linking of ontologies developed by different communities?”, “how can we speed up the expensive resource annotation process?”, etc. Although these are crucial questions, they are actually situated at the “semantic infrastructure” level: in other words, solving these problems is primarily important to enable us to focus on the really interesting issues related to inventing and deploying novel semantic web functionalities. So, assuming that the semantic web takes off and semantic services and resources become widespread, what are the interesting challenges for the community? My own personal interest is primarily in the seamless integration of semantic web technology in the normal tools of work. For instance, while I am writing this document, I would expect semantic agents to find me relevant information. Even more importantly, once this document is completed, rather than the standard “Save” button, I should be able to press a “Publish++” button, which may automatically annotate the document with semantic information (e.g., link it to relevant projects, research areas, application domains, technologies, etc.), may notify some of my colleagues (but only the ones who are going to be interested in the topic of the document!), may publish it in one or more semantically enriched repositories, which in turn allow users to locate relevant information for different classes of users, with different access rights, in a variety of work contexts to do with projects, research areas, application domains,

virtual teams, etc. Clearly, authoring is only one part of the picture. I also receive/send email, manage my own diary, access services on the web, belong to several virtual teams scattered around the world, etc. Can semantic web technology become ubiquitous and an enabler for individual and/or teamwork? This is, I believe, the greatest opportunity and the greatest challenge.

David Pearce

European Commission, Brussels, BE
David.Pearce@cec.eu.int

Short CV

David Pearce obtained his doctorate from the University of Sussex in 1980 with a thesis in the area of Logic and Scientific Method. In 1982 he joined the Free University Berlin as Assistant Professor in Philosophy and later, in 1988, as Heisenberg Research Fellow, establishing an interdisciplinary research group: Logic, Epistemics and Information. From 1994-2000 he was Senior Researcher at the German AI Research Centre, DFKI, in Saarbruecken, and is currently a Researcher in the Future and Emerging Technologies Unit of DG Information Society at the European Commission in Brussels. He has been Acting Professor at the Universities of Goettingen and Heidelberg and a Fellow of the Netherlands Institute for Advanced Studies. He has published widely in Philosophy of Science and Technology, especially in Intertheory Relations, in Logic and Epistemology, and in Formal Semantics for Logic Programming and Nonmonotonic Reasoning. His book on relations between disparate ontologies and conceptual schemes in science, 'Roads to Commensurability', was published by Kluwer in 1988.

Key words: Topic Maps, Addressability, Knowledge Networks, Semantic Connectivity.

Rudi Studer

Institute AIFB & FZI Research Center for Information Technologies, University of Karlsruhe, DE
http://www.aifb.uni-karlsruhe.de/WBS
http://www.fzi.de/wim
http://www.ontoprise.de

Short CV
Rudi Studer is Full Professor in Applied Computer Science at the University of Karlsruhe. His research interests include knowledge management, Semantic Web applications, knowledge portals, ontology engineering and knowledge discovery. He obtained a Diploma in Computer Science at the University of Stuttgart in 1975. In 1982 he was awarded a Doctor´s degree in Mathematics and Computer Science at the University of Stuttgart, and in 1985 he obtained his Habilitation in Computer Science at the University of Stuttgart.
From January 1977 to June 1985 he worked as a research scientist at the University of Stuttgart. From July 1985 to October 1989 he was project leader and manager at the Scientific Center of IBM Germany.
Rudi Studer is also leader of the Knowledge Management Group at the FZI Research Center for Information Technologies at the University of Karlsruhe as well as co-founder of the spin-off company ontoprise GmbH that develops technologies for the Semantic Web. He is a member of the steering committees for the conference series EKAW and SWWS as well as a member of the Executive Management Board of the IST project OntoWeb.

Key words:
Semantic Web technologies and applications, knowledge portals, knowledge management, ontology engineering, knowledge discovery

Abstract: Semantic Web for Generalized Knowledge Management
Knowledge is one of the most crucial success factors for enterprises. Therefore, Knowledge Management (KM) has been recognized as a very important strategy for enterprises. Clearly, KM is an interdisciplinary task, including human resource management, enterprise organization and culture as well as IT technology. However, there is a clear agreement, that IT technology plays an important role as an **enabler** for many aspects of a KM solution.

In the past, IT technology for knowledge management has focused on the management of knowledge containers using text documents as the main repository and source of information. Nowadays, Semantic Web technology, especially ontologies and relational metadata, pave the way to KM solutions that are based on semantically related knowledge pieces of different granularity: Ontologies define a shared conceptualization of the application domain at hand and provide the basis for defining metadata that have a precisely define semantics and are therefore machine-processable. Although first KM approaches and solutions have shown the benefits of ontologies and related methods there exists still a collection of open research issues that have to be addressed in order to make Semantic Web technologies a complete success for KM solutions:

- Industrial KM applications have to avoid any kind of overhead as far as possible. Therefore, a **seamless integration** of knowledge creation, e.g. metadata creation, and knowledge access, e.g. querying or browsing, into the working environment is required. Strategies and methods are needed that support the creation of knowledge as side-effects of activities that are carried out anyway.
- Access to as well as presentation of knowledge has to be **context-dependent** and **personalized**. Only then, information overload can be avoided, a prerequisite for keeping a KM application alive.
- In real life applications the manual engineering of ontologies is a crucial bottleneck. Furthermore, KM applications have to reflect their changing business environment, and so do ontologies. Cooperative ontology engineering environments that integrate tools for the manual construction of ontologies with machine learning and information extraction methods for generating parts of ontologies seem to be a promising approach. Yet, **ontology learning** is still a basic research topic.
- In the same way, the manual specification of metadata is not feasible in practice. Therefore, information extraction capabilities based on (simple) linguistic methods have to be combined with ontologies in order to semi-automatically **generate** the required **metadata**.
- **Semantically defined metadata** are not only required for text sources, but also for **multimedia sources**. Therefore, metadata standards for multimedia sources that are up to now more syntactically oriented, have to be combined with metadata standards that are available for text sources.
- Currently, ontology-based KM solutions are typically based on single ontologies. However, most real life scenarios will require the usage of **multiple ontologies**. Therefore, methods and tools are required for reusing ontologies that are stored in repositories somewhere on the web and for aligning them on the fly according to the given application scenario at hand. **Reuse and dynamic semantic alignment** of ontologies are both open research topics.
- Peer-to-Peer computing combined with Semantic Web technology will be an interesting path to get rid of the more centralized environments that are currently implied by ontology-based solutions. P2P scenarios open up the way to derive consensual conceptualizations in a bottom-up manner. However, the development of the methods and tools that are needed for achieving such **emergent semantics** in P2P environments is in the very beginning.
- KM solutions will in the near future be more and more based on a combination of intranet-based functionalities and mobile functionalities. Semantic Web technologies are a promising approach to meet the needs of the **mobile environments**, like e.g. **location-aware personalization** and adaptation to **presentation needs** of mobile devices, i.e. the presentation of the required information on an appropriate level of granularity.
- The development of knowledge portals serving the needs of companies or communities is still a more or less manual process. Ontologies and related metadata provide a promising conceptual basis for **generating** parts of such **knowledge portals**. Obviously, conceptual models of the domain, the users and the tasks are needed among others.
- **Virtual organizations** become more and more important in business scenarios that are characterized by decentralization and globalization. Obviously, semantic interoperability as well as trust is a must in inter-organizational business processes. Semantic Web technology provides a promising starting point for addressing these challenges.
- The **integration** of KM applications, e.g. skill management, with **eLearning** is an important field enabling a lot of synergy between these to areas. KM solutions and

eLearning have to be integrated from both an organizational and an IT point of view. Clearly, interoperabiltiy and/or integration of (metadata) standards are needed to realize such an integration.

Simon Buckingham Shum

Knowledge Media Institute, The Open University, UK
sbs@acm.org
http://www.kmi.open.ac.uk/sbs

Short CV

Dr Simon Buckingham Shum is a Senior Lecturer at the Knowledge Media Institute at the Open University in England, a 70-strong R&D centre which specialises in web, media and knowledge technologies. His first degree was in Psychology (University of York, 1987), followed by an M.Sc. in Ergonomics (UCL, 1988) and a PhD in Psychology/Human-Computer Interaction (York, 1991). As well as teaching on Knowledge Management technologies, Dr Buckingham Shum researches into the cognitive and group dynamics that influence whether tools for conceptual modelling are used effectively. He has worked since 1989 on concept and argument mapping tools, as applied to design rationale and organizational memory. He also works on the future of Net-based scientific publishing and discourse, as Co-Founder and Editor of the electronic *Journal of Interactive Media in Education* [www-jime.open.ac.uk] which pioneered interactive peer review debates, and also leading the *Scholarly Ontologies* project [www.kmi.open.ac.uk/projects/scholonto] for mapping research argumentation as semantic networks. He serves on the editorial board of the *International Journal of Human-Computer Studies*, and on the programme committees of conferences such as ACM Hypertext, Euro-CSCL, and BCS Human-Computer Interaction. He is co-editor of a forthcoming book (Springer) on Computer-Supported Collaborative Argumentation.

Key words: Human-Centred Design, Communities of Practice, Argumentation, Organizational Memory, Knowledge Media

Abstract:

"We're at the stage of figuring out how to mix tarmac that will hold together… it's going to be a while before we have good models for urban planning and traffic flow…" The Semantic Web seems to mean many things to many people. My own vision, drawing inspiration from Doug Engelbart's work, is of a global information infrastructure that people trust, and that augments intellectual work. Machine interoperability will doubtless assist this, but while we figure out how to 'get the tarmac to hold together', understandably, rather less is being said about how the people (for whom these services are being developed) will fit into and mould this infrastructure to their own ends. Like many others, I am developing technologies for people to create and manipulate conceptual structures, and envisage Semantic Web technologies as part of this. If so, Semantic Web research must begin to wrestle with a multitude of hard questions concerning the design of useful, usable, trusted formalisms and modelling techniques – questions that research communities have been wrestling with for years in hypertext, computer-supported collaborative work, human-computer interaction and knowledge-based systems. I will highlight questions that a human-centric perspective on formalisms and work practice raise.

- "Unnatural acts": the SemWeb will require vast amounts of knowledge acquisition and careful codification. Who will (be paid?) to do this? We know that normal people dislike creating metadata – that's why we have librarians. And we know that machines can extract metadata and populate ontologies at best in semi-automated manner.

- Distributed group working: metadata changes its meaning over time and across groups. How to manage this?
- Ontologies are all about classification. We know that taxonomies are embedded in perspectives, and often the subject of wars and compromises to establish a ‘standard’. How to represent the contextual factors that will support coherent maintenance and updating?
- How do we design SemWeb applications in a way that fully involves domain experts and end-users, when for instance, they don’t understand the technicalities of ontologies, nor how they may support inference?
- SemWeb applications model people and work more explicitly than previous technologies. How to make them trustworthy, and prevent them from getting out of sync in a fast changing world?

Dr. Bhavanai Thuraisingham

The MITRE Corporation + National Science Foundation, USA
bthurais@nsf.gov
http://www.interact.nsf.gov/cise/contact.nsf/53f032aef5157cbf852565db006fc19f/e74e3542ea4a2a0685256ae900445ab0?OpenDocument

Short CV

Dr. Bhavani Thuraisingham is the Program Director for Data and Information Management at the National Science Foundation. She is on IPA from the MITRE Corporation where she is the chief scientist in data management. She is the recipient of IEEE Computer Society's 1997 Technical Achievement Award for her research in secure data management. Her current research interests are in data mining and XML security.

Key words: Secure semantic web, XML security

Abstract: Secure Semantic Web

Semantic web is here to stay. Researchers are developing various technologies including agents, ontologies and information management to make the semantic web a reality. Since the semantic web will be the means of communication in the future, it is important that it be secure. This presentation will examine various web security and XML security efforts and focus on directions for secure access control to the semantic web. It will also focus on protecting then privacy of the individuals accessing the semantic web.

Frank van Harmelen

Vrije Universiteit Amsterdam, NL
Frank.van.Harmelen@cs.vu.nl

Short CV

Frank van Harmelen (1960) studied mathematics and computer science in Amsterdam. In 1989, he was awarded a PhD from the Department of AI in Edinburgh for his research on meta-level reasoning. After holding a post-doc position at the University of Amsterdam, he accepted a lectureship at the Vrije Universiteit Amsterdam. He is currently being appointed as full professor in the AI Department, where he heads the Knowledge Representation and Reasoning research group. He is author of a book on meta-level inference, and editor of a book on knowledge-based systems. He has published over 60 papers, many of them in leading journals and conferences. He has made key contributions to the CommonKADS project by providing a sound formal basis for the conceptual models. More recently, he has been co-projectmanager of the On-To-Knowledge project, and was one of the designers of OIL, which (in the form of DAML+OIL) is currently the basis for a W3C standardised Web ontology language. He is a member of the joint EU/US committee on agent markup languages (who are designing DAML+OIL), and a member of the W3C working group on Web Ontology languages.

Key words: Knowledge Representation, Approximate inference, web-based ontology languages

Abstract:

A strength of the current proposals for the foundational languages of the Semantic Web (RDF Schema, DAML+OIL), it that they are all based on formal logic. However, this reliance on logics is not only a strength but also a weakness. Traditionally, logic has always aimed at modelling idealised forms of reasoning under idealised circumstances. Clearly, this is not what is required under the practical circumstances of the Semantic Web. Instead, the following are all needed:

- reasoning under time-pressure;
- reasoning with other limited resources besides time;
- reasoning that is not ``perfect'' but instead ``good enough'' for given tasks under given circumstances;
- reasoning-algorithms that do not behave as yes/no oracles, but that instead display anytime behaviour;
- etc.

It is tempting to conclude that symbolic, formal logic fails on all these counts, and to abandon that paradigm. However, research in the past few years, to which members of my group and myself have contributed, has developed methods with the above properties while staying within the framework of symbolic, formal logic. These methods carry names such as approximate reasoning, anytime inference, knowledge compilation, etc.

FET - Future and Emerging Technologies

DIMACS — Center for Discrete Mathematics & Theoretical Computer Science

European Research Consortium for Informatics and Mathematics

ERCIM

www.ercim.org

This workshop is part of a series of strategic workshops to identify key research challenges and opportunities in Information Technology. These workshops are organised by ERCIM, the European Research Consortium for Informatics and Mathematics, and DIMACS the Center for Discrete Mathematics & Theoretical Computer Science. This initiative is supported jointly by the European Commission's Information Society Technologies Programme, Future and Emerging Technologies Activity, and the US National Science Foundation, Directorate for Computer and Information Science and Engineering.

More information about this initiative, other workshops, as well as an electronic version of this report are available on the ERCIM website at http://www.ercim.org/EU-NSF/