The Semantic Web:
from Concept to Percept

Ying Ding[1], **Dieter Fensel**[1] & **Hans-Georg Stork**[2,*]

[1] Institut für Informatik, University of Innsbruck, Technikerstrasse 25, 6020 Innsbruck, Austria

[2] European Commission, Directorate General Information Society,
rue Alcide de Gasperi, 2920 Luxembourg, G.D. Luxembourg

Abstract. We give a brief overview of a new and exciting research area: the Semantic Web. We introduce some of its main ideas and application domains (with an emphasis on e-business and e-commerce), and report on relevant public funding initiatives in Europe and the US. Finally, we discuss current challenges. The paper also contains a short bibliography as well as pointers to pertinent web sites and European Semantic Web projects.

1   The Concept

> *„The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in co-operation." – Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001*

The computer has been invented as a device for computation. Today it is also a portal to cyberspace, an entry point to a world-wide network of information exchange and business transactions. The Internet and its most popular application, the World Wide Web, have brought about this change. The World Wide Web in particular, is an impressive success story, both in terms of the amount of available information and the number of people using it. It has reached a critical mass very rapidly and now starts to penetrate most areas of our daily life as citizens and professionals. Its success is based largely on the simplicity of its underlying structures and protocols which make for easy access to all kinds of resources. However, this simplicity may hamper the further development of the Web. Indeed, as will be explained below, current web technology has severe shortcomings. While it works well for posting and rendering content of all sorts it can provide only very limited support for processing web contents. Hence, the main burden in searching, accessing, extracting, interpreting, and processing information rests upon the human user.

Tim Berners-Lee created the vision of a **Semantic Web** that enables automated information access and use, based on machine-processable semantics of data. In his informal 'Semantic Web Road Map' note[1] he outlined possible future directions for the evolution of the World Wide Web. These ideas, partly based on previous content and resource description activities, have met with growing enthusiasm of researchers and developers world-wide, both in academia and in industry (cf. [Fensel et al., 2002]). They encourage the integration of efforts that have been ongoing for some time in many R&D communities, involving specialists in various computer science disciplines. These efforts are aimed at capturing the semantics of digital content of all sorts and origins, and at devising ways of acting sensibly upon the formal knowledge representations thus gained.

The explicit representation of the semantics of data, grounded in domain theories (i.e., ontologies, see below), will enable a qualitatively new level of service. It will weave

---

[*] The views expressed in this note are those of the authors and do not necessarily engage the European Commission.
[1] http://www.w3.org/DesignIssues/Semantic.html

together a huge network of human knowledge, complement it with machine processability, and allow for automated services that support people (from all walks of life) in carrying out tasks that are contingent on the expedient use of information and knowledge. Access to these services may become as crucial as access to electric power (cf. [Fensel, 2001]).

Ontologies are the backbone technology for the Semantic Web and - more generally - for the management of formalised knowledge within the technical context of distributed systems. They provide machine-processable semantics of data and information sources that can be communicated between different agents (software and people). Many definitions of "ontology" have been given in the past. One that in our opinion, best characterises its essence has been proposed in [Gruber, 1993]: *An ontology is a formal, explicit specification of a shared conceptualisation.* A 'conceptualisation' refers to an abstract model of some phenomenon in the world which identifies the concepts that are relevant for that phenomenon. 'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine-readable. Different degrees of formality are possible. Large ontologies like WordNet ([Fellbaum, 1998]) comprise a thesaurus of over 100,000 natural language terms explained in natural language. At the other end of the spectrum is CYC ([Lenat, 1995]), that provides formal axiomatic theories for many aspects of common sense knowledge. 'Shared' reflects the notion that an ontology captures consensual knowledge: consensus is usually reached through co-operation and communication between different people who share the same or similar interests. People who agree to accept an ontology are said to *commit* themselves to that ontology. Basically, the role of ontologies in the knowledge engineering process is to facilitate the construction of a domain model. An ontology provides the required vocabulary of terms and the relations among them.

Ontologies and other technologies underlying the Semantic Web support access to unstructured, heterogeneous and distributed information and knowledge sources. They are now as essential as programming languages were in the 60's and 70's of the 20th century.

2   Application Areas

No technology can survive without convincing applications. In this section we will sketch three broad application areas. However, the reader should also be aware of the time span of innovation. For example, it took 30 years before the most prominent application of the Internet, the World Wide Web, came of age. Perhaps the kinds of application we are going to discuss will lead to major breakthroughs much faster.

2.1   Knowledge Management

The competitiveness of companies in fast changing markets depends largely on how they exploit and maintain their knowledge (their corporate memory). Most information in modern electronic media, be it on the Internet or on large company intranets, is rather weakly structured. Finding relevant information and maintaining it is difficult. Yet, more and more companies realise that their intranets could be valuable repositories of corporate knowledge. However, raw information does not by itself solve business problems; it is useless without an understanding of how to apply it effectively. Turning it into useable knowledge has become a major problem.

Corporate knowledge management is about leveraging an organisation's data and information sources for innovation, greater productivity, and competitive strength. Due to the globalisation and the impact of the Internet, many organisations are increasingly

geographically dispersed and organised around virtual teams. Such organisations need knowledge management and organisational memory tools that enable users to understand better each other's changing contextual knowledge, and that foster collaboration while capturing, representing and interpreting the knowledge resources.

Of course, knowledge management has dimensions that take it way beyond the needs of commercial enterprises. This concerns in particular scientists, scholars, educators and other professionals, and their specific knowledge resources (of all kinds of media). Yet the basic problems these *communities of practice* face when it comes to creating and exploiting their resources, are quite similar to those that "knowledge workers" in companies big and small have to tackle.

A fair number of knowledge management systems are already on the market, designed to deal with operations of relevance to the "knowledge lifecycle" within a given organisation or community of practice. However, these systems still have severe limitations, e.g.:

- **Searching information**: Existing keyword-based search retrieves irrelevant information due to term ambiguity, and misses information when material related to similar concepts is stored under quite different terms.

- **Extracting information**: Currently, people have to browse and read extensively in order to extract relevant information from textual or other representations. Software agents do not possess the common-sense knowledge required to assist effectively in tasks of this type, let alone automate them. Moreover, they fail to integrate information from different sources.

- **Maintaining** large repositories of **weakly structured text** is a difficult and time-consuming activity.

- Adaptation and dynamic reconfiguration of information repositories (e.g. websites) according to user profiles or other aspects of relevance, hinges on **automatic document generation** and is not yet fully mastered.

Semantic Web technologies and especially the use of ontologies, are expected to enable a much higher degree of automation and scalability in performing operations pertaining to the above mentioned tasks. For instance, in order to keep weakly structured collections consistent, or to generate information presentations from semi-structured data, the semantics of these collections and data must not only be machine-accessible but also machine-processable. In other words, the semantics must be represented based on formal ontologies.

2.2   Enterprise Application Integration

For a number of reasons the integration of data, information, knowledge, processes and applications within businesses becomes more and more important, e.g.:

- Company mergers often require large-scale integration of existing information technology (IT) infrastructures;

- within existing corporate IT infrastructures new software solutions often have to integrate existing legacy software;

- for reasons of cost and quality a company may decide to adopt products (e.g. for Customer Relationship Management/CRM and Enterprise Resource Planning/ERP) from different vendors; these products need to work together;

- companies are forced to adapt to ever changing IT standards.

Recent studies by Gartner and Forrester estimate that a significant share of future IT budgets will be spent on Enterprise Application Integration tasks. This may seriously hamper progress in IT: if most of a company's resources are spent on integrating existing solutions little is left to develop new approaches.

Up until now, many companies have been trying to meet their integration needs through adhoc projects. Adhoc integration, however, does not scale. Global integration platforms on the other hand, require major investments and are often likely to fall behind the current state-of-the-art very fast.

A successful integration strategy must combine the advantages of adhoc and global integration. It must be driven by business needs (identified in terms of business processes and available information sources) but also address the all important issues of extendibility and reusability:

- **Extendibility** can be achieved through the use of ontologies to prevent adhoc integration and to ensure that the integration effort can be extended in response to new and changing business needs. Ontologies provide the necessary controlled terminologies based on structured and well-defined domain theories.

- **Reusability** is greatly enhanced through the use of web service technology (see also Section 4) in combination with ontologies to meet further integration needs based on standardisation.

We expect that Semantic Web technologies will greatly benefit Enterprise Application Integration before they are successfully applied to tackling problems at the next higher level: the integration of several organisations, for instance in eCommerce environments.

2.3   eCommerce

eCommerce in business-to-business (B2B) relationships is not new. Initiatives to support electronic data exchange in business processes between different companies already existed in the 60's of the last century. To perform business transactions sender and receiver had to agree on common content formats and transmission protocols. In general, however, these arrangements did not live up to the expectations of their proponents: establishing an eCommerce relationship required a major investment and it was limited to a predefined number of trading partners, connected via a specific type of extranet.

Since then, the Internet and the World Wide Web have drastically increased the online availability of data and the amount of electronically exchangeable information. Internet-based electronic commerce now allows for more *openness, flexibility* and *dynamics*. This will help to improve business relationships in many ways, e.g.:

- Instead of implementing one link per supplier, a supplier can be linked to a marketplace with a large number of potential customers.

- Consequently, suppliers and customers can choose between a large number of business partners, and

- they can update their business relationships as the markets evolve.

In a nutshell, web-based eCommerce makes it possible to contact a large number of potential clients without running into the problem of having to implement as many communication channels. Hence, virtual enterprises can form in reaction to demands from the market and large enterprises can break up into smaller units, mediating their eWork relationship based on eCommerce relationships.

Achieving the desired level of openness and flexibility is not an easy task. The integration of various hardware and software platforms and the provision of a common protocol for information exchange might in fact be among the lesser problems to be solved. The real problems are in the openness, heterogeneity (in terms of *product*, *catalogue*, and *document* description standards) and dynamic nature of the exchanged content.

Openness of eCommerce cannot be achieved without standardisation, a lesson learnt from the success of the web. In eCommerce, however, the requirements on standardisation are much stricter: they extend to the actual content exchanged and thus go far beyond the requirement of standardising protocols and document layouts.

Flexibility of eCommerce cannot be achieved without multi-standard approaches. It is unlikely that a single standard acceptable to all vertical markets and cultural contexts, and covering all aspects of eCommerce will ever arise. And in any event a standard does not free us from the need to provide user-specific views on it and on the content it represents.

Dynamism of eCommerce requires standards to be like living entities. Products, services, and trading modes are subject to frequent change. An electronic trading arrangement must reflect the dynamic nature of the processes it is supposed to support.

Again, given these requirements, ontologies and other Semantic Web technologies are the most likely candidates to provide viable eCommerce solutions:

Ontologies span networks of meaning where heterogeneity is an essential feature. Tools for dealing with conflicting definitions, as well as strong support for interweaving local theories, are essential in order to make this technology work and scale.

Ontologies are used to exchange meaning between different agents. By definition (cf. Section 1) an ontology is itself the result of a social process. Therefore, it is not static. While ontologies are required in order to exchange meaning the very exchange of meaning may impact on an ontology. Ontologies evolve. Hence, capturing the time dimension is an essential requirement if ontologies are to be useful mediators of the information needs of eCommerce processes. It follows that ontologies must have strong versioning support and the underlying process models should cater for evolving consensus.

3   Making it happen

Commercial interest in applying Semantic Web technologies and in particular ontologies, to Knowledge Management, Enterprise Application Integration and eCommerce, is strong. Indeed, ever since Tim Berners-Lee, the director of the World Wide Web Consortium (W3C), set the ball rolling (see Section 1) work at the W3C has been gearing up on a range of pertinent recommendations for setting the formal framework of the Semantic Web (such as the XML and RDF families and, more recently, OWL, the web ontology language), and on evangelising them at conferences and major business events.

However, the technologies at issue are still at a pre-competitive stage. Their large scale deployment (on the Internet or on large private intra- and extranets) still requires substantial research. And whether or not the "Semantic Web" is going to repeat the Web's success story of its own accord is still an open question: Its underlying concepts are after all not so easy to grasp, and their potential benefits (e.g. in terms of creating mass markets, increased productivity, etc.) are not so easy to sell given the current perceived slump in the IT industry (and online business in particular).

Moreover, a critical mass problem has to be solved: Adding explicit semantics to content, processes and services does not pay off if no tools are available to make good use of it;

developing tools, on the other hand, does not pay off if there is little semantically-enriched content to work on.

So, although expectations are high and most players in the field agree on the enormous potential of these technologies, it is not clear whether commercial interest alone will bring about the momentum necessary for them to become a success.

This is the classical setting in which public funding can provide the incentives required to advance research and development up to a point where one can "let things take their course". (One may note that the Internet itself whose initial development depended largely on public funding, is a case in point.)

In this section we give a brief and non-exhaustive account of a number of public funding initiatives in Europe and the United States of America, that focus on the technologies at issue.

3.1 European initiatives

The European Commission's Information Society Technologies (IST) Programme was designed in 1997-1998 as part of the 5th Framework Programme for R&D in Europe, covering the period 1998 - 2002. One of its Key Action III ('Multimedia Content and Tools') modules was entitled 'Information access, filtering, analysis and handling' (IAF for short). Its objectives were to support the development of *"...advanced technologies for the management of information content to empower the user to select, receive and manipulate ... only the information required when faced with an ever increasing range of heterogeneous sources."* These technologies should lead to *"... improvements in the key functionalities of large-scale multimedia asset management systems (including the evolution of the World Wide Web) to support the cost-effective delivery of information services and their usage."*[2]

Picking up on the Semantic Web vision outlined in Section 1, the European Commission dedicated a specific action line of its IST Work Programme 2001 to 'Semantic Web Technologies' (Action Line III.4.1), thereby underlining the importance (in terms of research challenges and expected impact) of 'semantics issues' for achieving the declared goals of the IAF module of IST.

It offered four broad interrelated R&D tracks as an orientation for submitting project proposals:

- creating a usable formal framework in terms of formal methods, models, languages and corresponding tools for semantically sound machine-processable resource description

- fleshing out the formal skeletons by developing and applying techniques for knowledge discovery (in databases and text repositories), Ontology learning, multimedia content analysis, content-based indexing, etc.

- acting in a semantically rich environment, performing resource and service discovery, complex transactions, semantic search and retrieval, filtering and profiling, supporting collaborative filtering and knowledge sharing, etc.

- making it understandable to people through device-dependent information visualisation, semantics-based and context-sensitive navigation and browsing, semantics-based dialogue management, etc.

---

[2]Quoted from http://www.cordis.lu/ist/b-oj-en5.htm#KA3

This agenda provided some continuity with respect to previous Key Action III activities (notably on 'media representation and access' and digital libraries) and activities supported by other IST departments (for instance under action line 'Methods and tools for intelligence and knowledge sharing' of Key Action IV - Essential Technologies and Infrastructures, under Key Action II - New Methods of Work and Electronic Commerce - or under the Open Domain of FET - Future and Emerging Technologies).

But it also provided a sharper focus on the problems of creating and using knowledge representations, in the context of large-scale distributed systems, such as the World Wide Web. By not specifically addressing the activities supported directly by the W3C, however, it allowed for a wider scope than the original formal and informal Semantic Web notes issued by the W3C, might have insinuated. This scope included problems such as the automatic or semi-automatic creation of semantic annotation of all forms of content and resources (thus creating a link to multimedia resource description), or for instance, ontology learning in peer-to-peer systems.

Focus and scope were largely retained in Work Programme 2002 as part of Key Action III's 'Preparing for future research activities' action line (AL III.5.2). Moreover, Work Programme 2002, in one of its 'Cross Programme Activities (CPA)', took account of a new trend that has surfaced over the last couple of years: the application of Grid technologies (see also Section 4) to "knowledge discovery in ... large distributed datasets, using cognitive techniques, data mining, machine learning, Ontology engineering, information visualisation, intelligent agents..."[3], all more or less directly pertinent to the Semantic Web vision.

Calls for submission of proposals to these action lines were published in July (AL III.4.1) and November (AL III.5.2 & CPA9) 2001, respectively (Calls 7 and 8). Both calls drew altogether nearly one hundred submissions involving several hundred participating organisations. They resulted in a significant growth (by 17 projects) of a portfolio of projects that are all poised to contribute in one way or other, to making the "Semantic Web" happen (see the references in Section 5.3, including a few earlier and concurrent projects in Key Actions II, IV and in FET).

While at the time of writing the new projects have only just commenced, some of the older ones have already produced noteworthy results. It may suffice to mention projects *On-To-Knowledge* and *Ibrow*, probably the first Semantic Web projects ever to receive public funding (in Europe if not in the world). *On-To-Knowledge*[4] has become one of the birthing grounds of OWL, the proposed new Web Ontology language, currently under discussion at the W3C. *Ibrow (An Intelligent Brokering Service for Knowledge-Component Reuse on the World Wide Web)*[5] already started in 1997 when the terms "Semantic Web" and "Web Services" had not yet been coined or widely used. Perhaps one of its best known deliverables is UPML (Unified Problem-solving Method Development Language), a *"framework for developing knowledge-intensive reasoning systems based on libraries of generic problem-solving components"*.

In recognition of the central role ontologies are likely to play in building the Semantic Web', the European Commission, through its IST Programme, supports the 'Thematic Network' *OntoWeb*[6], a platform for fostering collaboration between industry and academia, on creating a 'semantic infrastructure' for applications in many different areas (e-business, Web services, multimedia asset management, community webs, etc.). Through OntoWeb,

---

[3] Quoted from WP2002, CPA9

[4] http://www.ontoknowledge.org/

[5] http://www.ibrow.org/

[6] http://www.ontoweb.org/

European researchers and practitioners also have an opportunity to make more targeted contributions to international standardisation activities and to the W3C process.

We note that the "Semantic Web Technologies" action line did not prescribe particular application domains. Its very title made this quite explicit. Yet, as explained above (Section 2), technologies must not be developed for the sake of developing technologies. Proposers were therefore advised to make sure their projects would not benefit a limited constituency only, or solve just one isolated problem. Rather, projects submitted under a generic action line should, in a final analysis, yield more widely applicable results, to be demonstrated through several showcases. This same 'principle of neutrality' regarding applications also holds *mutatis mutandis* for Key Action IV and FET (see above). It does not hold for Key Action II where projects were indeed required to focus on particular application domains, which could be broadly described as corporate knowledge management and eCommerce.

This being said, it is an interesting exercise to categorise the projects listed in Section 5.3 (and possibly more, of course) roughly along (at least) four dimensions: (i) generic problem class (such as: 'making semantics explicit' and 'acting upon explicit semantics'), (ii) technical solutions (e.g. automatic versus semi-automatic and interactive tools), (iii) type of content (e.g. text, corporate databases, multimedia objects, web pages, man-machine interaction records, etc.) and (iv) application domain.

A discussion of the first two of these can be found in [Stork, 2002]. Applications are in the areas broadly delineated in Section 2 of this note. They range from 'hard science' via engineering, education, training and infotainment (along the lines of Section 2.1), to enterprise application integration and eCommerce (as explained in Sections 2.2 and 2.3). Contents vary widely and are of course to some extent related to the application domains. The degree of multidisciplinarity of these projects depends on all of the above dimensions.

Apart from these more 'technical' dimensions there are political and economic ones. The 'European dimension' for instance, that can be expressed inter alia in terms of a project's perceived contribution towards achieving goals such as the ones proclaimed at the European Council summit in Lisbon[7], in early 2000; these goals have informed the eEurope Action Plan[8] that aims to turn Europe by the end of this decade into the world's most advanced knowledge-based society. And the IST programme is indeed seen as a key component of that plan.

While we are not prepared to delve into the political (and policy oriented) aspects in any detail it might be worth taking a look at the "economic" side. A preliminary assessment of all projects funded under the IST Programme 1998-2002 identified well over 350 projects emerging from the first three IST Calls, that address in one way or other the applications and some of the technologies this note is about (such as knowledge and information management, agent technologies, optimisation tools and decision support systems, supply chain management and generic organisational tools)[9].

Clearly, by far not all of these projects can be classified as "Semantic Web technologies and applications". Moreover, membership in that class is a fuzzy relation. Hence it is not easy to establish the total amount of EU funding allocated to it. A rather conservative estimate would be somewhere between 100 and 150 million Euro. We do have more precise figures though, for the focused Calls for Proposals under Action Lines III.4.1 (Semantic Web Technologies), III.5.2 and CPA9 (see above). The projects that emerged

---

[7] http://ue.eu.int/newsroom/LoadDoc.asp?MAX=1&BID=76&DID=60917&LANG=1

[8] http://europa.eu.int/information_society/eeurope/action_plan/pdf/ actionplan_en.pdf

[9] http://www.cordis.lu/ist/cpt/ippa.htm

from these Calls plus a number of pertinent projects (such as On-to-knowledge, Ibrow, Wonderweb and others) selected earlier or under other Action Lines (see the list in Section 5.3) represent an EU contribution of well over 45 million Euro. As EU monies usually cover only about 50% of the total cost of a project the true 'weight' of the projects listed is in the order of 90 million Euro.

It is expected that EU support of Semantic-Web-related R&D will continue under the forthcoming 6th Framework Programme[10] within the broader context of 'Knowledge Technologies', as part of the 'Priority Thematic Area' IST - Information Society Technologies. The overall agenda includes work *"on technologies to support the process of modelling and representing, acquiring and retrieving, navigating and visualising, interpreting and sharing knowledge"* and it addresses *"extensible knowledge resources and ontologies so as to facilitate service interoperability and enable next-generation Semantic-web applications"* [IST, 2002].

European national initiatives: The European Commission's research funding programmes encourage co-operation at the European level. Thus they fulfil an eminent role in realising the political vision of a United Europe. Yet one must not forget that the bulk of public European RTD funding is still managed by national authorities (of EU Member States and others) who are of course free to define their priorities according to their perceived needs. For many, the issues related to the Semantic Web do indeed rank high on their agenda, mostly implicitly but in some instances also explicitly. We give four examples:

Ireland: The *"Informatics Research Initiative"* (Irish National Informatics Directorate[11]), currently supports five projects in the 'Digital Media' and eight projects in the 'e-Business' domains; most of these also address issues discussed in this note. Moreover, a government funded research institute on Semantic Web Services will be set up shortly.

Germany: The IT2006[12] programme of the German federal government lists "Intelligent Systems and Knowledge Processing", "Knowledge Networking" and "Internet-based Business Processes" among its priority themes.

Austria: Through its *K-Plus* programme (for 'Competence Centres', that would improve academia-industry cross-fertilisation) the Austrian government has helped establishing the *Know-Centre*[13], for knowledge-based applications and systems. A more targeted Semantic Web initiative is currently in preparation. First projects could be scheduled to start in 2004.

United Kingdom: *CoAKTinG (Collaborative Advanced Knowledge Technologies in the Grid)*[14], one of the large *Interdisciplinary Research Collaborations (IRCs)* within the British *eScience*[15] programme, is poised to harness ontologies and other 'knowledge technologies' to enhance scientific collaboration. It may be worth noting that the *eScience* programme is very much centred around the notion of the 'Grid' with a clear emphasis on 'Grid services'. We shall pick up on this aspect again briefly in Section 4 of this article.

An excellent source of more information on Semantic Web related R&D in Europe (both at the national and European levels) is [Euzenat, 2002][16] containing reports on relevant projects, big and small, undertaken at publicly funded research centres throughout Europe.

---

[10] http://www.cordis.lu/fp6/
[11] http://www.nsd.ie/htm/home/home.php3
[12] http://www.it2006.de/
[13] http://www.know-center.at/en/about/about.htm
[14] http://www.aktors.org/
[15] http://www.research-councils.ac.uk/escience/
[16] see also: http://www.ercim.org/publication/Ercim_News/enw51/

The European Commission's 6th Framework Programme will provide new funding instruments (e.g. 'Networks of Excellence') that will help co-ordinating and focusing these national R&D activities, in order to create a truly European research area.

3.2   US Initiatives

'Funding landscapes' reflect to a large extent prevailing political and economic constraints. This holds for Europe with its multinational structure but also for the US where there are a number of agencies, linked to different government departments. It is therefore no surprise that several agencies in the US have launched or are going to launch initiatives to develop Semantic Web technologies, sometimes - as in Europe - without explicitly refering to the concept itself. Broader contexts such as 'knowledge management' as well as more specific or overlapping notions such as 'agent technologies' are often used instead. We mention but a few of these initiatives.

The most prominent one is certainly DAML[17], the DARPA Agent Markup Language' programme, designed and brought on its way in mid-2000 by James Hendler, an early pioneer of the Semantic Web. DAML could build on the earlier SHOE (Simple HTML Ontology Extensions) project, run by Hendler and his group at the University of Maryland. As its name indicates DAML is supported by DARPA, the US defense department's research funding agency. It is a cluster of some 20 grants which together with those allocated to the more theoretically oriented companion programme TASK (Taskable Agent Software Kit), add up to some US$ 70 million, over a period of five years. DAML is supposed to *"create the technologies so that software agents can dynamically identify, communicate and understand each other"*. The project is now moving into its second phase where 'blue-sky research' should be turned into practical applications. There are currently some 60 researchers in the US and Europe (!) involved[18].

In fact, already before the official start of the programme DAML researchers had actively sought contact with European colleagues in order to benefit from transatlantic synergies. A result of this collaboration is the joint Web ontology language proposal OWL that is based on DAML and OIL, the 'Ontology Inference Layer' developed by the EU project On-to-knowledge (see above).

At least two further related DARPA funding initiatives ought to be mentioned: the *High-Performance Knowledge Bases (HPKB)[19]* programme (now completed, cf. [Cohen et al., 1998]) and and its follow-up, the *Rapid Knowledge Formation (RKF)[20]* project. HPKB aimed to advance the ways computers can acquire, represent and manipulate knowledge. The key objective of RKF is *"to enable distributed teams of subject matter experts (SMEs) to enter and modify knowledge directly and easily, without the need for specialised training in knowledge representation, acquisition, or manipulation"[21]*. The applications envisaged will be commensurate with DARPA interests.

While DARPA programmes and projects tend to be rather focused the 'other' very large US funding agency, the National Science Foundation (NSF), offers more widely scoped opportunities. We note, however, that the latest ITR (Information Technology Research) Call (for Fiscal Year 2003) gives harnessing knowledge and information in large-scale distributed systems top priority on its list of research challenges. Indeed, it calls for ...

---

[17]http://www.daml.org/

[18] http://www.daml.org/researchers.html

[19] http://reliant.teknowledge.com/HKPB/

[20]http://reliant.teknowledge.com/RKF/

[21] http://reliant.teknowledge.com/RKF/proposals/rkf_pip.html

... advancing fundamental research and the technical state of the art of IT and assessing its impacts on other fields of science and engineering, including:

- *Extending the capability to process, manage, and communicate information on a global scale beyond what we imagine today. This includes new paradigms for communication, networking and data processing in large-scale, complex systems.*
- *Understanding how to extend, or scale up, the network infrastructure to include an extremely large number of computing and monitoring systems, embedded devices and appliances.*
- *Exploring new research directions and technical developments to enable wide deployment of pervasive IT through new classes of ubiquitous applications and creation of new ways for knowledge acquisition and management.*
- *Exploiting the power of IT and networking infrastructures to enable robust, secure and reliable delivery of critical information and services anytime, anywhere, on any device.*[22]

We expect this Call to yield also projects that will address Semantic Web related issues.

We also note the keen interest the NSF has shown in relevant EU sponsored activities. By the same token, EU researchers are greatly interested in putting their co-operation with US colleagues on a more stable footing. This mutual interest became manifest in two recent invitational workshops the first of which[23] was held in October 2001, in France, under the title *"Research challenges and perspectives of the Semantic Web"*. It was organised by ERCIM[24], under the auspices of the European Commission's IST programme and the Computer and Information Science and Engineering (CISE) directorate of the NSF. A second workshop[25] was held in April 2002, in Georgia, USA, sponsored by NSF-CISE and the EU-funded 'Thematic Network' project Ontoweb (cf. Section 3.1). It focused on *"Database and Information Systems Research for Semantic Web and Enterprises"* and provided a platform for a lively debate on future research directions in the area (cf. [Meersman& Sheth, 2002]).

Such workshops may indeed help to further EU-US co-operation, as part of the world-wide effort that is needed to turn the Semantic Web into a viable global infrastructure for accessing and integrating content and services.

As in the case of the European IST programme it is difficult to assess the extent (especially in terms of budget) to which Semantic Web technologies have been taken up in past or ongoing NSF funded activities, given that the technology boundaries are often not very clear-cut. However, the *digital libraries*[26] series of projects, that began in the mid-90s as a joint NSF/DARPA/NASA undertaking and that is now in its second phase, can with some justification be regarded as a 'forerunner' to and 'companion' of current Semantic Web activities. They are contributing greatly for instance to research on ontologically grounded metadata and many other issues pertaining to the concept of the Semantic Web. There has been EU/NSF co-operation in this area as well, mainly through joint workshops co-organised by DELOS[27], an EU funded 'Thematic Network' project that supports RTD work on digital libraries.

---

[22] http://www.nsf.gov/pubs/2002/nsf02168/nsf02168.htm

[23] http://www.ercim.org/EU-NSF/semweb.html

[24] http://www.ercim.org

[25] http://lsdis.cs.uga.edu/SemNSF/index.htm

[26] http://www.dli2.nsf.gov/

[27] http://www.delos.org

4 Prospects: Semantic Web Services

Software programs that can be accessed and executed via the web provide *"web services"*. A service can consist in giving plain information, for example a weather forecast, or it may have an effect in the real world, for instance when booking a flight, ordering a book or transfering money to someone's account. Thus web services turn the 'static web (of displays)' into a **'web of action'**, and bring the computer back as a device for computation.

In a business environment this could translate into automatic co-operation between enterprises if mechanisms were in place that allow for automatic discovery, selection, composition and execution of appropriate web services, according to whatever policies may apply to the business transactions at issue. A concrete example would be the supply chain relationships of a manufacturer of short-lived goods with his suppliers and buyers. Instead of having people constantly search for business partners a suitable web service infrastructure would make it possible to do this automatically under defined constraints (see also Section 2.3).

The great potential of web services puts them at the centre of attention of software developers world-wide, and recent standardisation efforts such as UDDI[28], WSDL[29], and SOAP[30], for advertising, describing and invoking them, aim at providing a more stable platform for their deployment and use.

However, service descriptions are still given in semi-formal natural language terms. Therefore, the human programmer must be kept in the loop and the scalability as well as the economy of web services are limited. Semantic Web technology is poised to remedy this situation by providing the required semantic elements for

- public process description and advertisement;
- discovery, selection and composition of services;
- delivery, monitoring and contract negotiation.

These elements would enable efficient inter-enterprise execution of web services [Bussler, 2001]. Any necessary mediation would be based on data and process ontologies and their automatic translation into each other.

First attempts have been made to apply Semantic Web technology to web services. [Trastour et al., 2001] examine the problem of matchmaking, highlighting the features a matchmaking service should exhibit and deriving requirements on metadata for describing services from a matchmaking point of view. And one of the outputs of the DAML programme (cf. Section 3.2) is **DAML-S** that *"supplies Web service providers with a core set of markup language constructs for describing the properties and capabilities of their Web services in unambiguous, computer-intepretable form. DAML-S markup of Web services will facilitate the automation of Web service tasks including automated Web service discovery, execution, interoperation, composition and execution monitoring"*[31]. [Ankolenkar et al., 2001] describe the overall structure of the DAML-S ontology, the service profile for advertising services, and the process model for the detailed description of the operation of services.

---

[28] Universal Description, Discovery and Integration of web services (http://www.uddi.org/)
[29] Web Services Description Language (http://www.w3.org/TR/wsdl)
[30] Simple Object Access Protocol (http://www.w3.org/TR/SOAP/)
[31] http://www.daml.org/services/

The **Web Service Modelling Framework (WSMF)** [Fensel et al., to appear] follows this line of research. It is a fully-fledged framework for describing the various aspects related to web services. It is centred around two complementary principles:

- strong *de-coupling* of the various components that realise an eCommerce application;

- strong *mediation* service enabling anybody to 'speak' with everybody in a scalable manner.

These principles are rolled out in a number of specification elements and an architecture describing their relationships. A joint EU/US committee has been set up recently to align the US DAML-S and the "European" WSMF initiatives.[32]

We expect many interesting future developments at the intersection of Semantic Web and Web Service technology. Indeed, we believe **Semantic Web enabled Web Services** may change our lives even more drastically than the current Web did. A rather promising avenue for instance, has opened up with the rapprochement of the Semantic Web and Grid 'movements'[33]. 'Grids' are high-performance computing platforms based on networked computational resources (cf. [Foster et al., 1998]). Realising them requires some of the same technologies (e.g. for resource description) that are underlying the Semantic Web. On the other hand Grids will provide a host of services. Webs would be the 'natural' interfaces to them. In fact, an *Open Grid Services Architecture (OGSA)[34]* has been proposed that is to be *"based on an integration of Grid and Web services concepts and technologies"*.

5 Resources: Web sites, books and projects

What started around five years ago has by now produced a wealth of material. Hence, what follows can only be a small sample of pointers.

5.1 Links to Web Resources

- http://www.w3c.org/ (World Wide Web Consortium)

- http://www.daml.org/ (DARPA Agent Markup Language programme)

- http://www.semanticweb.org/ (a portal to Semantic Web related resources, run by Stefan Decker, Stanford)

- http.//www.ontoweb.org/ (portal of Ontoweb, the EU funded 'Thematic Network' on Ontology-based Information Exchange for Knowledge Management and Electronic Commerce)

- http://www.diffuse.org/ (portal of Diffuse, an EU funded 'accompanying measure', providing up-to-date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information)

- http://www.ktweb.org (portal of Multiple, an EU funded 'accompanying measure', with links to European projects and people working in information access and knowledge technologies, in research, academia and business, and to relevant resources.)

---

[32] http://swsc.semanticweb.org/
[33] see also: http://www.semanticgrid.org/
[34] http://www.globus.org/ogsa/

5.2   Books on Semantic Web technologies and applications

- T. Berners-Lee: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web, 1999.

- J. Davis, D. Fensel, and F. van Harmelen: Towards the Semantic Web: Ontology-Driven Knowledge Management, Wiley, 2002.

- D. Fensel, B. Omelayenko, Y. Ding, E. Schulten, G. Botquin, M. Brown, and A. Flett: Intelligent Information Integration in B2B Electronic Commerce, 2002.

- D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster (eds.): Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press, Boston, 2002.

- D. Fensel: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001.

5.3   A selection of IST projects on semantics in distributed systems

see also: http://www.ktweb.org/pintro.cfm

COG Corporate ontology grid (CPA9) Collaborative ontology development in a corporate environment (automotive industries); automatic scripting for transformation and query; creating 'virtual views' of data from disparate sources. (http://www.cogproject.org/)

CoMMA Corporate Memory Management through Agents (KA2) Multi-Agent System, based on semantic enterprise and user models, and ontologies, applied to Corporate Memory Management, using techniques for learning from user behaviour. (http://www.ii.atos-group.com/sophia/comma/HomePage.htm)

ESPERONTO Application Service Provision of Semantic Annotation, Aggregation, Indexing and Routing of Textual, Multimedia, and Multilingual Web Content (KA3) Focuses on ``legacy'' Web content and develops ontologies to support multimedia and multilinguality. (http://esperonto.semanticweb.org)

FF-POIROT Financial Fraud Prevention-Oriented Information Resources using Ontology Technology (KA3) Interactive construction of multilingual ontologies through domain modelling, (automatic) text-mining and (semi-automatic) validation and alignment, as a basis for Semantic Web services for knowledge storage, management, retrieval and sharing. (http://www.starlab.vub.ac.be/research/projects/default.htm# Poirot)

GRACE Grid Search and Categorization Engine (CPA9) Develops a decentralized search and categorization engine for unstructured textual information; builds on-top of Grid-technology (peer-to-peer), uses locally computed indexes. (http://pertinax.cms.shu.ac.uk/projects/cmslb2/)

HtechSight: A knowledge management platform with intelligence and insight capabilities for technology intensive industries (KA2) Utilises standardised knowledge representation frameworks to facilitate interoperability, and contributes to the development of the next generation of Web-enabled Knowledge Management platforms. (http://banzai.etse.urv.es/~htechsight/).

IBROW An Intelligent Brokering Service for Knowledge-Component Reuse on the World-Wide Web (FET) Configures distributed, heterogeneous applications using pre-existing components (ontologies and problem solving methods - for information filtering, automatic classification and design problems) retrieved from distributed digital libraries. (http://www.swi.psy.uva.nl/projects/ibrow/home.html)

InDiCo Integrated Digital Conferencing (KA3) Develops semantics-based multimedia indexing and browsing methods for conference and distance learning applications. (http://indico.sissa.it/)

MONET Mathematics on the net (KA3) Develops and applies ontologies for mathematics services description, querying, explanation and use. (http://monet.nag.co.uk)

MOSES A modular and scalable environment for the Semantic Web (CPA9) Focuses on the scalability and linguistic aspects of ontology construction and evolution through content (mainly text) structure analysis.

OBELIX: Ontology-Based ELectronic Integration of compleX products and value chains (KA2) Develops an ontology tool suite for smart collaborative e-business, enabling scaleable integration and interoperability in dealing with complex products and services, supply chains and value networks, and associated electronic market transactions. (http://www.cs.vu.nl/~obelix/)

On-To-Knowledge Content-driven knowledge management through evolving ontologies (KA4) Design of languages (OIL) and implementation of tools for ontologies, for automatic derivation of semantics of semi-structured data (text-mining), knowledge maintenance, view definitions and agent supported information access. (http://www.ontoknowledge.com)

ONTO-LOGGING Corporate Ontology Modelling and Management System (KA2) Distributed formalisation of corporate ontologies; dynamic optimisation using agent technology for user modelling and category extraction. (http://www.ontologging.com)

SCULPTEUR Semantic and content-based multimedia exploitation for European benefit (KA3) Constructs a semantic layer enhancing search in distributed digital libraries of images of 3D objects, by linking low and high-level features; implementing agents for classification and search of structured and unstructured content. (http://www.sculpteurweb.org)

SEWASIE Semantic Webs and agents in integrated economies (KA3) Designs a distributed agent architecture for semantic search and inferencing based on multilingual ontologies. (http://www.sewasie.org/)

SPACEMANTIX Combining spatial and semantic information in product data (KA3) Enriches 3D graphics in product catalogues with semantic information (e.g. assembling instructions) for easy and natural access to and manipulation of 3D models. (http://www.agc.fhg.de/uniGoethe/forschung/SpacemantiX/)

SPIRIT Spatially-aware information retrieval on the Internet (KA3) Derives/extracts ontology-based geo-metadata from Web pages and digital map datasets, for spatially-aware search engines. (http://www.research-projects.unizh.ch/math/unit70600/area20/p2563.htm)

SWAD-Europe W3C Semantic Web advanced development for Europe (KA3) Informs W3C work on new ``Semantic Web'' recommendations, through research, open source implementation and testing (http://www.w3.org/2001/sw/Europe/)

SWAP Semantic Web and Peer-to-Peer (KA3) Realises a ``Semantic Web'' based peer-to-peer system, building on available Open Source peer-to-peer solutions, for sharing individual views on knowledge through emergent semantics. (http://swap.semanticweb.org)

SWWS Semantic Web enabled Web services (KA3) Develops semantic means for describing, recognising, configuring, combining, comparing and negotiating Web services, supporting Web service discovery and scalable mediation. (http://swws.semanticweb.org)

VICODI Visual Contextualisation of Digital Content (KA3) Provides mechanisms for contextualising distributed multilingual digital content (European history), taking into account topics (category, hierarchy), location and time, through semantic indexing and ontological markup and using neural classifiers; development of a suitable SVG-based visualisation interface.

WIDE Semantic Web-based information management and knowledge sharing for innovative product design and engineering (KA3) Integrates, using Semantic Web technologies, proprietary in-house databases, off-line and on-line catalogues, and the World Wide Web to support the information and knowledge sharing needs of industrial designers and product engineers. (http://www.cefriel.it/topics/research/index.xml?tid=27)

WISPER Worldwide intelligent semantic patent extraction & retrieval (KA3) Automates semantic mark-up of structured and multi-lingual digital content (patents), in support of searching and visualising search results.

WonderWeb Ontology infrastructure for the Semantic Web (FET) Analyses requirements for large-scale deployment of ontologies: ontology languages, semantic integration, migration, reconciliation and sharing of ontologies, foundational ontologies, tool support (for editing, integrating and extracting ontologies), ontology server architectures and services such as persistent storage and reasoning support. (http://wonderweb.semanticweb.org)

References

[Ankolenkar et al., 2001]
A. Ankolenkar, M. Burstein, T. Cao Son, J. Hobbs, O. Lassila, D. Martin, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, T. Payne, K. Sycara, and H. Zeng: DAML-S: Semantic Markup For Web Services, http://www.daml.org/services/daml-s/2001/10/daml-s.html.

[Bussler, 2001]
C. Bussler: The Role of B2B Protocols in Inter-enterprise Process Execution. In *Proceedings of Workshop on Technologies for E-Services (TES 2001) (in cooperation with VLDB2001)*. Rome, Italy, September 2001.

[Cohen et al., 1998]
P. R. Cohen, R. Schrag, E. Jones, A. Pease, A. Lin, B. Starr, D. Easter, D. Gunning, and M. Burke:. The DARPA High Performance Knowledge Bases Project, *Artificial Intelligence Magazine*, 19(4), pp.25-49, 1998.

[Euzenat, 2002]
J. Euzenat (Eds): ERCIM News - Special: Semantic Web; Newsletter of the European Research Consortium for Informatics and Mathematics, No. 51, October 2002.

[Fellbaum, 1998]
C. Fellbaum: WordNet : An Electronic Lexical Database. Cambridge, Mass: MIT Press; 1998.

[Fensel, 2001]
D. Fensel: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001.

[Fensel et al., 2002]
D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster (eds.): Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press, Boston, 2002.

[Fensel et al., to appear]
D. Fensel, C. Bussler, Y. Ding, B. Omelayenko: The Web Service Modeling Framework WSMF, to appear in *Electronic Commerce Research and Applications*.

[Foster et al., 1998]
Ian Foster and Carl Kesselman (Eds): The Grid - Blueprint for a New Computing Infrastructure; Morgan Kaufmann Publishers, 1998.

[Gruber, 1993]
T. R. Gruber: A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 5:199—220, 1993.

[IST, 2002]
COUNCIL DECISION of 30 September 2002 adopting a specific programme for research,technological development and demonstration: 'Integrating and strengthening the European Research Area' (2002-2006) (2002/834/EC); Official Journal of the European Communities L 294/1, 29.10.2002.

[Lenat, 1995]
Lenat, D. B. "Cyc: A Large-Scale Investment in Knowledge Infrastructure." *Communications of the ACM* 38, no. 11, 1995.

[Meersman & Sheth, 2002]
R. Meersman & A. Sheth (eds.), Special Issue of SIGMOD, 2002.

[Stork 2002]
Webs, Grids and Knowledge Spaces - Programmes, Projects and Prospects, Journal of Universal Computer Science, vol. 8, no. 9 (2002), 848-868.

[Trastour et al., 2001]
D. Trastour, C. Bartolini, and J. Gonzalez-Castillo: A Semantic Web Approach to Service Description for Matchmaking of Services. In *Proceedings of the Semantic Web Working Symposium*, Stanford, CA, USA, July 30 - August 1, 2001.