

Semantic Web Services: A Communication Infrastructure for eWork and eCommerce

Dieter Fensel

Institut für Computer Science (IFI), University of Innsbruck
Technikerstrasse 25, 6020 Innsbruck, Austria

Tel.: +43-512-5076485/8 Fax: +43-512-5079872

<http://www.google.com/search?q=dieter> or
<http://informatik.uibk.ac.at/nextwebgeneration>

“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' note¹ 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. 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.

¹ <http://www.w3.org/DesignIssues/Semantic.html>

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

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 by 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 comprise a thesaurus of over 100,000 natural language terms explained in natural language. At the other end of the spectrum is CYC, 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 cooperation 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.

Semantic web technology is still in its early stages. We are focusing on building its basic -- and mostly static -- infrastructure. The next step will be to produce active components that use this infrastructure to offer human users intelligent services. **Web services** aim to support information access and e-business. Examples include UDDI, a repository for describing vendors, products, and services. It uses WSDL to describe its entries, and SOAP as a protocol to define how they can be accessed. At present, none of these service description elements are based on semantic web technology. As a result, it requires tremendous human effort to perform such tasks as: searching for vendors, products, and services; comparing and combing products; and forming coalitions of vendors, etc. **Semantic web-enabled web services** can provide a higher level of service by mechanizing many of these aspects. Steps in this direction are being taken by projects, such as DAML and Ibrow. Within DAML, a service description language called DAML-S has been developed. This language allows formal competence descriptions that enable automatic inference as a means of selecting and com-

binning services. Ibrov developed a language called UPML that can be used to describe static and dynamic aspects of a semantic web. UPML offers features that describe Ontologies, heuristic reasoners (called problem-solving methods) and methods to interweave them. Based on these descriptions, an automated broker provides support in component selection, combination, and execution.

Web services described by WSDL are individual message exchanges. They can be synchronous or asynchronous one-way messages between a sender and a receiver or a pair of messages following a request/reply pattern between a sender and a receiver. While these two patterns are sufficient in many cases they are insufficient for more complex message exchange patterns (called public processes) like a purchase order (PO) and purchase order acknowledgment (POA) exchange whereby the PO message as well as the POA message are acknowledged individually by low-level message acknowledgements confirming the receipt of the message. WSDL is not able to define those public processes. If trading partners try to match their complex public processes in order to conduct business with each other they might encounter a mismatch of their public processes. For example, one trading partner expects message acknowledgements but the other trading partner does not provide them. Above mentioned languages do not support at all the compensation of these public process mismatches. In a peer-to-peer environment no third party mediator can be asked to compensate for the process mismatches. The trading partners have to do the compensation themselves in their environments. To implement the vision of mechanized support of public process integration the appropriate concepts as well as an approach for automatic and semantics preserving public process semantic web technology has to be applied to web services. Currently, all these service descriptions are based on semi-formal natural language descriptions. Therefore, the human programmer need be kept in the loop and scalability as well as economy of web services is limited. Bringing them to their full potential requires their combination with semantic web technology. It will provide mechanization in service identification, configuration, comparison, and combination. **Semantic Web enabled Web Services** have the potential to change our life in a much higher degree as the current web already did.

No technology can survive without convincing applications. Therefore, 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.

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 realize 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 organization's data and information sources for innovation, greater productivity, and competi-

tive strength. Due to the globalization and the impact of the Internet, many organizations are increasingly geographically dispersed and organized around virtual teams. Such organizations need knowledge management and organizational 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 concern 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 organization 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 Service 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.

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 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.

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 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.

Books covering the issues discussed above are: [1, 2, 3, 4, 5]

References

1. 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, 2003.
2. J. Davis, D. Fensel, and F. van Harmelen (eds.): *Towards the Semantic Web: Ontology-Driven Knowledge Management*, Wiley, 2002.
3. D. Fensel, B. Omelayenko, Y. Ding, E. Schulten, G. Botquin, M. Brown, and A. Flett: *Intelligent Information Integration in B2B Electronic Commerce*, Kluwer, 2002.
4. D. Fensel: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
5. D. Fensel: *Problem-Solving Methods: Understanding, Development, Description, and Re-use*, Lecture Notes on Artificial Intelligence, no 1791, Springer-Verlag, Berlin, 2000.