Foreword: Semantic web and semantic information management

B. Di Martino and F. Moscato*

Dipartimento di Ingegneria dell'Informazione
Faculty of Engineering
Second University of Naples, Italy
E-mail: beniamino.dimartino@unina.it
E-mail: francesco.moscato@unina.it
*Corresponding author

Biographical notes: Beniamino Di Martino is a Full Professor at the Second University of Naples (Italy). He is the author of five international books and more than 100 publications in international journals and conferences. He served as general and program chairman, and a member in program committees, of several international conferences, and as guest editor for several journals' special issues. He is an editorial board member and chair of international journals. He is Vice Chair of the Executive Board of the IEEE CS Technical Committee on Scalable Computing. His research interests include programming and compiler techniques for high performance and grid computing, mobile and intelligent agents, automated programme analysis and transformation, reverse engineering, semantic-based information retrieval, semantic web and semantic web services.

Francesco Moscato is a Researcher at the Second University of Naples (Italy). He graduated cum laude at the University of Naples 'Federico II' in Computer Science Engineering. He obtained his PhD in Electronic Engineering at the Second University of Naples. His research interests include high performance and grid computing, formal modelling of real-time systems, web and semantic web services, reverse engineering, multi-agent systems and embedded systems.

The World Wide Web (WWW) has made a huge amount of information electronically available, and is an impressive success story in terms of both available information and the growth rate of users (Fensel, 2003). The web has evolved not only from the point of view of networks and infrastructure, but also in terms of software applications. This success has been based mainly on its simplicity, giving software developers, information providers and users easy access to new contents. Nevertheless, the same simplicity that made the impressive expansion of the web possible has brought important, and in some cases critical, drawbacks that are conditioning its future development.

The WWW has drastically changed the availability of electronically accessible information. Currently billions of documents are used by more than 200 million users all over the world, and their number is growing astronomically. This makes it increasingly difficult to find, access, present and maintain information on the web.

The huge number of websites that have arisen in the last few years have also helped make the mentioned tasks difficult (Ding and Fense, 2002). This difficulty is because information content is presented primarily in natural language. Thus, a wide gap has emerged between the information available for the tools aimed at addressing the problems above and the information maintained in human-readable form. Furthermore, in the last decade Service Oriented Architecture (SOA) has been exploited to provide services to users on the web. Even if SOA aims to provide common standards for interoperability, the increasing number of web services has also introduced the problem of describing the provided services operations in order to classify and reuse services, if needed.

For this reason, *searching for information* on the web, and in general, among a (very) large number of digital documents is often a nightmare: text-based searches usually return huge amounts of material. Almost all of this material is irrelevant and it is not unusual to miss the really relevant material. Especially on the web, where the number of documents grows every minute, such searches are imprecise and often point to a large number of documents, which makes them impossible to read. Furthermore, the search process may be difficult and information may be obscured in some way. Several ranking algorithms may be used to order search results, making it possible to distinguish documents that are more relevant than others for a given user query, but the huge number of documents again reduces the benefits of those approaches.

Information retrieval does not end with documents retrieval, since usually only some parts of a document may satisfy the user needs. Furthermore, the same piece of knowledge in some documents may be presented in different contexts and be adapted to different users' needs and queries.

In such a context, two main problems arise:

1 *information search and presentation* – Even now, finding the right information in the huge amount of data available on the web is very difficult. Searches are imprecise, often yielding matches to many thousands of resources. Moreover, users face the task of reading and accessing the retrieved resources in order to extract from them the desired information. A related problem involves the maintenance of web sources, which has become very difficult. The *mission* of maintaining consistency among information on the web is by now almost impossible to accomplish, especially for human users. This has resulted in a vast number of sites containing inconsistent and/or contradictory information.

2 the *reuse of available web services for composition* – The high number of services provided on the web is appealing as it allows services designers to use existing web services in order to build composite (value-added) ones. This may be useful to reduce the development and testing times of new, complex services. The problem of having such service behaviours defined in a common way, not only by a syntactical definition of their interfaces (through WSDLs), is still open. Service behaviours, like information and data, are generally hard to define and, consequently, it is very difficult to retrieve the right service to use depending only on the *names* of their operations, inputs and outputs.

The large number of documents on the web leads to another problem: the redundant information on the web should be kept consistent, and incoherent or contradictory information should be managed properly.

Finally, information may be stored in different kinds of media and performing useful classical text-based searches on images, audio or video is often impossible.

There is an emerging awareness that providing solutions to these problems requires that there be a machine-understandable semantics for some or all of the information presented in the WWW, leading to a *semantic web* infrastructure.

The basics for the semantic web consist in: developing languages for expressing machine-understandable meta-information; developing tools and new architectures which use such languages to provide support for finding, accessing, presenting and maintaining information sources; realising applications that provide a new level of service to the human users of the semantic web.

In response to these problems, many new research initiatives and commercial enterprises have been set up to enrich the available information with machine-processable semantics. Such support is essential for "bringing the web to its full potential" (Berners-Lee *et al*., 2001). Director of the World Wide Web Consortium, referred to the future of the current WWW as the *semantic web*. The vision of the semantic web is to extend current web machine-readable information and automated services in order to extend far beyond current WWW capabilities (Berners-Lee *et al*., 2001).

The explicit representation of the semantics underlying the data, programs, pages, and other web resources will enable a knowledge-based web that provides a qualitatively new level of service. Automated services will improve in their capacity to assist humans in achieving their goals by 'understanding' more of the content on the web, and thus providing more accurate filtering, categorisation and searches of information sources. This process will lead to a large system that features various specialised reasoning services.

A key enabler for the semantic web is online ontological support for data, information and knowledge exchange. Given the exponential growth of the information available online, automatic processing is vital to the task of managing and maintaining access to that information. Furthermore, ontological information has to be associated with documents on the web. This leads to many problems related to automated annotations of digital documents. Finally, since web service behaviours, inputs and outputs can also be annotated with ontological information, composition techniques can be used with semantics-based approaches in order to build *composed* services for facing users' requests.

One of the most exploited ways to define meta-information is the use of ontologies. *Ontologies* serve as metadata schemas, providing a controlled vocabulary of concepts with explicitly defined and machine-processable semantics. By defining shared and common domain theories, ontologies help human users and machines to share information, concisely supporting semantics exchange. Hence, the success of the semantic web also depends on the proliferation of domain-specific ontologies which humans and machines can refer to in order to describe information in documents.

Although ontology-engineering tools have matured over the last decade, manual ontology acquisition remains a tedious, cumbersome task that can easily result in a knowledge acquisition bottleneck. For this reason machine-learning techniques for ontologies definition and for information extraction are the most active research topics in the semantic web.

Developing metadata languages, ontologies and tools which enable such techniques is a wide-ranging problem that touches on the research areas of a broad variety of research communities. Therefore, this special issue brought together colleagues from these different research communities, namely those in the areas of intelligent information integration, knowledge representation and management, information retrieval, natural-language processing, metadata and specific research areas such as electromagnetism or jurisprudence.

This special issue focuses on knowledge management in different domains and on the application of ontology-based reasoning in several contexts. Knowledge management in the legal domain is addressed in 'A semantic document management system for legal applications' (Amato *et al.*), where RDF statements are used for indexing, retrieval and long-term preservation of legal documents. Approaches based on ontologies reasoning are also described in 'Ontology-based modelling and execution of workflows for a virtual ISP' (Pop *et al.*) and in 'An ontology-based approach for providing multimedia personalised recommendations' (Brut *et al.*) respectively concerning web services composition and the deployment of *intelligent* web services. A case study from the medical domain is described in 'Design of a semantic person-oriented nurse call management system' (Ongenae *et al.*), while in 'A proposal for an electromagnetic ontology framework' (Esposito *et al.*), a case study from the electromagnetism domain is reported. Finally, an approach to manage e-learning systems by using ontology-based reasoning is discussed in 'An integrated system of data backup and supervision of the activities, with management of the natural language, in e-learning contexts'.

The papers of this issue are extended versions of the best works presented at the workshop *Web/Grid Information and Services Discovery* (WGISD-2008) and at the *International Conference on Complex, Intelligent and Software Intensive Systems* (CISIS-2008).

References

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The semantic web', *Scientific American*, Vol. 17, May.

Ding, Y. and Fense, D. (2002) 'Ontology library systems. The key to successful ontology reuse', *Proc. of 1st Semantic Web Working Symposium*, California, USA, pp.93–112.

Fensel, D. (2003) *Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd revision, Springer Verlag.