Report on EU-IST workshop semantic web technologies

Luca Botturi*

REPORT ON EU-IST WORKSHOP SEMANTIC WEB TECHNOLOGIES

(Jean Monnet Building, Luxemburg, November 22nd-23rd 2000)

On November 22nd-23rd 2000, the Jeanne Monnet Building in Luxembourg hosted the workshop *Semantic Web Technologies*, organized by the Directorate-General Information Society of the European Commission and enjoyed by more than a hundred people.

The workshop was part of the Key Action 3 (*Multimedia Content and Tools*), Information Society Technology program 1998-2002, area 4, *Information Access, Filtering, Analysis and Handling (IAF)*, which will promote the *Semantic Web Technologies* Action Line in the programme 2001[1].

The IAF Action Line «picks up on ideas that have been looming for a number of years and that have received their perhaps greatest push from the World Wide Web Consortium[2] (W3C)»[3], and that could be translated into «making content machine-understandable». This means to attach to digital content some formal description of itself — *metadata* — that could be automatically interpreted. Computers could perform some high-level tasks for their owners, such as searching and filtering relevant information on the Internet, negotiating business transactions, recognizing useful training modules, etc.. The goal is to provide huge content

* Facoltà di Scienze della comunicazione, Università della Svizzera italiana, Lugano (Switzerland), luca.botturi@lu.unisi.ch.
[1] The call has been announced for June 2001 — see the EU website http://www.cordis.lu/ist/ka3/iaf.
[2] Seminal documents of the workshop could be two informal papers by Tim Berners-Lee: *Semantic Web Road Map* (September 1998) and *Web Architecture: Describing and Exchanging Data* (June 1999). Both were published on the Consortium website http://www.w3c.org.
[3] Quotations are taken from the Workshop program.

repositories and the Internet itself with a semantic structure that would make them more useful and usable.

The workshop, chaired by the Department Head of Unit Franco Mastroddi and the Principal Scientific Officer Hans-Georg Stork, was the occasion for some preparatory work both for getting some feedback and planting some project seeds through fostering exchange and discussion among academic and corporate experts.

Invited Speakers and Short Talks

With six invited speakers and more than twenty short talks, coming from both the university and the business world, the three-session program was surely an intensive one. Many topics were brought into discussion, from ontology theory down to business programs and training applications, from RDF and XML up to wireless applications. In this magmatic situation, typical of any pioneering field, four main discussion threads are worth considering for a summary.

Formalization

The first technical step in making the web semantic is defining a formal language for expressing metadata. Two perspectives are currently under discussion.

The ontological[4] perspective was thoroughly discussed during the first session thanks to the contributions of Rudi Struder (University of Karlsruhe, invited speaker), Dieter Fensel (Free University of Amsterdam), Ian Horrocks (University of Manchester), Carole Globe (University of Manchester), Jérôme Euzenat (INRIA Rhône-Alpes), Pasqualino Assini (The Data Archive, University of Essex), Ralph Traphoener (Tecinno GmbH) and Atanas Kiryakov (OntoText Lab, Simma AI Ltd)[5]. This view promotes the use of complete ontologies for describing and «objectualizing» the web. This could be accomplished following two alternative paths: the so-called *hard ontologies*, more complex

[4] Ontology means technically a complete formal and computable definition of the objects (or entities) in a specific domain.
[5] The great number of contributions to the ontology debate probably occurred due to the fact that the ontology theory comes from the same area of intelligent agents which is the historically natural environment for semantic computing.

and efficient though not easy to define and introduce in an open and huge environment such as the Internet; or *light ontologies*, more flexible but less standardized and therefore less reliable in a multi-user environment. The OIL language, developed in the IST framework program last year as part of the *Ontoknowledge* project, was presented as a success story[6].

The other perspective is that of standard syntax such as XML (Extendible Mark-up Language), exploiting the formal structure offered by RDF (Resource Description Framework)[7]. Libby Miller (ILRT, University of Bristol), Bert Bos (W3C/INRIA), Eric van der Vlist (DYOMEDEA), and Martin Bryan (the SGML Center) provided insights for this discussion. The contributions focused on the syntax that could be used for metadata, without regard to the actual object definitions being used. The choice of a particular formal model and syntax obviously influences the conceptual definition of metadata and vice-versa. Moreover, establishing only a standard syntax (e.g., XML) means allowing any user to define his/her own metadata tags, thus potentially creating different systems, grammatically uniform but semantically incompatible[8].

Grounding

In order to achieve effective and robust semantic technologies it is important to ground any choice of a specific syntax or ontology system on data taken from real applications. The more the term *multimedia* is enlarged to include different formats and sources, the more the issue gets complex.

Wolfgang Klas, from the University of Vienna, and Lynda Hardman of W3C were the invited speakers who provided the guidelines along which other contributions developed: specific issues for visual content (Harold Boley, DFKI, and Jos De Roo, AGFA; Zdenek Mikovec, Czech Technical University), voice recognition (Christian Gronoff, Semiosphere) and other data structures (Paolo Avesani, ITC-IRST).

---

[6] See the project website http://www.ontoknowledge.org.
[7] Both RDF and XML are formal recommendations of W3C. More information about them can be found at www.w3c.org.
[8] RDF provides a common formal representation of metadata for calculating the semantic equivalence of expressions in different syntaxes; even if this makes the system more robust (it is not likely to have false synonims), it is not enough to make it complete (nothing insures that different users will develop compatible metadata schemas).

Although the speakers pointed out several relevant issues, a sound grounding for technological choices is still missing. Pushed by the faster and faster rhythm of the industry, and constrained by the intrinsic difficulties of the issue, researchers have begun to «first try, then fix» instead of proceeding from theoretical foundations to applications; moreover, the grounding of technological choices in this field is not only a technological concern, but also requires systematic interdisciplinary research by content experts, logicians, linguists, and communication experts.

Acting

Another thread was the applicative side of Semantic Web Technologies: what areas would get benefits from them, and for what services? Johan Hjelm of Ericsson (invited speaker) presented the current research and the future vision in personalized wireless services for cellular phones, insisting on information retrieval and filtering; other applications were presented in the field of e-commerce (Janne Saarela, invited speaker form Pro Solutions; Jürgen Nicklisch-Franken, Persits AG, and Heinz Schweppe, FU Berlin) and e-learning[9] (R.A. Poell, TNO; Mike Dewar, NAG Ltd.; Peter Fankhauser, GMD-IPSI; Luca Botturi, Università della Svizzera italiana; Heidrun Allert, University of Hannover).

The most suitable way of programming new semantic applications could be the intelligent agent paradigm, which was discussed in the final session thanks to David de Roure (invited speaker, University of Southampton) and other contributions (Stanislaw Ambroszkiewicz, Institute of Computer Science, Polish Academy of Sciences; Patrick De Causmaeker, KaHo Sint-Leven).

Interface

The problem of the interface has high priority: how should we interact with computers «semantically»? Moreover, the introduction of metadata requires a migration of non-semantic web applications to a new environment, which involves a great change in the information management process. How can the natural resistance against changes be overcome? A

---

[9] This latter field, much dependent on the personalization of multimedia content for effective learning, seems to be the nearest to an breakthrough thanks to new semantic technologies.

simple interface for all classes of users would be a powerful incentive.

The main idea proposed is that the shift should be transparent to the end-users: although complex, new web technologies must be easy to use — «simple and stupid» as the chair put it. The proposed path towards easy applications is that of visual interfaces, evolving the web paradigm to *conceptual browsers* (Mikael Nillson, Royal Institute of Technology; Mikael Jern, Advanced Visual Systems Aps and Linkoping University).

The Discussion

The vivid final discussion was no conclusion, but an opening and stimulus to new projects. Four main issues were raised, reported in the following paragraphs.

New Web Jobs Profiles

The applications of new theories and technologies should be designed according to realistic scenarios. What kind of users will navigate in cyberspace after the «semantic wave»? What are they looking for and why? Five profiles were sketched, with regard to their information needs:

1. *Enquirers* are users who browse the cyberspace in order to answer specific questions. They carefully plan their activity in the long-term and would like to get selected relevant information. They would benefit from one or more personal agents that help them in the job and that learn from experience. A personal publishing agent could distribute the enquirer's results to trusted and interested partners.

2. *Thinkers* are persons who navigate the web following a red-thread, but are still open to surprise and to analogy, in contrast with the rigid pattern of thought proper to machines. A looser information-filtering agent could help them both in finding specific information and at the same time provide them with new, unexpected stimuli for their work. Thinkers hope the web does not grow too semantic, and that some quire illogical link still remains available.

3. *Doers* are in charge of a complex task, which can be divided into smaller ones, personally accomplished or given in outsourcing; they are focused on short-term practical solutions and require highly specific information for each task in a short time.

4. Some will instead use the Internet as a medium — these are the *communicators:* they would like to have more flexible and sensible tools for interaction and collaboration.
5. Finally come the *players*, people who just browse around the web, with no specific professional goal, following the click-stream[10]. These could be interested in entertainment, learning, generic information, etc., which could be offered them according to their tastes, culture, etc.

These profiles could be taken as background for the requirements analysis of future semantic web applications and services.

Social Issues

The social aspect of technology also deserves attention. In this regard too, the discussion was particularly interesting: it presented some general issues for technologies, particularly urgent and up-to-date.

The first question raised during the discussion was about the changes in the job market that the semantic wave will bring about: will the semantic wave destroy job places or create new professions? Will there be new jobs for people creating and managing metadata, or will machines simply take the place of old-fashioned librarians? How to manage the shift? These questions are just the top of the iceberg of a complex issue: the relationships that these new technologies entail with the tradition they come from. The Internet is a new manifold instrument, which can only be understood and exploited in the context of archives, libraries, and traditional media. Specifically, one priority for future metadata systems and ontologies is to be open, namely, to consider also non-digital objects: all the manuscripts, books, videocassettes, existing archives, libraries, picture catalogues, and so on, that preserve the biggest part of our cultural heritage. Metadata definitions compatible with traditional cataloguing systems could be a first step towards a non web-centered semantic web.

Another social aspect of the shift to semantic technologies is motivation: why should webmasters and database managers begin to use metadata, adding a lot of short-term, pointless information to their products, making the whole management process slower? In one sentence, why and

---

[10] These were once called by the now old-fashioned term *web-surfers*.

how should people begin putting tags all over the Internet? One possible incentive to the process is granting visibility on the Web for those websites that offer metadata for search engines. More probably, specific needs in restricted domains will be the natural catalyst of a long-term shift. On the other side, this consideration pushes the research towards simple applications more than towards the theoretical study of ontologies and the definitions of syntax schemes.

Simple, Stupid, Cheap and Open

Franco Mastroddi summarized other general requirements for future applications with the slogan «keep it simple, stupid, cheap and open».

The complexity of semantic web applications, as mentioned before, must be encapsulated inside the system, being transparent for the end-users. Moreover, as new applications come to the market, they should be cheap — which means they should not require huge information or management restructuring for the company or institution adopting them. Finally, as underlined when talking about semantic technologies and traditional libraries, these systems should be as open as possible.

In one word, the semantic wave will have a future and influence our way to search for information, work, learn and live only through human-centered applications. For researchers, this can be translated as being focused on the users and their needs more than on technologies and their possibilities.

Although smart looking, this statement is tricky. Only a sound theoretical basis can support long lasting, scalable and evolving technologies; but if the effort is aimed just at producing applications in the short term, the aspect of such grounding necessarily fades in the background. A balance between the two aspects should be valued and supported.

Explain, Educate, Exploit

The history of technologies has taught us that the success of a new application does not depend only on its usefulness and reliability, but on the deployment strategy as well. The Chair used another effective slogan for summarizing insights on this point: «explain, educate, exploit».

Dealing with a new kind of technology, the first goal is to explain to the future users what it is all about, and why should they be interested in this new stuff. Users should then be educated to use them, i.e., the

processes that involve technologies should be revised and oriented towards a new shape, which must be perceived as «better» at least in some respect. Only after these two steps are made will it be possible to exploit new technologies in a complete way, so that they can be the basis for further development.

Conclusion

The EU Workshop on Semantic Web Technologies was a first rich and stimulating moment of discussion about the new semantic wave, on the guidelines of the W3C informal papers by Tim Berners-Lee. These technologies will probably revolutionize our professional and personal lives. Whether or not the time is ripe should be clear within a few years. The EU directorate surely bets on this area with the IST framework program. As was pointed out, much is still to be done, both in fundamental and applied research involving interdisciplinary study; many issues are at stake, both technological and social, both scientific and professional.

The overall impression is that the meaning of the key word — *semantic* — is not clear enough to let us imagine the future without some imagination. Is that just a ready-to-use formula for the magazines' hypes, or does it reveal an actual novelty? Is it just tagging resources with some kind of metadata, or does it involve more intimately the human understanding of information and knowledge? Surely, if taken seriously as the EU suggests, it could turn out to be a crucial issue for the near future.

The workshop report (introduction, contributions abstracts, and IST *Semantic Web Technologies* call text) and slides are available on the EU website under the URL http://www.cordis.lu/ist/ka3/iaf/swt_workshop.htm.