International Semantic Web Working Symposium (Stanford, July 30th - Aug. 1st, 2001)

Semantically Marked Up… Now What?

Simon Buckingham Shum, John Domingue and Enrico Motta

Knowledge Media Institute, The Open University
Milton Keynes, MK7 6AA, UK
{S.Buckingham.Shum, J.B.Domingue, E.Motta}@open.ac.uk

Broadening the vision

"Wired…now what?", was the challenge from Apple's Mark Miller to educationalists as the Internet took off. A new infrastructure was being laid, but even now we still barely know how to use it effectively for learning. We can expect a similar scenario to unfold for the semantic web. Currently, the semantic web community is necessarily preoccupied with infrastructure design (e.g., knowledge representation languages for the web, scaling issues, standardising terminologies and agent interoperability). But once this infrastructure is in place, more persistent questions will remain concerning the technology's relationship to the people expected to use it. Specifically, we ask: How can the semantic web be used to support *people* in *knowledge-intensive work*?

"Semantically marked up…now what?", is therefore a worthwhile question to balance prevailing discussion. In response, many offer machine-centric rather than human-centric visions: semantically aware search engines, integrated e-commerce, and self-stocking fridges. These are worthy and challenging goals, but there is little discussion of the *people* required to make these systems function. Is the semantic web only about interoperable machines, or is there a vision of "augmenting human intellect", to pursue Doug Engelbart's forty year mission? In other words, how will the semantic web support learning, collaboration and knowledge sharing? (We note in passing that this was also Berners-Lee's original goal in designing the web.)

Towards ubiquitous knowledge acquisition

When we talk about the semantic web becoming pervasive, the implication is that knowledge acquisition (KA) will become ubiquitous. It is increasingly the case that if one's work/profile is not on the web, it is invisible. In 1997 one of us co-authored a paper which stressed the need for new forms of literacy in order to maintain one's presence in the emerging semantic web space *[5]*. One obvious concern must be with the user interfaces to enable individuals to participate in the world of the semantic web, and the expertise required to use them. New methods for KA are required to support this codification process. However, we must go beyond user interfaces and training. A closer examination of the "interface" between such technologies and the workplace (academic or business) quickly leads one to issues of trust and ownership, since the tools now embody, more explicitly and richly than ever, ontologies codifying perspectives on how the world works. We are talking about the cognitive, social and political interface that a semantic web infrastructure unavoidably has with communities of practice.

In our experiences with designing and implementing semantic web technologies for the workplace, we are grappling with a different order of questions to those raised at the infrastructure level. When semantic web technologies embrace elements of *human* understanding and interpretation, with possibly suspicious, non-technical stakeholders, we encounter questions such as: "Who gets to define the ontology? (and can we trust them?)"; "How can we be involved in the design process, when we don't understand these formalisms?"; "We're all too busy. Who's going to encode material? (and can we trust them?)"; "An ontology requires consensus. This domain is in flux. Do you still have anything to offer?".

The semantic web for "augmenting human intellect"

These are big issues, which we are beginning to tackle in a number of projects, all of which are based on an underlying mission: *How do we harness the semantic web to become an enabler for "augmenting human intellect"?* We sketch below a number of strategies that seem to hold promise.

Support conflicting interpretations and perspectives. What does the semantic web have to offer in domains where there is little consensus? In the Scholarly Ontologies project [www.kmi.open.ac.uk/projects/scholonto], we are developing a semantic digital library server that seeks to provide services for researchers, whose business is, of course, the construction and debating of world views [1]. Our approach is to provide a discourse ontology for making, extending and challenging claims—an ontology for disagreeing as much as for stable knowledge. Although currently being applied to research literatures, the underlying approach could be extended to any domain where it is as important to capture principled *disagreement* as it is to capture consensus.

Personalise information sources. There is plenty of information out there: the challenge is to find the meaning. Filtering based on semantics is one obvious way to assist intellectual work. In the Advanced Knowledge Technologies consortium [www.aktors.org] we have developed ontology-based support for personalizing an electronic newsletter. This integrates web, language and

knowledge modelling technologies to enable readers of an electronic newsletter to easily define an interest profile according to several different dimensions, formalised in a shared ontology [3].

Provide active support for semantic markup to users with different levels of expertise. We are developing an annotation tool which provides different mechanisms for semantic markup by expert and naïve users. Expert users are given powerful editing/browsing facilities. Naïve users are supported by information extraction (IE) methods [6], which can be used directly from a library, or easily customised through the use of a simple interface for developing new IE methods. Finally, expert-level support for defining new IE methods are also provided.

Make semantic markup a by-product of work. The success of the web demonstrates that users are prepared to learn markup formalisms when the benefits are immediate and clear. It is as yet unclear whether the perceived benefits of semantic markup are so evident, and in any case manual markup will not scale. These issues are crucial in organizational contexts, where time and workload rule the day. However, if markup can occur as a by-product of ongoing work, sustained adoption will be more likely. Therefore our current work focuses on developing knowledge management solutions predicated on the above premise (i.e., transparent semantic markup). For example, we are currently developing technologies for our lab, which will automatically update a knowledge base (KB) when new reports are published, a new project is launched, new staff join, or other significant events take place. This KB will provide the main source for a query answering facility to allow visitors and staff to easily find information about the lab's work. The KB will integrate existing databases and will rely on information extraction methods to find information in relevant web pages. The process is ontology-driven, in the sense that the KB is based on an ontology which identifies the key concepts and relations required to describe academic life.

Another example of transparent markup is the *Compendium* approach to knowledge capture. The approach enables real time capture of knowledge arising in meetings, expressed by members of diverse communities of practice, integrating hybrid material into a reusable group memory. The shared, visual interface serves as a participatory front-end for eliciting information from domain experts who are not literate in knowledge representation. The resulting database can then be exported to knowledge-based tools, or to generate documents in requested formats [4].

Deliver customisable reasoning services to non-programmers. The semantic web is going to be about delivering sophisticated services to users. One project investigating this area is the *Internet Reasoning Service* [www.kmi.open.ac.uk/projects/irs] which provides a Web-based front-end enabling non-programmers to prototype knowledge-based applications quickly, using reusable components from distributed libraries. This approach will make artificial intelligence (AI) technology widely available, thus allowing non-AI experts to create knowledge-intensive services, either for experimentation purposes, for learning or to solve a problem. In addition, we also envisage that specialised configurations of the IRS will be developed for particular communities - e.g., we are working in the palaeontology domain and developing specialised ontologies and problem solving methods to support site interpretation.

Co-design ontologies with communities of practice. In this final example, we broaden the notion of "trust" on the semantic web beyond digital certification. Since significance resides in the interpretations placed on symbols by a given agent, semantic markup is always embedded in a perspective. Consequently, a semantic web application has integrity within a work setting only to the extent that its ontology is trusted. It is well established that systems are trusted when co-designed with the stakeholders expected to use them. A recent case study shows that when a community's perspective on an issue is stable, trusted knowledge services based on an ontology (embodying that perspective) can be co-constructed by knowledge engineers and stakeholders [2].

Further Reading

1. Buckingham Shum, S., Motta, E. and Domingue, J. ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse. *International Journal on Digital Libraries*, 3, 3, 2000, pp. 237-248 [www.kmi.open.ac.uk/projects/scholonto]
2. Domingue, J., Motta, E., Buckingham Shum, S., Vargas-Vera, M. and Kalfoglou, Y. Supporting Ontology Driven Document Enrichment within Communities of Practice. *Submission to K-CAP 2001: First Int. Conf. on Knowledge Capture*, Victoria, BC, Canada, 2001
3. Kalfoglou, Y., Domingue, J., Motta, E., Vargas-Vera, M. and Buckingham Shum, S. myPlanet: an ontology-driven Web-based personalised news service. *IJCAI'01 Workshop on Ontologies and Information Sharing*, Seattle, WA, 2001 [kmi.open.ac.uk/tr/papers/kmi-tr-102.pdf]
4. Selvin, A.M. and Buckingham Shum, S.J. Rapid Knowledge Construction: A Case Study in Corporate Contingency Planning Using Collaborative Hypermedia. *Journal of Knowledge and Process Management*, 2001, in press [kmi.open.ac.uk/tr/papers/kmi-tr-92.pdf].
5. Stutt, A. and Motta, E. Knowledge Modelling: An Organic Technology for the Knowledge Age. In: *The Knowledge Web: Learning and Collaborating on the Net,* Eisenstadt, M. and Vincent, T., (Ed.), Kogan Page: London, 1998, pp. 211-224
6. Vargas-Vera, M., Domingue, J., Kalfoglou, Y., Motta, E. and Buckingham Shum, S. Template-Driven Information Extraction for Populating Ontologies. *IJCAI'01 Workshop on Ontology Learning*, Seattle, WA, 2001 [kmi.open.ac.uk/tr/papers/kmi-tr-105.pdf]