# Semantically Annotate Everything?

Matthew Montebello

Software Engineering Process Improvement Research Group
Department of Computer Science and Artificial Intelligence
University of Malta
Malta

**Abstract.** Search as we know it will change and improve as new and enhanced methodologies to accommodate the ever expanding semantically annotated World-Wide Web will be developed and fruitfully employed. While a lot of basic infrastructure has already been developed to prepare for the onset of the Semantic Web, like meta-languages, editors, inference engines, and others, other critical developments need to be put in place too. This short paper discusses the issues as well as the best application areas to 'Semantically Annotate Everything' and transform on a mass scale the wealth of knowledge on the web into information as semantic metadata on the Semantic web.

## 1 Introduction

With the introduction of semantics in the second generation of the web as we know it, the Semantic Web, came the need to develop and make good use of tools and technologies in order to take advantage of such a richer resource. The Semantic web is strongly based on conceptual description of resources by means of semantic annotations, and the need to correctly map the available syntactic information onto a set of relevant conceptual entities is essential to correctly model the knowledge domain to which a resource belongs. This will have major repercussions primarily on searching as well as on related areas like indexing, retrieval and information extraction. The issue of describing all content semantically, thereby annotating everything over the World-Wide Web (WWW) is no simple task and so the need of upgrading the actual Web to the Semantic Web by means of automated semantic annotation strategies is very much desirable. The basic infrastructure has already been developed to prepare for the onset of this Semantic Web [1], like meta-languages (RDF [2], OWL [3]), editors [4], inference engines [5], and others, yet the main area that needs major input is the raw transformation of the mass of information over the existent WWW as well as the continuous flow of unstructured and semi-structured content appended automatically or humanly processed. The best path to the success of a technology is its flourishing use and universal adoption / application of the same technology. Any conceivable area within the existent WWW is a possible application area to semantically annotate the content and reveal the benefits and advantages such a technology can bring about, together with the functional strengthening and the intensified capabilities offered by existent technologies like search engines, information retrieval and data extraction / mining.

## 2 Conclusion

Such work is still in its infancy and plenty of work still needs to be done not just to get the technology perform up to its expectations, but also for its global acceptance and adoption. Several problems need still to be overcome among which stand out Ontology issues (mapping, versioning, instantiations), fully automatic annotation, meta-language expressivity and standardisation.

## References

1. Semantic Web - XML2000, http://www.w3.org/2000/Talks/1206-xml2k-tbl
2. Resource Description Framework, http://www.w3.org/RDF/
3. Web Ontology Language, http://www.w3.org/2004/OWL/ [4] OWL-S Editor, http://staff.um.edu.mt/cabe2/supervising/ndergraduate/owlseditFYP/OwlSEdit.html
4. OWL-S Editor, http://staff.um.edu.mt/cabe2/supervising/ndergraduate/owlseditFYP/OwlSEdit.html
5. F-OWL, http://ebiquity.umbc.edu/_file_directory_/papers/128.pdf
6. J.Broekstra, M. Klein, S. Decker, D. Fensel, F. van Harmelen, and I.Horrocks: ,. In Proceedings of WWW10), Hong Kong, May 2001.
7. D. Fensel, Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001.
8. Ontoknowledge, http://www.ontoknowledge.org
9. http://www.cs.vu.nl/ guus/papers/Hollink03b.pdf