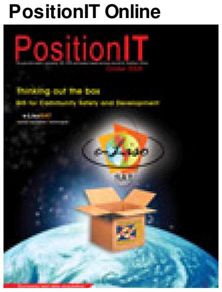
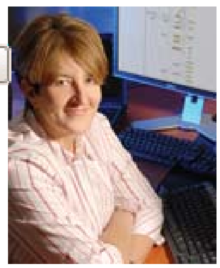
OCTOBER 2009



Ontologies – the new buzzword in modelling

by Dr. Aurona Gerber, Meraka Institute, CSIR

The complexity of modern information systems requires more extensive and advanced modelling technologies. Developers and analysts find it increasingly difficult to ensure that their models are complete, and they find it to be nearly impossible to comprehend all the consequences of assertions made in their models.



Dr. Aurona Gerber is a researcher in the Knowledge Systems Group at the Meraka Institute of the CSIR and the founder member of a TC Knowledge representation with specific focus on Ontologies, within the IEEE SMC in South Africa. Her interests are directed towards ontologies (specifically DL-based, formal ontologies), ontology engineering and modelling, and the intelligent information contained in information systems.

Generally, it is difficult enough to understand direct consequences of asserted statements, but understanding all the implications of these assertions throughout the model is humanly impossible. The implications of these statements cause unexpected results or system failure when models are implemented.

Recent developments within computer science, specifically with regards to description logic (DL) languages and their associated reasoning algorithms, are creating new opportunities to solve modelling problems [1]. DL) are described as a family of logic-based knowledge representation formalisms that can be used to represent, within an ontology, the conceptual knowledge of a specific application domain. Because DLs have formal semantics, several reasoning mechanisms and algorithms have been developed and implemented in reasoners. These implementations have yielded very promising results as is evidenced by the significant and pioneering work that was done by Horrocks [2] who developed some algorithms specifically tailored to medical applications. This work resulted in some of the most successful applications of DLs in the biomedical field [3, 4] such as the widespread use of SNOMED CT1. SNOMED CT is a large medical ontology, based on a DL, that comprises more than 300 000 concepts and more than twice as many relationships. Standard reasoning tasks have been performed on SNOMED CT, which took less than half an hour, a feat that provoked complete disbelief less than 10 years ago [5]. Other examples of biomedical ontologies are the gene ontology, and the class of ontologies maintained by the OBO Foundry2.

The word “ontology” has its roots in the branch of philosophy that deals with the study of the nature of reality. Within computer science it has adopted the meaning of “a formal model that captures domain knowledge” or, as formally defined by Gruber [6] “the specification of a conceptualisation”. Nowadays, the word “ontology” is also used to describe artefacts such as taxonomies and class diagrams, but a formal ontology uses at least one of the knowledge representation languages such as DLs as its modelling language of choice. Ontologies facilitate the formal representation of information in a domain of discourse [7], and the supporting DL technologies allow for reasoning over the domain with its associated information, thus enabling the computation of all the consequences of ontology. These new approaches substantially extend the representational and computational limits of traditional information repositories such as databases or other knowledge representation systems [8].

If we refer back to our original problem of understanding the consequences of modelling assertions, the use of formal ontologies allows analysts and modellers to determine all the consequences of their models, and this enables them to make sure that their models are sound and complete, and have the intended results.

- PositionIT Back Issues -

Ecological Circuits

- Ecological Circuits -

Researchers have started to apply the use of ontologies in a variety of domains. The ontology language OWL is DL-based and has been adopted by the W3C as a web standard [9] because web researchers realise the need to manage meta-data, or data about data, on the web. A problem that any present-day web user experiences is that of information overload, meaning that it is increasingly difficult for a user to ask a question and then to get timely access to correct and relevant results that can be verified to be trustworthy and from a reliable source. This problem is even more so for users in countries such as South Africa, who do not have access to the bandwidth and connectivity infrastructure of the developed world. It is foreseen that ontologies could capture the meta-data in domain models for information on the web so that web users can access information that is relevant to the domain of their query. In addition, these DL technologies and ontologies promise to help users make proper use of information available and also access only the relevant information for a specific task, thus requiring less bandwidth. It is because of this that ontology researchers often refer to the Semantic Web [10] as a good example of an ontology application.

At present several tools are available to help users adopt ontology technologies, such as the ontology editors Protégé3; and Swoop4, and the reasoners Pellet5 and FaCT++6. It is clear that the use of formal ontologies and associated reasoning will grow in a variety of application areas, and this is in line with the view of ontologies as the emerging technology driving the next generation web and the Semantic Web [10]. The current momentum and activity within this field holds the promise that IT researchers will meet ontologies again, possibly in an unexpected application domain.

Footnotes

1) www.ihtsdo.org/
2) www.geneontology.org/, www.obofoundry.org/
3) http://protege.stanford.edu/
4) http://code.google.com/p/swoop/
5) http://clarkparsia.com/pellet/
6) http://code.google.com/p/factplusplus/

References

[1] F Baader F, I Horrocks and U Sattler Description Logics as Ontology Languages for the Semantic Web 2003.
[2] Ian Horrocks. Optimising Tableaux Decision Procedures for Description Logics. PhD thesis, University of Manchester, 1997.
[3] Udo Hahn and Stefan Schulz. Ontological foundations for biomedical sciences. Articial Intelligence in Medicine, 39(3):179{182, 2007.
[4] Katherine Wolstencroft, Andy Brass, Ian Horrocks, Phil Lord, Ulrike Sattler, Robert Stevens, and Daniele Turi. A little semantic web goes a long way in biology. In Proceedings of the 2005 International Semantic Web Conference (ISWC 2005) , LNAI. Springer, 2005.
[5] Boontawee Suntisrivaraporn, Franz Baader, Stefan Schulz, and Kent Spackman. Replacing SEPTriplets in SNOMED CT using Tractable Description Logic Operators. In Proceedings of AIME 2007 , 2007.
[6] T R Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". In: International Journal Human-Computer Studies, 43(5-6):907-928, 1995.
[7] S Staab and A Maedche Ontology Engineering beyond the Modeling of Concepts and Relations. In Proceedings of the ECAI2000 Workshop 2000.
[8] D Calvanese, G D Giacomo, D Lembo, M Lenzerini, A Poggi and R Rosati Ontology- based database access. In Proceedings of the 15th Italian Conference on Database Systems (SEBD 2007) 2007. URL: www.inf.unibz.it/%7ecalvanese/papers/calv-etal-SEBD-2007. pdf.
[9] World Wide Web Consortium. URL: www.w3c.org
[10] T Berners-Lee, J Hendler and O Lassila. The Semantic Web. The Scientific American, vol. 5(1) 2001. URL: www.scientificamerican.com/2001/0501issue/0501berbers-lee.html.

Contact Dr. Aurona Gerber, CSIR, Tel 012 841-3595, agerber@csir.co.za