SEMANTIC MARKUP OF INFORMATION ON SANITATION INITIATIVES IN INFORMAL SETTLEMENTS

Esther Obonyo[1] and Alfred Omenya[2]

ABSTRACT

The fundamental problem with the ‘traditional’ Web is: information is designed for human consumption and even with the evolution of powerful search engines, information filtering still remains a task that requires significant human involvement. Consequently, any application that is designed to use Web-enabled technologies will eventually encounter problems related to ‘information overload’ and ‘digital anarchy’ (the latter problem emanates from lack of interoperability across systems especially in different organisations).

This paper describes a conceptual framework that addresses these issues using emerging Semantic Web Technologies. The framework will focus on demonstrating the potential benefits of using semantic markups and ontologies to augment Web content for Sanitation initiatives in informal settlements using the experience in Nairobi as a test case. It will address problems arising from the sheer scale and diversity of data on sustainable sanitation practice through the use of the Semantic Web technologies. The conceptual framework presented in this paper will be further defined through an iteration of workshops with the stakeholders and expanded into a platform for semantically matching a controlled taxonomy for Sustainable Sanitation for informal settlements.

KEY WORDS

Semantic Web, Sanitation, information management.

INTRODUCTION

The new millennium has been characterized by a great interest in the use of ICTs (Information and Communication Technologies) for Sustainable Development. As far as knowledge sharing is concerned, such interest has been largely motivated by information overload. A myriad of activities have been undertaken to mitigate the adverse impacts of human activities on the prosperity of the future generations. Over the past decade, many initiatives have been implemented across the globe with varying degrees of success as far as the social, economic and environmental metrics are concerned. Each of these initiatives leads to the generation of vast amount of useful information held in physically distributed and heterogeneous repositories.

[1] Assistant Professor, Rinker School of Building Construction, University of Florida, P.O. Box 115703, Gainesville FL 32611-5703, USA, Phone +1 352/273-1161, FAX 352/392-9606, obonyo@ufl.edu

[2] Doctoral Student, Dept of Architecture, University of Witwatersrand, Private Bag, 3, Wits 2050, SOurh Africa, Phone +27 072/569-0045, FAX +27 011/403-0936,aomenya2002@yahoo.co.uk

A significant proportion of knowledge that has been generated over the years remains inaccessible mostly because of being stored as unstructured information. This results in unsatisfactory hits being returned during querying. Further, even where search engines deliver sizable hits, the retrieved information has to be filtered by a human user to obtain the relevant content. The information that is being generated by various sanitation initiatives for low income communities in Kenya is not immune to the information overload problem. A number of organisations engaged in various activities in upgrading sanitation standards in the informal settlements have accumulated experience on the social, economic and impacts of their strategic decisions (Omenya et al 2005). However, such information has not been formally reviewed and documented in a common repository that can be used for accountability to and obtaining feedback from the local communities and governing bodies.

This paper describes the conceptual framework for the ontology creation (and reuse) and semantic query for Sanitation initiatives in informal settlements using the experience in Nairobi as a test case. The proposed framework will address problems arising from the sheer scale and diversity of data on sustainable sanitation practice through the use of the Semantic Web technologies. The framework will be refined through an iteration of workshops with the stakeholders and expanded into a platform for semantically matching a controlled taxonomy for Sustainable Sanitation for informal settlements.

RELATED WORK

Bett and Yang (2004) point out that existing IT applications for sustainable development within the construction industry are still in their infancy with the early applications being typically associated with the measurement and evaluation of sustainability targets. Such tools have been grouped by the IT Construction Best Practice Programme (ITCBP 2004) into:

- Energy modelling tools, for example, computer simulations of energy performance of buildings;
- Intelligent systems, for example, Intelligent and automated building control and operations;
- Assessment tools, for example, assessment of materials' embodied energy;
- Performance indicators Performance indicators, for example, buildings' environmental rating.

Among these tools, the closest match to the system proposed in this paper is the BEQUEST toolkit (Betevegna et al 2002). BEQUEST supports decision making in urban sustainability. The system is composed of one framework and four models of Protocol, Assessment Methods, Advisors and Glossary. It provides a general reference to sustainable design and urban development issues within a European context.

Outside the construction industry, Semantic Web technologies have been applied in the management of knowledge in sustainable development. A relatively advanced application in this domain used a semantic knowledge repository to support sustainable tourism development at the European regional and local level (Legrand, 2004). A significant amount of work has also been invested in developing ontologies in environmental management.Examples of ontologies in the environmental domain include:

- Semantic Web for Earth and Environmental Terminology (http://oceanesip.jpl.nasa.gov/sweet);
- National Institute for environmental eScience – Cambridge(http://www.niees.ac.uk/)
- Gene Ontology – 'dynamic controlled vocabulary that can be applied to all organisms' (http://www.geneontology.org/)
- Design and Creation of Ontologies for Environmental Information Retrieval-Telecordia (http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Kashyap1/kashyap.pdf);
- Ontologies of environmental applications for a digital library of scientific collections (http://dlforum.external.forth.gr:8080/papers/EnviroInfo.pdf), and;
- Waste water ontology – decision support for wastewater treatment processes (http://citeseer.nj.nec.com/ceccaroni00wawo.html).

There are many general computing applications that support the analysis urban development strategies. However, the scope of applying them across different organisations is limited because such applications are not interoperable. Schevers and Drogemuller (2005) proposed a system, which addressed the problem of interoperability using Semantic Web Technologies. They developed a proof of concept prototype that analysised simple characteristics of the urban master plan.

A number of researchers in the construction industry are also developing an interest the evolving Semantic Web technologies. One of the greatest challenges of ICT-enabled collaboration among diverse teams within the construction industry is the significant level of human intervention needed to process the differences in data representation. Mutis et al (2005) proposed a schema for modelling the processes which mediated among each actor's data representation within the construction industry. The proposed framework is not contextualized to actual firms or actors: it provides an 'umbrella' framework that can be used for more contextualized semantic implementations in the construction industry.

Pan and Anumba (2005) also defined a high level framework that demonstrated how the emerging Semantic Web technologies can be applied to project information management. In Chimay et al (2005), a more advance application of this concept was demonstrated through the development of a prototype that used semantic Web technologies in collaboration with wireless services. The deployed prototype used ontologies to structure document/knowledge domains and provided a commonly agreed understanding of the selected domain thus enabling the reuse and sharing of knowledge across groups.

Similar efforts have gone into deploying Semantic Web technologies on the infrastructure side of the construction sector. El-Diraby (2003) proposed a roadmap for a semantic process-oriented environment for knowledge exchange that supports integrated infrastructure development in metropolitan areas. The roadmap does not just focus on engineering design; it also encapsulates issues related to business, safety and sustainability. The integral components in this roadmap are:

- An infrastructure class platform of a platform of semantic objects depicting all common elements of infrastructure systems;

- A CAD-GIS continuum for project ‘data’ exchange depicting location (through GIS) and engineering details (through CAD) that can be extended to other dimensions such as time and cost;
- A UML-based model of the infrastructure development process;
- A Web-based environment for coordinated infrastructure decision-making.

El-Diraby’s proposed semantic platform is based on the e-COGNOS ontology (Lima et al 2003) in which in a Project, has a set of Actors, who use a set of Resources to produce a set of Products following certain Processes within a work environment and according to certain conditions. The proposed framework for Sanitation will build on ontologies that have been developed in the work highlighted in this section.

THE CHALLENGE OF WEB-ENABLED COLLABORATION

Since the 1990s many organisations in the various sectors have embraced the use of the Internet in their business processes convinced that Web-based communication technology would decentralize organizations by making information accessible to geographically distributed work teams, suppliers and partners. Evolving collaboration challenges in an era in which team working is paramount are being matched by new Web-enabled technologies and delivery mechanisms, which emphasise on flexibility and scalability.

Many leading researchers have in fact suggested that the Web, by supporting communication throughout the development process and across the different parties, holds the key to an integrated, and therefore a more productive, construction industry. Existing literature from leading researchers provided evidence of the potential of Web-based collaboration for the construction industry (Bjornsson 2000a and Björnsson 2000b).

The expansion in the use of the Web and its exponential growth are therefore well known facts. However, a problem has emerged from the very core of the success of the Web. The Web has evolved into a very large, unstructured but ubiquitous database. At the close of the 20^{th} century, it was estimated that just textual data on the Web is in the order of one terabyte (Baeza-Yates, 1998). Such information overload on the Web has resulted in substantial losses. KPMG (2000) established that two-thirds of 423 European and US organisations assessed in one survey suffered information overload and failed to share knowledge. Another study on the working habits of 6300 US employees established that on average, employees spent 8 hours a week (16% of their working week) searching for and processing external information. In the American business context, this translates into a financial cost of $107 billion a year (with each employee costing their firm $10,000 annually). Such substantial financial loses underscore the need for change.

The problems related to the use of the Web could also be assessed at the micro-level. During the past decade many organisations have invested in various components of information systems to support their business functions. There were great expectations that the use of Internet-enabled systems would improve productivity and reduce cycle times in various operations by collecting and providing the right information. A significant proportion of the forecasted gains is yet to be realised partly due to the lack of interoperability between systems (Madhusudan 2001). An extensive study conducted at the Center for Research in

Electronic Commerce (CREC), at the University of Texas-Austin to assess the collaborative capabilities of companies also arrived at the same conclusion (Samtami 2003). Collaboration has subsequently become more important for any company seeking to make the most out of e-business (Deloitte Research 2002).

The increase in inter-networked environments has led to more electronic information becoming available to people. Yet networks of information present the electronic information community with a fundamental problem: digitally held information exists in incompatible formats within heterogeneous repositories. Subsequently, systems holding digital information have become 'digital islands.' The problem is further exacerbated by the vast quantities of information that are available in these 'digital islands.' The total time spent locating and filtering through irrelevant information is quite substantial. There is a greater need for tools that help users to manage and navigate through the maze of digitally held information. Such tools must also be capable of integrating and communicating with disparate information repositories to minimise the effort spent re-keying in data. All these issues were quite correctly diagnosed by a group of researchers as the 'digital anarchy' problem (Radeke 1999). Clearly, there is a need for advanced technologies that enhance the use of the Web for information management and knowledge sharing through addressing the 'digital anarchy' and 'information overload' problems.

THE EMERGING SEMANTIC WEB

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Berners-Lee 2001). The fundamental problem with the 'traditional' Web is: information is designed for human consumption and even with the evolution of powerful search engines, information filtering still remains a task that requires significant human involvement. The W3C (World Wide Web Consortium) has championed the evolution of the Web into an instrument which would fulfill the original vision described by Berners-Lee (1989). His goal was to create a 'Web of data' rather than a 'Web of documents.' Work being carried out focuses on providing links to information contained in the documents displayed on the Web.

The W3C's Web of the future will hold machine-processable information. It will allow people to find, share, and combine information more easily (Hendler et al 2002). W3C has set out to define and link the Web in a way that it can be used for more effective discovery, automation, integration, and reuse across applications. The enabling technologies and standards for the Semantic Web have been represented in Figure 1.

The Unicode and URI layers establishes the use international characters sets and provides means for identifying the objects in the Semantic Web (Koivunen and Miller, 2001). Markup languages such as XML, RDF and DAML have been developed to allow the specification of Web in this manner. With RDF and RDF Schema it is possible to make statements about objects with URI's and define vocabularies that can be referred to by URI's.

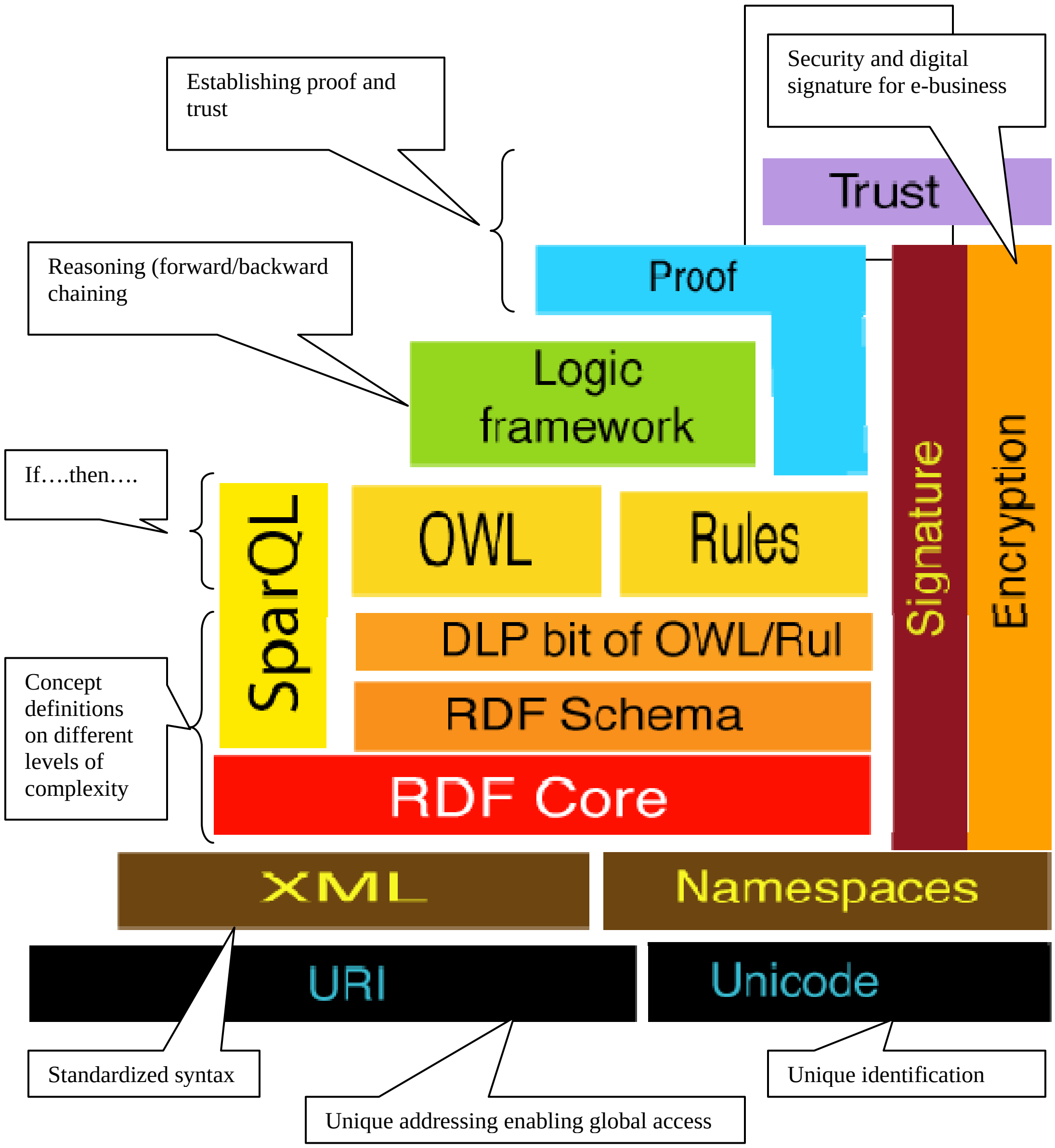

Figure 1:Semantic Web Layers (in Berners-Lee 2005)

This is the layer where users can give types to resources and links (Koivunen and Miller 2001). The next level is the ontology layer. In this layer, terms and their relationships to other terms are defined. The Ontology layer supports the evolution of vocabularies as it can define relations between the different concepts. SPARQL has been presented as part of the ontology layer in Figure 1. It is a query language for getting information from RDF graphs, which provides facilities to extract information in the form of URIs, blank nodes, plain and typed literals (Prud'hommeaux 2005). It also supports the extraction of RDF subgraphs and the construction of new RDF graphs based on information in the queried graphs.

The higher layers of Semantic Web Technologies depicted in Figure 1 are fairly advanced concepts; and since their deployment largely depends on the maturity of the 'base' technologies, the proposed work will initially focus on the use of the technologies in the lower layers.

McGuinness (2003) point out that the deployment of such technologies result in a Web that can be described as:

- Understandable term meaning and user background;
- Interoperable (can translate between applications);
- Programmable (thus agent operational);
- Explainable (thus maintains context and can adapt);
- Capable of filtering (thus limiting display and human intervention requirements);
- Capable of executing services.

THE PROPOSED SEMANTIC FRAMEWORK FOR SANITATION INITIATIVES

Computer applications can adeptly parse 'traditional' Web pages for layout and routine processing, for example, 'here a header,' 'there a link to another page.' However, such applications have no reliable way of processing the semantics such as: 'this link goes to the home page of a Community Based Organisation.' In the emerging semantic web, such markup information about the contents of a page is provided. The markup information describes what could be obtained and how it could be obtained (McGuinness 2002). Languages such as OWL, DAML+OIL (http://www.w3.org/2001/sw/WebOnt/, http://www.daml.org) capture background information and model user information for annotating web content. The proposed project will use DAML/OWL Language to extend the vocabulary of XML and RDF; thereby extending the semantic reach of current XML and RDF meta-data efforts.

The review of related work presented in a previous section established that it will not be necessary to build an entire ontology for the Sanitation Initiatives from first principle. The proposed work will exploit the emerging ontologies within the construction industry and augment these using ontologies developed in other domains, particularly the ones that have been deployed in environmental management. Other existing ontologies that will be suitable for reuse include upper level ontologies such as UNSPSC (www.unspsc.org) as well as Ontology libraries provided under DAML (see http://www.daml.org/ontologies/) and Ontolingua (http://www.ksl.stanford.edu/ontolingua).

The next phase of the project will involve an extensive review of the Sanitation domain through a series of workshops with the stakeholders. The findings will be mapped onto El-Diraby's (2003) proposed roadmap under the following components:

- Technical topics (boundary conditions) with concepts related to issues such as Productivity, Safety, Constructability, Value engineering and Partnering;
- Related domains (work environment domain), which is an umbrella for all related domains that have a bearing on the construction environment including the National

economic system to Unions, the Theory of architecture and Structural analysis practice;

- Processes including Administrative processes (like Accounting, Human resource management) and Engineering processes (like Design and Construction);
- Actors including Individuals and Organizations;
- Products including Basic products, Materials, Construction aids, Management Aids and Software;
- Resources including Materials, Labour force, Equipment, and Subcontractors;
- Projects;
- Relationships.

A conceptual representation of the proposed framework based on El-Diraby's adaptation of the e-COGNOS ontology is shown in Figure 2:

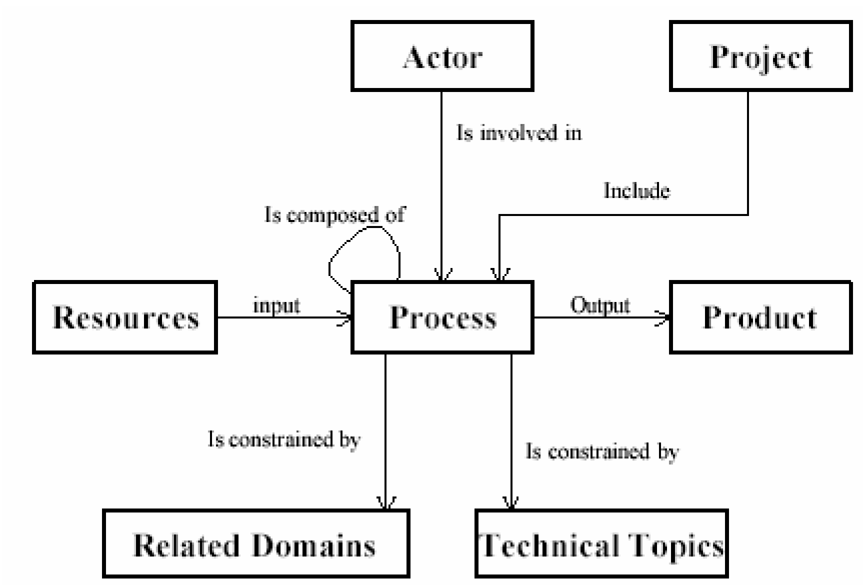

Figure 2: Semantic Framework

CONCLUSIONS

One of the greatest challenges in the developing economies' quest to sustainable strategies in community-based activities such Sanitation Initiatives is access having access to update information. Although a number of organisations have embraced the use of computing applications, many of these are specific systems and there are significant barriers to the seamless flow of information from one organisation to another. Even if the content were to be made available on the Web, without addressing the 'digital anarchy' and information overload problem, information that could potential improve the quality of decision making may never be accessed.

Emerging Semantic Web Technologies can be exploited to enhance knowledge sharing in Sanitation Initiatives. The use of these technologies results in the development of Web pages with machine-processable content and therefore reduce human intervention requirements. The use of such technologies also enables users to tag meaning into Web content using markup languages and ontologies thus enabling interoperability of applications across different organisations. Although the development of ontologies is a tedious task, there are a number of existing ontologies both within the construction industry and in other sectors, particularly environmental management that can be reused.

The next step in this project will be a comprehensive domain analysis through a series of workshops with the stakeholders. This analysis will investigate the existence of bottlenecks in knowledge sharing across different organisations in the selected domain. It will also explore the potential of using markup languages and ontologies to address the causes of such bottlenecks.

REFERENCES

Anumba, C.J., Ruikar, D., Aziz, Z., Carrillo, P.M. and Bouchlaghem, N. M., “Semantic Web and Wireless Services For Construction Collaboration.” In Soibelman, L and Pena-Mora, F. (editors). Proceedings, ASCE International Conference on Computing in Civil Engineering, Cancun, Mexico, 12th July [CD-ROM].

Baeza-Yates, R.A. (1998). “Searching the World Wide Web: Challenges and Partial.” In Coelho, H., editor. Progress in Artificial Intelligence - IBERAMIA 98, 6th Ibero-American Conference on AI, Lecture Notes in Computer Science, 1484 Springer, 39-51.

Bénédicte L. (2004). “Semantic Web Methodologies and Tools for Intra-European Sustainable Tourism.” JITT, May.

Betts, M., and Yang, J. (2004).The Significance and Priorities of IT in Sustainable Construction. Presented at the w78:2004, Integrated IT to Support Sustainable Construction, Toronto Canada

Bentivegna, V., Curwell, S., Deakin, M,. Lombardi, P., Mitchell, G., Nijkamp, P. (2002). “A vision and methodology for integrated sustainable urban development: BEQUEST.” Building Research & Information, 30(2), 83 – 94.

Berners-Lee, T. (2005). “Putting the Web back into Semantic Web.” Proceedings of the 4th International Semantic Web Conference, ISWC2005 ,Galway, Ireland

Berners-Lee, T, Hendler, J. and Lassila, O. (2001). “The Semantic Web.” Scientific American, May 2001, pp. 28-37.

Bjornsson, H. 2000, Impacts of Technology on the A/E/C Industry in the 21st Century. Presented at The Reading Construction Forum, London, October 31-November 1.

Bjornsson, H. 2000, "Emerging Business Models", Strategy (Japan), Vol. 80

Deloitte Research (2002). Directions in Collaborative Commerce, Managing the Extended Enterprise. A Deloitte Research Collaborative Commerce Viewpoint, Deloitte Research Report.

El-Diraby, T. E. (2003). "e-Infrastructure--a roadmap for semantic web-based development environment" Submitted to J. of IT in Construction.

ITCBP (2004). IT & Sustainability, The IT Construction Best Practice Programme website, (available at http://www.itcbp.org.uk/itsustainability/)

Koivunen, M. and Miller, E. (2001). "W3C Semantic Web Activity." Proceedings of the Semantic Web Kick-off Seminar in Finland.

KPMG Consulting (2000). Knowledge Management Research Report. (available at http://www.kmadvantage.com/docs/km_articles/KPMG_KM_Research_Report_2000.pdf)

Lima, C., El-Diraby, T., Fies, B., Zarli, A., and Ferneley, E. (2003). "The e-COGNOS project: current status and future directions of an ontology-enabled IT solution infrastructure supporting Knowledge Management in Construction." Construction Research Congress, ASCE, Honolulu, March 19-21.

Madhusudan, T. (2001). Enterprise Application Integration - An Agent-based Approach. Presented at the IJCAI Workshop on AI and Manufacturing, Seattle, WA, August.

McGuinness, D.L. (2003). "Ontologies Come of Age." In Fensel, D., Hendler, J., Lieberman, H. and Wahlster, W. (editors), Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2003.

McGuinness, D. (2002). The Semantic Web: Current and Future State and its potential for cyberinfrastructure for Environmental Research and Education. Invited talk for the National Science Foundation's Invitational Workshop on Cyberinfrastructure for Environmental Research and Education, National Center for Atmospheric Research, Boulder Colorado, October 30 - November 1.

Mutis, I., Issa, R. and Flood, I. (2005). "Semantic Schemas for specification Processes in the AEC Domain." in Soibelman, L and Pena-Mora, F. (editors). Proceedings, ASCE International Conference on Computing in Civil Engineering, Cancun, Mexico, 12th July [CD-ROM].

Omenya, A., Huchzermeyer, M., Kihato M. & Berrisford, S. (2004). "Slum upgrading in the complex context of policy change: the case of Nairobi." Proceedings of the The Perpetuating Challenge of Informal Settlements, 6th – 8th October, Johannesburg, SA.

Pan, J. and Anumba, C.J. (2005). A Framework for Project Information Management in A Semantic Web Environment, in Soibelman, L and Pena-Mora, F. (editors). Proceedings, ASCE International Conference on Computing in Civil Engineering, Cancun, Mexico, 12th July [CD-ROM].

Prud'hommeaux, E and Seaborne, A (editors)(2005). SPARQL Query Language for RDF, W3C Working Draft 23 November (available at http://www.w3.org/TR/2005/WD-rdf-sparql-query-20051123/)

Samtani, G. (2003): B2B integration, A Practical Guide to Collaborative E-Commerce, World Scientific Pub Co Inc, USA.

Schevers, H.A.J. and Drogemuller, R.M. (2005). "Semantic Web for an Integrated Urban Software System, In Zerger, A. and Argent, R.M. (editors). MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 170-176. ISBN: 0-9758400-2-9. (available at http://www.mssanz.org.au/modsim05/papers/ascough_1.pdf)