Fifth International Conference on Construction in the 21st Century (CITC-V)
"Collaboration and Integration in Engineering, Management and Technology"
May 20-22, 2009, Istanbul, Turkey

Exploring the Use of Semantic Web Techniques for Representing Knowledge about Sustainable Building Technologies

J.H.M. Tah, F.H. Abanda
Department of Real Estate and Construction, School of the Built Environment,
Oxford Brookes University, Oxford, OX3 0BP, UK
jtah@brookes.ac.uk, fabanda@brookes.ac.uk

Abstract
The global quest for sustainability in the exploitation of resources and the need for carbon foot-print reduction are generating a large number of innovations and a huge amount of knowledge on sustainable building technologies. Unfortunately, users are being overwhelmed with information overload in this area and it is difficult for them to make informed choices. The emergence of semantic Web technologies, the next generation of Web technologies, promises to considerably improve information representation, sharing and re-use to support decision making. The aim of the work presented in this paper is to explore the extent to which emerging semantic Web technologies can be exploited to both represent information and knowledge about sustainable building technologies and to facilitate system decision making in recommending appropriate choices for use in different situations. This paper presents an overview of emerging semantic Web technologies and emerging innovations in sustainable building technologies. A conceptual model for representing this information is presented. The use of this model to develop a prototypical ontology for representing sustainable building technology knowledge in the Photovoltaic system domain is discussed. The outcomes of the exploratory work that has been undertaken to identify and use various tools for the representation of knowledge and making inferences from the knowledge are discussed.

Keywords
Sustainable building technologies, Ontology, Knowledge representation, Reasoning, Knowledge management

1. Introduction

Society and governments around the world are encouraging the development and use of innovative sustainable building technologies to improve the performance of buildings and mitigate the effects of climate change. This has resulted in the development of a wide range of different innovations with a large amount of information and knowledge on sustainable building technologies. Information and knowledge about these innovations is being made available to users through the Web to facilitate accessibility and use. Although the attraction of the Web is its simplicity and ease of accessibility, the share wide ranging nature of these innovations means that internet searches often overwhelm individuals and practitioners with millions of pages that they have to browse through to identify suitable innovations to use on their projects. Users are therefore unable to make informed choices and have to rely on specialists with experience on a limited range of innovations for advice. It has been widely acknowledged that the solution to this problem is the use of a machine understandable language with rich semantics for some or all of the information on the Web. This has led to the emergence of the Semantic Web, the next generation of the Web, which promises to considerably improve information representation, sharing, re-use and automated processing by software agents to make inferences. Key to this is the use of a common language or an ontology for representing knowledge from different sources to facilitate decision making. The aim of the work presented in this paper is to explore the extent to which emerging semantic Web technologies can be exploited to both represent information and knowledge about sustainable building technologies and to facilitate system decision making in recommending appropriate choices for use in different situations. An overview of semantic Web technologies and sustainable building technologies is presented. A conceptual model, developed to facilitate abstraction and representation of this information and knowledge is presented. The outcomes of the exploratory work that has been undertaken to identify and use various semantic Web techniques and tools for the representation of knowledge and making inferences from the knowledge are discussed.

2. An Overview of Semantic Web Technologies

The emergence of the World Wide Web has brought exciting new possibilities in information access and electronic business. It has grown to be the largest distributed repository of information ever created. Web content consists largely of distributed hypertext and hypermedia, accessible via keyword-based search and link navigation. Simplicity is one of the Web's great strengths and an important factor in its popularity and growth. It is this simplicity that has fuelled its wide uptake and exponential growth but it is this same simplicity that is hampering its further growth and exploitation. The explosion in the range and quantity of Web content also exposes serious shortcomings in the hypertext paradigm. It is increasingly difficult to locate required content through existing search and browse methods. Finding the right piece of information is often challenging. Search engines can assist in finding content containing specific words, but it is very easy to get lost in the huge amounts of irrelevant material and miss that which is relevant. Searches are imprecise, requiring users to read through a large number of retrieved documents to extract the right information. There has emerged a view that the solution to this problem requires that there be a machine understandable semantics for some or all of the information on the WWW. The realisation of such a Semantic Web (Berners-Lee *et al*., 2001) requires developing languages for expressing machine-understandable languages (or ontologies) and making them available on the Web. According to the World Wide Web Consortium (W3C, 2009), the goal of the Semantic Web is to allow data to be shared effectively by wider communities, and to be processed automatically by tools as well as manually. The vision of a semantic Web is very ambitious and will require solving long-standing research problems in knowledge representation and reasoning, natural language computing, computer vision and agent systems. However, considerable progress is being made in the infrastructure required to support the semantic Web, particularly in the development of languages and tools for content annotation and design and deployment of ontologies. Although the realization of the semantic Web is still a long way off, our aim in the work presented in this paper is to explore the extent to which we can apply emerging developments in this area to provide decision support and recommendations of appropriate innovations in sustainable building technologies for use in a particular situation. An ontology language is a key feature of the semantic Web and is the subject of the ensuing section.

2.1 Ontology and Web Ontology Language

In the early 1990s, the semantic Web research community recognized the need for an ontology Web language and several proposals for new Web ontology languages emerged. These included Simple HTML Ontological Extensions (SHOE, 2009), the Ontology Inference Layer (OIL, 2009), and DAML+OIL (2009). The W3C (2009) set up a standardization working group in 2001 to develop a standard for a Web ontology language having recognized that an ontology-language standard is a prerequisite to developing the semantic Web. This resulted in the development of the Ontology Web Language (OWL) ontology language standard (W3C, 2009) in 2004 exploiting earlier work on OIL and DAML+OIL. A key feature of OIL is it's basis in Description Logics (DLs), a family of logic-based knowledge representation formalisms descended from Semantic Networks and KL-ONE (Brachman and Schmolze, 1985) but that have a formal semantics based on first-order logic (Baader *et al*., 2003). These formalisms all adopt an object-oriented model in which the domain is described in terms of individuals, concepts or classes, and roles or properties. For example in the Photovoltaic systems domain, a particular system on the market may be called PS-EMS1. In this case, the concept is “PhotovoltaicSystem”, the individual is “PS-EMS1”, and the role “isTypeOf” describes the relationship between the concept and the individual. In a strict object-oriented paradigm the terms classes is used for concepts, properties for roles, and instances for individuals. Descriptive Logics can be used to create a knowledge-base. The formal semantics allows for the development of reasoning algorithms that can be used to correctly make inferences and answer complex queries about a domain.

2.2 Ontology Applications

The Ontology Web Language (OWL) is being used in fields as diverse as biology (Sidhu *et al*., 2005), medicine (Shvaiko and Euzenat, 2005), geography (Goodwin, 2005), geology (Sweet, 2008), agriculture (Soerget *et al*., 2004), defense (Lacy *et al*., 2005), and construction (Beetza1 *et al*., 2009). The application of OWL is particularly common in the life sciences where it is used by developers of several large biomedical ontologies such as BioPAX (2009), GO (2009), SNOMED (2009), the Foundational Model of Anatomy (FMA, 2008) and the U.S. National Cancer Institute thesaurus on cancer terminology (USNCI (2008). Many ontologies are typically developed through collaborative endeavours within a given community aimed at facilitating information sharing and exchange. This is certainly the case with the ifcOWL developed by the building SMART (2009) initiative in the construction industry for application in the building information modelling domain. Many OWL ontologies are available on the Web. In the following section we discuss the sustainable building technologies domain and the prototypical ontology being developed to explore the potential application of the semantic Web techniques in this area.

3. An Overview of Innovations in Sustainable Building Technologies

There exists a broad range of products that fall within the sustainable building technology domain. There are many different characteristics which products may exhibit to demonstrate a degree of sustainability. Some products may be capable of being installed or assembled on site with minimum impacts on the environment. Some may be more energy efficient. Some may eliminate the use of fossil fuels. Some may include high proportions of re-cycled materials. Some may use natural sustainable materials. Some may promote water efficiency and some may reduce carbon emissions. These include materials, components, and systems for off-site manufacture (steel frame, timber frame, structurally insulated panels, etc.), natural building materials (such as straw-bale, strawboard, rammed earth, etc.), energy efficiency technologies including renewable energy technologies (such as wind turbines, photovoltaic, geothermal, air source heat pumps, etc.), water conservation and sustainable urban drainage systems (such as rainwater harvesting, grey water recycling, green roofs, porous pavements, etc), waste minimisation (recycled wall ties, recycled kitchen units, etc), etc. A wide range of innovative sustainable building technologies are emerging in response to the global quest for sustainability in the exploitation of natural resources and the need for carbon foot-print reduction in response to climate change. Many suppliers now boost of having a wide range of products in the order of several thousands in their catalogues. A great deal of knowledge and understanding of these innovations is required to be able to make the right choices to use in particular situations. This presents potential clients and users with significant challenges in selecting the right products to suit particular uses, especially in areas of emerging technologies. These catalogues have traditionally been produced in book form but are increasingly being made available through the web. The web has improved accessibility to users but even a simple search of a particular innovation will produce millions of pages through which a user needs to read. Even where a user has knowledge of a particular type of product, it is not possible for the user to issue even the simplest of queries to identify product instances that meet their particular requirement. This makes this particular area suitable for exploration for the potential application of semantic Web techniques to help users locate specific products that their particular need. This is the goal of this work. In a research endeavour of this nature with limited `resources, it is not possible to tackle the full range of products available. We will use a sub-set of the technologies available for exploratory reasons and hope to develop insights that can be applicable to the full range of sustainable building technologies. We have chosen to focus on Photovoltaic technologies as they are currently generating a lot of interest in the domestic building area with high public visibility.

3.1 Photovoltaic Systems

Photovoltaic (PV) cells are semiconductor devices which convert energy in sunlight directly into electricity. Individual cells only generate low voltages and currents, so they are usually grouped in rectangular 'modules' that comprise a transparent cover, a metal mounting frame and a backplate, thus forming a weatherproof enclosure. Modules are often grouped into arrays. PV cells can also be moulded into solar slates or solar tiles for integration into roofs, or bonded onto glass or metal sheets for incorporation into architectural glazing and fascia systems. Various types of PV technology use different semiconductor materials and manufacturing techniques. PV installations have a wide variation in outputs and are thus rated according to their peak power output (kWpeak or kWp). PV systems can be grid-connected or they can be stand-alone. A typical 'grid-connected' system allows the installation to put power into the building mains electricity supply in parallel with the local grid. When the building demands more electricity than the PV can provide, the grid provides the 'top-up'. When the PV is generating more energy than the building needs, the excess is exported to the grid. The key component in such a system is the inverter, which converts the PV generated direct current (DC) into alternating current (AC), and does so in synchrony with the mains. Grid connected systems require very little maintenance, generally limited to ensuring that the panels are kept relatively clean. The wiring and components of the system need to be checked regularly by a qualified technician. Stand-alone systems, i.e. those not connected to the grid, need maintenance on other system components, such as batteries. Prices for PV systems vary, depending on the size of the system to be installed, type of PV cell used and the nature of the actual building on which the PV is mounted. The size of the system is dictated by the amount of electricity required.

4. The PV Systems Ontology

A pre-requisite to developing a semantic Web-based application is the establishment of an ontology for the application domain. There are many groups developing ontologies in different application domains as previously indicated. Where ontologies already exist for a domain, it may be simply a case of adopting one or more of these and extending them where necessary. We are not aware of the existence of an ontology in the PV domain so we have had to develop one. Our approach was to start by undertaking an abstraction of the concepts that characterise a PV system. We consulted the literature on PV systems from various sources to undertake the abstraction and held discussions with a manufacturer of PV systems. We used the Unified Modelling Language (UML) to create a semantically rich class diagram to allow us to graphically represent and visualise the concepts and relationships between them. This class diagram is depicted in Figure 1. This indicates that a photovoltaic system is composed of photovoltaic and non-photovoltaic component parts. A photovoltaic component part can be a module or a combination of modules into a module panel, sub-array, array, or an array sub-field. The non-photovoltaic components consist of what is termed the Balance of Components which themselves are the electrical and mechanical components shown in the diagram. The diagram also shows that a PV system can be grid connected or non-grid connected. A non-grid connected system can be a stand alone DC, a standalone DC-AC, or a hybrid of both. We also extend this ontology to include concepts that would allow us to develop an application for recommending specific PVs for domestic use. For the sake of brevity, these concepts are not shown on the diagram. These concepts include household appliances that use the electricity generated and the organisations that supply PV systems. This then provided us with the basis to implement an initial prototypical ontology in an ontology development environment to allow us to undertake experimentation and exploration of semantic Web techniques as discussed in the ensuing section.

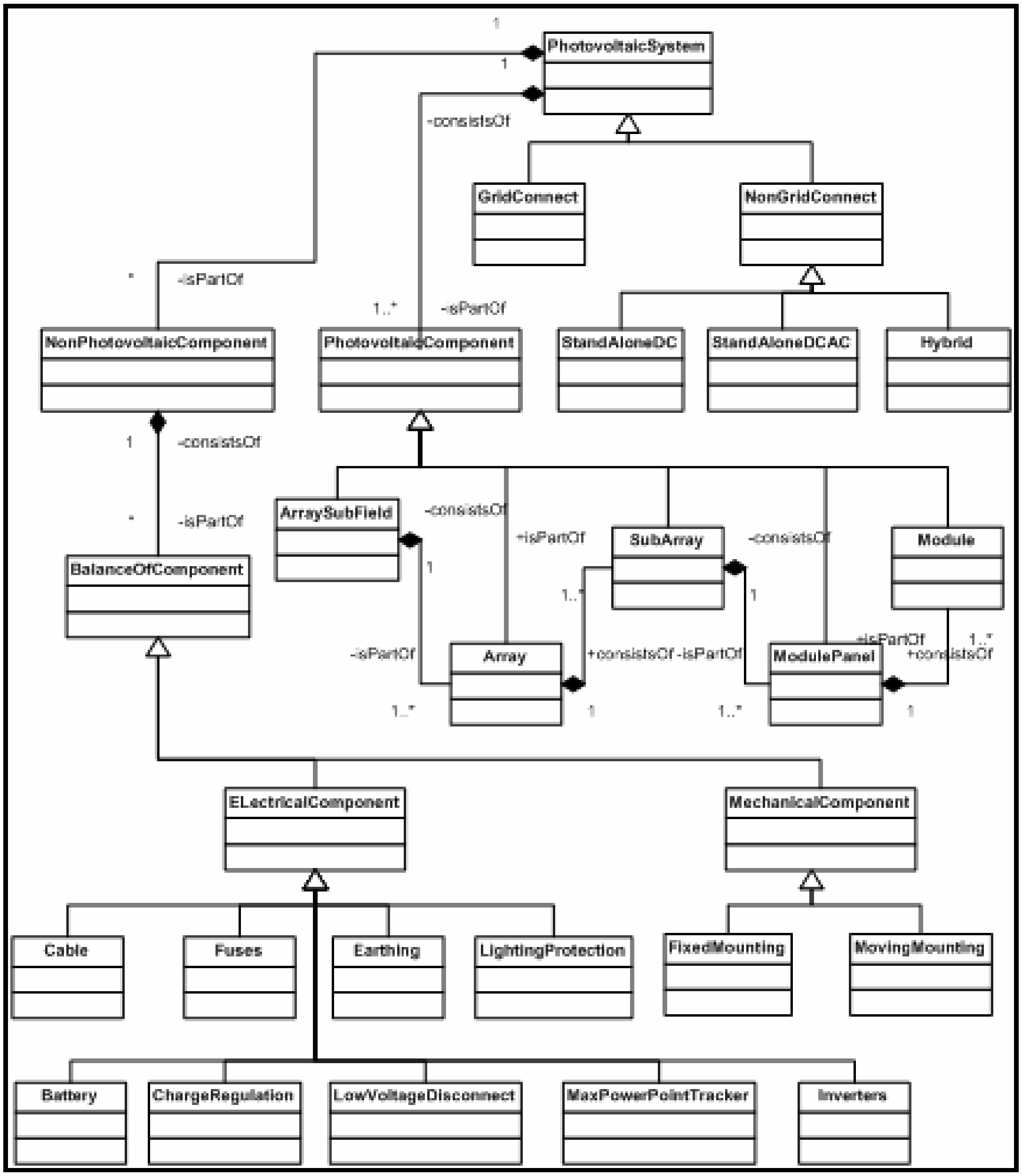

Figure 1: The PV System Class Diagram

4.1 The Initial Prototype Implementation of the PV System Ontology

The implementation of the prototype ontology required us to choose an appropriate ontology editor and development environment. The growing interest in the semantic Web has stimulated an exponential growth in the development of ontology editors based on different ontology development methodologies. There exist different ontology development methodologies (Corcho *et al.*, 2003) in the literature. The challenge lies in the choice of a methodology and hence tools used in building the ontology. Furthermore in making any choice, an ontology engineer is often guided by the prime purpose of the ontology. The PV ontology is intended to serve as a knowledge-base where intelligent searches about the various PV technologies, their properties, application, suppliers, installers and cost implications can be determined. In designing such a knowledge-base a balance must be established between the simplicity and complexity of the knowledge modelling facilities and the amount of information in the domain of discourse (Sure *et al*., 2005). There is a high degree of complexity in our problem domain, thus the knowledge modelling tools used have been carefully chosen to cope with the potentially large size of the PV and related ontologies, and to be able to provide reasoning capability. Based on the purpose of our PV ontology, initially reviewed in Abanda and Tah (2007) and on a review of various ontology editors by Corcho *et al*., (2003), the Protégé development platform which contains the Protégé-OWL ontology editor for the semantic Web was chosen for use in this study. It is compliant with the OWL Web Ontology Language standard recommendation of the W3C as indicated earlier. The OWL ontology language is available in three dialects supporting varying levels of expressive capability (Lacy, 2005). These are: OWL-Lite, OWL-DL and OWL Full. OWL-DL was chosen because of its richer expressive power than OWL-Lite, and it facilitates automatic reasoning than OWL-Full (Lacy, 2005). Furthermore, the availability of user-friendly plug-ins with a scalable architecture in Protégé-OWL is an added advantage. Three of the user-friendly plug-ins (OWLviz, FaCT++ and DL query plug-ins) have been used in our exploratory work. OWLviz enables the class hierarchies in an OWL ontology to be viewed and incrementally navigated, allowing for comparison of the manually constructed class hierarchy referred to as the asserted class hierarchy and the automatically computed class hierarchy referred to as the inferred class hierarchy. FaCT++ provides reasoning support for Descriptive Logic (DL) that underpins OWL (Horridge *et al.*, 2006; Horrocks, 2008). The DL query plug-in provides a powerful and easy-to-use feature for inferencing over ontology to make inferences on information about a particular class, property or individual in ontology.

4.2 Knowledge Representation, Ontology Validation and Reasoning

The Protégé editor was used to implement the PV concepts abstracted and the relationships between the concepts. We introduced further explicit concepts and attributes wherever possible to enrich the semantics of the PV ontology to facilitate autonomous reasoning. The representation of data type properties for units of measurement presented a challenge as Protégé does not currently offer an obvious way of representing this. For instance, it is not obvious to represent the fact that a PV array has an area of 1.2 m2 with the units of m2 included. This was overcome by creating Unit ontology and exploiting the annotation properties in describing extended knowledge about the data type properties and the units. The Protégé platform allowed us to use the FaCT++ plug-in to automatically validate the ontology. It automatically compares the manually constructed ontology concept taxonomy (called the asserted hierarchy) with the computer generated hierarchy (called the inferred hierarchy). If there are no differences between the asserted hierarchy and the inferred hierarchy before and after automatic computation by FaCT++, then the ontology has passed the validation test. Otherwise, areas of inconsistency are highlighted in the editor for examination and correction. We have experimented with the DL query facilities in Protégé to investigate the extent to which we can interrogate the PV ontology for decision support purposes. We have successfully developed a number of queries aimed at identifying the types of PV systems supplied by specific suppliers and restricting the results using constraints such as size and peak output. We are currently exploring how to construct composite and more sophisticated extended queries that can be used to make inferences to support far more substantive decision making.

5. Conclusion

A wide range of innovations in sustainable building technologies are emerging in response to the need to tackle the effects of climate change and promote sustainable development. Information and knowledge about these innovations is being made available to users through the Web to facilitate accessibility and use. Although the attraction of the Web is its simplicity and ease of accessibility, the share numbers and wide ranging nature of these innovations means that internet searches often overwhelm individuals and practitioners with millions of pages that they have to browse through to identify suitable innovations to use on their projects. Users are therefore unable to make informed choices and have to rely on specialists with experience on a limited range of innovations for advice. It has been widely acknowledged that the solution to this problem requires that there be a machine understandable language or semantics for some or all of the information on the Web. There is therefore a lot of on-going work on developing the next generation of the Web, called the Semantic Web. Key to this is the development of an ontology which is a semantically-rich representation of the key concepts and the relationships between them in a particular domain of interest. This represents information and knowledge about a domain in a language that can be automatically read and understood by computers which facilitate automated reasoning in support of decision making. We have explored the potential application of these techniques in the sustainable building technology domain. Due to the large scope of this domain, we have restricted the scope of the work reported in here to the Photovoltaic Systems domain. A prototypical ontology has been developed and validated. A prototype system has been implemented to explore the extent to which the ontology can be used in conjunction with other reasoning tools to support individuals in selecting PV systems for their applications. Simple queries have been developed to successfully interrogate the ontology and further work is ongoing to investigate the extent to which far more sophisticated reasoning mechanisms can be developed and used to support substantive decision making.

6. References

Abanda, F.H., and Tah, H.M. (2007). “Towards the development of a Knowledge-Repository of Emerging Innovations for Sustainable Housing Development”, *Proceedings of 6th International Postgraduate Research Conference*, University of Salford. pp. 345-353.

Baader, F., Calvanese. D., McGuinness, D., Nardi, D., and Patet-Schneider, P.F., Eds. (2003). The Description Logic Handbook: Theory Implementation and Applications. Cambridge University Press. Cambridge, U.K.

Beetza1 J., Jos van Leeuwenal, and Bauke de Vriesa1. (2009). “IfcOWL: A case of transforming EXPRESS schemas into ontologies”. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol. 23, pp. 89-101.

Berners-Lee, Tim, James Hendler and Ora Lassila. (2001). “The Semantic Web”. *Scientific American Magazine*, Vol. 284, No. 4, pp. 34-43. http://www.sciam.com/article.cfm?id=the-semantic-web&print=true. Retrieved on 26 March 2008.

BioPAX (2009). Biological Pathways Exchange. http://www.biopax.org/, accessed January 2009.

BuildingSMART (2009). Industry Alliance for Interoperability. http://www.iai.org.uk/, accessed January 2009.

Brachman, R.J., and Schmolze, J. (1985). “An overview of the KL-ONE Knowledge Representation System", *Cognitive Sci* , Vol. 9, No. 2.

Corcho, O., Fernández-López, Gómez-Pérez. (2003). “Methodologies, tools and languages for building ontologies. Where is the meeting point?”, *Data & Knowledge Engineering*, Vol. 46, pp. 41-64.

DAML+OIL. (2009). The DARPA Agent Markup Language. http://www.daml.org/, accessed January 2009.

FMA. (2008). (sig.biostr. washington.edu/projccts/fm/), accessed January 2009.

GO. (2009). Gene Ontology project. http://www.geneontology.org/, accessed January 2009.

Horridge, M., Tsarkov, D., and Redmond, T. (2006). Supporting Early Adoption of OWL1.1 with Protege-OWL and FaCT++. [Online] http://owl-workshop.man.ac.uk/acceptedLong/submission_15.pdf [20 September 2008].

Horrocks, I. (2008). “Ontologies and the Semantic Web”. *Communications of the ACM*, Vol. 51, No. 12, pp. 58-67.

Goodwin, J. (2005). “Experiences of using OWL at the ordnance survey”. *In Proceedings of the First OWL Experiences and Directions Workshop* (Galway, Ireland, Nov.). CEUR-WS, 2005; CEUR-WS.org/Vol-188/.

Lacy, L., Avites, G., Fraser. K., Gerber, W., Mulvehill. A., and Gaskilt, P. (2005). “Experiences using OWL in military applications”. *In Proceedings of the First OWL Experiences and Directions Workshop* (Gatway Ireland, Nov. 2005); CEUR-WS.org/Vol-188/.

Lacy, L. W. (2005). *OWL:Representing Information Using the Web Ontology Language*. UK: Trafford Publishing, UK.

OIL. (2009). Ontology Inference Layer. http://www.ontoknowledge.org/oil/, accessed January 2009.

SHOE. (2009). Simple HTML Ontology Extensions. http://www.cs.umd.edu/projects/plus/SHOE/onts/, accessed January 2009.

Shvaiko. P., and Euzenat, J. (2005). “A survey of schema-based matching approaches”. *Journal of Data Semantics IV*, Lecture Notes in Computer Science, 3730, pp. 146-171.

Sidhu, A.. Dillon, T., Chang. E., and Sidhu, O.S. (2005). “Protein ontology development using OWL”. *In Proceedings of the First OWL Expetiences ond Directions Workshop* (Galway, Ireland. Nov. 2005); CEUP-WS.org/Vol-188/.

SNOMED. (2009). SNOMED Clinical Terms, http://www.connectingforhealth.nhs.uk/ systemsandservices/data/snomed, access January 2009.

Soerget, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., and Katz, S. (2004).“Reengineering thesauri for new applications: The AGPOVOC example”. *Journal of Digita Information*, Vol. 4, No. 4.

Sure,Y., Bloehdorn, S., Haase, P. Hartmann, J., and Oberle, D. (2005). „The SWRC ontology - Semantic web for research communities”. *In Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005)*. Springer, Covilha, Portugal, December 2005.

Sweet. (2008). The Semantic Web for Earth and Environmental Terminology project, sweet.jpl.nasa.gov/

USNCI. (2008). The U.S. National Cancer Institute thesaurus (www.cancer.gov/cancertopics /terminologyresourceshow).

W3C. (2009). OWL Web Ontology Language [Online]. http://www.w3.org/TR/owl-features/, accessed January 2009.