Development of a Taxonomy for Indexing Web-Based Mining Safety and Health Research

A F Glowacki[1]

ABSTRACT

To solve the mining safety and health needs of the future, access to, and sharing of, timely and relevant information will be crucial. The next generation of the Web will provide tools (RDF, OWL, SPARQL, etc) to define and link content for more effective discovery, integration, and reuse across applications and organisations. A recent Gartner Research Report (Finding and exploiting value in Semantic Web technologies on the web, May 2007) predicts these technologies will begin to coalesce in ten years.

Implementing this *Semantic Web* requires structured web content described via metadata (information about information). If an industry or enterprise bases its metadata on a common vocabulary, it becomes possible to share and reuse information.

The National Institute for Occupational Safety and Health's (NIOSH) Mining web site provides access to more than 1500 NIOSH-developed publications, products, and topic pages on issues as diverse as ground control, respiratory diseases, and mine emergency management. As the site has grown, locating content has become increasingly difficult. To improve the searching process, a panel of internal experts representing all major research programs was established to develop a controlled vocabulary for indexing that content. The outcome is the NIOSH Mining Safety and Health Thesaurus, a faceted, hierarchical thesaurus containing hundreds of mining safety and health terms organised into 20 facets (eg diseases, injuries and disorders; hazardous events and exposures, equipment, engineering and administrative controls).

One immediate goal of this project is to reduce the complexity of locating information on the NIOSH Mining web site by implementing metadata and faceted information retrieval techniques. Another goal is to stimulate discussion in the international mining community regarding developing metadata standards so the industry is ready to adopt future Semantic Web technologies to improve access to mining safety and health information.

INTRODUCTION

Since its inception in the 1990s, the World Wide Web has grown to tens of billions of documents. While search engines such as Google and Yahoo do an impressive job of indexing this content, finding information is still difficult. Some problems include:

- a search for 'George Washington' (the US president) returns irrelevant content, such as pages about George Washington University or George Washington Bridge (false positives);
- a search on the word 'car' excludes pages where a synonym was used, such as automobile (false negatives);
- much information is locked away in databases that are not indexed by search engines; and
- a search returns tens or even hundreds of thousands of web pages, too much information for any human to process;

To address these and other problems, the World Wide Web Consortium (W3C) is developing technologies for a next-generation Semantic Web. The W3C is the primary international standards organisation for the Web. Founded by Tim Berners-Lee, creator of the first web browser and many other core web technologies (HTML, HTTP, etc), the W3C is represented by more than 400 member companies, universities, and other organisations.

As described by Berners-Lee, Hendler and Lassila (2001), the Semantic Web applies structure and meaning to web content, enabling software agents to automate the processing of data and perform sophisticated tasks without human intervention. It is an extension of the current Web:

> *... in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*

A scenario is presented where a hand-held web browser automatically schedules a physical therapy session, finding a provider in the patient's insurance plan, located within a specified driving distance, with an available appointment that does not conflict with the patient's calendar.

While this future Web is far from a reality, many of the necessary base technologies have been developed under the auspices of the W3C and are already being used to develop powerful web applications. This paper presents an overview of these technologies and their potential for revolutionising the Web. It also presents a thesaurus that is under development at the National Institute for Occupational Safety and Health (NIOSH) as a preliminary step toward applying semantic meaning to mining safety and health information.

OVERVIEW OF SEMANTIC WEB TECHNOLOGIES

The current Web is a collection of documents which can be individually retrieved and read by humans. The Semantic Web, on the other hand, focuses on data and information that can be manipulated by computers. Achieving this vision requires developing common frameworks and tools that allow information from diverse data sources to be integrated and combined, as if these information sources were one large database. Lam *et al* (2006) describes some of the necessary component technologies and tools for developing Semantic Web applications including:

- ontologies, or controlled vocabularies for representing information about a specific knowledge domain;
- standardised ontological languages for encoding ontologies to enable machine readability; and
- Semantic Web-aware tools, such as tools for querying and reasoning about information formatted in ontological languages.

Ontologies

Historically, the term *ontology* has been used to describe a wide range of classification systems of increasing complexity, including:

- Controlled vocabulary – a simple enumerated list of terms.
- Taxonomy – a controlled vocabulary organised into a hierarchical structure. The parent-child terms may have different types of implied relationships, including whole-part, genus-species and type-instance.
- Thesaurus – a taxonomy with associative relationships, in addition to the hierarchical ones. Terms can be defined as related, even if they do not have a parent-child relationship. Equivalence (synonym) relations are also permitted.

1. Computer Scientist, National Institute for Occupational Safety and Health, Pittsburgh Research Laboratory, 626 Cochrans Mill Road, Pittsburgh Pennsylvania 15236, USA. Email: aglowacki@cdc.gov

However, for the purposes of the Semantic Web, ontologies extend the concept of thesauri by allowing for even more detailed semantic relationships among terms. The goal is to represent not just facts but knowledge, and to support inference and reasoning by both humans and software. According to the W3C (2004a), ontologies must specify the following concepts:

- classes (concepts in a domain),
- subclasses (relate more specific classes to more general classes),
- clearly-defined relationships that can exist among things, and
- properties (or attributes) that those things may have.

Ontological languages

Ontological (or knowledge representation) languages codify the semantic knowledge about a particular domain. In many cases, they are constructed from existing taxonomies and thesauri. They are intended for processing by computers, not to be read by people. The following ontological languages, in increasing level of complexity, are approved W3C recommendations:

- Resource Description Framework (RDF) (W3C, 1999);
- RDF Schema (RDF-S) (W3C, 2004b); and
- Web Ontology Language (OWL) (W3C, 2004c).

RDF is a data model that provides limited semantic capability for describing resources and their relationships. A resource is anything that can be described, such as a book, photo, animal, person, or web page. RDF has roots in the Artificial Intelligence community. It is based on defining triplets consisting of a subject, predicate, and object, each of which is unambiguously defined. The means by which this is accomplished and the language syntax is beyond the scope of this paper. However, a simplified representation of two triplets is shown in Table 1. Based on this information, one can determine that the book with ISBN 0 307387 13 5 has the Title *No Country for Old Men* and is authored by Cormac McCarthy. Similarly, Cormac McCarthy is the creator of the book *No Country for Old Men* having ISBN 0 307387 13 5.

TABLE 1

Sample triplets that could be encoded in RDF format.

Subject	Predicate	Object
<isbn: 0307387135>	<hasTitle>	*No Country for Old Men*
<isbn: 0307387135>	<hasCreator>	Cormac McCarthy

RDF-S extends RDF with additional features, including the capability to define classes and subclasses of resources and instances of those classes. For example, the class-subclass relationships of a biological classification system (eg Species is a subclass of Genus, Genus is a subclass of Family, Family is a subclass of Order) can be represented. Members of a subclass inherit all attributes and relationships of the parent class. Therefore, defining a resource to be a member of a particular genus implies that it is also a member of all associated higher-level taxonomic groups.

OWL adds even more semantic capabilities. For example, cardinality constraints can be defined (eg a pet can have only one owner). OWL also provides the ability to define enumerated classes. For example, weekdays can be defined as {Monday, Tuesday, Wednesday, Thursday and Friday}. Transitive properties can also be defined. For example, if resource A has a specific relationship to resource B, and resource B has the same relationship with resource C, the implication is that resource A also has that relationship with resource C. One very powerful feature is the ability to denote that two classes or instances from different ontologies are the same thing (eg a 'person' in ontology A is the same as a 'human' in ontology B). This allows for sharing of information even when two organisations are not using the same ontology. Although OWL was approved as a W3C recommendation in 2004, extensions are underway.

Semantic Web tools

Additional tools needed to fully support a Semantic Web include a standard for searching and aggregating information in RDF format. That has been addressed by the W3C (2008) SPARQL Protocol and RDF Query Language recommendation. Also essential are rules and inference engines which can reason and make inferences (derive new information) from facts and relationships stored in RDF. As explained by Shadbolt, Hall and Berners-Lee (2006), OWL contains some inference support, but a core language is needed for defining more powerful rules. The W3C Rule Interchange Format Working Group is currently addressing this problem. Finally, trust and copyright protection is a concern. For example, since a Semantic Web would aggregate information without human intervention, methods are needed to assure users that retrieved information came from trusted sources. Much work remains to be done in these areas.

EVOLUTION OF THE SEMANTIC WEB

While many core Semantic Web technologies have been established, they are still in their infancy. The Semantic Web is far from a reality, and some doubt it will ever be fully realised. Feigenbaum *et al* (2007) acknowledge that the grand vision of software agents performing automated tasks has not occurred, but describe a number of enterprise-level successes, particularly in the life sciences and health care fields.

In a recent Gartner Research report, Cearley, Andrews and Gall (2007) estimate that the Semantic Web will evolve slowly over the next 20 years. It will begin with web sites implementing semantic hypertext, a simpler, though less powerful method of storing structured metadata directly within web pages. By 2012, 80 per cent of public web sites will use semantic hypertext and 15 per cent will use more extensive Semantic Web-based ontologies. Some domains will need more complex ontologies and extensive formal vocabularies, such as life sciences, health care, digital libraries, defence, government, energy (the oil industry) and financial services. By 2017, the Semantic Web will begin to materialise, as standards evolve and vocabularies and ontologies being developed by various industries begin to mature. Therefore, the authors recommend that enterprises adopt Semantic Web technologies in incremental steps. For most enterprises, the approach will begin with semantic hypertext, with RDF and OWL ontologies reserved for more complex needs.

SEMANTIC TECHNOLOGIES AND THE MINING INDUSTRY

The mining industry shares many similarities with domains that the Gartner report predicts will need more sophisticated semantic techniques, particularly health, digital libraries, government and energy. Assuming the Semantic Web materialises, it is not difficult to imagine how it could impact the mining industry. For example:

- Hendler (2003) argues that these technologies can facilitate scientific collaboration among scientists at different institutions and across disciplines, enabling scientists to more easily share information and to explore related research from entirely different scientific fields. However, this will require joint initiatives between research scientists and information technologists, such as the e-Science initiative in the United Kingdom.

- Spivack (2006) envisions more efficient marketplaces where software agents learn about products, services, vendors, transactions and market trends and can connect them in optimal ways.
- Semantic Web techniques could facilitate integrating information across government agencies, as well as enable governments to more effectively target information for different audiences (citizens, businesses, etc).

In many cases, the mining industry may be able to adopt ontologies and other formal vocabularies from other domains. For example, ANSI/NISO Z39.85-2007 (the Dublin Core Metadata Element Set) provides a standard for tracking bibliographic metadata for digital materials (publications, web pages, etc), expressible in RDF-S. However, it is unlikely that domains outside of the mining industry will develop vocabularies suitable for tracking more mining-specific content.

NIOSH is researching methods to better index and organise content on its Mining Safety and Health web site. The goals are to improve customer access to our content now, while preparing for integration with the next generation of the Web. The web site content currently includes more than 1500 publications, products, and web pages on topics as diverse as ground control, respiratory diseases, dust control, mine ventilation, explosions, explosives, hearing loss, diesel, electrical safety, ergonomics, emergency management, equipment design and education and training. The content is organised into topic pages covering these broad topics. However, that structure is ineffective for such a large amount of content. For example, the ground control area alone includes more than 300 publications. Furthermore, many publications relate to multiple research areas. A more flexible organisational structure is needed that allows content to be assigned to multiple, relevant categories.

NIOSH conducted a literature search for suitable taxonomies, thesauri, or ontologies for indexing mining content, with particular emphasis on safety and health topics. The hope was to reuse existing standards and vocabularies, rather than 'reinventing the wheel'. Table 2 shows some of the more relevant taxonomies identified, which either already existed or were under development.

While some of these classification systems contain useful terms for categorising mining safety and health-related web content, none are complete solutions for NIOSH. For example:

- The health industry has well-developed thesauri for indexing medical and health content, but most utilise advanced terms that are too complex for a non-medical audience.
- Several impressive occupational safety and health thesauri exist. However, they either don't address mining-specific issues, or they address them at a very general level. A preferable approach would provide high-level terms for classifying small collections of content, but more detailed terms for classifying larger collections.
- While NAICS categorises many sectors of the mining industry, it often groups them in ways that do not make sense from a health and safety standpoint. For example, commodities that do not have much in common with each other are placed in the same hierarchies.
- Many of these taxonomies are copyrighted or require a purchase fee, which limits their extensibility and reuse by the mining industry.

THE NIOSH MINING SAFETY AND HEALTH THESAURUS

Unable to find an existing taxonomy to meet its needs, NIOSH chose to develop a new one. While the focus is on developing a controlled vocabulary for indexing safety and health information, the taxonomy includes general-purpose mining terms that are applicable to other sectors of the mining industry.

TABLE 2

Sample taxonomies related to health, injury, safety and mining.

Category	Taxonomy
Health/injury	*International Classification of Diseases (ICD)* World Health Organization (2007)
	Occupational Injury and Illness Classification System Manual (OIICS) US Department of Labor (2007)
Safety	*Australian Occupational Health and Safety Thesaurus (AOHST)* National Occupational Health and Safety Commission (2003)
	Thesaurus of Injury Prevention Terminology Chang *et al* (2003)
	Injury Prevention and Safety Promotion Thesaurus (IPSP), under development Lawrence *et al* (2006)
	CIS Occupational Safety and Health Thesaurus International Occupational Safety and Health Information Centre (2001)
	Dictionary and Thesaurus of Environment, Health and Safety US Department of Energy (1992)
Mining, minerals and geosciences	*The Geoscience, Minerals and Petroleum Thesaurus (GeMPeT)* Edinger and Barker (2003)
	Multilingual Thesaurus of Geosciences (MULTHES) Gravesteijn *et al* (1995)
	WAND Minerals, Mining and Drilling Taxonomy Wand (no date)
Industry classifications	*North American Industry Classification System (NAICS)* US Census Bureau (2007)

Approach

The NIOSH Mining Safety and Health Thesaurus (NMSHT) is being developed as a faceted thesaurus compliant with ANSI/NISO Z39.19-2005 (Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies).

A thesaurus approach was selected as it provides greater flexibility over other types of classification systems such as controlled vocabularies and taxonomies. In particular, associative relationships can be defined between terms, allowing for more structured semantic descriptions. While a thesaurus falls short of the power of an ontology, it provides a simpler starting point with potential for future expansion to an ontology.

According to Batty (1998), a well-developed thesaurus is organised into one or more hierarchical structures, where concepts sharing a common characteristic are organised into families called *facets*. Adkisson (2005) describes how a faceted classification system might be implemented on an e-commerce web site that sells jewellery, allowing customers to browse products on more than one dimension, such as jewellery type or material of construction. Many current web sites use a faceted navigation approach, including amazon.com, overstock.com, and the recipe web site epicurious.com.

A panel of internal content experts representing all major research programs was established to develop the initial thesaurus facets and terms. In most cases, these were branch chiefs of the respective focus areas. Other terms are being obtained by reviewing existing classification systems used within NIOSH, keywords assigned to individual publications by their authors, and from compilations of keywords which visitors have typed into the NIOSH Mining web site search engine.

NIOSH Mining Safety and Health Thesaurus facets

The NMSHT currently contains 20 facets:

1. biological;
2. dusts, gases and emissions;
3. hazardous events and exposures;
4. diseases, injuries, and disorders;
5. measurement and analysis;
6. engineering and administrative controls;
7. education and training;
8. equipment;
9. geographic regions;
10. industries;
11. products and commodities;
12. mining methods;
13. populations and target audiences;
14. physical phenomena;
15. science;
16. tasks and activities;
17. work locations and environments;
18. media types;
19. types of materials; and
20. NIOSH program areas.

A portion of the *engineering and administrative controls* facet is shown in Figure 1. Much of the NIOSH Mining web site content discusses engineering and administrative solutions to specific safety and health problems. This figure shows some of the thesaurus terms for categorising content related to dust control, electrical safety, emission controls, equipment design, and explosives safety (a minus sign in front of a term indicates the hierarchy for that term is fully expanded; a plus sign indicates that child terms are not shown).

- Dust control
  - Air curtain systems
  - Filtration systems
  - Particle charging
  - Removal of deposited dust
  - Sealing of dust sources
  - Source wetting
  - Water spraying
  - Wetting of deposited dust
- Electrical safety
  - Circuit protection
    - Ground fault protection
    - Ground monitoring
    - Overload protection
    - Short circuit protection
    - Surge protection
    - Undervoltage protection
  - Electrical contact protection
    - Grounding
    - Lockout tagout
  - Electrical insulation
  - Static electricity elimination
- Emission controls
- Equipment design
  - Ergonomic design
  - Escape and rescue
- Explosives safety

FIG 1 - A section of the NMSHT *engineering and administrative controls* facet.

Figure 2 shows part of the *dusts, gases and emissions* facet. Three main terms of that facet (*dusts*, *gases and emissions*, and *fumes*) are expanded to show the child terms beneath them. Combining the facets in Figures 1 and 2, a web page or publication about a particular dust control method could be indexed by both the type of dust, and engineering control for that dust, covered within the content. A user browsing either facet on the web site would locate the same item of content.

- Dusts
  - Asbestos dust
  - Coal dust
  - DPM
  - Float dust
  - Grain dust
  - Limestone dust
  - Respirable dust
  - Silica dust
  - Wood dust
- Gases and emissions
  - Acetylene
  - Aerosols
  - Carbon dioxide
  - Carbon monoxide
  - Compressed gases
  - Diesel exhaust
  - Ethyl mercaptane
  - Ethylene
  - Ethylene oxide
  - Fumes
    - Blasting fumes
    - Welding fumes
  - Hydrogen
  - Hydrogen sulfide
  - Methane
  - Nitrogen
  - Nitrogen dioxide

FIG 2 - A section of the NMSHT *dusts, gases and emissions* facet.

Figure 3 shows a detailed display for the thesaurus term *pillar design*, a term from the *engineering and administrative controls* facet. The figure illustrates the types of relationships tracked by the thesaurus:

- Equivalence relationship ('use for', abbreviated UF) – three terms are considered equivalent, or synonyms, of the preferred term *pillar design*. In this case, the terms are not true synonyms, but are treated as such to simplify the thesaurus. Content is indexed using the preferred term. All references to the synonyms would be redirected to the preferred term. In a book index, this can be visualised as an entry to 'see pillar design'. On a web site with free text search capabilities, a search on a non-preferred term could be automatically redirected to instead search the preferred term.
- Hierarchical relationship ('broader term', abbreviated BT) – the preferred term has one broader term, indicating that *ground control* is the parent term that appears one level up in the hierarchy.

Current Term	Pillar design
UF	Pillar dimensions
	Pillar orientation
	Pillar spacing
BT	Ground control
RT	Room and pillar mining

FIG 3 - NMSHT entry for the term *pillar design*.

- Associative relationship ('related term', abbreviated RT) – the preferred term has one related term, *room and pillar* mining (a term from the *mining methods* facet). This can be interpreted as 'see also Room and pillar mining', to direct users to a related topic of possible interest.

While ANSI/NISO Z39.19-2005 and OWL permit more complex types of relationships between terms, the thesaurus development software package being used to develop the NMSHT only permits the ones described above. At this stage, that is not a serious limitation. The next step of this project will validate the thesaurus by attempting to use it to index content on the NIOSH Mining web site. Further revisions will likely be needed and the thesaurus will be improved incrementally.

CONCLUSION

When the Web emerged in the 1990s, few imagined the extent to which it would grow in less than 20 years, nor the myriad of applications that would be built upon it. Today, it provides the means for communication and collaboration, e-commerce, entertainment, text messaging, video downloading, and numerous other uses. It is not outside the realm of possibility that the next 20 years may bring even more astounding capabilities.

Many of the world's most prominent computer scientists are determined to make that happen. Spearheaded by the inventor of the original Web, the W3C is laying the groundwork for the next generation of the Web. The vision is a Semantic Web, where semantic capabilities and intelligent software agents automate many of the tasks currently performed by humans.

One key component of the Semantic Web is the development of ontologies for representing knowledge within specific domains. This paper discussed a thesaurus of mining-related terms under development at NIOSH. The NMSHT focuses on US mining safety and health concerns, is monolingual (English) and does not yet have the capabilities of a true ontology. However, it may serve as a starting point for discussion regarding the requirements of a mining ontology that addresses the needs of the international mining community. These discussions will be important if the industry is to address its technological needs in 2025 and beyond.

ACKNOWLEDGEMENT

The Australasian Institute of Mining and Metallurgy copyright does not extend to this article with the United States of America.

REFERENCES

Adkisson, H, 2005. Use of faceted classification [online]. Available from: <http://www.webdesignpractices.com/navigation/facets.html> [Accessed: 28 May 2008].

Batty, D, 1998. WWW – Wealth, weariness or waste: Controlled vocabulary and thesauri in support of online information access [online]. *D-Lib Magazine*. Available from: <http://www.dlib.org/dlib/november98/11batty.html> [Accessed: 28 May 2008].

Berners-Lee, T, Hendler, J and Lassila, O, 2001. The Semantic Web, *Scientific Am*, May, pp 34-43.

Cearley, D W, Andrews, W and Gall, N, 2007. Finding and exploiting value in semantic technologies on the Web, *Garner Research Report*, ID Number G00148725, 9 May.

Chang, C, Beghtol, C, Mackenzie, S, Maurice, P, Peck, S, Rogmans, W and Taliano, J, 2003. *Thesaurus of Injury Prevention Terminology*, 186 p (SMARTRISK: Toronto).

Edinger, J and Barker, T, 2003. *GeMPeT: The Geoscience, Minerals and Petroleum Thesaurus*, 379 p (Charted Information Services: Maylands).

Feigenbaum, L, Herman, I, Hongsermeier, T, Neumann, E and Stephens, S, 2007. The Semantic Web in action, *Scientific Am*, December, pp 90-97.

Gravesteijn, J, Kortman, C, Potenza, R and Rassam, G N (eds), 1995. *Multilingual Thesaurus of Geosciences,* second edition, 645 p (Information Today: Medford).

Hendler, J, 2003. Science and the Semantic Web, *Science*, 299(5606): 520-521.

International Occupational Safety and Health Information Centre, 2001. *CIS Thesaurus* [CD-ROM] (International Occupational Safety and Health Information Centre, International Labour Office: Geneva).

Lam, H Y K, Marenco, L, Clark, T, Gao, Y, Kinoshita, J, Shepherd, G, Miller, P, Wu, E, Wong, G, Liu, N, Crasto, C, Morse, T, Stephens, S and Cheung, K, 2006. Semantic Web meets e-neuroscience: An RDF use case [online], in *Proceedings International Workshop on Semantic e-Science (ASWC 2006)*, pp 158-170. Available from: <http://www.oracle.com/technology/industries/life_sciences/press/semantic_web_meets_eneuroscience.pdf> [Accessed: 28 May 2008].

Lawrence, D W, Guard, A, Meier, A and Laflamme, L, 2006. Developing the injury prevention and safety promotion thesaurus, international English edition: An interdisciplinary tool for indexing and searching for research literature, Progress report 1, *Safety Science,* 44(2006): 279-296.

National Information Standards Organization (NISO), 2005. ANSI/NISO Z39.19-2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (National Information Standards Organization Press: Bethesda).

National Information Standards Organization (NISO), 2007. ANSI/NISO Z39.85-2007, *The Dublin Core Metadata Element Set* (National Information Standards Organization Press: Bethesda).

National Occupational Health and Safety Commission, 2003. Australian occupational health and safety thesaurus: Guidelines, hierarchical structure of preferred terms, alphabetical list of preferred terms and non-preferred terms, third edition [online] Available from: <http://www.ascc.gov.au/NR/rdonlyres/5C2FCFD3-E03E-46F2-99A0-1738EFCE1C63/0/AOHST_e3.pdf> [Accessed: 28 May 2008].

Shadbolt, N, Hall, W and Berners-Lee, T, 2006. The Semantic Web revisited, *Institute of Electrical and Electronics Engineers (IEEE) Intelligent Systems*, 21(3):96-101.

Spivack, N, 2006. Minding the planet – The meaning and future of the Semantic Web [online]. Available from: <http://novaspivack.typepad.com/nova_spivacks_weblog/2006/11/minding_the_pla.html> [Accessed: 28 May 2008].

US Census Bureau, 2007. North American industry classification system (NAICS) [online]. Available from: <http://www.census.gov/epcd/www/naics.html> [Accessed: 28 May 2008].

US Department of Energy, 1992. *Dictionary and Thesaurus of Environment, Health and Safety*, 528 p (CRC Press, Inc: Boca Raton).

US Department of Labor, 2007. Occupational injury and illness classification manual [online]. Available from: <http://www.bls.gov/iif/oshoiics.htm> [Accessed: 28 May 2008].

Wand (no date). WAND minerals, mining and drilling taxonomy [online]. Available from: <http://www.wandtaxonomy.com/core/categoryMenu.aspx> [Accessed: 28 May 2008].

World Health Organization, 2007. International statistical classification of diseases and related health problems, tenth revision [online]. Available from: <http://www.who.int/classifications/apps/icd/icd10online/> [Accessed: 28 May 2008].

World Wide Web Consortium (W3C), 1999. Resource description framework (RDF) model and syntax specification, W3C recommendation 22 February 1999 [online]. Available from: <http://www.w3.org/RDF/> [Accessed: 28 May 2008].

World Wide Web Consortium (W3C), 2004a. OWL web ontology language use cases and requirements, W3C recommendation 10 February 2004 [online]. Available from: <http://www.w3.org/TR/webont-req/> [Accessed: 28 May 2008].

World Wide Web Consortium (W3C), 2004b. RDF vocabulary description language 1.0: RDF Schema, W3C recommendation 10 February 2004 [online]. Available from: <http://www.w3.org/TR/rdf-schema/> [Accessed: 28 May 2008].

World Wide Web Consortium (W3C), 2004c. OWL web ontology language overview, W3C recommendation 10 February 2004 [online]. Available from: <http://www.w3.org/TR/owl-features/> [Accessed: 28 May 2008].

World Wide Web Consortium (W3C), 2008. SPARQL query language for RDF, W3C recommendation 15 January 2008 [online]. Available from: <http://www.w3.org/TR/rdf-sparql-query/> [Accessed: 28 May 2008].