Thesauri, taxonomies and ontologies – an etymological note

Alan Gilchrist,
CURA Consortium and TFPL Ltd, London, UK

Received 30 June 2002
Revised 24 September 2002
Accepted 10 October 2002

Keywords *Internet, Data, Management*

Abstract *The amount of work to be done in rendering the digital information space more efficient and effective has attracted a wide range of disciplines which, in turn, has given rise to a degree of confusion in the terminology applied to information problems. This note seeks to shed some light on the three terms thesauri, taxonomies and ontologies as they are currently being used by, among others, information scientists, AI practitioners, and those working on the foundations of the semantic Web. The paper is not a review of the techniques themselves.*

Introduction

Some definitions from the *Oxford English Dictionary (OED)*:

Thesaurus
A "treasury" or "storehouse" of knowledge, as a dictionary, encyclopaedia or the like.
A collection of concepts or words arranged according to sense; also a dictionary of synonyms and antonyms.

Taxonomy
Classification, esp. in relation to its general laws or principles; that department of science, or of a particular science or subject, which consists in or relates to classification; especially the systemic classification of living organisms.

Ontology
The science or study of being; that department of metaphysics which relates to the being or essence of things, or to being in the abstract.

These three definitions will be familiar to many, but all three words have now been appropriated and subverted by contemporary workers in the information sciences. In the process, there tends to be, at times, significant overlap between, and even contradiction in, the three words as they are currently used. It would not be sensible to pontificate on the "correct" meanings of these words, but in trying to delineate the central characteristics of these three terminological constructs, it is hoped that the reader will gain a clearer understanding of their differences and similarities, as well as how the three might begin to be used more closely together.

Consultation of three dictionaries devoted to the topics of concern to library and information science (LIS) (Feather and Sturges,1997; Keenan and Johnston, 2000; Prytherch, 2000) showed that new uses of the word taxonomy had not yet been covered, and the word ontology neither in its *OED* sense quoted above, nor in its later sense. The word thesaurus was defined in both its earlier and

Journal of Documentation
Vol. 59 No. 1, 2003
pp. 7-18

0022-0418
DOI 10.1108/00220410310457984

LIS sense by *Harrods' Librarians Glossary* (Prytherch, 2000), and in its later sense only by the other two.

A brief scan through Library and Information Science Abstracts (LISANET) online at the British Library recorded 2,313 hits for thesaur, 285 for taxonom, and 163 for ontolog. The earliest reference to a new use of the word taxonomy appeared to be in 1997 in an article describing a technique employed by a search engine vendor, while the new use of the word ontology seemed not to be mentioned at all.

Similarly, little mention can be found in the last four volumes of *ARIST* (Williams, 1997, 1998, 2000, 2001). Thesauri are indexed nine times in relation to such aspects as automatic construction and machine learning; while taxonomies are indexed twice (both times in a new sense); and ontologies once in relation to the famous WordNet database, which will be mentioned later.

Wider searching shows, however, that the words taxonomy and ontology are being increasingly discussed in the published and grey literature, and are often likely to co-occur with such words as information retrieval, knowledge management, indexing, search and navigation, intranets and portals.

Thesauri

To the lay person, the word thesaurus will conjure up the name Roget (1852), whose *Thesaurus of English Words and Phrases* was first produced in 1852. The structure of this thesaurus was, according to Roget in his introduction, a "verbal classification ... the same as that which is employed in the various departments of natural history". Jones (1972) and McArthur (1986) have both discussed the work of Roget in the context of the previous thinking of such philosophers as Bacon and Descartes, and both trace a direct link between Roget and Wilkins who, sponsored by the Royal Society, set out to construct a new language to replace Latin, specifically for the advancement of seventeenth century science and which would remove ambiguities in universal communication. Wilkins' (1668) *Essay towards a Real Character and a Philosophical Language* was published in 1668, and included a decimal classification (referred to by McArthur as a taxonomy), ranging from God to "public relationships (civil, judiciary, naval, military, ecclesiastical)". Wilkins has his place, then, in a long line of epistemologists from Aristotle to Dewey and Bliss.

According to the *OED*, the third use of the term thesaurus, following the two quoted at the start of this paper, is attributed to Helen Brownson of the American National Science Foundation. In fact, Vickery quoted Brownson, at the 1957 Dorking Conference on classification research, as having conceived of the information problem being addressed by "the application of a mechanized thesaurus based on networks of related meanings" (Vickery, 1960). The *OED* definition has this meaning of thesaurus as "A classified list of terms, especially keywords, in a particular field, for use in indexing and information retrieval". It is interesting to note the use of the word "classified" for, despite the examples of Wilkins, Roget and Dewey, the early thesauri did not use

classification techniques, either in their compilation or for their display, and it was not until the appearance of the *Thesaurofacet* (Aitchison *et al.*, 1969) that the advantage for indexing and retrieval was clearly demonstrated in the combination of a faceted classification and thesaurus in the same volume, having a one-to-one correspondence between facets and descriptors.

This might now be seen as the heyday of the classical information retrieval thesaurus: a controlled vocabulary, used for indexing and retrieval; initially employing punched cards (edge-notched, peek-a-boo and Hollerith) but later supported by computers. Initially used by computers working in batch mode, the thesaurus survived into the era of online interrogation. The relative stability of this technology supported the creation of an international standard (International Organisation for Standardisation, 1986) and a manual (Aitchison *et al.*, 2000) that ran to four editions; and though the technology is still used today in many parts of the world in its original, but fairly varied forms, it has been largely overtaken by the relatively sudden and rapid spread of full text searching, and more recently by the prevalence of the Internet. Whereas information retrieval had hitherto relied on the indexer and the searcher following mirrored paths, both guided by a network of hierarchical, associative and equivalence relationships between terms, a new type of thesaurus began to appear, called loosely a "search thesaurus". Cochrane (1992) compared the two, pointing out that the search thesaurus could be composed of the merging of thesaurus lists; a construct produced, for example, by Knapp (1984) in BRS/TERM, and which had more than a passing resemblance to Roget in that its main characteristic was the clustering of synonyms.

Long before the advent of the search thesaurus, the problem of reconciliation between, or integration of, two or more thesauri had attracted the attention of researchers and practitioners. Possible solutions include mapping (e.g. Chaplan, 1995), the use of intermediate languages (e.g. BSC Panel, 1993), automatic switching (e.g. Busch and Petersen, 1995), or merging (e.g. Dextre Clarke, 1996; Hoppe, 1996). This problem has become more acute with the spread of electronic networks, making numbers of repositories of information available to searchers, but not necessarily making the content accessible. Consequently, a number of research projects are under way to ameliorate cross database searching (see e.g. http://hilt.cdlr.strath.ac.uk), while some practical success is being achieved, such as in the compilation of the Government Category List (GCL), a component part of the UK Government's "Metadata initiative" (www.govtalk.gov.uk). The GCL is a high level classification, containing some 400 terms and covering all aspects of government, onto which all public sector bodies are mandated to map their classifications and thesauri. This business of "mapping", to use the popular generic term, has been given some emphasis here because it is a problem that confronts those working with taxonomies and ontologies where, as will be seen later, some new techniques are becoming available.

In parallel with advances in thesaurus compilation and manipulation, work was being undertaken by workers in artificial intelligence in their construction

of expert systems. In this work they compiled what are called "semantic networks", described by Milstead (1995) as being "conceptually quite similar to a thesaurus, in that they show terms in the context of their semantic relationships. However, they offer different navigational capabilities, through graphic devices that represent multidimensional spaces, rather than through review of the cross references and scope notes of a thesaurus". In addition to drawing attention to the increasing importance of navigation and visualisation, this brief, but prescient paper, also opined that "the thesaurus may become almost invisible to most users". These are, in fact, some of the factors that gave rise to the interest in taxonomies, discussed in the next section, an interest particularly associated with "Enterprise information portals" being installed in large enterprises. Returning, though, to semantic networks, another major difference between them and thesauri was in the fact that the former went much further in defining the types of relationships between terms. For example, Ford (1991) in a textbook specially written for librarians and information scientists, presents a section of a semantic map devoted to cardiovascular illness, in which "heart disease" is shown as being "is-a" "cardiovascular illness" (is-a being a standard convention), which "has setting" "cardiovascular system"; while "heart disease" "has setting" "heart", which is "component of" "cardiovascular system". Two things arise from this more complex treatment. The first is that there is no intrinsic reason why a conventional thesaurus should not be extended and elaborated to include, for example, term definitions, notes on term usage, and more explicitly defined relationships. The second is that such enrichment allows the semantic network to be more easily manipulated by an inference engine, typically employing the IF . . . THEN operator. This enrichment of thesauri (see, e.g. Hazewinkel, 1997) is a feature of work in the area of ontologies, as will be shown in a later section.

Taxonomies

The fact that others might introduce old words with new meanings into previously relatively stable domains of discourse should make us think. What, then, have been the triggers that have given rise to an interest in "taxonomies"? A number of factors have been advanced, mainly at conferences and in the White Papers of software vendors, as the published literature on the new meanings of taxonomies is relatively small. These triggers include:

- *Information overload.* Conventional search engines are often now seen to be inadequate in dealing effectively with very large databases, and it is apparent that users need complementary search aids and filters.
- *Information literacy.* Research has shown that the majority of end-users have severe problems in knowing how to search for information, leading to wasted time and the missing of useful information.
- *Organisational terminology.* Published classifications and thesauri do not reflect the particular languages of organizations, in which, typically, 80 per cent of the information held has been created internally.

- *"Destructuring" of organisations.* Mergers and acquisitions have created cultural problems at the implementation stage. Similar problems are encountered in partnering through extranets, and in the establishment and operation of virtual communities, as different user groups begin to share information and knowledge within organisational contexts.

However, when one looks at the answers to these problems sheltering under this term taxonomy, it becomes apparent that it is a generic term covering a number of techniques and applications. A number of case studies were presented in a report of research conducted in 2000 (Gilchrist and Kibby, 2000), which showed a wide range of approaches being implemented in large organizations such as Glaxo Wellcome, and PricewaterhouseCoopers (as they then were), the US Postal Services and the British Broadcasting Corporation. From this, and subsequent, research it became apparent that the word taxonomy was being used with at least five separate meanings, though there is some overlap between them. These are:

(1) *Web directories.* These are commonly used on the Internet, and increasingly in intranets. They are, in fact, a form of classification. A menu of top terms is offered to the user. Clicking on a selected term will display a second level, and so on, for several more levels, arriving finally at perhaps some information or references, or the possibility of offering the last selected term to one or more search engines. Each level does not have to be hierarchical in the normally accepted sense, and terms may be repeated at different levels, providing alternative pathways for the searcher. A well-known example is the Open Directory Project (www.dmoz.com), which is reported to have over 350,000 terms in its taxonomy, organized by over 35,000 volunteer editors.

(2) *Taxonomies to support automatic indexing.* In a typical example on a commercial Web site, the user is presented with a two level classification, containing in all some 1,200 terms. What the user does not see is the taxonomy behind the classification which, for each term in the displayed classification, carries an algorithm comprising sets of words and phrases, synonyms and syntactic variations, weights and instructions. This rules base is used to automatically extract appropriate index terms, which may, or may not, be present in the documents. This approach is particularly attractive where throughputs are so large that manual indexing is economically impossible (see e.g. www.factiva.com; www.lexisnexis.com). The classification and taxonomy are manually constructed, and the taxonomy might be regarded as an enriched thesaurus.

(3) *Taxonomies created by automatic categorization.* A growing number of software packages are now available which purport to be able to analyse text, to automatically create categories from that analysis and to classify the analysed documents according to the categories created (see e.g.

www.gammasite.com). These categories may then be displayed in a similar format to Web directories (see above) and/or as two-dimensional maps, where related terms are linked to the selected term that appears in the middle of the map. Selection of a related term will then move that term to the centre and introduce a new set of related terms. As in the case of the Web directory, the result of the automatic categorization process is a classification and, often using software similar to that used in probabilistic retrieval, is based on the statistical analysis of the occurrence and co-occurrence of terms in the database, albeit augmented by human intervention.

(4) *Front end filters.* Here, a taxonomy is either created or imported and used in query formulation. Homographs can be disambiguated and synonyms clustered; while the searcher can also browse up and down hierarchies and jump to other related terms. The selected search term is then submitted to the Internet or intranet (see e.g. www.wordmap.co.uk). This taxonomy is a thesaurus formatted to enable easy navigation.

(5) *Corporate taxonomies.* The preceding examples of different taxonomies are used either at the processing stage of information input, or for displaying search terms and results. However, the modern enterprise faces the whole range of problems listed above and should consider all the new approaches being promoted by the suppliers. The prime need to make information easily accessible to its staff through the enterprise information portal or other channel is often compounded by large throughputs and legacy data. There is also a recognized need to provide maps and user guidelines to the repositories and their contents, which may be numerous and include both internal and external sources. This may involve the deceptively simple-sounding word "mapping" that, in fact, indicates the need for a huge amount of intellectual human effort. One example of this can be found at Glaxo Wellcome (now GlaxoSmithKline) wherein a number of existing thesauri have been loosely merged into a "megathesaurus", in which each term carries the addresses of any corporate information repositories using that term. Worldwide legal and packaging regulations are also added to terms where appropriate. This was a huge effort resulting in some 53,500 basic concepts (or lead terms), 201,750 synonyms (trade names and chemical equivalents used around the world), and 443,500 related terms, many of whose relationship types are defined. GlaxoSmithKline has now teamed up with a US software house specializing in ontologies, and claim that its program (for which a patent has been filed) is capable of assimilating a new thesaurus into the megathesaurus with a high degree of accuracy. While this could be regarded as a hybrid thesaurus/ontology, a new type of taxonomy is emerging that might be called a "corporate taxonomy". This is usually hand-crafted and results in a high level map on an enterprise portal, guiding staff to the intellectual capital of the

organization. It can encompass information about the organization itself: its business processes; methods, guidelines and standards; about the people within ("yellow pages") and associated with the organization ("blue book"); and about the content held in the various repositories. It can also be designed to support search engines and provide indexer templates; and can be tailored to provide global, national and even site perspectives. Such a taxonomy may be viewed as a master scheme from which a range of tools, including thesauri, can be derived.

Ontologies

Vickery (1997) was one of the first in the LIS field to draw attention to the emergence of the term ontology in knowledge engineering and in information science. In his 1997 paper he reviewed some of the more important ontologies of that time as well as reporting on the thinking of the leaders in the field. One of those leaders is Gruber (n.d.), who says: "An ontology can be defined as a formal, explicit specification of a shared conceptualisation".

> "Conceptualization" refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. "Explicit" means that the type of concepts used, and the constraints on their use are explicitly defined. "Formal" refers to the fact that the ontology should be machine-readable. "Shared" reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group. An ontology describes the subject matter using the notions of concepts, instances, relations, functions and axioms. Concepts in the ontology are organized in taxonomies through which inheritance mechanisms can be applied (Benjamins *et al.*, n.d.).

Two of the oldest and most widely-known ontologies are WordNet and CYC. WordNet (Fellbaum, 1998) was developed by the Cognitive Science Laboratory at Stanford University, and contains some 100,000 word meanings grouped into five categories: nouns, verbs, adjectives, adverbs and function words, and the meanings are related by synonymy, antonymy, hyponymy (the is-a relation), meronymy (part-of relationship between concepts) and morphological relations between word forms. WordNet is a lexical tool, available free on the Internet (www.wordnet.com), and has been used by a number of commercial organizations, including vendors of taxonomy software. The CYC ontology (Lenat and Guha, 1990; www.cyc.com) (produced by the company that is working with GlaxoSmithKline) provides a foundation for common sense reasoning by developing ontologies for a wide variety of domain-specific applications. Ontologies are manipulated by their own "languages" and, in the case of CYC, all of its knowledge is represented declaratively in a variant of predicate logic (a branch of logic dealing with propositions in which subject and predicate are separately signified), thus giving CYC more reasoning power than is available to WordNet.

A large number of domain-specific ontologies have been built for applications such as machine translation, enterprise modelling, knowledge re-use and information retrieval. Overviews of some of these projects, and the

tools used by them, can be found in Ding (2001), Ding and Foo (2002) and Fensel (2001).

Two areas where the use of ontologies are being most ambitiously promoted are in knowledge management and in the idea of the Semantic Web. Benjamins *et al.* (n.d.) discuss the application of ontology research to knowledge management, and organizational memories. Quoting other authors (Abecker *et al.*, 1998) they list three kinds of ontologies that they see as being useful for organizational memory systems:

- An organizational ontology describes the information meta-model, e.g. the structure and format of the information sources. This is the lowest level ontology.
- A domain ontology is used to describe the content of the information source (the area in which Benjamins *et al.* are working).
- An enterprise ontology is used for modelling business processes. Its purpose is to model the knowledge needs in a business process in order to describe a process context, which enables active knowledge delivering.

Other authors (Milton *et al.*, n.d.) are exploring the application of ontologies to the four transformation processes in information and knowledge transfer: personalization (knowledge sharing), creation/innovation, codification/discovery (information and knowledge retrieval) and capture/monitor (knowledge elicitation).

The Semantic Web has been proposed by Berners-Lee *et al.* (2001). Berners-Lee *et al.* describe the Semantic Web as being: "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 co-operation", in line with his original conception of the Web as a universal, decentralized system; and one which does not discriminate between "the scribbled draft and the polished performance, between commercial and academic information, or among cultures, languages, media and so on". The Semantic Web will be supported by the eXtensible markup language (XML) and by the resource description framework (RDF), which can be viewed for the moment as packages describing the information and data contained therein. However, as Berners-Lee points out, "two databases may use different identifiers for what is in fact the same concept" and the solution to this problem is to provide ontologies, which he defines as "a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules".

Conclusion

Wittgenstein said something to the effect that if you wanted to know the meaning of a word, you should look to see how it is used. Looking at the applications of thesauri, taxonomies and ontologies it is easy to see a progression of ideas that has resulted in some overlapping of detail. Clearly, this potential for confusion has been exacerbated by the presence of different players. In broad terms, one may say that the post-Roget thesaurus has been

the domain of information scientists; taxonomies appear to have been generated by a combination of information technologists and systems developers in corporate business together with software vendors; and ontologies have been adapted from the work of philosophers by people working in artificial intelligence (and there are cases of these two disciplines working together on ontologies). The progression alluded to above has been driven by a number of factors including:

- a growing trend in organizations to collate external and internal information (which has led to a convergence of special librarianship and records management);
- the huge amount of information now available (e.g. Microsoft has 3 million documents on its intranet, before accessing external material on the Internet);
- the availability of relatively inexpensive computing power.

Thus, where the thesaurus was mainly concerned with the processing of external information, some types of taxonomy are applied to dealing with large inputs of information (either external, internal, or both together in certain cases), while the "corporate" taxonomy may be applied to a graphic description of the organization itself, its personnel, and its content. It may support the research and development function, its customer relations management processes, its e-commerce capability, or the complete range of business processes. The quantity of information has rendered taxonomies necessary, and today's computing power has made their application possible, particularly with regard to the automatic analysis and processing of content. This computing power is even more important in the world of ontologies, where domain sizes and interconnections between them are likely to increase, thus requiring even more computing power to carry out the necessary reasoning. Another aspect of the progression can be seen in the widening areas of application, and the reader should be aware that the mentions of applications of ontologies in this paper represent only a small fraction of the activity in this area, both in research and in practice.

What is clearly common to thesauri, taxonomies and ontologies is that they all deal with natural language (at least in the context of this paper). If you have a number of objects, it is possible to arrange them into groups and apply labels to the groups – librarians have traditionally done this in classifying books for arrangement on shelves. If you are more deeply concerned with the information content of bibliographic material, you can devise a system of labels for the description of that content – and information scientists have devised thesauri for that purpose. When one realises that concepts can be treated as objects, then it becomes possible to group concepts and the thesaurofacet is born. Taxonomies use both classification and thesaurus techniques, and it is interesting to note how similar are some of the techniques used in automatic indexing and in automatic categorization, this being largely a matter of

granularity. Taxonomies may also use a combination of classification and thesaural techniques applied to a wider range of object types (and museums documentation and image retrieval may be mentioned here as areas where the object types pose particular problems, and where other techniques are being developed). Classification is made possible by analysis of the relations between concepts, and their associations with objects; and semantic networks, for example, have gone further than thesauri in defining relationships. But relationships may depend on context, and so it becomes necessary to add more information about the concept in the form of definitions and notes on usage. At the same time, it becomes feasible to deal with a wider range of word forms so that, contrary to the thesaural concentration on the noun form and the unit concept, other terminology systems can incorporate adjectives, verbs, morphological variations and syntactic strings beyond the phrase. These "enriched thesauri", which would be inappropriate at the user interface but may be used by the software in processing queries, are also a fundamental component of ontologies. The other component of the ontology is a manipulation layer derived from one of the many logics used by the AI community.

It may, therefore, be argued that it is wrong, as some librarians and information scientists do, to say that taxonomies are "just" classifications or "merely" thesauri; and wrong to say as Soergel (1999) does that " ... eventually ontology was used to designate any classification, particularly in the communities of linguistics, AI, and software engineering". Where Soergel is absolutely right is in pointing out that "The use of a different term is symptomatic of the lack of communication between scientific communities" and that "The vast body of knowledge on classification structure and on ways to display classifications developed around library classification and in information science more generally, and the huge intellectual capital embodied in many classification schemes and thesauri is largely ignored". Although things have improved to some extent since Soergel wrote these words in 1999, it still seems obvious that multidisciplinary teams will be needed if such dreams as the Semantic Web are to become reality.

References

Abecker, A., Bernardi, A., Hinckelmann, K., Kühn, O. and Sinteck, M. (1998), "Towards a technology for organizational memories", *IEEE Intelligent Systems*. Vol. 13 No. 3, pp. 40-8.

Aitchison, J., Gilchrist, A. and Bawden, D. (2000), *Thesaurus Construction and Use: A Practical Manual*, 4th ed., Aslib, London.

Aitchison, J., Gomersall, A. and Ireland, R. (1969), *Thesaurofacet: A Thesaurus and Faceted Classification for Engineering and Related Subjects*, English Electric Company, Whetstone, Leicester.

Benjamins, V.R. *et al.* (n.d.), "(KA)2: building ontologies for the Internet: a mid term report", available at: www.cs.vu.nl/~dieter/ftp/paper/km99.pdf

Berners-Lee, T., Hendler, J. and Lassila, O. (2001), *The Semantic Web*, available at: www.sciam.com/2001/0501issue/0501berners-lee.html

BSO Panel (1993), *Broad System of Ordering: Schedule and Index*. 4th rev. (on diskette), BSO Panel, St Albans.

Busch, J.A. and Petersen, T. (1995), "Automated mapping of topical special subject headings into faceted index strings using the art and architecture thesaurus as a machine-readable dictionary", *Advances in Knowledge Organization, Knowledge Organization and Quality Management, Proceedings of the 3rd International ISKO Conference, Vol. 4, Copenhagen, 20-24 June 1994*, Indeks Verlag, Frankfurt/Main, pp. 390-7.

Chaplan, M.A. (1995), "Mapping laborline thesaurus to library of congress subject headings; implications for vocabulary switching", *Library Quarterly*, Vol. 65, No. 1, January, pp. 39-61.

Cochrane, P.A. (1992), "Indexing and searching thesauri, the Janus or Proteus of information retrieval", in Williamson, N.J. and Hudon, M. (Eds), *Classification Research for Knowledge Representation and Organization*, Elsevier Science Publishers, Amsterdam.

Dextre Clarke, S. (1996), "Integrating thesauri in the agricultural sciences", *Compatibility and Integration of Order Systems. Research Seminar Proceedings of the TIP/ISKO Meeting, 13-15 September, Warsaw*, SBP, Warsaw, pp. 111-22.

Ding, Y. (2001), "A review of ontologies with the semantic Web in view", *Journal of Information Science*, Vol. 27 No. 6, 2001, pp. 377-84.

Ding, Y. and Foo, S. (2002), "Ontology research and development. Part 1 – a review of ontology generation" *Journal of Information Science*, Vol. 28 No. 2, pp. 123-36.

Feather, J. and Sturges, P. (1997), *International Encyclopaedia of Information and Library Science*, Routledge, London.

Fellbaum, C. (Ed.) (1998), *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA.

Fensel, D. (2001), *Ontologies: A Silver Bullet for Knowledge Management and Electric Commerce*, Springer, Berlin.

Ford, N. (1991), *Expert Systems and Artificial Intelligence*, Library Association Publishing, London.

Gilchrist, A. and Kibby, P. (2000), *Taxonomies for Business: Access and Connectivity in a Wired World*, TFPL, London.

Gruber, T. (n.d.), "What is an ontology?", available at: www-ksl.Stanford.edu/kst/what-is-an-ontology.html

Hazewinkel, M. (1996), "Enriched thesauri", in Thanos, C. (Ed.), *Proceedings of the First DELOS Workshop, March 1996*, ERCIM, pp. 27-32.

Hoppe, S. (1996), "The UMLS – a model for knowledge integration in a subject field", *Compatibility and Integration of Order Systems. Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September*, SBP, Warsaw, pp. 97-110.

International Organization for Standardization (1986), *Guidelines for the Establishment and Development of Monolingual*, ISO 2788, 2nd ed., ISO, Geneva.

Keenan, S. and Johnston, C. (2000), *Concise Dictionary of Library and Information Science*, 2nd ed., Bowker Saur, London.

Knapp, S.D. (1984), "BRS/TERM, a vocabulary database for searchers", *Database*, Vol 7 No. 4, December, pp. 70-5.

Lenat, D. and Guha, R.V. (1999), *Building Large Knowledge-based Systems*, Addison-Wesley, Reading, MA.

McArthur, T. (1986), *Worlds of Reference*, Cambridge University Press, Cambridge.

Milstead, J. (2000), "Invisible thesauri: the year 2000", *Online & CDROM Review*, Vol. 19 No. 2, pp. 93-4.

Milton, N., Shabolt, N.R., Cottam, H. and Hamersley, M. (n.d.), "Towards a knowledge technology for knowledge management", available at: www.psychology.nottingham.ac.uk/research/ktc/publics/ijhcs_web3.htm

Prytherch, R. (2000), *Harrod's Librarians Glossary*, 9th ed., Gower, London.

Roget, P.M. (1852), *Thesaurus of English Words and Phrases*, Longman, London and Harlow.

Soergel, D. (1999), "The rise of ontologies or the reinvention of classification", *Journal of the American Society for Information Science*, Vol. 50 No. 12, pp. 1119-20.

Sparck Jones, K. (1972), "Some thesauric history", *Aslib Proceedings*, Vol. 24 No. 7, July, pp. 400-11.

Vickery, B.C. (1960), "Thesaurus – a new word in documentation", *Journal of Documentation*, Vol. 16, December, pp. 181-9.

Vickery, B.C. (1997), "Ontologies" *Journal of Information Science*, Vol. 23 No. 4, pp. 277-86.

Wilkins, J. (1668), *An Essay towards a Real Character and a Philosophical Language*, London.

Williams, M.E. (Ed.) (1997), *Annual Review of Information Science and Technology*, Vol. 32, Information Today, Medford, NJ.

Williams, M.E. (Ed.) (1998), *Annual Review of Information Science and Technology*, Vol. 33, Information Today, Medford, NJ.

Williams, M.E. (Ed.) (2000), *Annual Review of Information Science and Technology*, Vol. 34, Information Today, Medford, NJ.

Williams, M.E. (Ed.) (2001), *Annual Review of Information Science and Technology*, Vol. 35, Information Today, Medford, NJ.