CHAPTER 1

Thesauri: Introduction and Recent Developments

This chapter introduces information retrieval thesauri and highlights some recent trends in the use of thesauri as search aids, in particular search and end-user thesauri. Addressed here are the differences among thesauri, taxonomies, and ontologies, along with the role that thesauri have played in the development of taxonomies and ontologies. This chapter also covers recent research trends that focus on the provision of semantic support for user interfaces provided by major search engines, areas such as faceted search, exploratory user interfaces, and dynamic term suggestion functionalities. The notion of social tagging is introduced, and a number studies that have compared controlled vocabularies and social tags are reviewed.

1.1 Thesaurus: A Brief History

The term *thesaurus* as a reference tool dates to the publication in 1982 of *Roget's Thesaurus*, and this, or some modern equivalent, is what most people have in mind when they think of a thesaurus (Broughton, 2006). Developed by Peter Mark Roget, *Roget's Thesaurus* is still the most widely used English language thesaurus, organizing words and their meanings in a systematic manner to assist people in identifying semantically related terms.

1.1.1 Information Retrieval Thesauri

The history of information retrieval thesauri can be traced back to the 1950s. Detailed accounts of the history of information retrieval thesauri can be found in Vickery (1960), Gilchrist (1971), and Aitchison and Dextre Clarke (2004). There is agreement that in the context of information retrieval, the word *thesaurus* was first used in 1957 by

Peter Luhn of IBM. The first thesaurus used for controlling the vocabulary of an information retrieval system was developed by the DuPont organization in 1959, and the first widely available thesauri were the *Thesaurus of Armed Services Technical Information Agency (ASTIA) Descriptors*, published by the Department of Defense in 1960, and the *Chemical Engineering Thesaurus*, published by the American Institute of Chemical Engineers (Aitchison and Dextre Clarke, 2004).

In the 1970s and early 1980s, commercial online database providers such as Dialog made use of thesauri alongside their bibliographic databases to enhance the quality of search. Chamis (1991) reported that in the 1980s about 30 percent of Dialog databases had either a printed or an online thesaurus. Many online databases now use thesauri for vocabulary control.

The introduction in 1974 of the first international standard for the construction of monolingual thesauri gave rise to the popularity of thesauri in various scientific and technological subjects. Several thesaurus construction standards have been developed during the past three decades: international standards (ISO 2788: 1986; ISO 5964: 1985); British standards (BS 5723: 1987; BS 6723: 1985); and UNISIST standards (UNISIST Guidelines, 1980, 1981). The U.S. standard on monolingual thesaurus construction, American National Standards Institute–National Information Standards Organization (ANSI/NISO) Z39.19, was published in 1993.

The advent of the web and the rapid growth of web-based information retrieval systems and services such as digital libraries, open archives, content management systems, and portals prompted international, U.K., and U.S. standards organizations to make revisions and changes to accommodate the demands of the electronic environment. The international standard ISO 25964-1 (2011), *Thesauri and Interoperability With Other Vocabularies*, revises, merges, and extends both ISO 2788 and ISO 5964 standards for the development of monolingual and multilingual thesauri. Guidelines for BS 5723 were replaced by BS 8723, *Structured Vocabularies for Information Retrieval*. BS 8723 was superseded by ISO 25964-1 in 2011. Details of the standard can be found at the British Standards Institution's website (www.bsigroup.com).

The new U.S. standard ANSI/NISO Z39.19, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, was published in 2005 and revised in 2010. Its new designation is ANSI/NISO Z39.19-2005 (R2010).

Major emphases in these changes and revisions were interoperability, electronic and web-based applications, thesaurus displays, and coverage of a wide range of vocabularies used in information retrieval systems and web-based services. In the field of information architecture, there is a firm belief in the advantages of staying close to the accepted standard. According to Morville and Rosenfeld (2007), these advantages are based on the following assumptions:

- "There's good thinking and intelligence baked into these guidelines.
- Most thesaurus management software is designed to be compliant with ANSI/NISO, so sticking with the standard can be useful from a technology-integration perspective.
- Compliance with the standard provides a better chance of cross-database compatibility so that when two companies merge, for example, it will be easier to merge their vocabulary sets." (p. 214)

1.1.2 What Is an Information Retrieval Thesaurus?

A thesaurus is a tool designed to support effective information retrieval by guiding indexers and searchers to consistently choose the same terms for expressing a given concept or combination of concepts (Dextre Clarke, 2001). Aitchison et al. (2000) define a thesaurus as "a vocabulary of controlled indexing language, formally organized so that *a priori* relationships between concepts are made explicit" (p. 1) that can be used in information retrieval systems ranging from the card catalog to the internet. The ANSI/NISO Z39.19 (2005) standard provides the following definition of a thesaurus: "A controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators." Some of the long-established and well-known thesauri are the Medical Subject Headings, also known as the MeSH Thesaurus, in the area of medicine and allied sciences, the Art and Architecture Thesaurus (AAT), and the Thesaurus of ERIC (Education Resources Information Center) Descriptors.

Standard thesauri incorporate three types of term relationships, namely, equivalence, hierarchical, and associative. Equivalence relationships are usually defined as relations between synonyms and quasi synonyms, for instance, between *computer languages* and *programming languages*. This type of relationship provides an alternative

access point for the user during searching. Equivalence relationships are shown by the notation UF (Used For).

Hierarchical relationships are assigned to terms that have various levels of specificity. For instance, the term *libraries* is a narrower term for *digital libraries*, while the term *user interfaces* is a broader term for *visual user interfaces*. These broader and narrower relationship types allow a user to semantically navigate in an information collection from terms that are general to more specific terms and vice versa. The boarder and narrower term relationships are shown by the notations BT (Broader Term) and NT (Narrower Term).

Associative relationships are designed to create relationships between terms that do not have equivalence or hierarchical relationships but would be conceptually or mentally related, for example, between *information overload* and *information filtering*. This type of relationship is represented by the notation RT (Related Term).

The following entry from the *ASIS&T Thesaurus of Information Science, Technology, and Librarianship* illustrates the various types of term relationships:

Internet
UF Cyberspace
  Information highway
  Information superhighway
BT Telecommunication networks
RT e-mail list servers
  ftp
  gophers
  Internet search systems
  National Research and Education Network
  Network computers
  Newsgroups
  telnet
  Web TV

Another characteristic of standard thesauri is their inclusion of scope notes. A scope note is a definition of the term or an explanation of its meaning and use in a specific database. The notation SN represents scope notes in thesauri.

1.1.3 Thesaurus Displays

There are several different methods of displaying thesauri on paper and on the computer screen:

- Alphabetical displays showing scope notes and equivalence, hierarchical, and associative relationships for each term
- Hierarchical displays generated from the alphabetical display
- Systematic and hierarchical displays showing the overall structure of the thesaurus and all levels of hierarchy
- Graphic displays of varying sorts (Aitchison et al., 2000) using arrows, family trees, or two- and three-dimensional visualization techniques (an extended discussion of user interfaces for thesauri appears in Chapter 5)

Guidelines for the design and construction of thesauri are beyond the scope of this book. Readers interested in this area should consult the practical manuals developed by Aitchison et al. (2000) and Broughton (2006).

1.1.4 Thesauri as Knowledge Organization Systems

The literature of indexing, thesaurus construction, and subject access and information representation categorizes thesauri as controlled vocabularies. Thesauri have also been classified as *knowledge organization systems* (KOSs) (Hodge, 2000; Broughton et al., 2005), a term coined by the Networked Knowledge Organization Systems Working Group (NKOS) at its initial meeting at the Association for Computing Machinery Digital Libraries 1998 conference in Pittsburgh, Pennsylvania. Hodge (2000) explains the use of thesauri and other types of KOSs on the web in these terms:

> Knowledge organization systems are used to organize materials for the purpose of retrieval and to manage a collection. A KOS serves as a bridge between the user's information need and the material in the collection. With it, the user should be able to identify an object of interest without prior knowledge of its existence. Whether through browsing or direct searching, whether through themes on a web

> page or a site search engine, the KOS guides the user through a discovery process. (p. 3)

NKOS is devoted to the discussion of the functional and data models for enabling KOSs—such as classification systems, thesauri, gazetteers, and ontologies—to function as networked interactive information services that support the description and retrieval of diverse information resources through the internet. The American and European NKOS groups have held annual workshops in conjunction with the Joint Conference on Digital Libraries and the European Conference on Digital Libraries, providing a venue for research, development, and evaluation of KOSs on the web. Thesauri and their applications have been the focus of many presentations and publications in these workshops.

1.1.5 Uses and Functions of Thesauri

A thesaurus may be employed as an indexing tool, a searching aid, or a browsing and navigation function. As an indexing tool, a thesaurus can be used to assign indexing terms to a given document collection. Many bibliographic and commercial database providers use a thesaurus for indexing purposes.

As a searching tool or a query formulation support feature, thesauri can be used as an interactive term suggestion tool or as an automatic query expansion support functionality.

In the interactive term suggestion approach, users are presented with a list of terms to choose from. This can be the result of matching an initial query term with the thesaurus terms to provide synonyms or semantically related terms for the user's guidance. In the case of automatic query expansion, a thesaurus can be used to automatically add terms from it to the query terms a user has initially submitted in order to improve or enhance the retrieved results. Thesauri can provide a browsing user interface in which thesaurus terms and their relationships are presented on the user interface to assist users by making term selection a more engaging and interactive process. An extended discussion of thesauri as supporting tools for query formulation and expansion is provided in Chapter 3.

All of these uses and functions have been adopted by several generations of information retrieval systems, from traditional indexing and abstracting commercial databases to current web-based digital libraries, portals, repositories, and open archives. Aitchison et al.

(2000) note that thesauri may be used for both indexing and searching, for indexing but not searching, and for searching but not indexing. These uses are associated with the ways in which a thesaurus can be developed and incorporated into an information representation and retrieval system.

Additional uses of a thesaurus as noted by Broughton (2006) are as a source of subject metadata and query formulation and expansion, and as a browse and navigation tool. In his discussion of the functions of thesauri, Soergel (2003) comments that they can facilitate the combination of multiple databases or unified access to multiple databases in the following ways:

A. Mapping the users' query terms to the descriptors used in each of the databases

B. Mapping the query descriptors from one database to another (switching)

C. Providing a common search language from which to map to multiple databases

Another useful and interesting function that he refers to is document processing after retrieval, for instance, the meaningful arrangement of search results and the highlighted descriptors responsible for retrieval.

1.1.6 Types of Thesauri

The types and uses of thesauri depend largely on the ways in which they are constructed and incorporated into an information retrieval system. The well-known types of thesauri can be categorized as follows:

1. Standard, manually constructed thesauri: These are standard subject-specific thesauri with equivalence, hierarchical, and associative relationships, used in the indexing and retrieval of print and digital collections. Some databases and information retrieval systems use these thesauri for indexing purposes only, while others present these tools more explicitly to end users to support their search term selection.

2. Search thesauri: Search thesauri, also referred to as end-user thesauri and searching thesauri, are defined as a category of tools enhanced with a large number of entry terms that are synonyms, quasi synonyms, or term variants that assist end users in finding alternative terms to add to their search queries (Perez, 1982;

Piternick, 1984; Bates, 1986; Cochrane, 1992). Aitchison et al. (2000) note that the role of thesauri here is usually to assist users in searching free-text databases by suggesting search terms, especially synonyms and narrower terms. A number of searching thesauri have been designed and developed (Anderson and Rowley, 1991; Lopez-Huertas, 1997; Knapp et al., 1998; Lykke Nielsen, 2001) and have been evaluated in query expansion research (Kristensen and Jarvelin, 1990; Kristensen, 1993; Kekäläinen and Jarvelin, 1998). A searching thesaurus can also provide greater browsing flexibility. It can allow users to browse part or all of a thesaurus, navigating the equivalence, hierarchical, and associative relationships. Terms (or the combination of preferred and variant terms) can be used as predefined or "canned" queries to be run against the full-text index. In other words, a searching thesaurus can become a true portal, providing a new way to navigate and gain access to a potentially enormous volume of content. A major advantage of the searching thesaurus is that its development and maintenance costs are essentially independent of the volume of content. On the other hand, such thesauri put much greater demands on the quality of equivalence and mapping (Morville and Rosenfeld, 2007).

3. Automatically constructed thesauri: These thesauri are constructed with computer algorithms and are not as semantically well-structured as standard manually created thesauri. A wide range of statistical and linguistic techniques have been developed to build such thesauri. Unlike hand-crafted thesauri, corpus-based thesauri are constructed automatically from the corpora or information collection, without human intervention. There are two different methods of extracting thesaural relationships from text corpora, namely, co-occurrence statistics and grammatical relations (Mandala et al., 2000).

4. Linguistically and lexicographically focused thesauri: The well-known examples of these thesauri are *WordNet* and *Roget's Thesaurus*. *WordNet* is a manually constructed thesaurus, available electronically, and has been used in many information retrieval experiments for query expansion purposes. It is a general purpose thesaurus and therefore lacks the domain-specific relationships found in standard thesauri. *Roget's Thesaurus* is also available in electronic format and has been used in information retrieval experiments.

1.1.7 Knowledge Organization Trends

Several researchers have studied research and development trends associated with knowledge organization in general and thesauri in

particular. In her review of knowledge organization research between 1998 and 2003, McIlwaine (2003) highlights thesauri initiatives as one of the recent trends along with such topics as terminology, internet, search engines, resource discovery, interoperability, visual presentation, and universal classification systems. Williamson (2007) notes that, currently, controlled vocabularies of various kinds (e.g., thesauri and taxonomies), as well as other kinds of information structures, are deemed to have an important role to play. She says it is clear that thesauri have now assumed a role as a search tool. She provides a discussion of the application of thesauri on the web between 1997 and 2006 with a particular focus on their role in searching, browsing, and navigation.

Recent developments in the use of thesauri highlight how pre-web applications and standard tools such as thesauri are being used to make metadata more usable. As the organization of knowledge and information continues to evolve in the digital environment, it seems evident that the relevance of core principles of knowledge organization will remain high, despite shifting trends. These principles will most certainly help enhance both the browsability and searchability of emerging web-based environments, such as digital libraries, content management systems, institutional repositories, and virtual learning environments (Saumure and Shiri, 2008).

Subject analysis in general and the use of thesauri in particular enjoyed a flurry of interest in the 1970s and have recently become a focus of attention again. The scholarly community carrying out work in this area has become more diffuse and grown to include new groups such as information architects (Schwartz, 2008). The need to improve users' browsing, navigation, and experience in digital information spaces has brought both controlled vocabularies and thesauri to the center of attention.

1.1.8 Emergence of Thesauri Search Tools

With the development of the web, the use of thesauri is coming to the forefront of knowledge organization studies. New trends in developing thesauri have also been emerging since the advent of the web (Saumure and Shiri, 2008).

Over the past 15 years, numerous researchers have discussed the status, suitability, importance, and diversification of the function of thesauri in the new information environment. Aitchison et al. (2000) have noted that the role of thesauri is changing but that they are likely

to remain an important retrieval tool. This shift in the functions of thesauri is viewed as an expansion, including a role for thesauri not only in performance enhancement in full-text systems but also as tools for use on websites; in intranets; and for indexing, search statement expansion, and visual organization. While initial proposals for the use of thesauri focused on their ability to ensure consistent analysis of documents during input to information retrieval systems, these tools have increasingly become vital as aids to effective retrieval. Indeed, in the near future, it appears likely that thesauri will be used more during retrieval than at input. Thesauri can complement full-text access by aiding users in various ways: by focusing their searches, by supplementing the linguistic analysis of the text search engine, and even by serving as one of the analytic tools used by the linguistic engine (Milstead, 1998).

To reassess the functions and capabilities of thesauri in the digital age, any revisions to thesaurus construction standards should take into account at least four essential areas: 1) the nature and function of thesauri in full-text databases, 2) term definition and all types of term relationships, 3) dynamic and interactive display of thesauri in the digital environment, and 4) thesauri as support for the internet (Williamson, 2000). In a discussion of the importance of providing browsing capabilities for thesauri and subject headings, Olson (2007) notes that in many abstracting and indexing services, users are forced to switch between the thesaurus and the database in order to form an understanding of the references and relationships between terms and to make effective use of thesauri in support of searching. To make knowledge structures such as thesauri more browsable, she suggests that emphasis needs to be placed on the references and relationship types and on their visibility to searchers.

Shiri and Revie (2000) note that although there are few operational information retrieval systems that have effectively incorporated thesauri as search and retrieval aids, we are witnessing an increased enthusiasm among thesaurus developers to make their tools available on the web for potential applications. The reasons for this enthusiasm and the increasing availability of online thesauri are closely linked to five key issues associated with the emergence of the web:

1. The colossal growth of information resources, demanding better subject identification
2. The migration of traditional information resources to the web, calling for more consistent subject approaches

3. An urgent need for resource description and discovery through reuse of existing information management tools such as controlled vocabularies
4. Problems associated with the quality of unstructured information retrieved from the web
5. The need to provide users with knowledge structures such as thesauri for rapid and easy access to better-organized information

Shiri and Revie introduce some of the early developments associated with the use of thesauri on the web, such as thesauri incorporated into web-based databases, stand-alone thesauri, thesauri in multithesaurus search systems, and thesauri in subject gateways.

Miller (2003) argues that, as the use of the web becomes widespread, the problem of semantic organization of information will become more and more urgent. To address this problem, he suggests that a thesaurus should be constructed on the basis of the maximum possible number of terms and their synonyms, objective relations between terms, multiple languages, and receptivity to new terms. Lykke Nielsen (1998) suggests that future thesauri should also function as search tools to support users in analyzing and conceptualizing their information needs, in locating and choosing appropriate access points, and in refining requests as well as queries. However, today's pressures for intuitive end-user access and seamless flows of information from one system into another compel new thinking about ways of designing, implementing, and presenting vocabulary search tools (Aitchison and Dextre Clarke, 2004).

Thesauri have been used to develop organizational taxonomies for library and information science (Wang et al., 2008). Gilchrist (2003) comments that taxonomies use both classification and thesaurus techniques, and it is interesting to note how similar some of the techniques are in automatic indexing and 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; museums documentation and image retrieval may be mentioned here as areas in which the object types pose particular problems and in which other techniques are being developed. Faceted classification techniques can be used to provide a framework on which taxonomies can be built. The focus on noun forms and unit concepts popular in thesauri can be adopted to provide a more consistent

approach to taxonomy construction. In a discussion of the past 50 years of knowledge organization, Dextre Clark (2008) writes as follows:

> As the taxonomy buzz-word spread around, many information professionals seized a different opportunity. They rescued their existing home-grown thesauri, subject heading schemes and classification schemes, dusted them off a little, and re-branded them "taxonomy." The controlled vocabulary had now become more popular than ever before! (p. 433)

These developments suggest that the terms *thesaurus* and *taxonomy* have been loosely and interchangeably used and that some people who have used the term *taxonomy* were unaware of the long-standing research and development behind thesauri and their construction standards.

Gruber (2009) notes that "an ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modeling a domain and the specification takes the form of the definitions of representational vocabulary (classes, relations, and so forth), which provide meanings for the vocabulary and formal constraints on its coherent use" (p. 1,964).

A quick analysis shows that there are a number of similarities between *ontologies* and *thesauri*, namely, in their treatment of concepts, classes, and relationships. Therefore, it is not surprising that these two terms have been used interchangeably, and confusingly, in the literature. A very good example of this confusion can be found in the terms used to refer to *WordNet*, a large lexical tool for the English language. It has been called a *thesaurus* in numerous information retrieval studies during the past decade, but it has also been called an *ontology* by the World Wide Web Consortium and a *taxonomy* by some researchers.

However, one of the key characteristics of ontologies is that they provide a more formal and detailed set of conceptual constructs and relationships than do thesauri, and the formalization lends itself very well to the web environment. As Gruber (2009) suggests, ontologies are used "to exchange data among systems, provide services for answering queries, publish reusable knowledge bases, and offer services to facilitate interoperability across multiple, heterogeneous systems and databases" (p. 1,965.)

An analysis of these functions shows that they are common to both thesauri and ontologies. Therefore, development of any high-level, sophisticated, and machine-processable ontology can benefit from the conceptual and semantic structures inherent in various existing thesauri. Gilchrist (2003) suggests that the main characteristic that thesauri, taxonomies, and ontologies have in common is that they all address natural language. Soergel (1999) refers to a recent interest in ontologies as classification tools in such areas as artificial intelligence, linguistics, and software engineering and notes that "indeed, once these communities increased their awareness that there is not only a problem of classification but also of terminology, 'ontologies' included lead-in vocabularies as well, and became full-fledged thesauri"(p. 1,120.)

His argument points to the fact that scholarly communities outside library and information science identified the need for classification and used the term *ontology* without actually benefiting from the long-standing research, development, and standardization forming the basis of numerous well-structured controlled vocabularies such as thesauri and classification schemes. He calls for collaboration among these various communities to create better information access systems.

From an information architecture point of view, Morville and Rosenfeld (2007) comment that thesauri are expected to be more widely used in the coming years as they become a key tool for dealing with the growing size and importance of websites and intranets. One advantage of thesauri is their tremendous power and flexibility to shape and refine the user interface over time. Not all of the capabilities can be exploited at once, but one can user-test different features, learning and adjusting incrementally as one proceeds.

A review of the literature on thesauri and their applications and functions in the new digital information environment identifies a wide range of ways in which thesauri can be made more suitable for the new search environment. Some of the more common approaches are as follows:

- Revising thesaurus construction standards to facilitate the development and use of thesauri. The British and U.S. thesaurus construction standards have recently been revised to reflect current changes and development in the areas of thesauri and other types of controlled and structured vocabularies.

- Using a wide range of user-based and document-based techniques for thesaurus construction, including bibliometric approaches, term co-occurrence analysis, word association tests, transaction logs, and data-mining and web-mining technologies.
- Enriching thesauri by incorporating a larger number of terms and relationships so as to provide a vast entry vocabulary to support users' initial interaction with the information retrieval system. Search thesauri are one example of these tools that may support free text searching.
- Enhancing the semantic structure of thesauri, such as expanding the relationship types within a thesaurus or covering a broader range of relationships among terms.
- Constructing more-sophisticated user interface features and functionalities. Many information retrieval systems and databases have a thesaurus but do not provide seamless, straightforward access to the thesaurus to support end users in their search process. This kind of access can be designed in such a way as to make thesaurus structures more explicitly visible for browsing, searching, and navigation purposes. Interface design techniques and strategies that combine browsing and searching can be adapted to provide more dynamic and interactive interfaces.
- Using thesauri for interactive (visible) or automatic (invisible) query formulation or expansion to support users' information interaction.
- Using thesauri as sources of subject metadata. Many thesauri are now being adapted to provide consistent subject description in well-known metadata standards such as Dublin Core.
- Using existing thesauri to organize and visualize web-based information systems and services. Examples are websites, intranets, content management systems, portals, and subject gateways.
- Using existing thesauri to develop simplified or more sophisticated knowledge structures for organizing and

representing disciplinary or multidisciplinary web-based applications.

- Employing multilingual thesauri for web-based cross-lingual information retrieval.
- Bringing into play user evaluation of thesauri and their usefulness within the context of web-based information systems and services in order to provide insight into the ways in which thesauri may support users' search behavior.

1.2 Thesauri and Information Architecture

The Information Architecture Institute (2005) defines *information architecture* as the art and science of organizing and labeling websites, intranets, online communities, and software to support usability and findability. Rosenfeld and Morville (1998), in the first edition of *Information Architecture for the World Wide Web*, were among the first authors to introduce the information architecture community to thesauri and controlled vocabularies. They note that the relationships in standard thesauri can be useful for determining the labeling of the different levels of a website.

While the terms of a thesaurus can be adapted, however, the website designer needs to remember that the narrower and the more specific its vocabulary, the better the thesaurus terms will perform for the website. For example, if the site users are computer scientists, a computer science thesaurus will "think" the same way that its users do. In choosing a labeling or KOS, the authors particularly emphasize the importance of taking into account the types of users and their information search habits.

A successful website will have a well-organized knowledge structure that accommodates users' search and interaction behavior. Constructing and using a controlled vocabulary impose an important degree of consistency that supports search and browsing. A thesaurus on the back end can enable a more seamless and satisfying user experience on the front end (Morville and Rosenfeld, 2007). Even though the first thesauri were developed for libraries, museums, and government agencies long before the advent of the web, Morville and Rosenfeld believe that information architects can draw on these decades of experience.

Designing labeling and organization structures for websites and intranets can benefit from the characteristics and features of thesauri. Synonym management is the most important function of a thesaurus used as part of a website. The mapping of many synonyms or word variants onto one preferred term or concept is an important feature allowing users to deal with the ambiguities of language during their searching and finding experience (Morville and Rosenfeld, 2007).

Thesauri have come back into our everyday life via the web. More than a tool to get more and better words, thesauri are used to create a web of interconnected terms to help people find information (Wodtke and Govella, 2009).

The Argus Center for Information Architecture polled its membership about subject matters with which information architects are concerned. Based on the responses of 241 participants between February 9 and 21, 2001, survey results showed that some 54 percent of respondents felt that controlled vocabularies and thesauri were among the subject areas with which information architects are concerned (Zhang et al., 2002).

Thesauri, taxonomies, and topic maps have been compared and discussed as tools that assist information architects to develop better user interfaces for their websites and intranets. Thesauri provide a much richer vocabulary for describing terms than taxonomies do and so are much more powerful retrieval tools. As can be seen, using a thesaurus instead of a taxonomy would solve several practical problems in classifying objects and also in searching for them (Garshol, 2004). Other researchers have demonstrated that all the characteristics of standard thesauri, such as broader, narrower, and related terms, as well as scope notes and synonymous terms, can be effectively used to create topic maps and well-structured taxonomies (Ahmed, 2003).

Pastor-Sanchez et al. (2009) discuss the advantages of thesaurus representation in Simple Knowledge Organization System format, a World Wide Web Consortium standard to promote the use of KOSs in support of the semantic web. They suggest that the conceptual structures of thesauri allow 1) the possibility of establishing lexical relationships adapted to the terminological reality of each language; 2) the indexing of webpages with a thesaurus to present queries without users' having to perform a predictive selection of terms; 3) the development of organization schemes; and, 4) the possibilities of expanding and redefining searches, showing references to documents with

content related to that of directly retrieved documents, and suggesting new search terms.

In the context of information retrieval, BS 8723 for *Structured Vocabularies for Information Retrieval* (2005) suggests this:

> It is inappropriate to use the classical definition of taxonomy as the science of classification, or to be concerned with its long-standing adaptation to the classification and naming of organisms. BS8723 deals in general with vocabulary tools designed as retrieval aids, hence the definition of taxonomy used in this standard, as a structured vocabulary using classificatory principles as well as thesaural features, designed as a navigation tool for use with electronic media. The standard also notes that the term taxonomy is used differently.

Therefore, many of the taxonomies that have been used in websites and portals are not used for vocabulary control or do not follow thesaurus construction standards to serve as information retrieval tools. There are practical examples of web-based tools and services that have made use of thesauri for designing their information architecture. The SMETE (Science, Mathematics, Engineering, and Technology Education) Digital Library in the U.S. makes use of a thesaurus developed by the Mathematics Association of America that contains mathematical concepts (Dong and Agogino, 2001).

In the absence of user learning, and with no easy way for users to exploit thesaurus relationships, attention has recently turned to what has come to be called guided navigation. It is one result of the intersection between information architecture and library and information science. As designers of web user experiences, information architects need to find ways to help users, especially online shoppers and corporate employees, navigate through large information spaces containing objects with many potentially searchable attributes (Schwartz, 2008).

Beeson and Chelin (2006) note that if one scans the burgeoning literature on information architecture that is associated with the spread of applications on the web, one finds theories for organizing and searching information, as well as methods for creating metadata, controlled vocabularies, and thesauri—all of which could have come from a textbook on information science.

Almost all the books on information architecture have a chapter on controlled vocabularies and thesauri and the ways in which these tools can be used to properly organize content, as well as to effectively assist users in their information access and retrieval.

1.3 Faceted Search User Interfaces

1.3.1 Facet Analysis

S. R. Ranganathan (1967) proposed the idea of facet analysis, which he used in his faceted Colon Classification scheme. The basic idea was that any component, aspect, or facet of a subject can fit into one of five categories, namely, personality, matter, energy, space, and time.

This technique has been widely used in the design and development of classification schemes and thesauri. The first thesaurus constructed on the principles of facet analysis was *Thesaurofacet*, developed by Jean Aitchison in the 1960s. Examples of thesauri developed on the basis of the facet analysis technique are the AAT and the *ASIS&T Thesaurus of Information Science, Technology, and Librarianship*.

Aitchison et al. (2000) emphasize that faceted classification is useful in thesaurus construction in several ways. First, it provides a tool for the analysis of subject fields and for determining the relationships among concepts. Second, the resulting faceted classification may be used as the systematic display in a thesaurus. Third, facets may be added to terms in existing vocabularies, in order to further define the meaning and role of such terms.

Figure 1.1 shows one of the key facets used by the AAT. As can be seen, the *styles and periods* facet has a rich and detailed hierarchy consisting of sub-facets such as *styles and periods by general area* and *styles and periods by region*. This type of arrangement provides a useful browsing structure for users, who can refine or specify a certain category of style period on the basis of the faceted structure.

Figure 1.2 shows the facet *knowledge and information* and the subfacet *knowledge organization systems* in the *ASIS&T Thesaurus of Information Science, Technology, and Librarianship*. The detailed view provided by this type of faceted structure not only allows users to gain a complete overview of each facet and its scope but also makes browsing and navigating around the thesaurus a more easily understood process.

Application of facet analysis and faceted thesauri has become prevalent among information retrieval user interface designers,

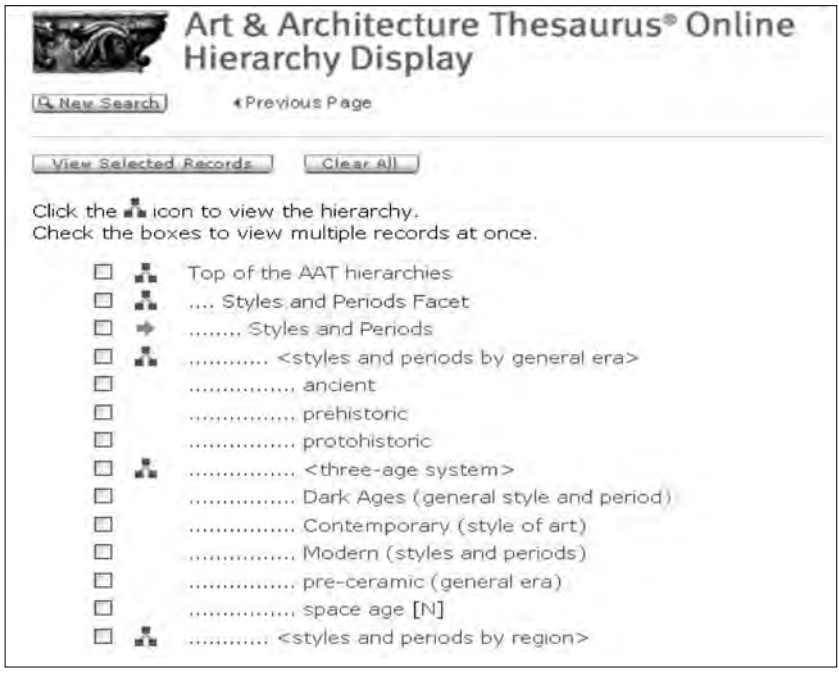

Figure 1.1 Display of the *styles and periods* facet in the Art and Architecture Thesaurus

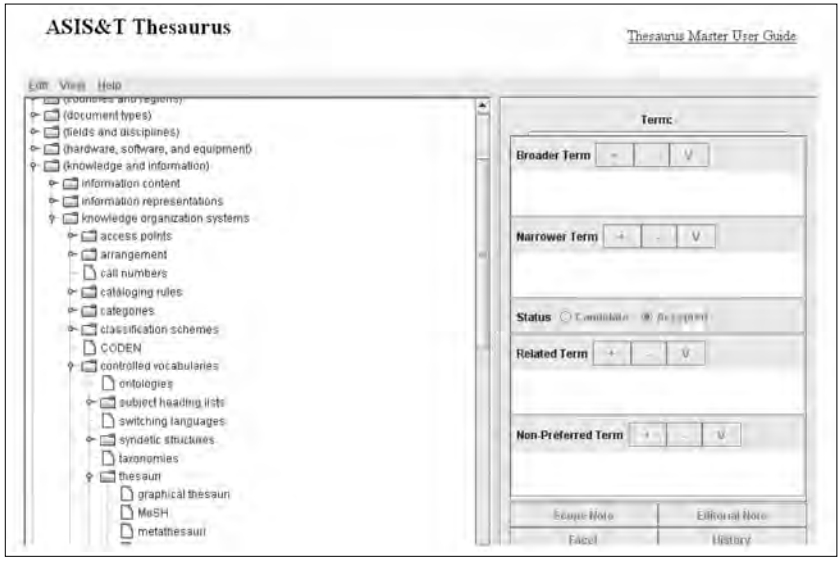

Figure 1.2 The faceted structure of the *ASIS&T Thesaurus of Information Science, Technology, and Librarianship*

information architects, and web developers of based services. Such applications and web interfaces tend toward a broader view of facets than the traditional library focus on document subjects, incorporating various metadata elements such as commodity price or scalar properties of an object. This can include facets that are essentially pick lists, and there is usually little notion of the semantics of combining facets.

Nonetheless, this simple facet treatment can yield attractive browsing interfaces for websites (Tudhope and Binding, 2008). The FACET (Faceted Access to Cultural hEritage Terminology) project investigated the potential of multifaceted semantic query expansion in controlled vocabulary indexed applications. Query expansion was based on a faceted thesaurus, the AAT. In FACET, such expansion provides an option to include closely related concepts in search. Results are ranked in order of decreasing relevance to the initial query, based on the number of matching query terms and the degree of match between concepts.

1.3.2 Faceted Search

The world of the web is beginning to realize that the tools of facet analysis can build robust, dynamic, mutable, and responsive systems (La Barre, 2004). The term *facet* is widely used in the information science community, but in other disciplines similar concepts are referred to as *attribute*, *dimension*, *metadata*, *property*, or *taxonomy* (Dumais, 2009).

The terms *faceted search*, *faceted navigation*, *faceted metadata*, and *faceted browsing* have been used interchangeably, and sometimes loosely, in the literature. In part, this is because of the increasing popularity of integrated searching and browsing in faceted search interfaces. Also called *guided navigation* and *faceted search*, the faceted navigation model leverages metadata fields and values to provide users with visible options for clarifying and refining queries. Faceted navigation is arguably the most significant search innovation of the past decade (Morville and Callender, 2010). It features an integrated, incremental search and browse experience that lets users begin with a classic keyword search and then scan a list of results.

Dumais (2009) outlines the key components of faceted search interfaces and suggests that most systems show the query, the facet structure, the subset of results currently specified, and, sometimes, a

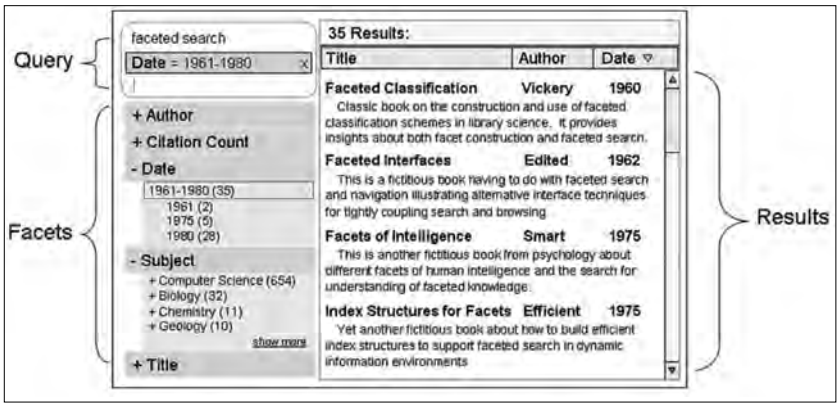

Figure 1.3 Example of a faceted search interface

detailed view of an individual item. Figure 1.3 depicts three main areas of a faceted search interface, namely, query, facets, and results. The interface demonstrates a combined approach to searching and browsing by presenting both the query box and the facets. Each facet can be collapsed and viewed.

One of the early examples of using facet-based user interfaces was HIBROWSE (High Resolution Interface for Database Specific Browsing and Searching), developed by Pollitt et al. (1994). They designed a series of user interfaces for several bibliographical and multilingual databases. An example of such an interface is shown in Figure 1.4; the interface is developed for hotels based on such categories as name, city, number of rooms, rating, and so forth.

In a discussion of user interface design for faceted navigation, Hearst (2008) comments that faceted navigation is a proven technique for supporting exploration and discovery within an information collection. Faceted classification and faceted navigation are now widely used in website search and navigation.

In research on the Flamenco project, Hearst and colleagues (Hearst, 2000; Hearst et al., 2002; Yee et al., 2003; Hearst, 2006) describe the importance of faceted classification systems for website navigation; they have also designed and studied a series of user interfaces to support faceted navigation for everyday users. The overarching design goals of the Flamenco project were to support the following:

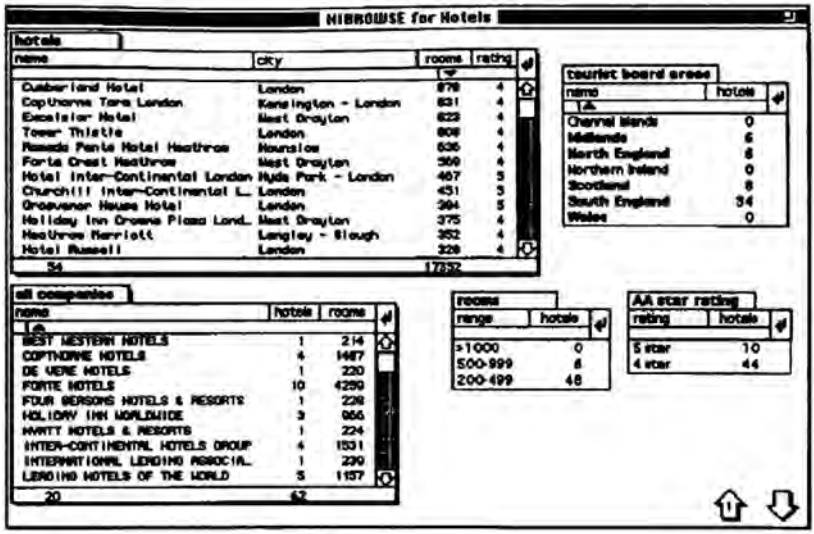

Figure 1.4 HIBROWSE user interface for hotels

- Flexible navigation
- Seamless integration of browsing with directed (keyword) search
- Fluid alternation between refining and expanding
- Avoidance of empty results sets
- User control and understanding at all times

Hearst also notes that another of the Flamenco project's goals was to promote the idea of faceted navigation in online systems, both as an alternative to the hierarchical focus of website structure and in response to the failure of subject searching in online catalogs.

Figure 1.5 shows the Flamenco user interface developed for the University of California–Berkeley Architecture Visual Resources Library, which is organized and represented using such facets as *people*, *periods*, *locations*, *styles*, and *view types*. The interface also allows users to browse and navigate subcategories within each facet.

Based on the idea of faceted search, Cutrell et al. (2006) developed Phlat (Figure 1.6), a user interface to facilitate and improve personal information management (PIM). The interface combines searching and browsing with facets provided as a sorting mechanism.

Figure 1.5 Flamenco user interface

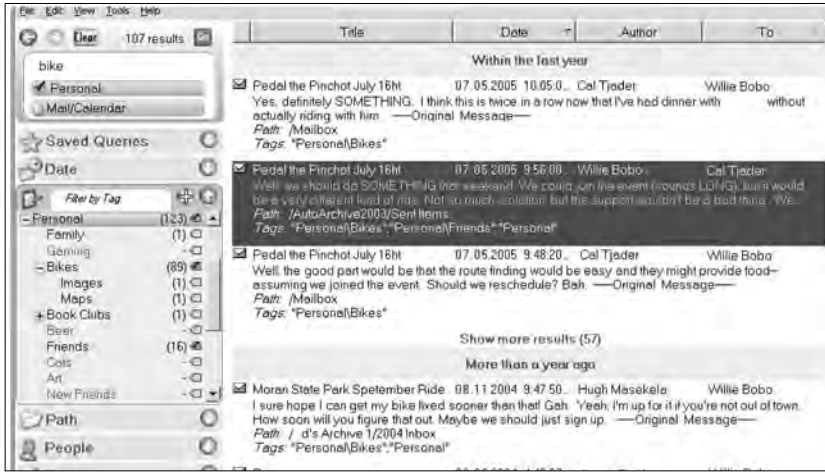

Figure 1.6 The Phlat interface

Several commercial and ecommerce websites make use of faceted browsing and navigation. Two very well-designed examples of faceted navigation are eBay Express and Yelp. In eBay Express (Figure 1.7), a search for *perfume* will provide the user with several facets to navigate, such as *gender and age*, *fragrance name*, *brand*, *condition*, *type*, *location*, and *buying format*. Visitors to the website can refine or reformulate their initial query using these facets and sub-facets.

Yelp is a social networking, user review, and local search website that provides location-aware information services in the U.S., Canada, and the U.K. A search for *organic grocery stores* in Edmonton

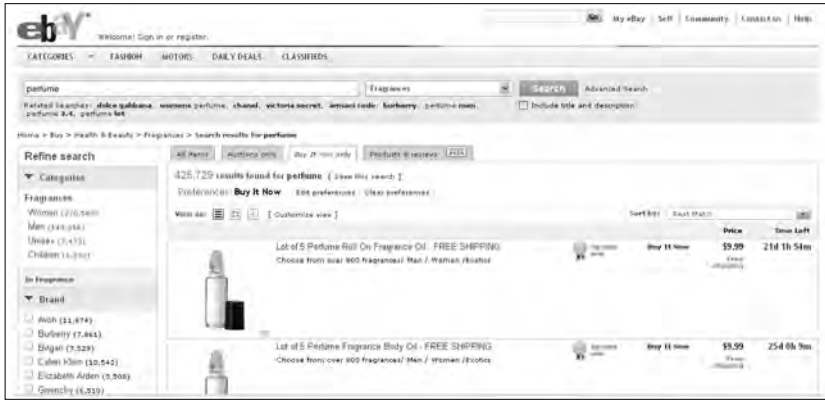

Figure 1.7 eBay Express

Figure 1.8 Yelp user interface

(Figure 1.8), for example, provides users with the facets *distance*, *features*, *price and category*, *highest rated*, and *most reviewed*.

Online library catalogs have rich metadata, and many have recently started using the metadata to provide faceted navigation of their collections. Faceted navigation enables new ways of and approaches to resource discovery in library catalogs. Figure 1.9 shows a search for *information retrieval* in WorldCat, the world's largest network of library content and services, with the user able to browse and employ various metadata elements such as *author*, *format*, *year*, *audience*, and *language*.

Figure 1.9 Faceted navigation in WorldCat [*Copyright owned by the Online Computer Library Center, Inc., and screenshot used with its permission.*]

Faceted searching, including browsing and navigation, is a promising area now widely used on the web. However, faceted search interfaces are not widely available for general web search as facet values are available only for a small portion of the web. Key determinants of successful application of faceted search methods for web content are 1) understanding which facets are most important to support the varieties of information needs for which people use the web and 2) handling large-scale dynamic collections (Dumais, 2009). Morville and Callender (2010) suggest that faceted navigation is a master search pattern impacting all search and navigation patterns, together with the information architecture as a whole.

1.4 Exploratory Search Interfaces

The term *exploratory search* can be used to describe an information-seeking problem context that is open-ended, persistent, and multifaceted. It can also describe information-seeking processes that are opportunistic, iterative, and multitactical. In the first sense, exploratory search is commonly used in the context of scientific discovery, learning, and decision making. In the second sense, exploratory tactics are used in all manner of information seeking in order to reflect seeker preferences and experiences as much as their information seeking goal (Marchionini, 2006).

Highly interactive and dynamic user interfaces for exploratory browsing and searching of digital information collections have been the focus of some recent research. White et al. (2006) suggest that in exploratory search, users generally combine querying and browsing strategies to foster learning and investigation. Marchionini (2006) points out that to engage people more fully in the search process and put them in continuous control, researchers are devising highly interactive user interfaces. He proposes that exploratory search consists of "look up," "learn," and "investigate" activities in which examining and comparing results and reformulating queries to discover the boundaries of meaning for key concepts, as well as serendipitous browsing, take place. His view of exploratory search focuses on user interface functionalities that support a combination of browsing and searching, as well as providing the user with a conceptual space for exploration and comprehension of concepts and ideas.

In exploratory search, people usually submit a tentative query to navigate proximal to relevant documents in the collection and then explore the environment to better understand how to exploit it, all the while selectively seeking and passively obtaining cues about their next steps. Examples of exploratory search systems include visualization systems, document clustering and browsing systems, and intelligent content summarization systems (White and Roth, 2009).

Thesauri, as semantic tools and knowledge structures, have the potential to support exploratory searches and can be incorporated into exploratory search interfaces to assist users in the exploration and comprehension of concepts and ideas. As Marchionini (2006) notes, helping searchers to understand data structures and infer relationships among concepts is an important step in exploring and discovering the boundaries of meaning for key concepts. Thesauri, with their rich semantic relations, are capable of facilitating exploratory search activities through allowing the user to form a conceptual map of a particular subject area and to create a context for search and exploration.

Faceted search interfaces combine querying and browsing, allowing people to quickly and flexibly find information based on what they remember about the information they seek. Faceted search interfaces can also help people avoid feelings of being lost in the collection and make it easier for them to explore.

White and Roth (2009) suggest the following set of principles that support exploratory search activities:

- "Support querying and rapid query refinement: Systems must help users formulate queries and adjust queries and views on search results in real time.
- Offer facets and metadata-based result filtering: Systems must allow users to filter and explore results through facet selection and document metadata.
- Leverage search context: Systems must leverage available information about their user, their situation, and their current exploratory search task.
- Offer visualization to support insight and decision making: Systems must present customizable visual representations of the collection being explored in order to support hypothesis generation and trend spotting.
- Support learning and understanding: Systems must help users acquire both knowledge and skills by presenting information in ways amenable to learning, given the user's current knowledge or skill level." (p. 41)

A review of these principles suggests that both thesauri and facets can support some of these exploratory activities through the provision of semantic and conceptual maps of digital information collections. Exploratory search principles may be used to enhance the utility and usefulness of many existing thesauri and faceted classification schemes and structures.

It is interesting to observe the gradual convergence of several lines of current research, namely, exploratory search, faceted search, metadata-based search, and information architecture. All of them share a common aim: to improve and enhance users' access to digital information via similar principles developed over the past four decades. In fact, faceted search interfaces and exploratory search interfaces share similarities to the point that some of the former have also been introduced as the latter.

Figure 1.10 shows mSpace Explorer, a multifaceted, column-based client for exploring large data sets. The mSpace Explorer runs on top of the mSpace framework, an exploratory search system that allows users to choose predefined facets within a broad topic and dynamically modify results in real time. It also assists users in filtering information based on any categories that have been defined as the facets of the mSpace "slice," for example, as shown in the image, categories

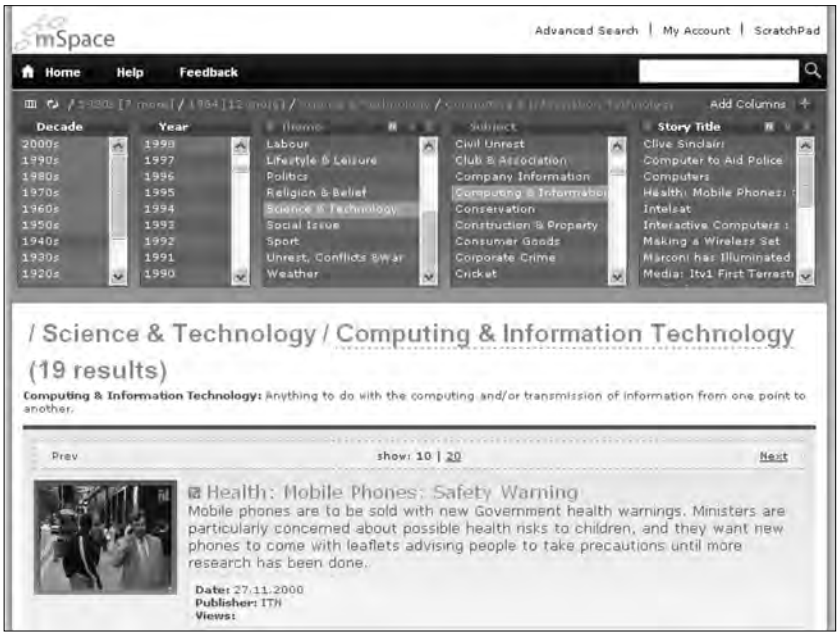

Figure 1.10 mSpace Explorer user interface

such as *year*, *theme*, *subject*, and *storyline*. Another feature of this interface lies in its integration of query and browsing.

Another example of an exploratory user interface is Relation Browser, developed by researchers at the University of North Carolina across a series of projects (Zhang and Marchionini, 2005). Figure 1.11 shows an example of the Relation Browser developed for the U.S. Bureau of Labor Statistics (Capra and Marchionini, 2007). It is designed as a tool for understanding relationships between items in a collection and for exploring an information space (i.e., a set of documents). The interface is highly interactive and tightly couples searching and browsing, allowing users to view facets and results at the same time. The results can be dynamically updated and viewed. Users can filter results using such high-level facets as *topic*, *genre*, *region*, and *format*. Figure 1.11 shows the user interface features of Relation Browser.

1.5 Dynamic Term Suggestion Systems

Query formulation is a challenging, yet key, stage in the information retrieval process. One of the strategies to engage users in the search

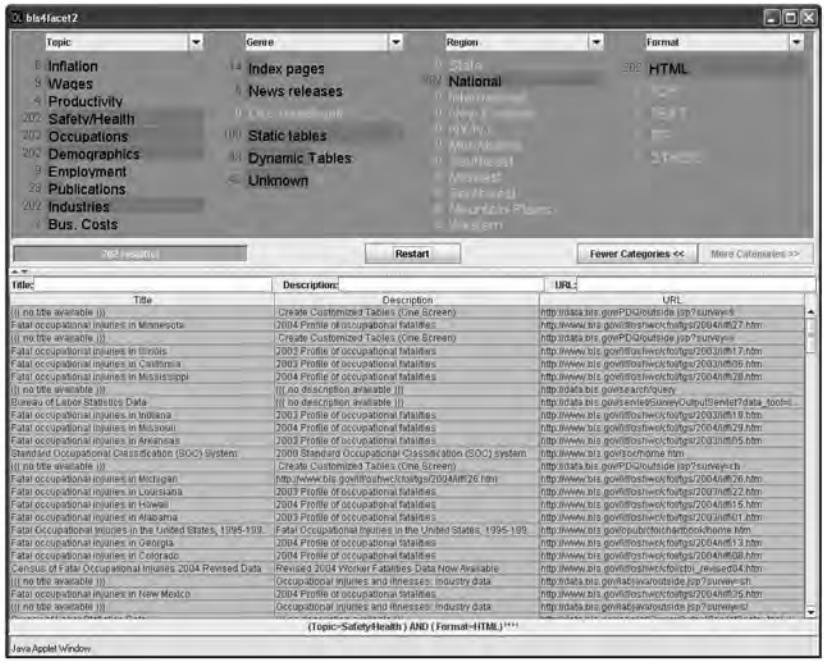

Figure 1.11 Relation Browser user interface

process and support them in formulating better approaches is to suggest search terms. Recently, a number of web search engines and information retrieval systems have incorporated new user interface features that support search term suggestion.

In the literature of search and information retrieval, these features have been called interactive, dynamic, or automated term suggestion mechanisms. These search term suggestion features aim to assist users in query formulation through suggestions of alternative terms and phrases for allowing users to refine or expand their initial search terms.

The advantage of term suggestion is that it helps users to formulate a particular query and, at the same time, form a quick understanding of what the information collection contains on that term or similar terms. As Hearst (2009) notes, the suggestion terms may come from several different sources, including the characteristics of the collection; terms derived from the top-ranked results; a combination of both; a domain-specific, hand-built thesaurus; query logs; or a combination of query logs with navigation or other online behavior.

Recently, numerous search engines, commercial databases, ebusiness websites, and online public access catalogs (OPACs) have started to incorporate term suggestion features into their systems and user interfaces. For example, the Yahoo! Search interface offers search term suggestions as a user starts typing in keywords. Figure 1.12 shows a search for the term *search engines*.

One of the early applications of thesaurus-enhanced interactive term suggestion can be attributed to Schatz et al. (1996), who developed a user interface for the University of Illinois Digital Library Initiative. The interface makes use of the Inspec Thesaurus to suggest terms to the user. Figure 1.13 shows an example of a search for *deductive databases* from the prototype developed by Schatz et al. Displayed are several terms for users to browse through or to select for refinement or reformulation of their initial search.

Other researchers have used mapping and matching techniques to design interactive term suggestion facilities. For instance, Gey et al. (2001) have studied the interactive suggestion to users of subject terms by means of probabilistic mapping between the user's natural language and the technical classification vocabularies. This occurs through a methodology called Entry Vocabulary Indexes. Other researchers have made use of thesauri to suggest terms and query refinement strategies to the user as well.

An interesting and efficient example of incorporating a thesaurus into a search user interface to support interactive term suggestions is the International Atomic Energy Agency (IAEA) digital collection.

Figure 1.12 Yahoo! term suggestion interface [*Reproduced with permission of Yahoo! Inc. ©2011 Yahoo! Inc. YAHOO! and the YAHOO! logo are registered trademarks of Yahoo! Inc.*]

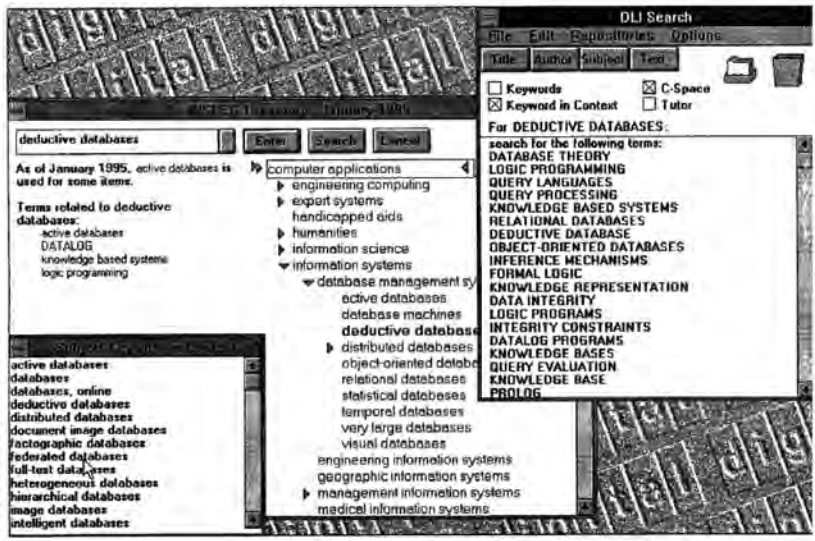

Figure 1.13 Interactive term suggestion interface developed by Schatz et al. (1996)

When a user searches for a term in the collection, the main search page shows the results for the term, and a list of suggested terms for narrowing down the search appears on the right-hand side of the interface. For example, a search for *pollution* retrieves 48,200 results, as indicated in Figure 1.14.

The user can then click on the narrower terms shown to reduce and refine the retrieved results to a more specific set of documents. In this example, if the user decides to narrow down the search using one of the narrower terms, say, *air pollution monitoring*, the number of retrieved results decreases to 3,240. The user can further narrow down the search by choosing another narrower term from the right side of the interface, as shown in Figure 1.15.

Recently, Gray et al. (2010) have developed a system that uses multiple astronomical thesauri to assist users in finding the right term in their search process. As part of the system, Gray et al. created Vocabulary Explorer, which allows users to search and browse the various thesauri. Detailed information about any matched term will be shown in order to help the user identify the right term.

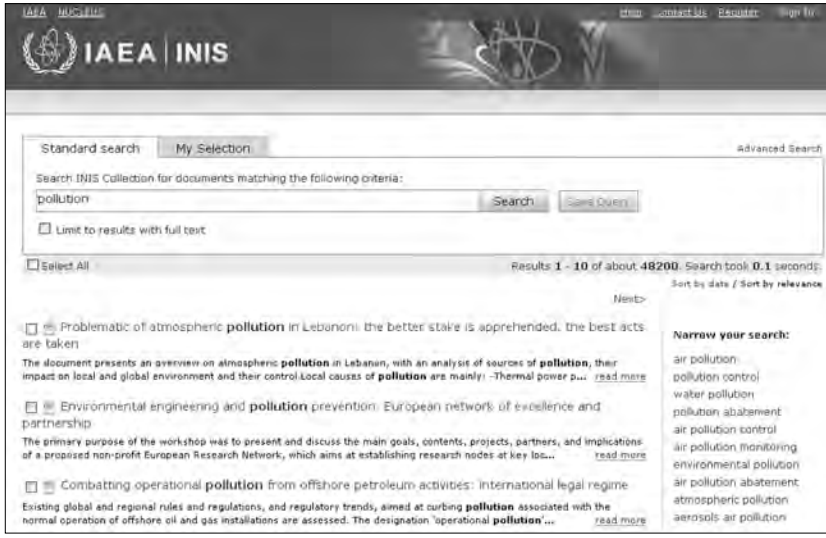

Figure 1.14 IAEA digital collection search term suggestion based on the International Nuclear Information System Thesaurus

Figure 1.15 Narrowing down the search in the IAEA digital collection using thesaurus-based term suggestions from the International Nuclear Information System Thesaurus

1.6 Thesauri and Social Tagging

Social tagging, sometimes referred to as *social bookmarking*, is defined variously as the classification of resources "by the use of informally assigned, user-defined keywords or tags" (Barsky and Purdon, 2006, p. 66) and elsewhere as the classification of resources "using free-text tags, unconstrained and arbitrary values" (Tonkin, 2006). In addition to *social bookmarking*, quasi-synonymous terms for social tagging include *collaborative tagging*, *folksonomy*, *folk categorization*, *communal categorization*, *ethno-classification*, *mob indexing*, and *free-text tagging*.

Social tagging emerged in popular practice around 2003, at the same time as social networking websites, and constitutes an important part of the interactive, democratic nature of Web 2.0 because the responsibility for the classification of web resources is placed squarely in the hands of the users. Tonkin (2006) proposes a two-part taxonomy of social tagging systems: "'broad,' meaning that many different users can tag a single resource, or 'narrow,' meaning that a resource is tagged by only one or a few users."

Shiri (2009) provides a comparative examination of a typology of social tagging systems that encompasses social networks, social bookmarking, video blogging and sharing, photo sharing, academic bookmarking, and slide sharing. He notes that some social tagging services, such as Technorati, Flickr, Bubbleshare, YouTube, and MySpaceTV, require users to organize their posted items in predetermined categories imposed by the service (generally anywhere from five to 20 categories). These categories represent a thesaurus-like hierarchical structure and often serve as a complement to tagging activities. For example, a YouTube user posting a video must put it in a category such as entertainment, comedy, or news, as well as describe it with appropriate tags.

A number of studies have discussed the comparison and reconciliation of controlled vocabularies with social tagging and folksonomies. Macgregor and McCulloch (2006) provide a succinct review of early debates about controlled vocabularies and collaborative tagging. Most of the difficulties associated with social tags and folksonomies (e.g., low precision, lack of collocation and consistency) originate from the absence of those properties that have come to characterize controlled vocabularies. Macgregor and McCulloch speculate that, ultimately, the coexistence of controlled vocabularies and collaborative tagging systems will emerge, with each appropriate

for use within the following distinct information contexts: formal (e.g., academic tasks, industrial research, corporate knowledge management) and informal (e.g., recreational research, PIM, exploration of exhaustive subject areas prior to formal exploration).

Spiteri (2007) evaluated tags against Section 6 (choice and form of terms) of the NISO guidelines for the construction of controlled vocabularies and found that the folksonomy tags correspond closely to the NISO guidelines pertaining to the types of concepts expressed by the tags, the predominance of single tags, the predominance of nouns, and the use of recognized spelling. She suggests that folksonomies could serve as a very powerful and flexible tool for increasing the user-friendliness and interactivity of public library catalogs.

Hastings et al. (2007) report the findings that various studies have in common on people's image tagging and descriptions: 1) tags assigned to groups of images and individual images differ in terms of their level of abstraction, 2) image tagging specificity and exhaustivity levels differ greatly among individuals, and 3) the accordance between existing controlled vocabularies and tags varies in terms of image attributes.

In a user-centered study of authors and readers of digital collections, Golub et al. (2009) investigated how social tags can be enhanced by the use of controlled vocabularies such as classification schemes and thesauri. Their findings showed the importance of controlled vocabulary suggestions for both indexing and retrieval in order to accomplish several functions: help produce ideas of tags for users, make it easier to find focus for the tagging, ensure consistency, and increase the number of access points in retrieval. The quality of the suggestions from the controlled vocabularies was found to be a key factor.

In a series of studies comparing social tags and controlled vocabularies, Kipp (2010) and Lu and Kipp (2010) concluded that there is continuity between conventional indexing and user tagging, and that this continuity could form the basis for a complementary system of subject access that would enrich conventional indexing and support its continued utility.

These studies suggest that social tagging and controlled vocabularies have their own advantages and disadvantages, but that social tags do not replace the latter; rather, social tags complement controlled vocabularies and provide additional access points for users. To afford better user experiences, information access and retrieval systems should use a combination of controlled vocabularies and social tags in order to create more-inclusive user interfaces. The ways in which

combined use of controlled vocabularies and tags can be achieved depend, to a large extent, on the nature of the target audience, on the content and context of the information collection, and on the information search tasks that the system is designed to support.

1.7 Conclusion

This chapter has provided a brief history of information retrieval thesauri, along with the associated standards. Functions, uses, and types of thesauri were introduced. It was noted that the advent of the World Wide Web facilitated much greater use of thesauri on the web and in a variety of search environments.

Developments related to web technologies and web-based services and systems provide an opportunity for the reusing and repurposing of thesauri as networked KOSs.

The information architecture community benefits from various applications of thesauri as searching, browsing, and navigation tools.

Faceted and exploratory search systems and interfaces have adopted thesauri to expand and enhance the search horizon through semantic and conceptual structures embedded in thesauri, thus facilitating the exploration of digital collections and the performance of effective searches.

Thesauri have long been used as search strategy support mechanisms to suggest terms to users in a dynamic and interactive mode, with the goal of encouraging and engaging users in the search process. All of these developments suggest that thesauri have an increasingly major role to play in powering search in the new information environment.

References

Ahmed, K. (2003). Topic map design patterns for information architecture. *XML Europe, Londra 2003*, pp. 5–8. Retrieved from www.techquila.com/tmsinia.html (accessed May 1, 2012).

Aitchison, J., and Dextre Clarke, S. D. (2004). The thesaurus: A historical viewpoint. With a look to the future. *Cataloguing and Classification Quarterly*, 37(3/4), 5–21.

Aitchison, J., Gilchrist, A., and Bawden, D. (2000). *Thesaurus construction and use: A practical manual*, 4th ed. London: Aslib.

Anderson, J. D., and Rowley, F. A. (1991). Building end-user thesauri from full-text. In: Barbara H. Kwasink and Raya Fidel (Eds.), *Advances in classification*

research (Proceedings of the 2nd ASIS SIG/CR classification research workshop, pp. 1–13). Medford, NJ: Learned Information.

ANSI/NISO Z39.19: 1993. (1993). *Guidelines for the construction, format, and management of monolingual thesauri*. Bethesda, MD: National Information Standards Organization Press.

ANSI/NISO Z39.19: 2005. (2005). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda, MD: National Information Standards Organization Press.

Barsky, E., and Purdon, M. (2006). Introducing Web 2.0: Social networking and social bookmarking for health librarians. *Journal of the Canadian Health Libraries Association*, 27(3), 65–67.

Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science,* 37 (6), 357–376.

Beeson, I., and Chelin, J. (2006). Information systems meets information science. *ITALICS,* 5(2). Retrieved from www.ics.heacademy.ac.uk/italics/vol5iss2.htm (accessed May 1, 2012).

Broughton, V. (2006). *Essential thesaurus construction*. London: Facet.

Broughton, V., Hansson, J., Hjørland, B., and López-Huertas, M. J. (2005). Knowledge organization. In: *European curriculum reflections on library and information science*, 133–148. Retrieved from www.webcitation.org/5Vl9HJpm1 (accessed May 1, 2012).

BS 5723: 1987. (1987). *Guide to establishment and development of monolingual thesauri*. London: British Standard Institutions.

BS 6723: 1985. (1985). *Guidelines for the establishment and development of multilingual thesauri*. London: British Standards Institution.

BS 8723: 2005. (2005). *Structured vocabularies for information retrieval: Guide. Part 2. Thesauri*. London: British Standards Institution.

Capra, R., and Marchionini, G. (2007). Faceted browsing, dynamic interfaces, and exploratory search: Experiences and challenges. In: *Workshop on human-computer interaction and information retrieval: Workshop proceedings* (pp. 7–9). Retrieved from projects.csail.mit.edu/hcir/web/hcir07.pdf (accessed May 29, 2012).

Chamis, A. Y. (1991). *Vocabulary control and search strategies in online searching*. New York: Greenwood Press.

Cochrane, P. A. (1992). Indexing and searching thesauri, the Janus or Proteus of information retrieval. In: N. J. Williamson and M. Hudon (Eds.), *Classification research for knowledge organization*, FID, pp. 161–178.

Cutrell, E., Robbins, D. C., Dumais, S. T., and Sarin, R. (2006). Fast, flexible filtering with Phlat: Personal search and organization made easy. In: R. E. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Geffries, and G. Olson (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 261–270). Montreal, Canada.

Dextre Clarke, S. D. (2001). Thesaural relationships. In: C. A. Bean and R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 37–52). Boston: Kluwer.

Dextre Clark, S. D. (2008). The last 50 years of knowledge organization: A journey through my personal archives. *Journal of Information Science*, 34(4), 427–437.

Dong, A., and Agogino, A. M. (2001). Design principles for the information architecture of a SMET Education Digital Library. In: E. Fox and C. Borgman (Eds.), *Proceedings of the ACM/IEEE joint conference on digital libraries 2001* (pp. 314–321). New York: ACM Press.

Dumais, S. (2009). Faceted search. In: L. Liu and M. T. Özsu (Eds.), *Encyclopedia of database systems*. New York: Springer.

Education Resources Information Center. ERIC thesaurus. Retrieved from www.eric.ed.gov/ERICWebPortal/thesaurus/thesaurus.jsp (accessed May 1, 2012).

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, 30(4), 378–391.

Gey, F., Buckland, M., Chen, A., and Larson, R. (2001). Entry vocabulary: A technology to enhance digital object search. In: J. Allan (Ed.), *Proceedings of the first international conference on human language technology* (pp. 91–95). Stroudsburg, PA: ACM Press.

Gilchrist, A. (1971). *The thesaurus in retrieval*. London: Aslib.

Gilchrist, A. (2003). Thesauri, taxonomies, and ontologies: An etymological note. *Journal of Documentation*, 59(1), 7–18.

Golub, K., Jones, C., Lykke Nielsen, M., Matthews, B., Moon, J., Puzon, B., and Tudhope, D. (2009). EnTag: Enhancing social tagging for discovery. In: F. Heath, M. L. Rice-Lively, and R. Furuta (Eds.), *Proceedings of the joint conference on digital libraries (JCDL)* (pp. 163–172). New York: ACM.

Gray, A. J. G., Gray, N., Hall, C. W., and Ounis, I. (2010). Finding the right term: Retrieving and exploring semantic concepts in astronomical vocabularies. *Information Processing and Management*, 46(4), 470–478.

Gruber, T. (2009). Ontology. In: L. Liu and M. T. Özsu (Eds.), *Encyclopedia of database systems*. New York: Springer.

Hastings, S., Neal, D., Rorissa, A., Yoon, J., and Lyer, H. (2007). Social computing, folksonomies, and image tagging: Reports from the research front. Panel presentation. In: *Proceedings of the 2007 American Society for Information Science & Technology 70th annual meeting* (Vol. 45, pp. 1026–1029). Milwaukee, Wisconsin.

Hearst, M. A. (2000). Next generation web search: Setting our sites. *IEEE Data Engineering Bulletin*, 23(3), 38–48.

Hearst, M. A. (2006). Design recommendations for hierarchical faceted search interfaces. In: A. Z. Broder and Y. S. Maarek (Eds.), *Proceedings of the 29th annual international ACM SIGIR conference on research and*

development in information retrieval (SIGIR'06) workshop on faceted search (pp. 26–30). Seattle, Washington.

Hearst, M. A. (2008). UIs for faceted navigation: Recent advances and remaining open problems. In: *The workshop on human computer interaction and information retrieval, HCIR 2008*. Redmond, Washington.

Hearst, M. A. (2009). *Search user interfaces*. Cambridge, UK: Cambridge University Press.

Hearst, M. A., English, J., Sinha, R., Swearingen, K., and Yee, K. P. (2002). Finding the flow in web site search. *Communications of the ACM*, 45(9), 42–49.

Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Washington D.C.: Digital Library Federation. Retrieved from www.clir.org/pubs/reports/pub91/contents.html (accessed May 1, 2012).

Information Architecture Institute. (2005). Retrieved from www.iainstitute.org (accessed May 1, 2012).

International Atomic Energy Agency (IAEA). International Nuclear Information System (INIS) Collection. Retrieved from inis.iaea.org/search/default.aspx (accessed May 1, 2012).

ISO 2788: 1986. (1986). *Guidelines for the establishment and development of monolingual thesauri*. International Organization for Standardization.

ISO 5964: 1985. (1985). *Guidelines for the establishment and development of multilingual thesauri*. International Organization for Standardization.

ISO 25964-1: 2011. (2011). Information and documentation. *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. International Organization for Standardization.

Kekäläinen, J. and Jarvelin, K. (1998). The impact of query structure and query expansion on retrieval performance. In: W. B. Croft et al. (Eds.), *Proceedings of the Association for Computing Machinery Special Interest Group on Information Retrieval (ACM/SIGIR) 21st annual international ACM SIGIR conference on research and development in information retrieval 98* (pp. 130–137). Melbourne, New York: ACM Press.

Kipp, M. E. I. (2010). Convergence and divergence in tagging systems: An examination of tagging practices over a four year period. In: *Proceedings of the 2010 annual meeting of the American Society for Information Science and Technology*. Pittsburgh, Pennsylvania.(Conference Poster)

Knapp, S. D., Cohen, L. B., and Judes, D. R. (1998). A natural language thesaurus for humanities. *Library Quarterly*, 68 (4), 406–430.

Kristensen, J. (1993). Expanding end-users' query statements for free text searching with a search-aid thesaurus. *Information Processing and Management*, 29 (6), 733–744.

Kristensen, J., and Jarvelin, K. (1990). The effectiveness of a searching thesaurus in free text searching of a full-text database. *International Classification*, 17 (2), 77–84.

La Barre, K. (2004). Adventures in faceted classification: A brave new world or a world of confusion? In: I. C. McIlwaine (Ed.), *Advances in knowledge organization: Knowledge organization and the global information society* (Proceedings of the eighth international ISKO conference; pp. 79–84). Würzburg, Germany: Ergon Verlag.

Lopez-Huertas, M. J. (1997). Thesaurus structure design: A conceptual approach for improved interaction. *Journal of Documentation*, 53 (2), 139–177.

Lu, K., and Kipp, M. E. I. (2010). An experimental study on the retrieval effectiveness of collaborative tags. In: *Proceedings of the 2010 annual meeting of the American Society for Information Science and Technology*. Pittsburgh, Pennsylvania.

Lykke Nielsen, M. (1998). Future thesauri: What kind of conceptual knowledge do searchers need? In: W. M. El Hadi, J. Maniez, and S. Pollitt (Eds.), *Structures and relations in knowledge organization* (Proceedings of the 5th international ISKO conference; pp. 153–160). Würzburg, Germany: Ergon Verlag.

Lykke Nielsen, M. (2001). A framework for work task-based thesaurus design. *Journal of Documentation*, 57 (6), 774-797.

Macgregor, G., and McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5), 291–300.

Mandala, R., Tokunaga, T., and Tanaka, H. (2000). Query expansion using heterogeneous thesauri. *Information Processing & Management*, 36(3), 361–378.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 41–46.

McIlwaine, I. C. (2003). Trends in knowledge organization research. *Knowledge Organization*, 30(2), 75–86.

Miller, U. (2003). Thesaurus and new information environment. In: M. Drake and M. N. Maack (Eds.), *Encyclopedia of library and information science*, 2nd ed. Boca Raton: Taylor & Francis Group.

Milstead, J. L. (1998). Use of thesauri in the full-text environment. Retrieved from www.bayside-indexing.com/Milstead/useof.htm (accessed May 1, 2012).

Morville, P., and Callender, J. (2010). *Search patterns*. Sebastopol, CA: O'Reilly.

Morville, P. and Rosenfeld, L. (2007). *Information architecture for the World Wide Web: Designing Large-Scale Web Sites,* 3rd ed. Sebastopol, CA: O'Reilly.

mSpace Explorer. Retrieved from research.mspace.fm/projects/explorer (accessed May 1, 2012).

Networked Knowledge Organization Systems/Services (NKOS). Retrieved from nkos.slis.kent.edu (accessed May 1, 2012).

Olson, H. A. (2007). How we construct subjects: A feminist analysis. *Library Trends*, 56(2), 509–541.

Pastor-Sanchez, J. A., Martinez, F. J., and Rodriguez, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14(4), paper 422. Retrieved from InformationR.net/ir/14-4/paper422.html (accessed May 1, 2012).

Perez, E. (1982). Text enhancement: Controlled vocabulary vs. free text. *Special Libraries*, 73(July), 183–192.

Piternick, A. (1984). Searching vocabularies: A developing category of online searching tools. *Online Review*, 8(5), 441–449.

Pollitt, A. S., Ellis, G. P., and Smith, M. P. (1994). HIBROWSE for bibliographic databases. *Journal of Information Science*, 20(6), 413–426.

Project ISO 25964. (2012). Thesauri and interoperability with other vocabularies. Retrieved from www.niso.org/workrooms/iso25964 (accessed May 1, 2012).

Ranganathan, S. R. (1967). *Prolegomena to library classification*. New York: Asia Publishing House.

Rosenfeld, L., and Morville, P. (1998). *Information architecture for the World Wide Web: Designing Large-Scale Web Sites*. Sebastopol, CA: O'Reilly.

Saumure, K., and Shiri, A. (2008). Knowledge organization trends: A comparison of the pre- and post-web eras. *Journal of Information Science*, 34(5), 651–666.

Schatz, B. R., Johnson, E. H., and Cochrane, P. A. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In: E. Fox and G. Marchionini (Eds.), *Proceedings of the 1st Association for Computing Machinery international conference on digital libraries* (pp. 126–133). Bethesda, MD: ACM Press.

Schwartz, C. (2008). Thesauri and facets and tags, oh my! A look at three decades in subject analysis. *Library Trends*, 56(4), 830–842.

Shiri, A. (2009). An examination of social tagging interface features and functionalities: An analytical comparison. *Online Information Review*, 33(5), 901–919.

Shiri, A. A., and Revie, C. (2000). Thesauri on the web: Current developments and trends. *Online Information Review*, 24(4), 273–279.

Soergel, D. (1999). The rise of ontolgoies or the reinvention of classification. *Journal of the American Society for Information Science*, 50(12), 1119–1120.

Soergel, D. (2003). Functions of a thesaurus/classification/ontological knowledge base. Retrieved from ontolog.cim3.net/file/work/Ontologizing Ontolog/TaxoThesaurus/SoergelKOSOntologyFunctions2—Dagobert Soergel_20060616.pdf (accessed May 1, 2012).

Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, 4(2), Article 41. Retrieved from www.webology.org/2007/v4n2/a41.html (accessed May 1, 2012).

Tonkin, E. (2006, April 30). Folksonomies: The fall and rise of plain-text tagging. *Ariadne*, (47). Retrieved from www.ariadne.ac.uk/issue47/tonkin (accessed May 1, 2012).

Tudhope, D., and Binding, C. (2008). Faceted thesauri. *Axiomathes*, 18(2), 211–222.

UNISIST (1980). *Guidelines for the establishment and development of multilingual thesauri*, rev. ed. Paris, UNESCO.

UNISIST (1981). *Guidelines for the establishment and development of monolingual thesauri,* 2nd ed. Paris, UNESCO.

U.S. National Library of Medicine. (2011). Medical Subject Headings (MeSH). Retrieved from www.nlm.nih.gov/mesh (accessed May 1, 2012).

Vickery, B. C. (1960). Thesaurus—A new word in documentation. *Journal of Documentation*, 16(4), 181–189.

Wang, Z., Chaudhry, A. S., and Khoo, C. S. (2008). Using classification schemes and thesauri to build an organizational taxonomy for organizing content and aiding navigation. *Journal of Documentation*, 64(6), 842–876.

White, R. W., Kules, B., Drucker, S. M., and Schraefel, M. C. (2006). Supporting exploratory search. *Communications of the ACM*, 49(4), 36–39.

White, R. W., and Roth, R. A. (2009). *Exploratory search: Beyond the query-response paradigm*. San Rafael, CA: Morgan & Claypool.

Williamson, N. (2000). Thesauri in the digital age: Stability and dynamism in their development and use. In: C. Beghtol, L. C. Howarth, and N. Williamson (Eds.), *Proceedings of the sixth international ISKO conference* (pp. 268–274). Germany: Ergon Verlag.

Williamson, N. (2007). Knowledge structures and the internet: Progress and prospects. *Cataloging & Classification Quarterly*, 44(3/4), 329–342.

Wodtke, C., and Govella, A. (2009). *Information architecture: Blueprints for the web*. Berkeley, CA: New Riders.

WorldCat. Retrieved from www.worldcat.org (accessed May 1, 2012).

Yee, K., Swearingen, K., Li, K., and Hearst, M. (2003). Faceted metadata for image search and browsing. In: G. Cockton and G. Korhonen (Eds.), *Proceedings of the ACM conference on human factors in computing systems* (pp. 401–408). New York: ACM Press.

Yelp. Retrieved from www.yelp.com (accessed May 1, 2012).

Zhang, J., and Marchionini, G. (2005). Evaluation and evolution of a browse and search interface: Relation Browser++. In: L. Delcambre and G. Giuliano (Eds.), *Proceedings of the 2005 national conference on digital government research* (pp. 179–188). Marina del Rey, CA: Digital Government Society of North America.

Zhang, X., Strand, L., Fisher, N., Kneip, J., and Ayoub, O. (2002). Information architecture as reflected in classrooms. In: *Proceedings of the American Society for Information Science and Technology annual meeting* (pp. 78–82). Philadelphia, Pennsylvania.