Generation and application of faceted vocabulary for knowledge discovery on the Web: an exploratory study

Kiduk Yang, & Elin K. Jacob
Indiana University
1320 East 10th Street, Bloomington, Indiana, USA
kiyang@indiana.edu, ejacob@indiana.edu

Key Words

Faceted Vocabulary
Knowledge Discovery
Automatic Classification
Concept Search
Concept Relationships

Abstract

We explored generation and application of faceted vocabulary as a potential mechanism to enhance knowledge discovery on the Web. Generation of faceted vocabulary involved examination of concept pairs extracted from an existing hierarchical classification structure, while application of faceted vocabulary involved both user interface design as well as fusion of retrieval and classification approaches. This paper describes our approach to leveraging existing classification hierarchy to construct and utilize a flexible classification scheme, where users can engage in knowledge discovery process via search and traversal of dynamic concept relationships.

Introduction

John Dewey, an American philosopher, said, “Knowledge is classification.” Langridge (1992) states that without classification there could be no human thought, action, and organization. Classification, in a broad sense, is a mechanism for both organizing and utilizing information by representing knowledge as a set of concepts and relationships. In that light, transformation of information into knowledge via classification is a comprehensive process that involves not only the construction of classification structure (i.e. a set of concepts and relationships) and categorization of information units (e.g. cataloging, automatic classification), but also the utilization of classification data for knowledge discovery.

Not surprisingly, librarians for centuries have been utilizing various classification approaches (e.g. LCC, DDC) to bridge the library patron’s information need and the library collection. The success of computerized full-text search dampened the popularity of classification-based approach to information discovery for the past few decades, but classification has once again become a prevalent concept that is reflected across the Web landscape, from Web directories (e.g. Yahoo!) and metadata initiatives to Semantic Web and Digital Library movements. There exists, however, many challenges in applying classification-based approaches to the Web. For instance, it is difficult to “organize” the whole Web due to its massive size and diversity. Even if such a feat were possible, most clustering approaches, which are not incremental, and text categorization approaches, which are based on a static classification scheme, will not be able to deal with the dynamic nature of the Web corpus. Consequently, methods of organizing Web information need to be efficient, flexible and dynamic. Moreover, post-retrieval organization of retrieved documents may be a more desirable as well as realistic approach than trying to organize the entire Web.

Based on our belief that dynamic and adaptive nature of faceted classification is well-suited for the Web, we conducted an exploratory study that investigated generation and application of faceted vocabulary as a potential approach for knowledge discovery on the Web. Construction of faceted vocabulary, which involves examination of concept pairs extracted from an existing hierarchical classification structure[1], as well as application of faceted vocabulary and concept relationships for retrieval and knowledge discovery, which involves both user interface design and fusion of retrieval and classification approaches, are a part of ongoing research to lay the groundwork for the Classification-based Search and Knowledge Discovery (CSKD) project[2]. In this paper, we describe our work-in-progress explorations of classification-based approach to knowledge discovery. The rest of paper is organized as follows. First, we start with a general discussion of faceted systems of representation, which is followed by the description of our faceted vocabulary construction process and the presentation of the faceted search interface.

Faceted Systems of Representation

A traditional classification scheme is a system of representation that attempts to enumerate all the knowledge of a given domain -- past, present and future -- within a fixed set of static and predefined classes typically arranged as a hierarchical tree structure. Such an enumerative scheme is a well-constructed language (Foucault, 1970) because it both defines and orders the domain of interest: it defines the domain by standardizing a representational vocabulary that

---

[1] The concept pairs were extracted from the EPA Topics (http://www.epa.gov/epahome/topics.html), which is a classification scheme used by the United States Environmental Protection Agency (EPA).
[2] CSKD is a long term project at Indiana University that investigates methods of leveraging both the ontological and link-structural knowledge embedded in classified corpora of Web documents for searching and organizing the Web.

prescribes for each class term both an *intension* (a definition) and an *extension* (a set of referents); and it orders the domain by establishing relationships between entities, whether explicitly through the medium of definition or implicitly through the construction of the hierarchical structure. Because it creates a one-to-one relationship between a class term and the individual concept the term represents, the traditional classification scheme functions as a controlled vocabulary and facilitates the transfer of knowledge across time and space without loss of information (Jacob, 1994). Furthermore, in its role as a controlled vocabulary, the enumerative scheme serves to legitimize and reify a single ideological or sociopolitical perspective of the domain (Jacob & Albrechtsen, 1997).

There is growing recognition among classification theorists that an enumerative classification scheme may not be the most effective access system for all users. On the one hand, because the users of a resource collection will frequently represent different discourse communities, knowledge domains or levels of expertise, they may not share the same understanding of the scheme's vocabulary (Hjørland & Albrechtsen, 1995; Jacob & Albrechtsen, 1997; Williamson, 1998). On the other hand, given the increasingly interdisciplinary nature of knowledge -- an interdisciplinarity "characterized by instability, lack of predictability, and spontaneous response to politically, socially and environmentally based issues" (Williamson, 1998, p. 118) -- the application of a single, universal system of organization may actually inhibit access to resources. Because the relationships between classes in an enumerative scheme are predetermined, the rigidity inherent in such a structure is incapable of representing meaningful associations between phenomena that are conceptually related but do not occur within the same branch of the hierarchy. As Williamson (1998) points out, when the hierarchical structure of the classification scheme is organized by discipline, aspects of a subject are frequently scattered across the class structure; but when the structure is organized by phenomena, the disciplines themselves are scattered.

The inability to represent relationships other than those created by the nesting structure of the hierarchy renders traditional enumerative schemes such as the Library of Congress Classification less than effective for organizing resources in the diverse and multidisciplinary environment of the World Wide Web. A highly variable and dynamic environment such as the Web demands an organizational approach that not only provides flexibility of representation but can also accommodate the dynamic nature of human knowledge itself. At the same time, such an approach should be able to respond to the immediate information needs of a highly diverse and increasingly interdisciplinary population.

Recognizing the inherent rigidity of traditional enumerative structures, Ranganathan (1944, 1945) proposed a more flexible approach to classification. His idea was to identify the various aspects (or *facets*) of a domain so as to derive a set of independent concept hierarchies that would represent the range of characteristics relevant to that domain, including, for example, such characteristics as activities, operations, space and time. Each of these concept hierarchies would be populated by the set of possible values (or *isolates*) that could be applied to describe that particular characteristic. Isolates from these modular concept hierarchies could then be combined to generate only those classes required to represent the actual materials in the collection. And the combination of isolates according to an established citation order would assure collocation of related resources within a dynamically-generated hierarchy. Thus, in contrast to enumerative classification schemes, construction of a faceted representational system neither prescribes a finite set of classes. nor predetermines the relationships among classes. Rather, it establishes control over the formal semantics underlying the scheme and, in so doing, provides a conceptual basis both for formation of classes and for establishment of relationships among the classes that will comprise the faceted classification scheme.

The fundamental organizing principle underlying the development of a faceted system is the grouping of that which is related and the separation of that which is unrelated. Unlike the fixed structure of classes produced by enumeration, faceting provides for the organization of concepts in modular hierarchies by splitting (separating) unrelated or dissimilar concepts and lumping (grouping) related or similar concepts. Relevant concepts are identified by partitioning domain terminology into mutually-exclusive baseline facets (Priss & Jacob, 1999) whose internal composition and external relationships reflect the theoretical and/or practical perspectives of the domain being modeled. These baseline facets may then be combined to form higher-order facets. This bottom-up approach to construction of the faceted system provides for consistency and structure within each facet while assuring that the connections established between facets remain flexible and adaptable to context and usage. Values from these modular hierarchies can then be combined using a predetermined citation order -- an algorithm establishing the order for combining concept values that allows for the creation of only those classes which are required by a particular collection of resources while maintaining the consistency across class representations that is necessary to generate a hierarchical tree structure.

A faceted system can be used to create a classificatory structure that is tailored to the intellectual content of the resource collection. However, because it prescribes a fixed combinatorial order, a faceted classification scheme, like an enumerative scheme, decrees both the set of established access points and an explicit relational structure which may not serve the information needs of individual users. It is important to recognize, then, that the value of a faceted system is in the creation of a representational vocabulary whose flexibility and adaptability is inherent in the very modularity of the vocabulary itself. Having constructed a faceted system, application of the vocabulary is not limited to the creation of a classification scheme. Rather, a faceted vocabulary can be adopted as a post-coordinate indexing language or can be used to develop pre-coordinate indexing chains and subject headings. A faceted scheme can also be adapted to provide dynamic class structures capable of responding to the individual's immediate information needs by supporting user-generated ordering of facets. This offers the potential for flexible reconfiguration of the organizational structure capable of identifying new relationships between resources and thus accommodating a broader range of information needs than traditional enumerative schemes or faceted schemes using a fixed citation order (Jacob & Priss, 1999). For example, an educational scheme whose facets consist of *grade level*, *subject* and *instructional method* might adopt the citation order *subject* + *instructional method* + *grade level*, thereby providing access only through specific subject areas -- access that is of little use to teachers interested in the specifics of a particular instructional method or the various instructional methods that are best applied at a given grade level. But an access system that allows for user-generated combination of facets would provide access based on subject, instructional method or grade level while supporting dynamic reconfiguration of the organizational structure that could lead to the identification of new knowledge through the creation of new relationships among these concepts.

The typical approach to developing a faceted vocabulary involves identification and grouping of relevant values in an iterative process of inductive or bottom-up clustering of concepts (Batty, 1989). Initial clusters are aggregated, where feasible, into progressively more comprehensive groupings that will identify general concepts and provide the initial set of baseline facets (Priss & Jacob, 1999). Ideally, the internal composition of these baseline facets and the external relationships among them will reflect the basic ontological commitments of the domain being modeled -- the entities, activities, properties and attributes that characterize a domain's intellectual content and the relations between them that are most relevant to the domain. These baseline facets may then be combined to form superordinate facets. At each level in the hierarchy, each facet is internally consistent and conceptually complete; but it is flexible and open to modification, both in its composition and in its external relationships with other facets. Thus, for example, the facet color may be conceptually complete in that it includes values (isolates) for all colors of relevance to a particular domain: for example, *red*, *blue*, *green* and *yellow*. Should changes in the domain require introduction of *grue* as an isolate of the facet *color*, it would simply be inserted in the internal ordering of the facet between *blue* and *green*. Further changes in representation of the domain might necessitate the specification of color for objects whose color was previously negligible or irrelevant, a modification that could be readily accommodated by simply inserting the facet *color* in the appropriate slot in the citation order. In this manner, a faceted structure of concepts and concept values provides consistency of representation and coherence of structure within individual facets, while connections between facets remain adaptable to context and usage (Priss & Jacob, 1998).

Although it is difficult, if not impossible, to harness the inherent flexibility of faceted systems outside a computer-based medium, the application of faceted vocabularies holds potentially significant implications for the extension of user access within the Web environment. Unfortunately, these systems have not been exploited because of the perceived complexity of the faceted system itself. Thus widespread development and application of faceted systems of representation and organization has been forestalled by the intellectual effort required both to generate the faceted vocabulary and to index resources using a faceted approach. However, the very fact that a faceted system can be used to generate subject heading systems and chained index strings points to the possibility of employing existing representational systems to simplify and potentially automate the process of constructing faceted vocabularies. This paper reports on the development of a methodology for using an existing representational structure to seed the semi-automatic construction of a faceted vocabulary that can then be superimposed on the original structure to enhance retrieval performance.

Constructing the Faceted Vocabulary

The process of constructing a faceted classification scheme is generally described as *analytico-synthetic*. That is, the terms that populate a domain of knowledge are analyzed to determine the central concepts that will provide the initial set of facets -- the entities, activities, properties and attributes that characterize the domain's intellectual content and comprise the baseline facets (Priss & Jacob, 1999). Having specified a set of baseline concepts by grouping similar values (paradigmatic variables or *isolates*) into mutually-exclusive sets, all relevant values are identified for each facet (Priss & Jacob, 1998). These baseline facets are then grouped into more comprehensive concept hierarchies, where possible. With this faceted structure in hand, a representation can then be created (or *synthesized*) for each resource in the collection by combining the appropriate isolates from these concept hierarchies according to a predetermined ordering of elements which may or may not correspond to specific facets in the scheme. For example, in the citation order Personality-Matter-Energy-Space-Time adopted by Ranganathan, the first three elements do not correspond to specific facets but indicate, instead, the functional role of the facet value (the isolate) within the resource representation.

Because construction of a faceted structure is assumed to begin with the collection and subsequent grouping of linguistic terms specific to a given domain (Batty, 1989), the process is generally described as "bottom-up" in order to distinguish it from the "top-down" process of division employed in the creation of enumerative classification schemes. The creation of a faceted system necessarily begins with analysis of the linguistic vocabulary of the associated domain; but such analysis may not be effective if executed within a vacuum. Rather, analysis of domain content should be carried out within the existing conceptual framework of the domain, utilizing both inductive "bottom-up" and deductive "top-down" strategies (Priss &

Jacob, In press). Because the conceptual framework of the domain – the underlying paradigm or ontological commitment of the domain -- influences the way in which domain phenomena are to be conceptualized and subsequently organized, the semantic value of an individual term-as-concept cannot be fully appreciated without understanding its relationship(s) to the range of terms-as-concepts which comprise the domain. By beginning with a top-down investigation of the conceptual framework of the domain, subsequent analysis of domain terms and term relationships will be better able both to identify the most relevant concepts that will constitute the initial set of baseline facets and to establish the meaningful relationships that obtain between those facets (Priss & Jacob, 1999). Thus the first step in creation of a faceted vocabulary is necessarily "middle-out" in that it combines bottom-up acquisition of the linguistic base with top-down analysis of the domain's conceptual framework.

To assess whether this middle-out approach could be adapted to automate the process of facet generation from an existing enumerative system, it was decided to begin the process of constructing a faceted vocabulary by identifying a lexicon of concepts from a representational system currently used to index a collection of Web documents. The representational system selected for the current project was EPA Topics (http://www.epa.gov/epahome/topics.html), an indexing scheme used by the United States Environmental Protection Agency [EPA] to provide access to a collection of high-quality resources dealing with a range of environmental issues. EPA Topics is composed of 825 separate categories nested within a three-level hierarchical tree structure. The first level consists of 17 main categories (e.g., *Air*, *Cleanup*); the second level consists of 167 categories (e.g., *Air Pollutants*, *Air Pollution*); and the third level consists of 641 categories (e.g., *Aerosols*, *Asbestos*). The primary collection of resources indexed by EPA Topics includes 3080 documents that have been manually assigned to one or more of the categories in the EPA Topics structure[3]. In addition, over 37,000 additional resources have been indexed with manually-assigned descriptors. It is important to note that EPA Topics is not a true classification scheme: not all categories are mutually exclusive and any concept or category may be nested within more than one branch of the hierarchical tree structure. However, this representational system does provide a systematic ordering of nested categories with each category represented by a chain of descriptors (single terms and noun phrases) that indicate its position within the larger conceptual structure.

The first step in generating the faceted scheme involved the inductive, bottom-up creation of a primary lexicon base consisting of all unique, information-bearing terms in the set of EPA Topics descriptors used as category labels. For example, the category label *Compliance and Enforcement* is a descriptor that consists of the unique terms *Compliance* and *Enforcement*. To assess the conceptual framework of the domain and its influence on how domain phenomena were conceptualized and subsequently organized, all pairs of descriptors were generated in order to establish the broader context within which each unique term occurred. For example, the third-level category *Administrative Penalties* is nested within the second-level category *Civil Enforcement*, which is nested within the first-level category *Compliance and Enforcement*. This nested structure yields three pairs of representational strings: *Administrative Penalties* and *Civil Enforcement*; *Civil Enforcement* and *Compliance and Enforcement*; and *Administrative Penalties* and *Compliance and Enforcement*.

The process of identifying descriptor pairs was executed automatically and generated 1368 unique pairings. These 1368 pairs were then parsed to identify the 658 unique descriptors, composed of either single words or noun phrases, which occurred in the union of all descriptors. The analysis of unique terms within the context of their associated descriptors and descriptor pairs serves to identify unique concepts by establishing the context within which each term occurs. For example, the concept *penalty* is represented by two unique terms

[3] EPA category statistics are based on the crawl of EPA Topics site on 11/20/2003.

(*penalty* and *penalties*) that are found in association with only five other unique terms (*Enforcement*, *Compliance*, *Administrative*, *Civil* and *Settlement*) but occur in seven unique descriptors (*Administrative Penalties*; *Civil Penalties*; *Compliance Penalties*; *Compliance Penalty*; *Enforcement Penalties*; *Enforcement Penalty*; and *Settlement Penalty*).

Manual analysis of the automatically generated lexicon base within the conceptual framework provided by the associated terms and concept pairs allows specification of the context within which an individual concept occurs. Manual analysis also points to the lack of consistency in the existing representational system that, undetected, might undermine efforts to construct the faceted vocabulary. However, the more important implication of this analysis is the need for a hybrid approach to the construction of faceted vocabularies -- an approach that combines automated specification of context with the conceptual rigor provided by manual analysis of terms and term relationships in the identification of concepts and their subsequent grouping into facets.

Applying the Faceted Vocabulary

Despite the success and popularity of Web search engines in recent years[4], full-text search does not adequately address some basic needs of information seekers. Information retrieval approach of the full-text search focuses on finding information that are likely to be “relevant” to a given query, but provides little assistance in bridging the gap between the “anomalous state of knowledge” (Belkin, 1982) and formulation of an effective query. Furthermore, retrieval-based approach to information discovery does not identify important concepts, let alone concept relationships, that may be useful in satisfying a more complex information need than finding specific information, thus leaving the task of knowledge discovery mostly to the user.

One of the ways classification-based approach to information discovery can address the shortcomings of retrieval-based approach is by letting the user browse a classification structure representing important concepts and relationships, where he/she can quickly gain an overview of information landscape that may facilitate the information discovery process. The static hierarchical structure of conventional classification approach, however, can sometimes hinder information discovery process by confusing, misleading, or restricting the user with its rigidity and complexity. Faceted classification structure, on the other hand, facilitates flexible representation of knowledge using a set of concepts and relationships that can be structured dynamically to accommodate individual user’s needs and perspectives. In addition, faceted vocabulary can be used in conjunction with full-text search as well as in complementing conventional classification-based approach. For example, facets can help both query refinement and focused taxonomy traversal by restricting the information domain.

While we investigated the methods of faceted vocabulary construction, we explored in parallel the utilization of classification data for knowledge discovery by experimenting with a “concept search” application that can effectively combine the strengths of full-text search, concept hierarchy and faceted vocabulary. The focus of concept search application is in efficient and effective identification of important concepts and concept relationships that are useful in fulfilling user’s information need. Fully implemented concept search application will consist of a Web directory harvester, which crawls and harvests an existing Web document classification structure, a faceted vocabulary construction component, which constructs a faceted vocabulary from the harvested classification structure, an automatic facet classifier, which maps faceted vocabulary to Web pages, a fusion retrieval module, which

---

[4] In February 2003, there were 670 million Web searches per day reported by six major search engines (Search Engine Watch, http://www.searchenginewatch.com).

integrates text- and classification-based retrieval, and a concept search interface, which combines searching and browsing approaches to information discovery. Unfortunately, we were not able to fully implement the concept search application for this paper due to the exploratory nature of our study that called for incomplete faceted vocabulary construction, which is to be iteratively refined in subsequent studies. Consequently, what follows is a description of partially implemented concept search application that prototypes the main idea of knowledge discovery by integrated approach to concept and concept relationship identification.

Concept Search Application

EPA, whose hierarchical classification structure was used as data source in our study, employs both browsing and searching (Figures 1a, 1b, & 2) to utilize its classification data. The browse interface, called "Browse EPA Topics", implements a conventional hierarchical tree navigation similar to Yahoo!, and the search interface, called "Try-Me-First Search", implements a layered search that returns up to four sets of search results returned from matching the query to different sources of evidence (i.e. Topics category labels, classified EPA pages, controlled vocabulary metadata, full-text index)[5]. The concept search application developed in our study combines browsing and searching in an integrated interface that emphasizes concept and concept relationship identification (Figures 3a, 3b, & 3c).

The initial search screen of the concept search (Figure 3a) allows the user to search either single word or phrase concept pairs extracted from EPA Topics, as well as providing query refinement option by facets (e.g. Environment, Location)[6]. The presence of facets in the initial search screen can also serve as a guide to identifying key concept attributes (e.g. *air*, *soil*, *water* for "pollution"). The initial concept search result display consists of concept pairs that include query terms, EPA Topic hierarchy where those concept pairs occur in, and related concept terms and their occurrence frequencies in EPA Topics (left two columns of Figure 3b). Clicking the "related concepts", which are extracted from matched concept pairs and displayed in a descending order of frequency, will execute an initial concept search using the clicked concept to display the results in the rightmost column. Both concept pair displays, the original query results in upper left frame and related concept query results in upper right frame, provide checkbox options that allow the user to select concept pairs of interest and display associated Web pages (Figure 3c).

The following example demonstrates a potential concept search session. A user, who heard about the dangers of carbon monoxide emission from fireplace but has only a vague recollection, enters "pollution" in the initial search box. After examining the concept search results, where the "Location" facet isolates are displayed at the top of the *Related Concepts* column, or simply after viewing the isolates in the dropdown box of the "Location" facet in the initial search screen, the user submits the "pollution" query refined with the "Location" facet value of "air". User then sees the "Indoor Air Pollution" ranked third in *Related Concepts* and clicks it to display a concept pair list of indoor air pollution in the right column (Figure 3b), where he selects "Fireplaces" and "Carbon Monoxide" to display the documents of interest (Figure 3c). As illustrated in the previous example, concept search employs faceted vocabulary to assist query formulation and refinement process as well as leveraging existing classification structure to provide context for query terms and identify concept relationships, thus facilitating both the search and knowledge discovery process.

---

[5] See http://www.epa.gov/epahome/searchglossary.html for more information on Try-Me-Search.

[6] Only two facets are integrated in the initial concept search for demonstration purposes.

Discussion

In this study, we explored generation and application of faceted vocabulary as a potential mechanism to enhance knowledge discovery on the Web. Our faceted vocabulary construction process revealed some heuristics that can be refined in follow-up studies to further automate the creation of faceted classification structure, and concept search application demonstrated the utility and potential of integrating classification-based approach with retrieval-based approach. Despite its exploratory nature, our study outlined an approach to leveraging existing classification hierarchy to construct and utilize a flexible classification scheme, where users can engage in knowledge discovery process via search and traversal of dynamic concept relationships. In follow-up studies, we will streamline the faceted vocabulary construction process by iteratively refining the facet identification heuristic, fully implement the concept search application as described in the previous section, and conduct a user study to evaluate both the faceted vocabulary construction and application processes.

Integration of text- and classification-based methods as outlined in this paper combines the strengths of two vastly different approaches to information discovery by constructing and utilizing a flexible information organization scheme from an existing classification structure. Concept search application, which is designed to highlight important concepts and identify relationships among them, focuses on concept rather than document discovery, and help users make sense of the document collection by dynamically organizing concepts according to individual user's information needs and requirements.

Figure 1a: Browse EPA Topics Interface (Homepage)

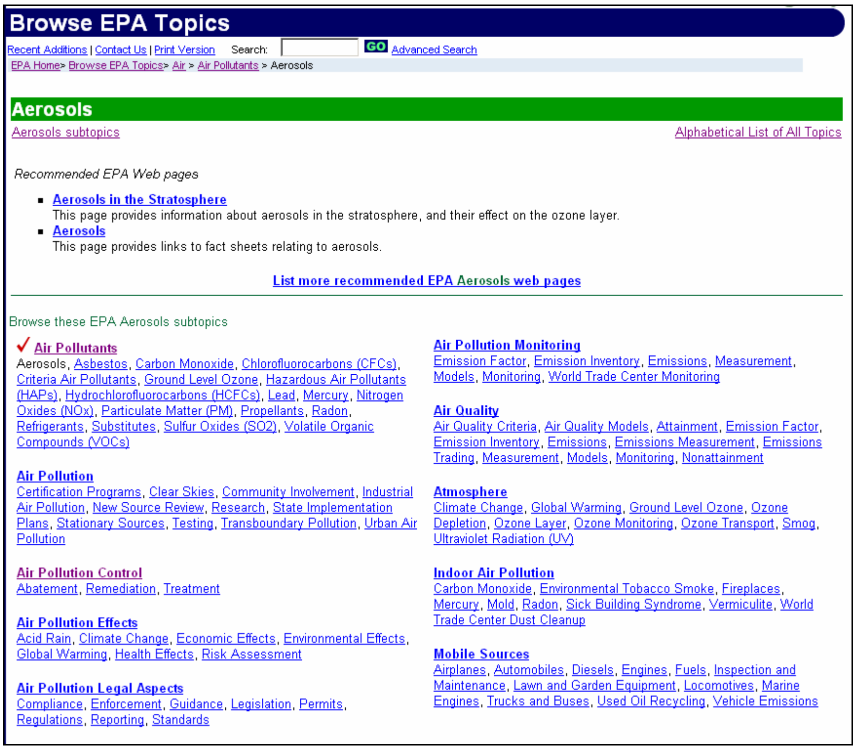

Figure 1b: Browse EPA Topics Interface

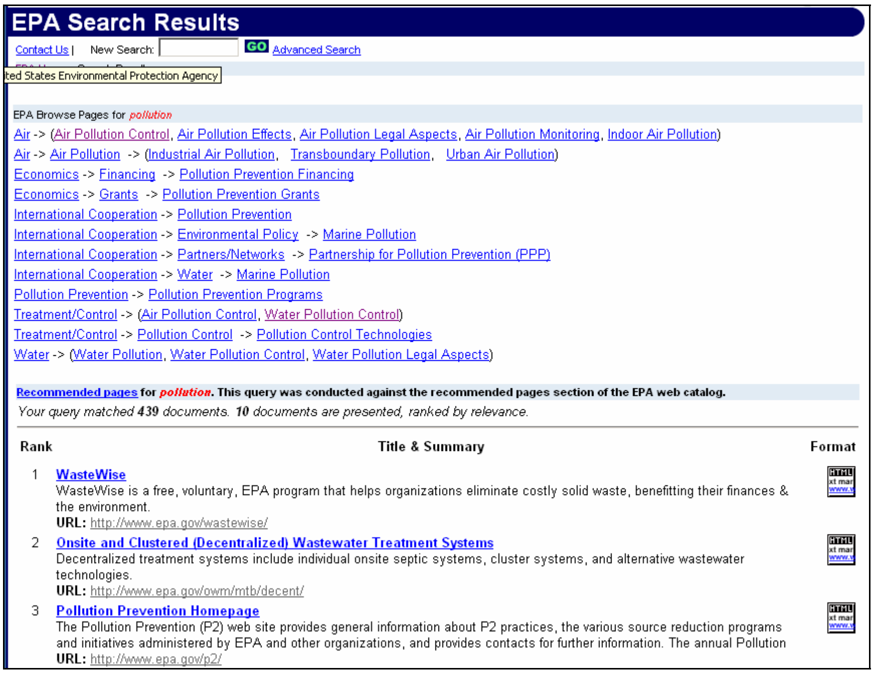

Figure 2: EPA Search Interface

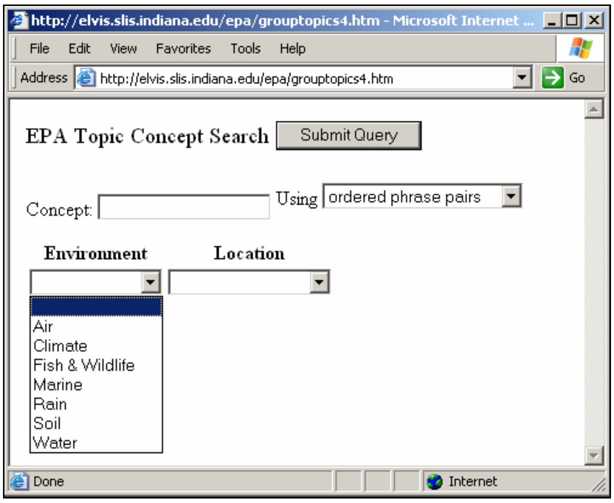

Figure 3a: Concept Search Interface (Homepage)

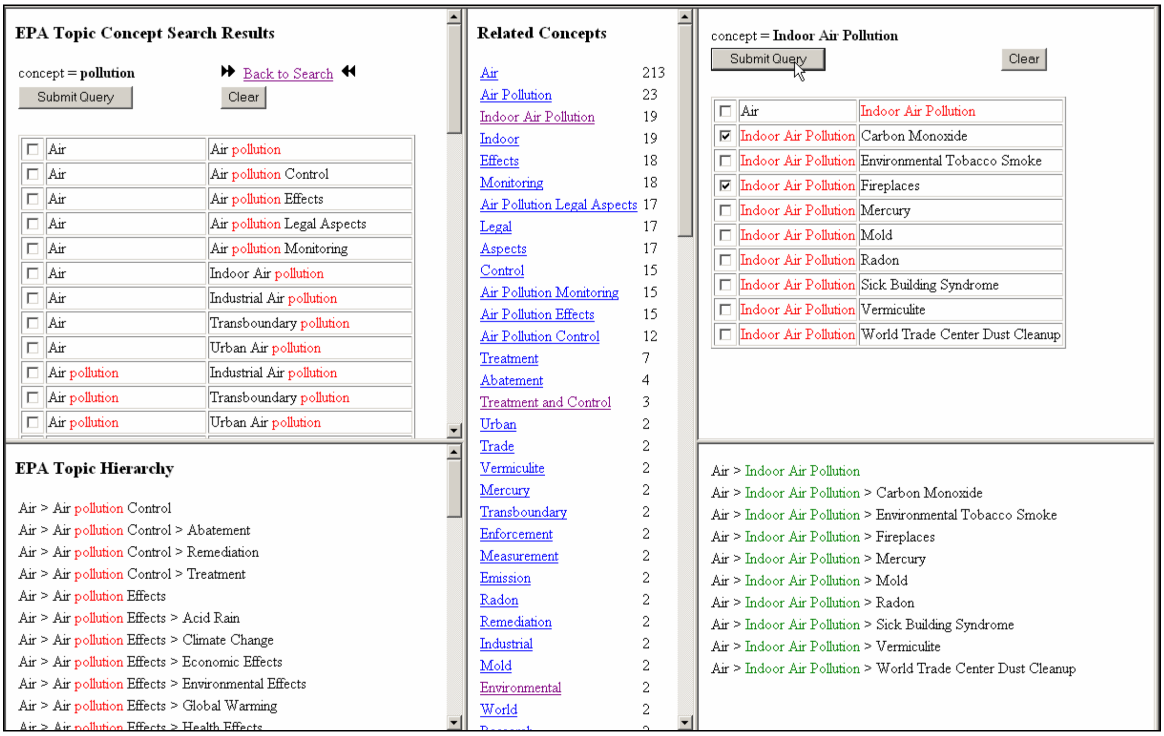

Figure 3b: Concept Search Interface (Concepts Display)

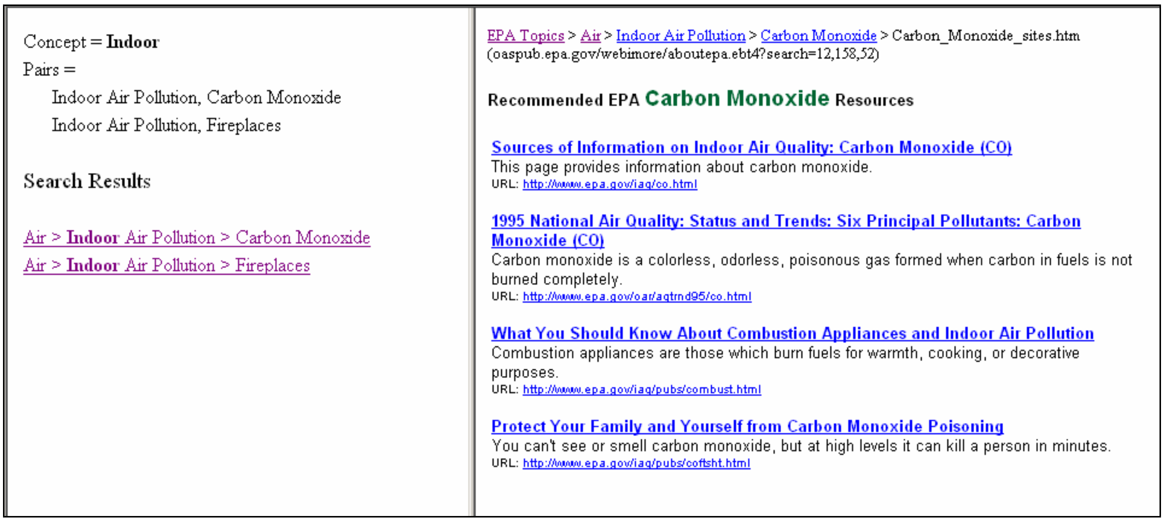

Figure 3c: Concept Search Interface (Search Result Display)

References

Albrechtsen, H., & Jacob, E.K. (1998). Classification systems as boundary objects in diverse information ecologies. In E. N. Efthimiadis (Ed.), *Advances in classification research, vol. 8* (pp. 1-16). Medford, NJ: Information Today for the American Society for Information Science.

Batty, D. (1989). Thesaurus construction and maintenance: a survival kit. *Database 12* (1), 13-20.

Belkin, N. J. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, *38*, 61-71.

Foucault, M. (1970). *The order of things: an archaeology of the human sciences*. New York: Vintage Books.

Han, E. & Karypis, G. (2000). Centroid-Based Document Classification: Analysis & Experimental Results. *Proceedings of the 6th ACM SIGKDD International Conference Knowledge Discovery and Data Mining*.

Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science, 46*(6), 400-425.

Jacob, E. K. (1994). Classification and crossdisciplinary communication: breaching the boundaries imposed by classificatory structure. In H. Albrechtsen and S. Oernager (Eds.), *Knowledge organization and quality management: Advances in knowledge organization, vol. 4* (pp. 101-108). Frankfurt/Main: Indeks Verlag.

Jacob, E.K., & Albrechtsen, H. (1997). Constructing reality: the role of dialogue in the development of classificatory structures. In I. C. McIlwaine (Ed.), *Knowledge organization for information retrieval*: *Proceedings of the 6th International Study Conference on Classification Research, 14-16 June 1997, London* (pp. 42-50). The Hague, Netherlands: International Federation for Documentation.

Jacob, E.K. & Priss, U. (1999). Application of faceted classification structures in electronic knowledge resources. In H. Albrechtsen and J.-E. Mai (Eds.), *Proceedings of the 10th ASIS SIG/CR Classification Research Workshop* (pp. 87-106). Silver Spring, MD: American Society for Information Science.

Jacob, E.K., & Priss, U. (2001). Non-traditional indexing structures for the management of electronic resources. In H. Albrechtsen & J.-E. Mai (Eds.), *Advances in classification research, vol.10* (pp. 73-90). Medford, NJ: Information Today for the American Society for Information Science.

Langridge, D.W. (1992). Classification: Its kinds, elements, systems, and applications. London: Bowker Saur.

Priss, U., & Jacob, E.K. (1998). A graphical interface for faceted thesaurus design. In E.K.

Jacob (Ed.), *Proceedings of the 9th ASIS SIG/CR Classification Research Workshop* (pp. 107-118). Silver Spring, MD: American Society for Information Science.

Priss, U., & Jacob, E.K. (In press). Conceptual representation of faceted thesauri. In E.K. Jacob (Ed.), *Advances in classification research, vol.9.* Medford, NJ: Information Today for the American Society for Information Science.

Ranganathan, S. R. (1945). *Elements of library classification: based on lectures delivered at the University of Bombay in December 1944*. Poona: N.K. Publisheing House. [Kaikhushru Taraporevala memorial series in library science, No. 4].

Ranganathan, S.R. (1944). *Library Classification: fundamentals and procedures with 1008 graded examples and exercises*. Madras: Madras Library Association.

Williamson, N. (1998). An interdisciplinary world and discipline based classification. In W. M. el Hadi, J. Maniez, & S. A. Pollitt (eds.), *Structures and relations in knowledge organization: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France* (p. 116-124). Würzburg, Germany: Ergon Verlag.