No (e-)Democracy without (e-)Knowledge

Giovanni M. Sacco

Dipartimento di Informatica, Università di Torino, Corso Svizzera 185,
10149 Torino, Italy
sacco@di.unito.it

Abstract. Citizens have never had complete and up-to-date information on all the laws, regulations and opportunities that concern them. Although the law does not excuse ignorance, information-publishing techniques, usually Official Gazettes on paper, make effective knowledge and awareness virtually impossible. One of the greatest opportunities of e-government is to overcome this information gap and to supply timely and complete information to everybody. The electronic availability of information is but a minor aspect of this problem. Rather, effective, timely and accurate ways of disseminating information must be found. We discuss several solutions, ranging from text retrieval to ontologies and agents, and focus on dynamic taxonomies, a model recently proposed for the intelligent exploration of heterogeneous information bases, that can provide guided browsing and personalized exploration for complex laws and regulations.

1 Introduction

Ignorantia legis non excusat - ignorance of the law does not excuse - is a centuries-old criminal law maxim familiar to everybody: all are presumed to be familiar with all the laws that concern them or face the costs of their ignorance. This principle usually extends to civil law and regulations in the large. Nowadays, European laws and regulations exist at several levels: at the town, province, region, nation level up to the European Union level. Despite many promises of reduction, the efficiency of the legislative machinery is excellent and corpora are constantly growing. In addition to law and regulations that limit the rights of individuals and corporations, there are legislative actions to grant new rights to special categories (e.g. handicapped persons) and financial aids or opportunities that are targeted to specific areas/subjects.

No real democracy or participation exists in practice, if the citizen is not fully informed of all the rights, duties, opportunities and law-making in-progress that concern her. Sir Cecil Carr remarked many years ago that "*as a collection, our statute books might be summed up as beyond the average citizen's pocket to purchase, beyond his bookshelves to accommodate, beyond his leisure to study and beyond his intellect to comprehend*". Internet and digital storage make purchasing and accommodating this vast amount of data a trivial endeavor. Finding what is needed is another story.

The problem studied in this paper is how to make citizens aware of all the laws, regulations and opportunities that concern them. Traditionally, access paradigms have focused on retrieval of data on the basis of precise specifications: examples of this approach include queries on structured database systems, and information retrieval. However, most search tasks, and notably accessing large legal repositories, are exploratory and imprecise in essence: the user does not usually know precisely what he wants (e.g. a specific law), but rather he needs to explore the information base, find relationships among concepts and thin alternatives out in a guided way. Traditional access methods are not helpful in this context, so that new access paradigms are required. We show how dynamic taxonomies, a model recently proposed for the intelligent exploration of heterogeneous information bases, can provide guided browsing and personalized exploration for complex laws and regulations.

2 Traditional information access methods

Since the vast majority of normative material is essentially textual and unstructured in nature, information retrieval techniques [20] were extensively used in the past both in pull and push strategies [16]. These techniques are quite appealing because, at least in principle, they require almost no editorial or manual processing of information. In addition to low costs, this also means the immediate availability of new material. Normally, full text is augmented by manually inserted metadata, such as year of publication, language, type of document (e.g. EUR-lex, http://europa.eu.int/eur-lex/en/index.html). Since laws and regulations quite often refer to and/or amend previous norms, many systems use some kind of hypertext linking [3] to make navigation through references viable for the user.

The limitations of commercial information retrieval systems have been known for some time. Blair and Maron [1] report a study on the use of IBM Stairs in a legal environment: they found that almost 80% of relevant documents were not retrieved. Although IBM Stairs is somewhat outdated, most search engines used to manage laws and regulations are even less powerful. The major problem in text retrieval is the extremely wide semantic gap between the user model and the system model. The user works at a conceptual level and is interested in retrieving complex concepts, while the retrieval system usually works at an extremely low level and usually only understands strings of characters. During the years, a number of improvements have been proposed: from inflectional normalization, to intelligent thesauri such as WordNet [2], to relevance ranking [6]. Despite these attempts, the semantic gap is still very large. Any user of search engines experiences low precision and recall for all but the most concrete queries.

In addition to these semantic problems, information retrieval systems have other shortcomings. First, they are extremely poor from the point of view of user interaction: the user has to formulate his query with no or very little assistance. This is usually difficult because the user often does not know precisely what the information base contains. Second, results are presented as a flat list with no systematic organization. Browsing the infobase is usually difficult or altogether impossible. The success of Yahoo as an alternative to search engines is a convincing proof of this statement.

Hypermedia [3] mainly addresses the problem of browsing/exploring, but it has a number of serious drawbacks. First, there is no systematic picture of relationships among infobase components. Second, exploration is performed one-document-at-a-time, which is quite time consuming. Third, building and maintaining a complex hypermedia network can be extremely costly.

The success of Yahoo makes traditional taxonomies appealing. Here, a hierarchy of concepts can be used to select areas of interest and restrict the portion of the infobase to be retrieved. Taxonomies support abstraction and are easily understood by end-users. However, they are not scalable for large databases. In fact, a normal taxonomy can be used for discrimination just down to the lowermost level of the hierarchy (terminal concepts, which are no further specialized): at that point, the list of items associated with the selected terminal concept must be inspected manually. The major problem here is that if the infobase is large, the average number R of the documents to be manually inspected will be too large for manual inspection. In fact, R = D/T, where D is the number of documents in the infobase and T is the number of terminal concepts in the index [9]. A small infobase of 100,000 documents with a reasonable subject index of 1,000 terminal concepts, already produces 100 documents per terminal concept on the average, too many for manual inspection. Traditional taxonomies are monodimensional: an item can be classified under one and only one concept. If we relax this assumption, and allow an item to be classified under j several concepts, things get worse because the number of documents to be manually inspected increases by a factor of j [9].

Solutions based on semantic networks have been proposed in the past [14] and are now addressed again in the current effort on ontologies and Semantic Web. This approach is quite expensive in terms of design and maintenance of complex conceptual schemata. General ontologies are more powerful and expressive than plain taxonomies, but they are better suited to programmatic access and much more difficult to understand and manipulate by the casual user. Usually, user interaction must be mediated by specialized agents. This increases costs, time to market and decreases generality and flexibility of user access. The solution we propose in the following can be seen as a complement, in addition to being an alternative, to complex ontologies: in section 5, we provide a preliminary discussion of how our taxonomy-based model can provide a user-understandable view on complex semantics.

3 Dynamic taxonomies

Dynamic taxonomies [7], [8] are a general knowledge management model for complex, heterogeneous information bases. It has been applied to very diverse areas, including news archives, encyclopedias, multimedia databases [10], electronic commerce [11], and medical guidelines [19]. The intension of a dynamic taxonomy is a taxonomy designed by an expert. This taxonomy is a concept hierarchy (directed acyclic graph taxonomies modeling multiple inheritance are supported but rarely required) going from the most general to the most specific concepts. A dynamic taxonomy does not require any other relationships in addition to subsumptions (e.g., IS-A and PART-OF relationships).

In the extension, items can be freely classified under several topics at any level of abstraction (i.e. at any level in the conceptual tree). This multidimensional classification is a departure from the monodimensional classification scheme used in conventional taxonomies. Besides being a generalization of a monodimensional classification, a multidimensional classification models common real-life situations. First, an item is very rarely classified under a single topic. One reason is that items are very often about different concepts: for example a funding opportunity could be classified as Agriculture>corn, Agriculture>soybean, Location>Southern Italy, Company>Turnover>less than 1 million euro. Second, items to be classified usually have different independent features (e.g. Time, Location, etc.), each of which can be described by an independent taxonomy. These features are often called *perspectives* or *facets*.

By taking a "nominalistic" approach (concepts are defined by instances rather than by properties), a concept C is just a label that identifies all the items classified under C. Because of the subsumption relationship between a concept and its descendants, the items classified under C (*items(C)*) are all those items in the *deep extension* [17] of C, i.e. the set of items identified by C includes the *shallow extension* of C (i.e. all the items directly classified under C) union the deep extension of C's sons. By construction, the shallow and the deep extension for a terminal concept are the same.

There are two important consequences of our approach. First, since concepts identify sets of items, logical operations on concepts can be performed by the corresponding set operations on their extension. This means that the user is able to restrict the information base by combining concepts through the normal logical operations (and, or, not).

Second, dynamic taxonomies can find all the concepts related to a given concept C: these concepts represent the conceptual summary of C. Concept relationships other than IS-A are inferred through the extension only, according to the following *extensional inference rule*: two concepts A and B are related iff there is at least one item D in the infobase which is classified at the same time under A (or under one of A's descendants) and under B (or under one of B's descendants). For example, we can infer a (unnamed) relationship between Michelangelo and Rome, if an item that is classified under Michelangelo and Rome exists in the infobase. At the same time, since Rome is a descendant of Italy, also a relationship between Michelangelo and Italy can be inferred. The extensional inference rule can be seen as a device to infer relationships on the basis of empirical evidence.

The extensional inference rule can be easily extended to cover the relationship between a given concept C and a concept expressed by an arbitrary subset S of the universe: C is related to S iff there is at least one item D in S which is also in items(C). Hence, the extensional inference rule can produce conceptual summaries not only for base concepts, but also for any logical combination of concepts. In addition, dynamic taxonomies can produce summaries for sets of items produced by other retrieval methods such as information retrieval, etc. and therefore access through dynamic taxonomies can be easily combined with other retrieval methods.

Dynamic taxonomies work on conceptual descriptions of items, so that heterogeneous items of any type and format can be managed in a single, coherent framework. Finally, since concept C is just a label that identifies the set of the items classified under C, concepts are language-invariant, and multilingual access can be easily supported by maintaining different language directories, holding language-specific labels for each concept in the taxonomy.

Dynamic taxonomies can be used to browse and explore the infobase in the following way. The user is initially presented with a tree representation of the initial taxonomy for the entire infobase. Each concept label has also a count of all the items classified under it (i.e. the cardinality of items(C) for all C's). The initial user focus F is the universe (i.e. all the items in the infobase).

In the simplest case, the user can then select a concept C in the taxonomy and *zoom* over it. The zoom operation changes the current state in two ways. First, concept C is used to refine the current focus F, which becomes

$$F=F\cap \text{items}(C) \quad \textbf{(1)}$$

items not in the focus are discarded. Second, the tree representation of the taxonomy is modified in order to summarize the new focus. All and only the concepts related to F are retained and the count for each retained concept C' is updated to reflect the number of items in the focus F that are classified under C'. The reduced taxonomy is a conceptual summary of the set of documents identified by F, exactly in the same way as the original taxonomy was a conceptual summary of the universe. In fact, the term *dynamic taxonomy* is used to indicate that the taxonomy can dynamically adapt to the subset of the universe on which the user is focusing, whereas traditional, static taxonomies can only describe the entire universe.

The retrieval process can then be seen as an iterative thinning of the information base: the user selects a focus, which restricts the information base by discarding all the items not in the current focus. Only the concepts used to classify the items in the focus, and their ancestors, are retained. These concepts, which summarize the current focus, are those and only those concepts that can be used for further refinements. From the human computer interaction point of view, the user is effectively guided to reach his goal, by a clear and consistent listing of all possible alternatives.

Dynamic taxonomies can be integrated with other retrieval methods in two basic ways. First, focus restrictions on the dynamic taxonomy can provide a context on which other retrieval methods can be applied, thereby increasing the precision of subsequent searches. Second, the user can start from an external retrieval method, and see a conceptual summary of the concepts that describe the result. These two approaches can be intermixed in different iteration steps during a single exploration. The integration of dynamic taxonomies with information retrieval is especially important in the present context because dynamic taxonomies can be used to describe abstract, conceptual queries and information retrieval can be used to define concrete queries (for instance proper names). By offloading conceptual manipulation to the dynamic taxonomy component, the information retrieval component needs not be very sophisticated, and in fact, an efficient text retrieval system is usually adequate.

4 An example

Giovanni, a farmer located in Cuneo, Piedmont, Italy is interested in opportunities (funding, etc.) in agriculture. These opportunities may arise from different sources (Cuneo, Piedmont, Italy, EU), may involve different crops, may have different requirements (perhaps, a minimum turnover), etc. With a conventional text retrieval system, finding the relevant documents is quite a difficult task. A query for agriculture is at the same time too broad (because it will retrieve all documents about agriculture, possibly thousands) and too restrictive (it will fail to retrieve tobacco-growing, if agriculture as a word is not mentioned in the document). How can Giovanni be sure he considers every relevant document? He can't, and he will probably find himself querying for a broad term, hoping it is all-inclusive but with no guarantee it is, and wade through a very long list of items that are mostly irrelevant to his needs.

Now assume that a dynamic taxonomy exists, whose top level is organized as:

- *Sector*, a facet describing sectors of activity (agriculture, chemistry, etc.)
- *Location*, a facet describing the location(s) to which a specific document applies
- *Subject*, a facet describing the subject(s) (e.g. persons or companies, public companies, etc.) to which a specific document applies
- *Document type*, a facet describing the type of document (law, regulation, opportunity, etc.)
- *Issuer*, a facet describing the issuer (town, country, etc.)

Each facet can be as articulated as required; the schema above is obviously quite simplified.

With this simple schema, Giovanni can explore opportunities by selecting *Sector>Agriculture* and then *Document type>Opportunity*, if this concept exists in the reduced taxonomy: if it does not, no opportunities for agriculture exists, and Giovanni is done. After this compound focus is set, the reduced taxonomy will show all the concepts related to it: specific Locations, Subjects, etc. Browsing is completely symmetric: the same result is obtained if he selects *Document type>Opportunity* and *Sector>Agriculture.*

Giovanni is also helped to discover new opportunities which involve changing something in his business: if he grows corn, he might discover that better opportunities exist for soybean crops, or that setting up a company increases his opportunities, etc. The simple translation of the labels for the concepts in the taxonomy, allows Jean, Johannes, John and Ivan to find relevant information in exactly the same way.

Note that a conventional taxonomy would not provide sufficient discrimination in thinning the result set: selecting *Sector>Agriculture* retrieves all the documents for Agriculture, including laws, regulation and other irrelevant material. Conversely, the selection of *Document type>Opportunity* retrieves opportunities in all sectors, not just Agriculture. In fact, result thinning in conventional taxonomies can occur only by specializing the current concept, and concepts in different branches in the taxonomy are not available.

5 Dynamic taxonomies applied to laws and regulations

The advantages of dynamic taxonomies over traditional methods are dramatic in terms of convergence of exploratory patterns and in terms of human factors. Sacco [9] provides analytical evidence that three zoom operations on terminal concepts are sufficient to reduce a 1,000,000-item information base described by a compact taxonomy with 1,000 concepts to an average 10 items. Experimental data on a real newspaper corpus of over 110,000 articles, classified through a taxonomy of 1100 concepts, reports an average 1246 documents to be inspected by the user of a traditional, static taxonomy vs. an average 27 documents after a single zoom on a dynamic taxonomy.

Dynamic taxonomies require a very light theoretical background: namely, the concept of a subject index (i.e. the taxonomic organization) and the zoom operation, which seems to be very quickly understood by end-users. Hearst et al. [4] and Yee et al. [21] conducted usability tests on a corpus of art images. Despite an inefficient implementation that caused slow response times, their tests show that access through a dynamic taxonomy produced a faster interaction and a significantly better recall than access through text retrieval. Perhaps more important are the intangibles: the feeling that one has actually considered all the alternatives in reaching a result.

Differently from most previous research, dynamic taxonomies cleanly separate the process of classifying documents from the use of the classification information in the browsing system. Obviously, the classification system and the design of the taxonomy must take into account the way classification is used, i.e. the extensional inference rule. First, dynamic taxonomies actually perform concept association mining. This simplifies index creation and maintenance since concept associations, which are often quite dynamic in time, need not be forecasted and accounted for in schema design. At the same time, the user is presented with associations the schema designer might not even be aware of (discovery). In traditional approaches, relationships among concepts must be explicitly described in the conceptual schema. Since only these relationships will be available to the user for browsing and retrieval, the schema designer must anticipate all the possible relationships among concepts: a very difficult if not helpless task. On the one hand, the designer will define relationships that do not actually occur in the corpus and are useless. On the other hand, some relationships will not be defined at all, either because they are not interesting for the designer (but they might be for the user) or because they have not occurred before and are unexpected. All these problems are solved because concepts relationships are automatically derived from the actual classification.

Second, since dynamic taxonomies synthesize compound concepts, these need usually not be represented explicitly. This means that the main cause of the combinatorial growth of traditional taxonomies is removed. Sacco [8] developed a number of guidelines that produce taxonomies that are compact and easily understood by users. Some of these guidelines are similar to the faceted classification scheme by Ranganathan [5], at least in its basic form: the taxonomy is organized as a set of independent, "orthogonal" subtaxonomies (facets or perspectives) to be used to describe data. As an example, a compound concept such as *Agriculture in Southern Italy* need not be accounted for, because it can be synthesized from its component concepts: *Sector>Agriculture* and *Location>Southern Italy*. Thus, one of the main causes of complexity in the design of comprehensive taxonomies is avoided: by synthesizing concepts, we avoid the exponential growth due to the description of all the possible concept combinations, and the resulting taxonomy is significantly more compact and easier to understand. In addition to minimizing the concepts in the taxonomy, breaking compound concepts into their base components allows the user to easily correlate concepts and explore such correlations. In the example, the user focusing on *Sector>Agriculture* will immediately find all the relevant locations related to agriculture (which include *Southern Italy*). If compound concepts were used, correlation cannot be carried out automatically, but it would require the manual inspection of labels. In addition, the excellent convergence of dynamic taxonomies allows the designer to define taxonomies that are much simpler and smaller than traditional ones.

Although the discussion above indicates that "minimal" taxonomies are desirable, there are situations in which additional concepts or facets actually improve user interaction. One of the problems of traditional taxonomies is that they require the user to perceive the world through the same concepts and understandings as the schema designer. However, as one of the referees remarked, a farmer does not probably think along the same lines as the legal expert coding the documents. With the inflexible monodimensional classification of traditional taxonomies, this problem has no solution. In dynamic taxonomies, additional facets can easily accommodate different, alternate perceptions of conceptual organizations, providing powerful personalization capabilities. In this case (farmer vs. legal expert), the two facets are not orthogonal and would not fit in a strict faceted classification [5]: just another proof that faceted classification systems are a subset of the multidimensional classification scheme proposed by dynamic taxonomies.

Both personalization and push strategies can be supported by dynamic taxonomies. In both cases, they can be implemented by using boolean expressions on the concepts in the taxonomy. In the case of personalization, such an expression defines a user profile, and can be automatically added to user queries in a way similar to query modification in relational database systems [15]. Farmer Giovanni's profile, for instance, could be *Location>Europe>Italy>Cuneo AND Sector>Agriculture*. User-friendly front-ends that shield Giovanni from the complexities of writing boolean expressions can be easily devised. Giovanni's profile will be considered as the initial context or focus, instead of the universe of discourse: when the initial reduced taxonomy is presented, only the concepts under which there are documents satisfying Giovanni's profile are preserved. Thus, Giovanni will not see documents and concepts for Latvia, nor documents or concepts for the automotive sector. This same user profile (or multiple versions of it) can be used to implement push strategies, since it provides an accurate statement of interests. In this way, the system acts in a proactive way and informs the user whenever new relevant material is available. Sacco [12] describes an efficient algorithm for dynamic taxonomy-based push strategies.

Two research topics are currently being investigated. The first one is the semantics of classification, and is especially important in the context of personalization and push strategies: what additional classifications can be inferred by the fact that a document d is classified under a concept C? Sacco [8] shows that the inclusion constraint implicit in subsumptions requires a *backward inheritance*, i.e. that a document classified under a concept C is also classified under all of C's ancestors. As an example, a document classified under *Rome* is also classified under *Lazio*, *Italy*, and *Europe*. Most systems based on dynamic taxonomies allow classifying documents under terminal concepts only. If we relax this limitation and allow a document d to be classified under a non-terminal concept C, backward inheritance still holds, but there might be implications on the descendants of C. For example, assume that d is a document describing funding for Agriculture and is consequently classified under *Agriculture*. Since d is not about some specific aspects of agriculture, it also applies to any of Agriculture's descendants, from corn to tobacco. Hence classifying d under C also implies that d is also classified under each of C's descendants. We call this type of inheritance *forward inheritance*, to distinguish it from the standard backward inheritance discussed above. Top-down (i.e. forward) inheritance is the standard inheritance rule for properties and methods in object-oriented systems. Extended inheritance (i.e., backward plus forward inheritance) also applies in a similar form to hierarchical structured material, such as video stories [13]. This notwithstanding, forward inheritance does not always apply. Often, especially in PART-OF relationships, a specific document may apply to a concept, but not to its components. A trivial example is a license plate that applies to a car, but not to its engine.

Forward inheritance is important when a) documents can be classified under non-terminal concepts, and b) when conceptual expressions are used to define a context for personalization or push strategies. In fact, with forward inheritance, Giovanni's profile *Location>Europe>Italy>Cuneo AND Sector>Agriculture* will retain all the documents classified under Italy, but not specifically under Cuneo: if only backward inheritance were used, these documents would be lost.

The second area of investigation is the automatic derivation of taxonomies from complex ontology schemata. The goal is to use dynamic taxonomies as a user-friendly front-end to complex information that is also available in a richer semantic form for agents and programmatic access. Sacco [12] shows that relational views can be translated into dynamic taxonomies by considering each tuple as a document, and transforming all the attributes of the view into facets. We are currently investigating how this mapping can be extended to cover schemata that are semantically richer than relational ones. However, we believe that in most practical cases, dynamic taxonomies coupled with information or database retrieval will be used as the unique access path to complex information. We feel that the current emphasis on ontologies and agents for search tasks is overstressed and that dynamic taxonomies often provide an efficient and effective alternative, which is easier to build and maintain, and much more transparent and intuitive for the casual user.

6 Conclusions

Dynamic taxonomies represent a dramatic improvement over other search and browsing methods, both in terms of convergence and in terms of full feedback on alternatives and complete guidance to reach the user goal. For these reasons, and because of easy personalization, multilingual access and push strategies, they give that kind of interactive knowledge management that is required for political awareness and participation. Systems based on dynamic taxonomies, such as Knowledge Processors' Universal Knowledge Processor [18], are currently available and offer real-time operations even on large information bases.

References

1. Blair, D. C., Maron, M. E: An evaluation of retrieval effectiveness for a full-text document-retrieval system. Comm. ACM, 28:3 (1985)
2. Fellbaum, C. (ed.): WordNet-An Electronic Lexical Database. MIT Press (1998)
3. Groenbaek K., Trigg R. (eds.): Hypermedia. Comm. ACM 37, vol. 2 (1994)
4. Hearst, M., et al.: Finding the Flow in Web Site Search, Comm. of the ACM, 45, 9 (2002)
5. Ranganathan, S. R.: The Colon Classification. Rutgers University Press (1965)
6. Robertson, S. E., Sparck Jones, K.: Relevance weighting of search terms. J. of ASIS. 27 (1976)
7. Sacco, G.M.: Navigating the CD-ROM, Proc. Int. Conf. Business of CD-ROM, (1987)
8. Sacco, G. M.: Dynamic Taxonomies: A Model for Large Information Bases. IEEE Transactions on Knowledge and Data Engineering 12, 2 , p. 468-479 (2000)
9. Sacco, G. M.: Analysis and Validation of Information Access through Mono, Multidimensional and Dynamic Taxonomies, Tech Report Dept. of Informatica, Univ. of Torino (2002)
10. Sacco, G. M.: Systematic Browsing for Multimedia Infobases, in Proc. CISST2003, Int. Conf. on Imaging Science, Systems, and Technology (2003)
11. Sacco, G. M.: The Intelligent E-Sales Clerk: the Basic Ideas, Proc. INTERACT'03 Int. Conf. on Human-Computer Interaction (2003)
12. Sacco, G. M.: Dynamic taxonomy process for browsing and retrieving information in large heterogeneous data bases. US Patent 6,743,349 (2004)
13. Sacco, G. M.: Uniform access to multimedia information bases through dynamic taxonomies, Proc. IEEE Int. Symposium on Multimedia Software Engineering (2004)
14. Schmeltz Pedersen, G.: A Browser for Bibliographic Information Retrieval, Based on an Application of Lattice Theory. Proc. ACM SIGIR Conf. (1993)
15. Stonebraker, M.: Implementation of integrity constraints and views by query modification. Proc. ACM SIGMOD Conf. (1975)
16. Story, G. A., et al.: The Rightpages Image-Based Electronic Library for Alerting and Browsing. IEEE Computer, 25:9 (1992)
17. Straube, D. D. & Ozsu, M. T.: Queries and query processing in object-oriented database systems, ACM Trans. on Information Systems, 8: 4 (1990)
18. Universal Knowledge Processor, www.knowledgeprocessors.com
19. van Rijsbergen, C. J.: Information Retrieval. Butterworths (1979)
20. Wollersheim, D., Rahayu, W.: Methodology For Creating a Sample Subset of Dynamic Taxonomy to Use in Navigating Medical Text Databases. Proc. IDEAS 2002 Conf. (2002)
21. Yee, K-P., et al.: Faceted Metadata for Image Search and Browsing, Proc. ACM CHI Conf. (2002)