Browsing Image Collections with Representations of Common-Sense Activities

Andrew S. Gordon

Institute for Creative Technologies, 13274 Fiji Way, Marina del Ray, CA 90292.
E-mail: gordon@ict.usc.edu

To support browsing-based subject access to image collections, it is necessary to provide users with networks of subject terms that are organized in an intuitive, richly interconnected manner. A principled approach to this task is to organize the subject terms by their relationship to activity contexts that are commonly understood among users. This article describes a methodology for creating networks of subject terms by manually representing a large number of common-sense activities that are broadly related to image subject terms. The application of this methodology to the Library of Congress Thesaurus for Graphic Materials produced 768 representations that supported users of a prototype browsing-based retrieval system in searching large, indexed photograph collections.

Browsing Interfaces for Image Retrieval

The use of subject terms in the cataloging of images has been one of the real success stories in the information sciences. The growth of large indexing thesauri, notably the Library of Congress Thesaurus for Graphic Materials (LCTGM) (Library of Congress Prints and Photographs Division, 1995) and the Art and Architecture Thesaurus (Peterson, 1990), has been coupled with a rich theoretical understanding of subject term assignment (Orbach, 1990; Shatford, 1986; Shatford Layne, 1994; Svenonius, 1994). Together, these tools have supported the cataloging and organization of millions of photographs and other images in our most valued collections, including those of the Library of Congress, the New York Public Library, and the Chicago Historical Society, among many others. When assisted by reference librarians who are knowledgeable of the indexing vocabulary, users of these collections have been able to meet a wide variety of retrieval needs, including those of historical research, publishing, and advertising.

In the online environment, where users are not assisted by knowledgeable reference librarians, subject access to image collections has not faired as well. The general prob-

lem stems from the difficulty in matching the vocabulary that people use to describe their retrieval needs to the way that collection materials are cataloged. In investigating what they referred to as the *Vocabulary Problem*, Furnas, Landauer, Gomez, & Dumais (1987) found that people use vastly different words to refer to the same item, despite their best efforts to do otherwise. In a set of studies, these researchers found that if a single archivist assigns a label to a particular item, other untutored users of the archive will fail to access the item on 80 to 90% of their attempts. Throughout its history, the information retrieval research community has primarily addressed this problem by matching against the full text of documents, expanding users' queries or catalogers' indexes, and filtering based on collaborative or content statistical information.

Browsing-based retrieval systems take a different approach. Rather than designing better algorithms for matching users' textual queries to documents in a collection, these systems attempt to organize the materials into a coherent browsing space that can be traversed by users to find what they are looking for. A growing theoretical understanding of the browsing process has begun to emerge (Bates, 1989; Chang & Rice, 1993). In addition, application prototypes have become increasingly similar in design and in the manner that users explore a collection (Allen, 1994; Cooper & Byrd, 1997; Johnson & Cochrane, 1995; Schatz, Johnson, Cochrane, & Chen, 1996). Rather than traversing a network of collection materials directly linked to each other, users of these systems are presented with a network of the subject terms that are used to index the collection—a browsing space that can remain constant as the collection size changes.

In adapting this style of interaction for image retrieval, we must reconsider the character of the browsing space of indexes that the user navigates. Browsing spaces for the retrieval of text documents can reasonably contain words that have little or no correlation to traditional subject terms from controlled indexing thesauri. Indeed, some of the most innovative approaches extract the indexes of the browsing space from the text collection itself using a variety of

statistical and lexical analysis techniques (Chen, Yim, Fye, & Schatz, 1995; Cooper & Byrd, 1997). For the retrieval of images from previously indexed collections, browsing spaces must be limited to the terminology that is present in the controlled indexing thesauri used. In following this approach, our problem immediately becomes: *How can we construct effective networks of image subject terms to support browsing-based retrieval?*

This article describes an approach to solving this problem by elaborating the relationship between image subject terms and common-sense knowledge related to their corresponding concepts. We argue that image indexes are best organized in a manner that parallels our common-sense understanding of human activities in the everyday world. This article presents a methodology for identifying a set of common-sense activities that have broad coverage over the domain represented by an image thesaurus, and describes how this methodology was used to identify 768 activities related to the terms in the LCTGM. These representations were used to create an intuitive and richly interconnected browsing space integrated into a browsing-based photograph retrieval system. Evaluations of this retrieval system with large photograph collections are reviewed.

Relationships Between Terms in the LCTGM

Identifying appropriate links between subject terms has become an expected activity of any thesaurus authoring effort (American National Standards Institute, 1993). However, if we examine the quantity of links between thesaurus terms for image indexing, it is difficult to argue that these efforts have led to richly interconnected browsing spaces. A recent release of LCTGM contained 5,760 terms, linked by 4,421 taxonomic links (Broader and Narrower term pairs) and 5,888 associative links (Related term pairs). Added together, these 20,618 one-directional links offer only 3.58 links, on average, for each of the terms in the thesaurus.

The key question to ask here is: How many browsing links are enough? In the ideal case, we would like each term in the thesaurus to be linked to all of the other terms that would be useful in supporting end-user navigation of the thesaurus. However, for any given term in the LCTGM, the links that are provided by its authors seem inadequate. Consider the LCTGM term *School Children*, which has one Broader Term, *Children*, and four Related Terms, *Classrooms*, *School Recesses*, *Schools*, and *Students*. While these five links are more than are given to most terms in the LCTGM, there are an enormous number of equally well-related terms in the thesaurus, including some that seem absolutely essential, for example, *Teachers*. Other, less essential but useful links include those to *School discipline*, *School meals*, *Physical education, Religious education, School safety patrols, Sunday schools, Teaching methods,* and *Children reading & writing*. What is needed by the authors of the LCTGM is a principled and efficient method of identifying additional appropriate links.

The theoretical discussion of the nature of relationships between thesaurus terms has enjoyed a long history (Green & Bean, 1995; Maniez, 1988; Neelameghan & Maitra, 1974; Soergel, 1967; Willets, 1975), largely grounded in linguistic theory. However, we argue that this discussion would be better placed within the theoretical context of cognitive science. Put simply, if we are interested in building browsing spaces that are understandable to users, we are better off thinking about conceptual relationships between terms than linguistic relationships between words.

Neelameghan and Maitra (1974), in listing 39 types of relationships between subject terms, noted one particular type of relationship that was just beginning to become a major focus of attention in cognitive science research at that time. They identified an important relationship: "Two Persons interacting in a special context," such as between *Teachers* and *Students*, and *Doctors* and *Patients*. In cognitive science research, these *special contexts* have been referred to as *Scripts*, which describe the relationship between people, places, things, and actions that had roles in commonly understood, stereotypical human activities (Abelson, 1975, 1981). The script-based relationship between *Doctors* and *Patients* is that these terms indicate people who play a role in our common-sense understanding of the activities of (1) going to a doctor's office for a physical examination, (2) performing surgery in an operating room, and (3) working in an emergency care center, among others.

Although it is unclear whether activity contexts can play a role in organizing other types of subject thesauri, they seem appropriate as a theoretical tool for organizing and linking subject terms for images. The scope of the terms in one of the major image thesauri, the LCTGM, is largely limited to indexes that refer to people, places, things, and activities. Moreover, this thesaurus was primarily designed to index photographs, which are largely taken in specific activity contexts that are broadly understood, capturing both instances and exceptions to our expectations of the world around us.

Representing Activities using LCTGM Terms

Taking the theoretical position that shared activity contexts are the backbone by which image subject terms may be linked, a new methodology for identifying appropriate links was developed. The main task in this methodology is to identify all of the activity contexts that are widely shared among thesaurus terms. The goal is to find a large set of activities that provide a reasonable amount of coverage over the breadth of the subject vocabulary, and which are widely understood as common sense in the user community at large. These activities should include commonly experienced activities like taking a subway train and taking a shower, as well as commonly understood but less frequently experienced activities like giving birth to a baby or taking shelter during a tornado.

The methodology, which is best completed by a knowledge engineer who is familiar with the subject terms, requires a manual analysis of every term that is in the thesaurus. For each thesaurus term, a knowledge engineer attempts to identify and represent a set of activities in which the term plays an expected, conceptual role. This is done by first generating a candidate list of common-sense activities, identified by a short textual phrase authored by the knowledge engineer. Then, for each of the candidate activities, the knowledge engineer attempts to map the other expected roles of the activity on to other subject terms that exist in the thesaurus. To aid in the identification of other vocabulary terms, the knowledge engineer organizes the search based upon a simple representation frame for activities consisting of five slots: Events, Places, People, Things, and Miscellaneous Ideas. The result of this analysis, then, is a representational frame consisting of categorized lists of image subject terms, and a descriptive title.

For example, the LCTGM term *Caves* would elicit a set of common-sense activities including "Exploring a cave" and "Working as a miner in an Underground Mine." In representing the activity of "Exploring a cave," LCTGM terms for Events would include *Crawling & creeping*, *Discovery & exploration*, and *Erosion*. Places would be *Cave dwellings*, *Caves*, and *Passageways*. An appropriate term for the People slot is *Explorers*. Things include *Bats*, *Cave drawings*, *Helmets*, *Lanterns*, *Rock formations*, *Shadows*, and *Stalactites & stalagmites*. Miscellaneous Ideas, a catch-all category for abstract concepts, include *Air Quality* and *Cold*.

Two constraints are placed on the knowledge engineer during iteration through the thesaurus terms. First, when no suitable term can be located in the thesaurus that adequately refers to the desired concept, then the concept is not included in the representation. For example, in representing the activity of exploring a cave, there are many concepts and terms that do not appear in the representation. This activity might also contain concepts of spelunking, flashlights, and underground waterways. However, these concepts do not match terms that exist in the LCTGM, and are therefore not included in the representation. Second, if the first constraint yields a representational frame that is very sparse (few terms could be found that corresponded to concepts in the activity), then the frame is removed from the final set of represented activities.

Although these constraints may be viewed as a hindrance to the development of a rich set of common-sense representations, it effectively constrains the knowledge representation task to only the work that will be useful in developing a rich browsing space of subject terms. The primary goal of this methodology is to identify only the activities that are widely related to terms in the thesaurus, and that are within the domain that the thesaurus represents. There certainly exists a much larger set of common-sense activities, which includes those that could only partially be represented by thesaurus terms as well as those that are unrelated to these terms at all.

This methodology was used by the author to identify a set of common-sense activity contexts that have broad coverage over the terms of the LCTGM. One pass of the 5,760 terms in the LCTGM (in alphabetical order) required half-time work over the course of 4 months to complete. The resulting set of activities consisted of 768 representations (published in Gordon, 1999). In all, subject terms were used 9,417 times to represent concepts in activities, with an average of 1.63 activities per subject term and 12.26 subject terms per activity representation.

Applications and Evaluations

The representations of activity contexts that are produced using this methodology provide the basis for generating large numbers of intuitive links between image subject indexes. In the simplest approach, any two terms that participate in the same activity representation can be linked. For example, a representation of "Taking a dog for a walk" can be used to fully interconnect the LCTGM terms *Parks*, *Collars*, *Dogs*, and *Fire Hydrants*, among others.

The representations identified by the methodology described above can be used to produce a rich network of LCTGM terms for use as an end-user browsing space. The 768 representations constitute 100,894 unique one-directional links between subject terms (as well as 13,416 redundant links due to overlap in the representations of different activities). Dividing this by the number of terms in the LCTGM (5,760), there were an average of 17.52 unique one-directional links produced using this methodology (compared to 3.58 links, on average, provided by the authors of the LCTGM).

Although it may be reasonable to simply include these generated links in the LCTGM as additional Related terms, the added value of the textual label and representational frame add a level of comprehensibility that can be exploited in the design of a browsing-based retrieval interface. To explore this possibility, a retrieval application called Déjà Vu was built to provide subject access to collections of photographs, as well as other media collections indexed using LCTGM subject terms (Gordon & Domeshek, 1998).

The Déjà Vu retrieval application was designed to aid users in locating thesaurus terms that best meet their retrieval needs and that index materials that are available in a particular collection. At any given moment during the browsing process, the user is focused on one term in a particular thesaurus. Links are provided to the user to other terms in the thesaurus, which can be clicked to change the current focus term. In addition to the standard thesaurus links (Broader, Narrower, and Related terms), each of the activity representations (referred to as *Expectation Packages* in previous publications) that include the focus term are also displayed, providing direct links to each of the other terms in each of these representations as well. Rather than being presented as merely a list of words, the representations are displayed in the manner that they were authored, which includes the textual title and an organizing frame

(with headings for each of the five slots used in the representations: Events, People, Places, Things, and Miscellaneous Ideas).

Every term that is displayed to the user indicates whether or not it is being used to index available materials using a visual cue. When a user identifies an appropriate term that is being used as an index in the collection, he or she selects the term to retrieve all of the materials to which it has been assigned. When the number of retrieved materials is too large, the users can progressively narrow this set by continuing to browse for terms that can be conjoined with the current selection and still retrieve a nonempty set of results.

The Déjà Vu retrieval application and the 768 LCTGM activity representations were evaluated as a photograph-finding aid in two separate library site installations. At the Library of Congress Prints and Photographs Division the application was evaluated as a finding aid to more than 25,000 photographs in their largest LCTGM-indexed collection. At the North Dakota Institute for Regional Studies at North Dakota State University, the application was used as the primary finding aid to the institute's collection of over 11,000 photographs over a 6-month evaluation period (Gordon, 1999).

The results of these evaluations, which were conducted as part of an iterative design process rather than as a controlled study, revealed several necessary improvements that should be included in the design of future, production-quality systems. Above all, it was found that subject access alone is not sufficient to meet the retrieval needs of photograph collection users. Even when users are primarily concerned with a particular subject content, they normally need to be able to further constrain their searches by specifying a particular photographer, time period, geographic location, the names of photographed people, or visual composition, among other things. In short, they need access to all of the other information that is usually specified in library records of photographic materials. As most of these search criteria can be specified in current query-based systems, the main challenge in developing future systems is seamlessly integrating query and browsing-based search techniques.

Although these evaluations provide insight on the design of effective browsing-based retrieval interfaces and demonstrate the value of a particular interface over other search applications, they do not adequately assess the value of activity representations as an organizing principle for image browsing spaces. As we expect research on the design of networks of thesaurus terms to continue, it will be necessary to develop evaluation methods to judge the quality of the browsing space itself, separate from the idiosyncratic context of a particular retrieval interface.

Conclusions

The influx of computer scientists into research areas traditionally populated by library scientists has generally changed the values and character of the work that is done. In general, library scientists rarely shy away from making ontological commitments, developing systems of classification, or judging the relevance of materials to subject topics—all by hand and in large scale. In contrast, computer scientists generally value solutions that are completely automated over those that require any amount of manual analysis or knowledge engineering effort. It is generally argued that this shift in approach is necessary due to the explosive growth of online text, although the growing availability of text-analysis algorithms has surely played a role as well.

In furthering research on image retrieval, we must be more selective about which problems are best tackled by automated solutions. The focus of this article, developing a rich browsing space of image subject terms, is a problem that seems best solved by a thorough, manual analysis of the subject terms. Executing the methodology that is presented in this article required a significant amount of difficult work; however, this effort pales in comparison to the two decades of work that went into developing the LCTGM in the first place, or the countless years of work spent by catalogers using the thesaurus to index millions of photographs.

That being said, it would be advantageous if automated thesaurus-linking techniques could produce organizations of subject terms that were equally intuitive and comprehensible as those produced by manual analysis. One could easily imagine that effective automated techniques could help reduce the subjective and idiosyncratic nature of links authored by only one or a handful of knowledge engineers. Ideally, these approaches would also aid in the construction of browsing spaces that are specifically tailored to the common-sense knowledge of particular social and organizational cultures.

The primary direction for future work is to further explore techniques for using common-sense knowledge to connect users' visual retrieval needs to available materials. It is reasonable to believe that there are other types of common-sense knowledge besides activity contexts that can be used to organize networks of subject terms. Likewise, there may be a number of ways to take advantage of activity representations in other information retrieval interaction styles, including query-based search and more proactive information delivery systems. By studying information needs and common-sense reasoning in parallel, we hope to further align complimentary research in the information and cognitive sciences.

References

Abelson, R. (1975). Concepts for representing mundane reality in plans. In D. Bobrow & A. Collins (Eds.), Representing and understanding: Studies in cognitive science (pp. 273–309). New York: Academic Press.

Abelson, R. (1981). The psychological status of script. American Psychologist, 36, 715–729.

Allen, R. (1994). Navigating and searching in hierarchical digital library catalogs. Proceedings of the first annual conference on the theory and practice of digital libraries (Digital Libraries '94). College Station, TX.

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

Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13, 407–424.

Chang, S., & Rice. R. (1993). Browsing: A multidimensional framework. In M. Williams (Ed.), Annual review of information science and technology (ARIST), 28. Medford, NJ: Learned Information.

Chen, H., Yim, T., Fye, D., & Schatz, B. (1995). Automatic thesaurus generation for an electronic community system. Journal of the American Society for Information Science, 46, 175–193.

Cooper, J., & Byrd, R. (1997). Lexical navigation—Visually prompted query expansion and refinement. Proceedings of the second ACM international conference on digital libraries, Philadelphia, PA (pp. 237–246). New York: ACM Press.

Furnas, G., Landauer, T., Gomez, L., & Dumais, S. (1987). The vocabulary problem in human–systems communication. Communications of the ACM, 30(11), 964–971.

Gordon, A. (1999). The design of knowledge-rich browsing interfaces for retrieval in digital libraries. Doctoral dissertation, Northwestern University, Evanston, IL.

Gordon, A., & Domeshek, E. (1998) Deja vu: A knowledge-rich interface for retrieval in digital libraries. Proceedings of the 1998 international conference on intelligent user interfaces, San Francisco, CA. New York: ACM Press.

Green, R., & Bean, C. (1995). Topical relevance relationships, II. An exploratory study and preliminary typology. Journal of the American Society for Information Science, 46, 654–662.

Johnson, E., & Cochrane, P. (1995). A hypertextual inteface for a searcher's thesaurus. Proceedings of the second international conference on the theory and practice of digital libraries (Digital Libraries '95), Austin, TX.

Library of Congress Prints and Photographs Division. (1995). Thesaurus for graphic materials. Washington, DC: Library of Congress Cataloging Distribution Service.

Maniez, J. (1988). Relationships in thesauri: Some critical remarks. International Classification, 15, 133–138.

Neelameghan A., & Maitra, R. (1974). Non-hierarchical associative relationships among concepts: Identification and typology. International Federation for Documentation Committee on Classification Research (FID/CR) Report No. 18. Bangalore, India: Documentation Research and Training Centre.

Orbach, B. (1990). So that others may see: Tools for cataloging still images. In R. Smiraglia (Ed.), Describing archival materials: The use of the MARC AMC format. Binghamton, NJ: Haworth Press.

Peterson, T. (1990). Art & architecture thesaurus. New York: Oxford University Press.

Schatz, B., Johnson, E., Cochrane, P., & Chen, H. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. Proceedings of the 1st ACM international conference on digital libraries, Bethesda, MD. New York: ACM Press.

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. Cataloging and Classification Quarterly, 6, 39–62.

Shatford Layne, S. (1994). Some issues in the indexing of images. Journal of the American Society for Information Science, 45, 583–588.

Soergel, D. (1967). Some remarks on information languages, their analysis and comparison. Information Storage and Retrieval, 3, 219–291.

Svenonius, E. (1994). Access to nonbook materials: The limits of subject indexing for visual and aural languages. Journal of the American Society for Information Science, 45, 600–606.

Willets, M. (1975). An investigation of the nature of the relation between terms in thesauri. Journal of Documentation, 31, 158–184.