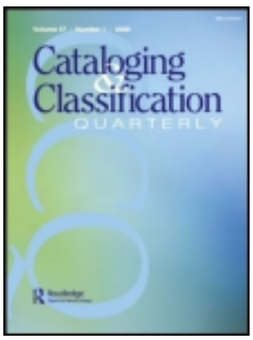


Cataloging & Classification Quarterly

ISSN: 0163-9374 (Print) 1544-4554 (Online) Journal homepage: https://www.tandfonline.com/loi/wccq20

User Comprehension and Searching with Information Retrieval Thesauri

Jane Greenberg

To cite this article: Jane Greenberg (2004) User Comprehension and Searching with Information Retrieval Thesauri, Cataloging & Classification Quarterly, 37:3-4, 103-120, DOI: 10.1300/J104v37n03_08

To link to this article: https://doi.org/10.1300/J104v37n03_08

User Comprehension and Searching with Information Retrieval Thesauri

Jane Greenberg

SUMMARY. While information retrieval thesauri may improve search results, there is little research documenting whether general information system users employ these vocabulary tools. This article explores user comprehension and searching with thesauri. Data were gathered as part of a larger empirical query-expansion study involving the *ProQuest® Controlled Vocabulary*. The results suggest that users' knowledge of thesauri is extremely limited. After receiving a basic thesaurus introduction, however, users indicate a desire to employ these tools. The most significant result was that users expressed a preference for thesauri employment through interactive processing or a combination of automatic and interactive processing, compared to exclusively automatic processing. This article defines information retrieval thesauri, summarizes research results, considers circumstances underlying users' knowledge and searching with thesauri, and highlights future research needs.

KEYWORDS. Thesaurus, thesauri, information retrieval, automatic processing, interactive processing

---

Jane Greenberg is Assistant Professor, School of Information and Library Science, University of North Carolina at Chapel Hill (SILS/UNC-CH). Her research and teaching activities focus on metadata and classification problems.

[Haworth co-indexing entry note]: "User Comprehension and Searching with Information Retrieval Thesauri." Greenberg, Jane. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 37, No. 3/4, 2004, pp. 103-120; and: *The Thesaurus: Review, Renaissance, and Revision* (ed: Sandra K. Roe, and Alan R. Thomas) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2004, pp. 103-120. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

http://www.haworthpress.com/web/CCQ

Digital Object Identifier: 10.1300/J104v37n03_08

INFORMATION RETRIEVAL THESAURI

Information retrieval thesauri, also identified as structured thesauri (Greenberg, 1998, 2001a), present rich semantic networks of vocabulary terms. These tools (hereafter referred to as *thesauri*) are constructed to support document indexing and retrieval. Among several significant features defining information retrieval thesauri and distinguishing them from other vocabulary tools are the following:

- They are created according to an established set of standards. The ANSI/NISO Z39.19-1993, *Guidelines for the Construction, Format, and Management of Monolingual Thesauri* (1994), is one of the most frequently used standards.
- They generally encode equivalent, hierarchical and associative relationships among vocabulary terms (Aitchison et al., 1997, p. 47-66; ANSI//NISO Z39.19 1994, p. 15-21; Lancaster, 1986, p. 35-49).
- They are produced by human processes. Initial construction may stem from automatic processing of electronic discipline-specific text(s), although human intervention is required to encode distinct types of lexical-semantic relationships (e.g., equivalent, hierarchical, and associative relationships).
- They are generally domain-specific tools. For example, the *Thesaurus of ERIC Descriptors* (2001) includes language in the education domain; and the *ASIS Thesaurus of Information Science and Librarianship* (Milstead, 1998) includes language in the library and information science domain.
- Their construction is guided by the principle of *literary warrant* in that thesaurus terminology corresponds to the language used in the published literature of a selected discipline (or disciplines) (Aitchison et al., 1997, p. 47-66, p. 123; ANSI/NISO Z39.19, 1994; and Lancaster, 1986, p. 24-26). Their construction is also guided by *end-user warrant* in that authorized headings are the terms most commonly used by the community(s) for which the thesaurus is designed (Lancaster, p. 26-27).
- They are distinct from both algorithmic or similarity thesauri, which are generated via statistical techniques based on term frequencies, co-occurrence equations, and weighting techniques (e.g., Chen et al., 1995). They are also distinct from general-purpose thesauri, such as *Roget's Thesaurus of English Words and Phrases* (1990), which distinguish grammatical treatments of

words (e.g., nouns, adjectives, adverbs, and verbs) and different senses of these grammatical treatments.

- Finally, they are distinct from subject heading lists because they have been primarily designed to support "post-coordinate searching," whereas subject heading lists contain subject heading strings (e.g., Chinese Americans–Education), inverted subject headings (e.g., Art, Roman), and many more multi-term headings (e.g., Women in Business) because they were initially designed to support "pre-coordinate searching" (Dykstra, 1988).[1]

The features outlined here identify thesauri as unique tools that facilitate the organization and access of information. Studying the current and potential application of thesauri is important–particularly as people increasingly search information systems from the comfort of their own home, wireless network connections, such as the campus coffee shop, or other places without assistance from an information professional. This paper explores questions concerning current thesauri/user relationships in an effort to improve the use of these tools.

THESAURI RESEARCH AND USERS

Studying the interaction between "users and thesauri" is a growing trend. Two key factors motivating this growth include: (1) intensified research efforts in human computer interaction and information seeking behavior, and (2) increased public access to thesauri-supported information systems via the World Wide Web.

The User

A common focus of human computer interaction and information seeking behavior research is the *general information system user* (hereafter referred to as the *user*). This type of user is generally without any professional training in online searching. It's possible that the user may have learned about information retrieval thesauri or controlled vocabularies by participating in a library's bibliographic instruction session. Even so, such users lack the information professional's experience and knowledge garnered from searching thesaurus-supported information systems daily to help clients solve problems. Library and information science researchers want to circumvent general users' searching limitations, increase thesaurus use, and ultimately improve users' retrieval results.

Research Trend-Setters

Setting the "thesauri/user" research trend is a growing body research on thesaurus interface design, end-user warrant, and processing options.

Most *thesaurus interface design* research is construction-oriented, testing new technologies and techniques. The overriding goal is to design user-friendly interfaces that invite and encourage thesaurus use. Hypertext supports easy navigation of terminological semantic relationships encoded in thesauri (e.g., broader term [BT], narrower terms [NT], related terms [RT]) and appears to be one of the most favored technologies (e.g., Shapiro & Yan, 1996). Researchers experimenting with innovative technologies are also developing prototype thesauri with interfaces representing semantic term-relationships via graphical visualization techniques (Rorvig et al., 1999; Ramsey et al., 1999), in three-dimensional space (Hemmje et al., 1994; LyberWorld, 1999), and through animated spatial maps (*Plumb Design Visual Thesaurus*, 1999). *"End-user" warrant* studies assess the degree of matching between users' search terms and thesauri or controlled vocabulary terms (e.g., Carlyle, 1989; Greenberg, 2001a; Humphreys, McCray, and Cheh, 1997). A high degree of matching indicates that a thesaurus is effective, whereas a poor degree of matching leads to questions about thesaurus currency and functionality (Aitchison et al., 1997, p. 1-22; Lancaster, 1986).

Bates' (1986, 1990) influential work on the *end-user thesaurus* advances our thinking in the area of end-user warrant. Through writings, Bates proposes the design of an end-user thesaurus containing "rich" entry vocabulary, allowing users to easily connect to thesaurus terminology. Bates explains this idea with the phrase of "hitting the side of the barn" (1986, p. 365), whereby a user's search need only to map to the thesaurus, but need not exactly match an authorized term. Additionally, end-user thesaurus entry vocabulary may grow over time. In Bates' discussions, the end-user thesaurus supports a much greater degree of mapping compared to an indexing thesaurus (a thesaurus with authorized terms used by an indexer). The end-user thesaurus allows users to benefit from professional indexing without having to learn about the rules and restriction inherent in the indexing thesaurus.

The final thesauri/user trend to note in this literature review is the exploration of *user preferences for automatic* or *interactive thesaurus processing*. Automatic processing requires users to input their initial search term(s). User search terms are then seamlessly mapped to "authorized" thesaurus terminology for search execution. This method appeals to users practicing the *principle of least effort* (Mann, 1993)–that

is, the user not wanting to put time in to the search process and not necessarily worried about retrieving the best documents. Automatic processing can be frustrating for users wanting to understand how their search is manipulated, particularly users wanting to have some control over the search process. Interactive thesaurus processing requires users to select thesaurus terms when initiating a search, or more commonly select additional thesaurus terms through means of query expansion (an iterative process) after evaluating retrieval results of an initial search. A good example of interactive searching with controlled vocabulary is provided with the Okapi experimental system underlying the online catalog at City University, London (Beaulieu et al., 1996; Beaulieu, 1997). Interactive processing appeals to users who want to have some control over their search, although research has shown that users give up during interactive term selection due to the labor intensity surrounding their involvement (Drabenstott & Weller, 1996).

It's likely that the most effective thesaurus processing will come from research exploring both automatic and interactive processing methods. One example in this area is the DARPA Unfamiliar Metadata Project (http://metadata.sims.berkeley.edu/GrantSupported/unfamiliar.html) in the School of Information Management & Systems (SIMS) at the University of California at Berkeley. A part of this project explores "automatic" mapping algorithms for the initial search and cross-language information retrieval. "Interactive" and "automatic" processing methods are then combined for the relevance feedback term selection activities (Buckland et al, 1999). This research also includes statistical means for matching terminology. Research like this seeks to exploit the strengths and minimize the weaknesses of both automatic and interactive processes.

Research Summary

Research highlighted here discusses thesauri and users and addresses thesauri processing for information retrieval. Research needs to also address fundamental questions about user comprehension and employment of thesauri. For example: Do users actually know what an information retrieval thesaurus is and how it operates? Given basic knowledge about thesauri, will users search with a thesaurus? Do users prefer that thesauri be employed via automatic and/or interactive processes during a search activity? Examining these questions will complement existing thesauri/user research, advance our knowledge on this topic, and help improve thesauri implementation in information retrieval systems.

RESEARCH QUESTIONS

The research presented in this article explores user comprehension and searching with thesauri. The research is part of a recent study that focused on automatic and interactive query expansion via lexical semantic relationships encoded in the *ProQuest Controlled Vocabulary and Classification Codes* (1997) (hereafter referred to as the *ProQuest Thesaurus*). This article reports on the portion of the study that explored users' thesauri knowledge, desire to employ a thesaurus when searching, and the preferred processing methods for working with thesauri. Research questions explored were:

1. What knowledge base do general information system users (users) have of information retrieval thesauri (thesauri)?
2. Given basic thesauri knowledge, and assuming thesaurus availability, will users employ a thesaurus when searching?
3. Given basic thesauri knowledge, and assuming thesaurus availability, what thesaurus processing methods do users prefer, if any?

RESEARCH METHODS

The survey method was the primary means for examining the above research questions. Two surveys were implemented: A "Participant Profile Survey," and a "Thesaurus Use Survey," which was a subset of a larger study's Post-evaluation Questionnaire. The surveys used are presented in Appendix A-1 and A-3 in Greenberg (2001a). The examination was also supported by a brief thesaurus introduction. The study took place in an operational setting defined by the real users, real queries, the ABI/Inform database and its underlying thesaurus, the *ProQuest Thesaurus*.

PROCEDURES

Graduate students at Katz Graduate School of Business, University of Pittsburgh, who intended to search ABI/Inform, were recruited for the study. Potential participants had to submit a query where at least one search term mapped to the *ProQuest Thesaurus* via a series of mapping rules. Searches were limited to topical terms and could not include *named entities* (e.g., names of corporate, personal, geographical, or architectural entities). Qualification and mapping rules for participants

are found in Appendix B-1 and B-2 of Greenberg (2001a). The Participant Profile Survey was implemented as part of the recruitment process to gather data on participants' backgrounds, online searching experiences, and thesauri knowledge.

Participants' queries were mapped to the *ProQuest Thesaurus* (at least the one term that matched) and then searched using semantically related thesauri terminology and an extended Boolean algorithm. Participants were then asked to evaluate the relevancy of the retrieval results.

After the relevance evaluations for the retrieval results, participants were given a "brief thesaurus introduction." The introduction included presenting participants with a vocabulary list from the *ProQuest Thesaurus* in the immediate vicinity of their mapped search terms. That is, the *ProQuest Thesaurus* "BT" (broader term), "NT" (narrower term), "RT" (related term), "Use," and "UF" (use for) terms directly connected to their mapped search term(s). These *ProQuest Thesaurus* terms were classified and alphabetically arranged by each participant's original search terms. Common thesaural identifiers of BT, NT, RT, Use, and UF were eliminated from this introduction in order to emphasize the thesaurus as a searching vocabulary. Section A-1 of the Appendix illustrates the presentation of *ProQuest Thesaurus* terms in a classified alphabetized fashion without thesaural identifiers. Thesaurual identifiers are removed because it was speculated that their inclusion and distinguishing authorized headings would distract users from learning about thesauri as basic searching vocabulary tools. For comparison purposes, Section A-2 of the Appendix shows the actual thesaural identifiers (the reference structure) for the terms listed in Section A-1. Participants did *not* see this second display. As an extension of the presentation of thesaurus terms, participants were asked to select any terms that they thought would retrieve additional useful documents. This request was part of a larger interactive query expansion study (Greenberg, 2001b). The thesaurus introduction was followed by the Thesaurus Use Survey, which gathered data about users' desire to employ thesauri and preferred processing methods for working with these tools.

DATA ANALYSIS

Data gathered provided information on participants' backgrounds and online searching experience, thesaurus experience, and preferred processing methods for working with thesauri.

Participants' Backgrounds and Online Searching Experience

Forty-two M.B.A. students participated in the study. The majority of participants (83.4%, 35) were in the first of six required modules (similar to semesters). English was the native language for slightly over half of the participants (52.4%, 22 participants). Comfort level with the English language for non-native English-speakers was recorded on a semantic differential scale ranging from "1" (not very comfortable) to and "5" (very comfortable). Nineteen (95.0%) of the 20 non-native English-speaking participants gave a score 4.0 or higher, 15 of which gave a score of 5.0–the highest value. One participant gave a score of 3.5, which was still above average. It was concluded that the participants' comfort level would not interfere with interpreting of the study's results.

Participants identified themselves as fairly regular searchers of online information retrieval systems (hereafter referred to as *information systems*), as indicated in Table 1.

Searching comfort level for each participant was recorded on a 5-point semantic differential scale ranging from "1" (not very comfortable) to "5" (very comfortable). The mean comfort level score was 4.1, with a mode of 5.0 (45.2%, 19 participants selected 5.0) demonstrating a high degree of comfort in searching information systems.

About one-fifth (8 participants, 19.0%) had searched ABI/Inform prior to participation in the larger query expansion study. Four of these participants indicated they search ABI/Inform one to three times a month. The other four participants selected the "other" option for frequency. Two of these participants indicated that their ABI/Inform searching varied, depending on current assignments and information needs; one participant

TABLE 1. Use of Information Systems*

Search frequency per month	No. of participants
8 or more times	26 (61.9%)
4 to 7 times	9 (21.4%)
1 to 3 times	4 (9.5%)
Varied per month	2 (4.8%)
Never	1 (2.4%)

* *Information systems* were defined as PITTCAT (the University of Pittsburgh's online library catalog), newspaper indexes, digital library resources, but excluded larger Internet databases underlying commercial search engines, such as Yahoo!, Google, Lycos, etc.

noted that ABI/Inform was used at a previous job; and one participant indicated that ABI/Inform is used at a current job, although not on any regular basis and not recently. A 5-point semantic differential scale ranging from "1" (not very comfortable) to "5" (very comfortable) was used to record participants' ABI/Inform comfort level. Only eight responses could be examined for this question. Responses ranged from 1 to 4, with a mean of 2.8. Comfort level corresponded with participants' use frequency in that the participants who searched ABI/Inform more had provided higher comfort-level scores for this system.

Experience with Thesauri

Data gathered on participants' experiences with thesauri provided insight into their knowledge of these tools. Only six participants (14.3%) indicated that they had previously used a thesaurus to aid online searching. Searching thesauri identified by participants are recorded in Table 2. Column three in Table 2 provides a verification note for each identified thesaurus.

Two participants indicated that they had used *Roget's Thesaurus of English Words and Phrases* (hereafter referred to as *Roget's Thesaurus*) for online searching. Although *Roget's Thesaurus* is available online through the University of Pittsburgh Digital Library, it is not available

TABLE 2. Thesaurus Use and Availability Verification

No. of Participants	Thesauri "Named" by participants	Verification Note
2	*Roget's Thesaurus [of English Words and Phrases]*	Thesaurus is available online via the University Library's digital resources webpage, although it is not attached to any online information system.
1	*Wilson Marketing Database Thesaurus*	Thesaurus could not be verified. Participant might be referring to the controlled vocabulary underlying the H. W. Wilson's™ Business Periodicals database.
1	*Web Thesaurus for Business Books*	Thesaurus could not be verified.
2	Could not recall thesaurus used	N/A

as a searching device for any of the online information systems to which the University subscribes. It is conceivable that these two participants have looked up a term in *Roget's Thesaurus* before searching an online information system, but it is more likely that these participants recorded the title "*Roget's Thesaurus*" because it was a thesaurus with which they were familiar. The other two thesaurus titles provided by participants could not be verified, and two participants could not recall the thesaurus previously used.

Only two participants (4.8%) indicated that they were aware that ABI/Inform had a thesaurus. These two participants were among the six participants who indicated they had previously used a thesaurus for online searching. A discrepancy was found as one of these two participants indicated the larger query expansion experiment was their first time using ABI/Inform. In a follow-up question, the participant indicated that this answer was based on an assumption that "many [information] systems have a thesaurus." It is possible that the presentation of this question after the basic thesaurus introduction and term selection activity influenced the participant's response.

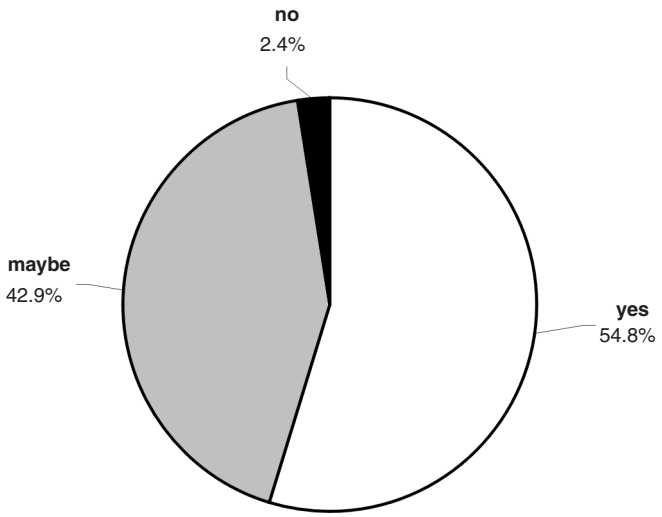
Preferred Processing Methods for Working with Thesauri

The third research question posited in this research explored the preferred processing methods for working with thesauri. The Thesaurus Use Survey asked participants if they would consider using a thesaurus next time they searched ABI/Inform. This inquiry took place after participants completed the relevancy evaluations from larger query expansion study and partook in the thesaurus introduction. Figure 1 presents participants' desire to employ a thesaurus when searching ABI/Inform next time.

A little over half of the participants (54.8%, 23 participants) selected the "yes" option, 42.9% (18 participants) selected "maybe," and one participant (2.4%) selected "no."

Following the thesaurus employment question, participants were asked to select the preferred processing methods for selecting additional search terms from a thesaurus to aid searching. The researcher was on hand to explain the different methods, and participants were encouraged to ask for clarification if they did not understand the different options. A little over ten percent of the participants (5 participants, 11.9%) asked for clarification. Results for preferred processing methods are presented in Table 3.

FIGURE 1. Future Thesaurus Employment During ABI/Inform Searching



Results indicate that the majority of participants favor automatic processing as long as they have the option to manually interact and select terms. Preference for some control is further evidenced by the fact that 15 participants (33.3%) favored exclusively interactive term selection over the four participants (9.5%) that favored exclusively automatic processing. Only two participants (4.8%) didn't indicate a preference.

DISCUSSION

This study gathered data on participants' backgrounds and online searching experience, thesauri experience, and preferred processing methods for working with thesauri. The data gathered permitted exploration of the research questions posited above and provide insight into the topic of thesauri and users.

Users' Thesauri Knowledge

Results found that participants' thesauri knowledge was extremely limited. This was evidenced by the fact that only 14.3% of the participants (6 participants) said they had used a thesaurus for searching before

TABLE 3. Preferred Thesaurus Processing

Options	No. of participants
Automatically.	4 (9.5%)
By giving you a list to choose from.	15 (33.3%)
Automatically and also by giving you a list to choose from.	22 (52.4%)
I don't care.	2 (4.8%)

participating in the study. Only three different thesauri were identified by four participants, and none of the thesauri could be verified as devices associated with any specific information system (see Table 2).

To some extent these results are surprising, given participants' education level and frequency of and comfort with online searching. All participants were pursuing graduate education at the M.B.A. level–a program that requires an earned undergraduate degree. It's likely that these participants had more library experience and interaction with library information systems compared to general "users" at less-advanced educational levels. In turn, it is likely that various information systems used throughout their educational tenure had thesauri, although clearly system interaction doesn't confirm thesaurus use. Nevertheless, the *thesaurus use expectation* for this study was further influenced by the fact that over half of the participants (61.9%, 26 participants) search information systems eight or more times a month. This figure excludes searching databases underlying commercial search engines (e.g., the Yahoo! database). Additionally, close to half of the participants (45.2%, 19 participants) selected the very highest comfort-level score of "5" for online searching, with an average of "4.1" on a 5-point scale. In other words, it was not unreasonable to anticipate greater interaction with thesauri, given the fairly frequent use and high comfort-level found with users' online searching. Contrary to these results, it's possible that access to information systems with thesauri was limited during participants' college education. Data on participants' age, place of undergraduate education, and year of completion were not gathered in this study.

Plans for Thesaurus Employment When Searching ABI/Inform

This study found that a little over half of the participants (54.8%, 23 participants) plan on thesauri searching next time they work with ABI/In-

form and that 18 participants (42.9%) might consider thesaurus searching in ABI/Inform given their response of "maybe" (see Figure 1). It's likely that the basic thesaurus introduction, whereby users were presented with *ProQuest Thesaurus* terms related to their search term, contributed to the positive responses. To reiterate, the thesaurus introduction emphasized the thesaurus as a searching vocabulary, and eliminated thesaurus identifiers and the identification of authorized headings. The results suggest that a good way to encourage thesaurus use is to present searchers with thesauri terms related to their search after they have evaluated an initial set of retrieved documents. Spink (1997) found that search term modification as part of the relevance feedback process improved search results, although participants' search terms were not mapped to a thesaurus in this earlier research. The results of the study reported on here have implication for educating users about thesauri and relate to the following discussion on thesaurus processing methods.

Preferred Processing Methods for Working with Thesauri

Exploring the preferred processing methods for working with thesauri is the final research topic reported on in this paper. Slightly over half of participants (22 participants, 52.4%) were in favor of combining automatic and interactive processing methods. Participants in this group may have been generally pleased with the initial retrieval results from the automatic processing, although they may have wanted to retrieve a greater number of documents–and may have thought that they could have by interactive term selection.

There are very few operational systems supporting both automatic and interactive thesaurus processing. The Okapi system underlying City University's online catalog, which was mentioned above, supports both methods in a unique way. Controlled vocabulary terms are extracted from bibliographic records that the user judges relevant, and these terms are then presented to the user for interactive term selection (Beaulieu et al., 1996; Beaulieu, 1997). Although this system deals with subject headings, it provides a potential model for an information system that could facilitate both automatic and interactive thesaurus processing.

A strong preference for both automatic and interactive processing was followed by a third of the participants (15 participants, 33.3%) favoring *only* interactive processing. This group of participants may have been less pleased with the initial retrieval results from the automatic processing. It's likely that this group of 15 participants perceived a greater potential to retrieve additional relevant documents via interactive processing.

It's possible that some participants in this group and the group preferring both processing methods may have been aware of the advantages associated with human/manual indexing compared to the weaknesses of automatic indexing. Data supporting this possibility were not gathered.

Reasons given for participants' preferences are difficult to verify. This is in part because the results reported on here stem from a larger query expansion study. Participants' relevance evaluations were based on records retrieved via both their initial search terms and *ProQuest Thesaurus* terms semantically related to their initial search terms. Participants were therefore exposed to thesaurus terms at this early phase of the experiment without being informed.

Despite the limitations noted here, the complete set of results show that 37 participants (85.7%) favor some form of interactive thesaurus use, compared to only four participants (9.5%) exclusively in favor of automatic processing. The population of 37 participants is based on the 22 participants endorsing automatic and interactive processing together and the 15 participants endorsing, exclusively, interactive processing of thesauri. Those preferring only automatic processing may have perceived this as the easiest means of getting information and were likely pleased with their initial retrieval results. Additionally, the two participants (4.8%) with no preference may have predicted that both methods produced equally satisfactory results, although it's conceivable that they didn't understand or take time to think about the difference between both processes. Regardless of the exact reasons behind preferences found, the high percentage of participants supporting some level of interactive processing is a *significant* finding that researchers need to take into account in designing thesauri applications.

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The study explored user comprehension of information retrieval thesauri, thesaurus searching, and preferred processing methods for working with thesauri. Participants were educationally advanced students pursing M.B.A. degrees, with business queries that were mapped to the *ProQuest Thesaurus*. The results suggest that:

- Users' thesauri comprehension is extremely limited.
- Given a basic thesaurus introduction, users indicate a desire to employ these tools.
- Given a basic thesaurus introduction, users favor either interactive or a combination of automatic and interactive thesaurus processing compared to completely automatic processing.

The results of this study are useful in that they provide insight into user comprehension of thesauri, their desire to use these tools, and processing preferences. The results are, however, limited by the participant population, the nature of queries, ABI/Inform's contents, and the *ProQuest Thesaurus*. Other limitations are attributed to the fact that the study was a peripheral part of a larger query expansion study. Even so, the study provides a framework for future thesauri/user examinations. Future investigations addressing the topics underlying this study might be improved by gathering data about participants' previous information system use during college or other advanced education, and by specifically asking for feedback about the thesaurus processes preferences.

This study raises questions about the emphasis thesaurus developers and researchers place on producing systems seamlessly linking user search terms to thesaurus terms during information retrieval activities. Indeed, these efforts are important because they aim to eliminate user burdens associated with learning about thesauri intricacies. The emphasis in this area has, however, limited the attention given to different types of users, particularly users who may prefer some degree of thesaurus interaction. In efforts to fully take advantage of thesauri, research needs to also consider how to stimulate user exploration of thesauri and meet the needs of users like the participants in this study.

Future thesauri research needs to also consider current system design limitations and user behavior. Current thesaurus-supported systems often fail to adequately highlight a thesaurus search option. Information systems may include the word "Thesaurus" on a navigation bar or as a hypertext button, but the explanation of how this feature can assist with the selection of search terms may be hidden. Additionally, systems that include a thesaurus often provide confusing interfaces. They use thesaural identifiers like "BT" and "NT" or phrases like "broader term," which may not be clear to a user (non-professional searcher), and they have limited tutorials or explanations of the thesaurus feature.

In the area of user behavior, attention needs to be given to increasing participation in bibliographic instruction sessions and also to creating and improving thesauri tutorials that invite user exploration. Additional research may help determine what factors influence a user's decision to retain knowledge about information retrieval thesauri. For example, how might ease of thesaurus access, a good or intuitive thesauri tutorial (online or personal), improved results during first-time use, or repetitive use impact a user's acquisition and ability to retain thesauri knowledge? Answers to these questions may improve thesauri employment by users.

Thesauri are intellectual creations. They are valuable for indexing and retrieving information. Increasingly, information systems with thesauri are being linked to web portals. Digital libraries and other web initiatives are adopting thesauri for organizing information. People are searching these and other information systems that include thesauri twenty-four/seven (around the clock), without assistance from an information professional. Thesauri developers and researchers need to better understand the current thesauri/user relationship, and highlight the splendid nature of these tools in order to improve user thesauri employment.

NOTE

1. This distinction is becoming less of an issue, as *subject heading lists* go through a process of *thesaurification*. For example, string headings are disconnected, inverted headings are reversed, multi-term headings are deconstructed, and thesaural abbreviations are added (e.g., BT, NT, etc.) to identify semantic relationships. At the same time, online systems increasingly support post-coordinate searching of individual terms still found in subject headings strings, inverted subject headings, and multi-term concepts.

REFERENCES

Aitchison, Jean, Alan Gilchrist, and David Bawden. *Thesaurus Construction and Use: A Practical Manual*, 3rd ed. London: Aslib, 1997.

ANSI/NISO. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. Bethesda, MD: NISO Press, 1994. ANSI/NISO Z39.19-1993.

Bates, M. J. "Design for a Subject Search Interface and Online Thesaurus for a Very Large Records Management Database." In *Proceedings of the 53rd ASIS Annual Meeting, Toronto, Ontario, November 4-8, 1990*, edited by Diane Henderson, 20-28. Medford, N.J.: Published for the American Society for Information Science by Information Today, 1990.

Bates, M. J. "Subject Access in Online Catalogs: A Design Model." *Journal of the American Society for Information Science* 37, no. 6 (1986): 357-376.

Beaulieu, M. "Experiments of Interfaces to Support Query Expansion." *Journal of Documentation* 53, no. 1 (1997): 8-19.

Beaulieu, M., S. Robertson, and E. Rasmussen. "Evaluating Interactive Systems in TREC." *Journal of the American Society for Information Science* 47, no. 1 (1996): 85-94.

Buckland, Michael, Aitao Chen, Hui-Min Chen, Youngin Kim, Byron Larn, Ray Larson, Barbara Norgard, and Jacek Purat. "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies." *D-Lib Magazine* 5, no. 1 (1999). Available online at URL: http://www.dlib.org/dlib/january99/buckland/01buckland.html.

Carlyle, A. "Matching *LCSH* and User Vocabulary in the Library Catalog." *Cataloging & Classification Quarterly* 10, no. 1/2 (1989): 37-63.

Chen, Hsinchun, Tak Yim, David Fye and Bruce Schatz. "Automatic Thesaurus Generation for an Electronic Community System." *Journal of the American Society for Information Science* 46, no. 3 (1995): 175-193.

Drabenstott, Karen Markey and Marjorie S. Weller. "Failure Analysis of Subject Searches in a Test of a New Design for Subject Access to Online Catalogs." *Journal of the American Society for Information Science* 47, no. 7 (1996): 519-537.

Dykstra, M. "LC Subject Headings Disguised as a Thesaurus." *Library Journal* 113, no. 4 (1988): 42-46.

Greenberg, Jane. "An Examination of the Impact of Lexical-Semantic Relationships on Retrieval Effectiveness During the Query Expansion (QE) Process." Completed in partial fulfillment of University of Pittsburgh, School of Information Sciences, 1998.

Greenberg, Jane. "Automatic Query Expansion via Lexical-Semantic Relationships." *Journal of the American Society for Information Science and Technology* 52, no. 5 (2001a): 402-415.

Greenberg, Jane. "Optimal Query Expansion (QE) Processing Methods with Semantically Encoded Structured Thesauri Terminology." *Journal of the American Society for Information Science and Technology* 52, no. 6 (2001b): 487-498.

Hemmje, Matthias, Clemens Kunkel, and Alexander Willet. "LyberWorld–A Visualization User Interface Supporting Fulltext Retrieval." In *Proceedings of 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1994 July 3-4, Dublin, Ireland*, 249-259. Berlin: Springer-Verlag, 1994.

Humphreys, B. L., A. T. McCray and M. L. Cheh. "Evaluating the Coverage of Controlled Health Data Terminologies: Report on the Results of the NLM/AHCPR Large Scale Vocabulary Test." *Journal of the American Medical Informatics Association* 4, no. 6 (1997): 484-500.

Lancaster, F. W. *Vocabulary Control for Information Retrieval*, 2nd ed. Arlington, VA: Information Resources Press, 1986.

LyberWorld homepage. (2002): http://www.darmstadt.gmd.de/~hemmje/Activities/Lyberworld/.

Mann, Thomas. *Library Research Models: A Guide to Classification, Cataloging, and Computers.* New York: Oxford University Press, 1993.

Milstead, Jessica L., ed. *ASIS Thesaurus of Information Science and Librarianship*, 2nd ed. Medford, NJ: Published for the American Society for Information Science by Information Today, 1998.

Plumb Design Visual Thesaurus. (1998). http://www.visualthesaurus. com/index.jsp.

ProQuest® Controlled Vocabulary and Classification Codes. Ann Arbor, MI: UMI, 1997.

Ramsey, Marshall C., Hsinchun Chen, Bin Zhu, and Bruce R. Schatz. "A Collection of Visual Thesauri for Browsing Large Collections of Geographic Images." *Journal of the American Society for Information Science* 50, no. 9 (1999): 826-834.

Roget's Thesaurus of English Words and Phrases. New York: Portland House, 1990. [Originally authored by Dr. Peter Mark Roget in 1852.]

Rorvig, M. E., C. H. Turner and J. Moncada. "The NASA Image Collection Visual Thesaurus." *Journal of the American Society for Information Science* 50, no. 9 (1999): 794-798.

Shapiro, Celia D. and Puck-Fai Yan. "Generous Tools: Thesauri in Digital Libraries." *National Online Meeting Proceedings–1996, Proceedings of the 17th National Online Meeting, New York, May 14-16, 1996*, 323-332. Medford, NJ: Information Today, 1996.

Spink, Amanda. "Information Science: A Third Feedback Framework." *Journal of the American Society for Information Science* 48, no. 8 (1997): 728-740.

Thesaurus of ERIC Descriptors, 14th ed. James E. Houston, editor/lexicographer. Phoenix, AZ: Oryx Press, 2001.

APPENDIX. Term Selection Test (Part of the Thesaurus Introduction)

A-1

Classified-alphabetized lists presented to a participant.*

The query submitted by the participant was: "Entertainment Industry AND Market Potential."

Dear Participant:
Please circle the search terms that appear to be, or that you think would have been useful to your search.

Entertainment Industry
Amusement industry
Amusement parks
Broadcasting industry
Casinos
Celebrities
Entertainers
Entertainment technology & design
Home entertainment industry
Motion picture industry
Music industry
Radio broadcasting
Recording industry
Reservation systems
Sports & recreation clubs
Television broadcasting
Tickets
Video industry

Market Potential
Commercial markets
Commercialization
Demand analysis
Market research
Market saturation
Market strategy

*The presentation of these terms introduced participants to thesauri as a source for additional searching terminology. The request for the selection of additional search terms was part of a larger interactive query expansion study (Greenberg, 2001b).

A-2

***ProQuest*® *Controlled Vocabulary* (1997) reference structure for the terms listed in A-1.**

Entertainment Industry
UF: Amusement industry
NT: Amusement parks
Broadcasting industry
Casinos
Home entertainment industry
Motion picture industry
Music industry
Radio broadcasting
Recording industry
Reservation systems
Video industry
RT: Celebrities
Entertainers
Entertainment technology & design
Sports & recreation clubs
Television broadcasting
Tickets

Market Potential
RT: Commercial markets
Commercialization
Demand analysis
Market research
Market saturation
Market strategy