# Hierarchical Navigation: An Exploration of Yahoo! Directories

Hur-Li Lee* and Hope A. Olson**

*/**University of Wisconsin-Milwaukee, School of Information Studies,
P.O. Box 413, Milwaukee, Wisconsin 53201-0413, USA,
*hurli@uwm.edu **holson@uwm.edu

Hur-Li Lee is an Associate Professor in the School of Information Studies at the University of Wisconsin-Milwaukee. Her major research interests include the organization of information; users' interactions with structures of information systems; and social and cultural aspects of information and information services. She and Dr. Hope A. Olson are currently principal investigators for a major grant from the Institute of Museum and Library Services (U.S.) to recruit diverse students to specialize in the organization of information.

Hope A. Olson is a Professor in the School of Information Studies at the University of Wisconsin Milwaukee. Dr. Olson's research addresses classification theory and problems of bias in subject access to information using feminist, poststructural, and postcolonial perspectives. Among her current interests are the cultural specificity of classificatory structure, structures of indigenous knowledge, searcher navigation of hierarchical structures (with Dr. Hur-Li Lee), and the nature of consistency (with Dr. Dietmar Wolfram). She has published: The Power to Name: Locating the Limits of Subject Representation in Libraries (Kluwer Academic, 2002); Information Resources in Women's Studies and Feminism as editor (K.G. Saur, 2002); and Subject Analysis in Online Catalogs 2nd ed. with John J. Boll (Libraries Unlimited, 2001) and was editor-in-chief of the journal Knowledge Organization from 2000 to 2004.

Lee, Hur-Li, and Hope A. Olson. **Hierarchical Navigation: An Exploration of Yahoo! Directories.** *Knowledge Organization, 32(1).* 10-24. 37 refs.

**ABSTRACT:** Although researchers have theorized the critical importance of classification in the organization of information, the classification approach seems to have given way to the alphabetical subject approach in retrieval tools widely used in libraries, and research on how users utilize classification or classification-like arrangements in information seeking has been scant. To better understand whether searchers consider classificatory structures a viable alternative to information retrieval, this article reports on a study of how 24 library and information science students used Yahoo! directories, a popular search service resembling classification, in completing an assigned simple task. Several issues emerged from the students' reporting of their search process and a comparison between hierarchical navigation and keyword searching: citation order of facets, precision vs. recall, and other factors influencing searchers' successes and preferences. The latter included search expertise, knowledge of the discipline, and time required to complete the search. Without a definitive conclusion, we suggest a number of directions for further research.

## 1. Introduction

Bibliographic classification has long been a standard device for organizing libraries' print collections on the shelves. However, the classified approach seems to have given way to the alphabetical subject approach in retrieval tools widely used in libraries. Until the nineteenth century, classified catalogs were the norm in most libraries. With Charles Cutter's *Rules for a Printed Dictionary Catalogue* in 1876, the shift to alphabetical subject catalogs became entrenched in the United States. Libraries in other parts of the

world maintain a mixture of classified and alphabetical subject dictionary catalogs, although the latter are prevalent, perhaps because of the globalization of an American model through the distribution of cataloging copy. A few indexing and abstracting databases utilize classifications as do some individual journals. However, the most widespread use of classification-like arrangements for information retrieval, other than material arrangements on library shelves, is probably in the directory functions of some internet search engines. Yahoo! (http://www.yahoo.com/) began in 1994, using a hierarchical browsing structure exclusively, and still has one of the deepest hierarchies. For some time, Yahoo! has not been alone in utilizing this type of structure. Many other search engines also have hierarchical directories, including Google, the most popular search engine in the United States, Yahoo! being second (Sullivan 2004). It makes one wonder: Why do successful enterprises like Yahoo! and Google invest in something akin to the classified catalogs that American libraries abandoned over a century ago?

When online searching first became readily available, popular debate pitted verbal controlled vocabulary against natural language (Svenonius 1986). That debate was more or less dissolved by the widespread provision of both options in many applications. Keyword searching of both natural language and, where available, controlled vocabulary has now evolved as a common approach from Internet search engines like Yahoo! to library online catalog interfaces that frequently make keyword searching the default basic search.

The hierarchical navigation enabled by classificatory structures provides a third search mechanism often viewed as an alternative to keyword approaches to retrieval whether through controlled or natural language (see Mai 2004). These two options, hierarchical navigation and keyword searching, have sometimes been said to represent different paradigms (Hildreth 1995) or even "distinct patterns of human behavior" (Dodd 1996, 281). Hierarchical navigation may use traditional classifications or it may use classification-like schemes such as Yahoo! directories or the structure of files within folders within folders on one's computer. Not only does a classificatory structure gather the resources on a particular subject, it puts them in logical proximity to resources on closely related subjects in the same class. Such structures help to fulfill the collocation function in organizing information and allow searchers to browse in a way that verbal systems do not.

A classificatory structure also facilitates navigation up and down a hierarchy. Charles Cutter (1904, 79) recognized the shortcomings of verbal systems to allow this type of browsing:

> Subject-entries, individual, general, limited, extensive, thrown together without any logical arrangement, in most absurd proximity – *Abscess* followed by *Absentee-ism* and that by *Absolution*, *Club-foot* next to *Clubs*, and *Communion* to *Communism*, while *Bibliography* and *Literary history*, *Christianity* and *Theology*, are separated by half the length of the catalogue – are a mass of utterly disconnected particles without any relation to one another, each useful in itself but only by itself.

He also noted that catalog users of his time tended to search for broader topics than they actually sought (Cutter 1904, 67). While Cutter attributed this tendency to his public's previous use of classified catalogs, more recent research suggests that it is still the case where users have no prior experience with classified catalogs (Cousins 1992; Drabenstott and Vizine-Goetz 1994; Larson 1991). It seems logical to deduce that classification-like arrangements continue to be useful.

The question is: Do searchers consider classificatory structures a viable alternative or supplemental approach to information retrieval? That is the general question that leads to this specific study. More specifically: How do searchers use these directories in this age of keyword searching? Answers to this question can inform our understanding of how people utilize classification as an instrument of information retrieval; and this knowledge, in turn, can assist in future system design.

The study reported here is an exploratory study intended to address the following research questions:

1. Do searchers use hierarchical directories successfully? Are they able to navigate the hierarchies?
2. Do searchers prefer hierarchical navigation or keyword searching?
3. What factors potentially influence searchers' success and preferences?
4. What characteristics of classification are germane to searchers' performance?

This initial project analyzes the work of library and information studies students assigned to compare directory and keyword searching on a prescribed topic

in Yahoo! The students, in the first level of their study of organization of information, are aware of basic information retrieval concepts, but have not yet attained a professional level of expertise. Therefore, they combine some remaining naïveté in regard to organization of information with the ability to describe their search processes and perceptions.

## 2. Literature Review

Jens-Erik Mai (2004, 93-94) notes that most research on improving access to Web documents through organization of information focuses on descriptive aspects such as author and title and what little research has been done on the use of classification relates primarily to the use of traditional library classifications. He suggests that "the community of bibliographic knowledge organization theory and practice has not been able to make its knowledge available to the Web directory community." Conversely, we have found that little use has been made of the Web as a laboratory for understanding classificatory structures as retrieval tools. Alan Wheatley (2000, 140) notes that "subject trees [in Web search engines] have not been examined by performance testing such as that applied to many other information retrieval techniques." In fact, the entire body of research on how people navigate using classification or classification-like schemes is scant.

The lack of research on navigation using classificatory structures may be at least partly attributable to the information retrieval research model established by the Cranfield tests as discussed by David Ellis and Ana Vasconcelos (2000, 110). Further, the first Cranfield study, in comparing the application of the *Universal Decimal Classification*, an alphabetical subject catalogue, a faceted classification scheme, and Uniterms, found that simple indexing systems such as Uniterms can be effective. The second Cranfield study affirmed that "natural language, with slight modifications of confounding synonyms and word forms, combined with simple coordination, can give a reasonable performance" (Cleverdon and Keen 1966, 263). This emphasis on natural language has drawn attention away from controlled vocabularies, including classification.

Fortunately, some research in navigating library catalogs through classifications has been conducted during the intervening decades. Karen Markey (1986) explored the use of the *Dewey Decimal Classification* (*DDC*) in a large scale experiment using catalogs in four major libraries. She found that "class number searching was a successful subject searching strategy" and made suggestions for its effective implementation (1986, 150). Unfortunately, today's OPAC interfaces have not fulfilled the promise of searching via classification in spite of IFLA's invocation to "display the hierarchical relationship between a classification number and the entire classification" (principle 22, p.47, in its *Guideline for OPAC Displays*, Yee 1998), which implies the efficacy of hierarchical navigation. Christine Borgman and others (1995) tested a *DDC*-based hierarchical browsing system for children's use with similar success. The results of these two studies suggest that further examination of users' navigation using hierarchical structures is potentially fruitful.

More recent research has explored hierarchical Web directories, but not in relation to navigation or searching. David G. Dodd (1997) compared the main classes in directories with those in library classifications and compared the results of keyword searches to hierarchical navigation using examples of both known-item searches and topic searches. His data are not extensive, but do suggest that hierarchical navigation using directories was more effective for topic searches, at least in the late twentieth-century Web. Marthinus S. Van der Walt (1997 and 1998) looked at the hierarchical structures of Web directories including specificity and citation order. Michèle Hudon (2003) made more in-depth comparisons in the field of education. Van der Walt's and Hudon's findings show both similarities and differences between the structure of the directories and that of library classifications, suggesting that transfer of findings between the two should be approached with caution.

Most recently, a study by Said Mirza Pahlevi and Hiroyuki Kitagawa (2005) demonstrated the potential of hierarchical directories in the retrieval process. Pahlevi and Kitagawa developed and tested a mechanism that combined natural language search terms with directory categories, probed the directories for relevant material, derived search modifiers from that material, and then combined the derived modifiers with the original natural language, thus creating a refined search that can be run in any search engine. Their purpose in deriving search modifiers from directory categories was to develop more precise searches. Searchers can take advantage of the hierarchical navigation to identify a category, and the resultant modifiers provide a context apart from the human-indexed categories. Ellis and Vasconcelos (1999) agree that the context provided by directories

is important in Web searching and eases the need to generate search terms.

Both the research and non-research literatures endorse the use of hierarchical navigation. Nancy Williamson (1997, 24) noted that: "classification aids in retrieval through the use of logical structure and helpful order to facilitate browsing and filtering of large quantities of data. It also has the potential to make possible multilingual access and improved interoperability with other services." The ease of hierarchical navigation in library catalogs was borne out by Borgman et al. who discovered that children abandoned keyword searches more frequently than hierarchical ones because of the need to generate and spell search terms (1995, 682). Bob Ainsbury (2002) suggests the same advantages for corporate portals, as does Martin White (2001) in the context of intranets.

The main process that facilitates retrieval using hierarchical navigation is browsing. As Julian Warner (2000, 37) noted: "The value of an information system could then be the ability it offers discriminatingly to follow 'paths and tracks, however slight.' Classification schemes themselves … can then be received not as fixed models of stable entities but as valuable exploratory devices." The value of browsing is recognized through the DESIRE (Development of a European Service for Information on Research and Education) project. Traugott Koch, Michael Day, and others (1997) elaborate on enabling the broadening and narrowing of topics and identification of context in hierarchical browsing. For DESIRE, the potential for classification to accommodate multilingual access is also important. Others have also suggested classification as a potential switching language in multilingual contexts (McIlwaine 2003).

Lois Mai Chan, Xia Lin, and Marcia Zeng (1999) noted that hierarchical browsing "improves precision by first defining and narrowing the domain for searching." It does so by collocating (inclusion) and partitioning (exclusion). They elaborated on earlier recognition of hierarchical Web directories as precision devices by Mary Micco (1996), Greg R. Notess (1997), Thomas Pack (1999), and others.

The directory used in this project is from Yahoo! It has been based on literary warrant with categories established as sites on new topics were received (Callery 1996; Steinberg 1996). It is, as Bella Hass Weinberg (1999) points out, an alphabetico-classed system. So although it benefits from a hierarchical structure, it does not include a notation and, therefore, has the limitations of the alphabet within a given category. Yahoo! is generally second only to Google in use and popularity as documented throughout Danny Sullivan's *Search Engine Watch* website. Yahoo!'s directory is the most specific of the widely used search engines (Wheatley 2000, 137). Hence, it is a reasonable environment in which to explore the use of hierarchical navigation.

## 3. Research Method

The purpose of this study was to identify approaches taken by students in navigating across Yahoo! directory hierarchies and to compare navigation with keyword searching. Because of its exploratory nature, the researchers chose to conduct the study with a small, convenient sample. The results, therefore, are limited by the nature of the study and its participants. Future research will be required to establish firm conclusions.

In 2003, one of the researchers asked students in two online classes to complete an assignment independently over a two-week period by using the Yahoo! services. Both classes were sections of a required course, Organization of Information, in the Master's in Library and Information Science (MLIS) program at the University of Wisconsin-Milwaukee. Some of the students were beginning MLIS students and others had taken other MLIS classes in previous semesters. According to information about themselves provided by students in both classes, none of them had had any prior formal training or knowledge in information organization in general or in classification theory specifically. Although as a group they might be more experienced searchers than average users, their experience with the Internet was overwhelmingly limited to keyword searching; few indicated that they had used Yahoo! directories in a substantive way.

When they received instructions for the assignment, students were given information about the study explaining to them that participation in the study was voluntary, would not affect their grades, and did not require any extra work. An additional step taken to ensure confidentiality was to ask those volunteering to participate to send their consent forms to the other researcher who was not the instructor of the classes. Students were assured that the instructor did not know which students would or would not participate in the study before grading and that data extraction would begin only after all grades were submitted. Out of a total of 44 students, 26 agreed to participate in the study. Twenty-four of the submissions were used in the study and two were

excluded because the two students did not follow the assignment instructions. When the texts were extracted for data analysis, all students' names were replaced by numbers that were assigned in no particular order.

The assigned task was to find five national or international scholarly or professional organizations in the field of microbiology by conducting two types of searches in Yahoo!: navigating across directories and searching by keyword (the text of the assignment is appended to this article). Due to the fact that these students were beginners in classification theory and that they came from varied backgrounds, the topic assigned was intentionally straightforward, not requiring any subject knowledge or advanced skill in using classification. To simplify grading, students were asked to use "Science" as the starting category. The assignment instructions clearly specified that navigation using the Yahoo! directories should be completed first because it would be more interesting to see students' own paths in this kind of search. The hierarchical paths along the directories indicated by Yahoo! in keyword search results might influence students' judgment in navigating the hierarchies and thus the keyword search had to be conducted after the navigational search. Although the results of navigation might influence terms chosen in the keyword search, this influence was considered inconsequential in this instance because selection of keywords was not a central concern in either the assignment or the study. Students were then asked to write a two-page comparison of the two distinct searches. Those comparative narratives provided the data for this study.

Data analysis was of a qualitative nature. Categories were developed by the researchers as derived from the texts of students' papers. Categories were not stipulated in the assignment (see appendix). Each researcher independently coded a subset of assignments, developing categories on the basis of the data. These categories were then merged and differences of interpretation were resolved. The researchers then coded all of the data, recording the preliminary subsets.

## 4. Results

### *4.1. Hierarchical Paths Taken*

As indicated in Table 1, 21 of the 24 students participating in the study took the same path in navigating the Yahoo! directories: "Science" to "Biology" to "Microbiology" to "Organizations" (Path A). Out of these 21 students, 15 did not attempt any other paths. However, it cannot be concluded that the majority of the students saw this as the *only* logical path, because once they successfully completed the task required in the assignment there was no reason for them to explore other options. On the other hand, three others tried more paths, mostly other categories at one particular level of the hierarchy, after they examined the results at the end of Path A; of these, 2 simply wanted to see more and 1 gave a specific interpretation of the task and determined that the sites listed at the end of Path A were not completely satisfactory. The other 3 students exploring alternatives to Path A considered Path A to be one of the options that were all equally logical. Of the remaining 3 students who did not take Path A, 1 took Path H, 1 took Path I, and the last did not indicate the path taken.

| Path Taken | Number of Students | No. of Students Who Took This Path Exclusively |
|---|---|---|
| A. Science→Biology→Microbiology→Organizations | 21 | 15 |
| B. Science→Biology→Microbiology→Web Directories | 1 | 0 |
| C. Science→Biology→Parasitology→Organizations | 1 | 0 |
| D. Science→Biology→Organizations | 1 | 0 |
| E. Science→Biology→Organizations→Microbiology | 3 | 0 |
| F. Science→Biology→Organizations→Professional | 1 | 0 |
| G. Science→Medicine | 1 | 0 |
| H. Science→Medicine→Microbiology and virology→Organizations | 1 | 1 |
| I. Science→Organizations→Biology→Microbiology | 3 | 1 |
| J. Science→Research | 1 | 0 |
| Not indicated | 1 | N/A |

*Table 1. Hierarchical paths taken by the students.*

| Keyword | Number of Students |
|---|---|
| Microbiology | 24 |
| Organization(s) | 23 |
| Professional | 8 |
| Association(s) | 4 |
| National | 1 |
| Scholarly | 1 |
| Society | 1 |

*Table 2. Keywords used by the students.*

| Number of Keywords | Number of Students |
|---|---|
| 2 keywords | 12 |
| 3 keywords | 11 |
| 5 keywords | 1 |

*Table 3. Number of keywords used*

| Search Techniques | Number of Students |
|---|---|
| Boolean | 3 |
| Quotation marks | 3 |
| Advanced search | 1 |
| Not indicated | 19 |

*Table 4. Keyword search techniques applied*

| Advantage | # of Students | Disadvantage | # of Students |
|---|---|---|---|
| Quick, easy, and not time-consuming | 13 | Requires more steps | 3 |
| High precision | 11 | Low recall | 11 |
| Showing relationships between topics | 8 | Limited to the structure developed by Yahoo | 3 |
| Useful for experts or someone familiar with the topic | 3 | Requires knowledge of a particular subject hierarchy | 8 |
| No need to generate search terms | 2 | Subjectivity in categories and assignment of sites to categories | 1 |
| No need to know advanced search techniques | 1 | Requires knowledge of hierarchical search | 4 |
| | | Updated slowly | 1 |

*Table 5. Advantages and disadvantages of hierarchical navigation.*

### *4.2. Keywords and Search Techniques Applied*

Table 2 lists the words used by the participating students in the keyword search. All students but one entered either "microbiology organization" or "microbiology organizations." The last student typed in "microbiology associations," instead. Three students tried both "organization(s)" and "association(s)"; another student used both "organization" and "society." The data did not indicate why few of the students considered synonyms. Three of the keywords (national, scholarly, and society) came up only once each. According to Table 3, one half of the students applied only 2 keywords, 11 others used 3 keywords, and only 1 student typed in 5 keywords. Among the 3-keyword searchers, 7 used "microbiology," "organization(s)," and "professional". The person who used 5 keywords included the string "national scholarly professional organizations microbiology".

Among the 24 participants, only 5 consciously applied one or more complex techniques in the keyword search (Table 4). Judging from other students' search results (typically in tens or hundreds of thousands), it was clear that they simply typed in the chosen keywords as one character string and clicked the search button. The two most frequently applied advanced search techniques were Boolean ("and" and "or") and quotation marks to search as a phrase. Only one student took the step to go to the advanced search page.

### *4.3. Advantages and Disadvantages of Hierarchical Navigation*

In comparing the two types of searches, the students offered 6 advantages of hierarchical navigation (Table 5). First, possibly due to the fact that the assigned topic was simple and straightforward, 13 students in-

dicated that navigating through the Yahoo! directories was quick, easy, and not time-consuming. Some of them clearly understood that other more complex topics might require lengthy exploration of numerous unfamiliar categories. Eleven students mentioned the high precision rate in the search results: with the exception of one broken link, all sites retrieved in this search were relevant. To 8 of the students, Yahoo! directories presented hierarchical relationships between topics and thus could assist users in identifying broader or narrower topics that were potentially more appropriate. Three students believed that this type of search was useful for people who were able to follow easily the hierarchical path because of their familiarity with the general area of the search topic. On the other hand, 2 other students thought that this type of search was useful for novices if they were unable to come up with good keywords or were unfamiliar with advanced keyword search techniques.

The disadvantages of hierarchical navigation are listed in Table 5. Nearly half (11) of the students noticed the low recall rate in search results. In other words, this type of search did not retrieve many of the relevant sites they found using the keyword search. Eight students thought that this type of search would require the searcher to have a certain level of knowledge about the hierarchical structure in the subject area and 4 students believed that it would require a basic understanding of the concept of a hierarchical arrangement. Their view was that the average person would have difficulties with hierarchical navigation. Interestingly, 3 students said that the structure of categories was developed by Yahoo! and that users were limited by this home-grown structure. Another student also expressed a concern over the subjectivity in Yahoo's categories and in the way that sites were assigned to individual categories. Three students complained that it took many more steps to go through the categories as compared to a keyword search that took only one step: entering keywords. None of these students noted that in addition to steps taken in conducting the search, steps taken to evaluate retrieved sites should also be considered. The last disadvantage, according to 1 student, was the slowness of Yahoo! in updating its directories so that new sites would most likely be excluded.

### *4.4. Advantages and Disadvantages of Keyword Searching*

As shown in Table 6, high recall rate in search results was the number one advantage of keyword searching mentioned by half of the students (12). They drew this conclusion because they identified many more relevant organizations through keyword searching than through hierarchical navigation. A second factor seen as an advantage by 9 students was access to other directories, searching aids, and pathfinders. In addition, 4 other advantages of keyword searching were listed by 1 or 2 students: user-friendly, more up-to-date, no need to have subject knowledge, and more useful for an experienced searcher or some one with subject knowledge. The last one was quite interesting. The student argued, "through key word searching one often gets many false drops, but if the searcher is an experienced searcher or has subject experience, he or she may know subject terms to aid in the search thus eliminating many non-relevant terms that may result in false drops."

A majority of the students (16) saw its time-consuming nature as a significant disadvantage of keyword searching (Table 6). The second most

| Advantage | # of Students | Disadvantage | # of Students |
|---|---|---|---|
| High recall | 12 | Low precision | 10 |
| As access to others' directories, searching aids, or path finders | 9 | | |
| Time-saving | 2 | Time-consuming | 16 |
| User-friendly | 2 | | |
| More up-to-date | 1 | | |
| Best if the user had little subject knowledge | 1 | Need to know useful keywords | 6 |
| Good for experienced user | 1 | Need to know advanced search techniques to be effective | 5 |
| | | Possibly missing relevant sites if they are not on first few pages | 2 |

*Table 6. Advantages and disadvantages of keyword search.*

common complaint about keyword searching given by those students (10) was the overwhelming number of sites retrieved, a large percentage of which were irrelevant (i.e., low precision rate). Six students pointed out that users needed to come up with useful keywords on their own and not many users would be able to do so. In addition, 5 students thought that a keyword search would only be effective if the user knew how to apply advanced search techniques to reduce the size of the retrieved results. Lastly, 2 students mentioned the possibility of missing relevant sites that did not show up on the first few screens because few users would go beyond that point.

### *4.5. Preference Between The Two Methods*

Table 7 shows that in completing the assignment, 6 students preferred hierarchical navigation, 3 preferred keyword searching, 8 indicated that it would depend on a user's need or situation, and 7 did not give any preference.

| | **Preference** |
|---|---|
| Hierarchical navigation | 6 |
| Keyword search | 3 |
| Depending on needs/situations | 8 |
| Not indicated | 7 |

*Table 7. Students' preferences regarding search methods applied.*

## 5. Discussion

The data collected in this study raise four main factors in comparing hierarchical navigation and keyword searching: the hierarchical structure itself, including issues of specificity and citation order; the influence of these two approaches on precision and recall; time consumed in retrieval; and knowledge and skill required for retrieval.

### *5.1. Hierarchies and Categories*

Possibly because the assignment was intended to be uncomplicated, most of the participants had little difficulty navigating Yahoo! directories in completing the required task. The majority of them (21 out of 24) followed the hierarchical Path A to find the desired results. Nevertheless, the descriptions of the paths attempted and the students' reflections on this type of hierarchical navigation provide possible insight into general issues related to the use of hierarchies and categories in searches. These issues included (1) the hierarchical relationship; (2) the principle of specificity; and (3) citation order.

First, the ability to understand a hierarchical relationship depends on the knowledge of the concepts involved in the relationship. It is difficult for a person to navigate this relationship when he or she has inadequate knowledge of the concepts. A useful example is the relationship between parasitology and microbiology. *McGraw-Hill Encyclopedia of Science & Technology* (2003) defines parasitology as "a branch of biology which deals with those organisms, plant or animal, which have become dependent on other living creatures" and microbiology as "the science and study of microorganisms, including protozoans, algae, fungi, bacteria, viruses, and rickettsiae." The organisms studied in parasitology may or may not be microorganisms and the microorganisms studied in microbiology may or may not be dependent on other living creatures. The relationship between the two fields of study can be expressed as a Venn diagram (Figure 1).

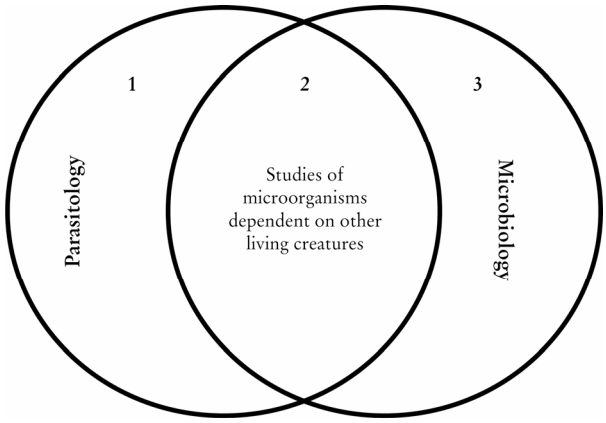


*Figure 1. Illustrated relationship between parasitology and microbiology.*

The circle on the left represents the field of parasitology and the circle on the right microbiology. Area 2, the overlapping area between the two circles, represents the branch of science that studies microorganisms dependent on other living creatures. Therefore, navigating through Path C, Science-Biology-Parasitology-Organizations, in Yahoo! would lead to organizations dealing with not only studies within the scope of microbiology (i.e., Area 2) but also those outside microbiology (i.e., Area 1). In other words, the student who followed Path C made a mistake in choosing the subcategory "Parasitology".

A similar mistake was following Path H, Science-Medicine-Microbiology and virology-Organizations.

Again two overlapping circles may be drawn, one representing medicine and the other microbiology. Each circle has an area that is outside the other circle. When the student took Path H, the only organizations found at the end of the path were those dealing with microbiology and virology as a subfield of medicine (the overlap of the two circles). Although this time no irrelevant items were retrieved, the results were limited, leaving out many relevant organizations that dealt with microbiology outside medicine.

The second issue pertains to the level of specificity – a chief principle in subject analysis at least since Cutter's 1876 codification. The case mentioned in the above paragraph was one example of being too specific. As a result, the searcher was only able to identify organizations in a sub-area of microbiology. Another student had a difficult time finding any relevant sites at first because she did not reach the most specific level, the form subdivision "Organizations," in following Path A. The student went from "Science" to "Biology" to "Microbiology" and examined all sites under "Microbiology" before discovering that "Organizations" was a subcategory under "Microbiology." Conversely, a third student attempted 4 other less specific paths (D, F, G, and J) because she was dissatisfied with the sites listed at the end of Path A. None of the 4 less specific paths turned out to be helpful.

"Organizations" was a problematic concept in terms of its placement in the hierarchy. The conventional citation order of facets usually places a general form facet (such as "Organizations") after topical ones. In both Paths E and I taken by the students in the study, "Organizations" came between topical facets. Two students thought that the 3 paths were equally valid:

A. Science→Biology→Microbiology→*Organizations*
E. Science→Biology→*Organizations*→Microbiology
I. Science→*Organizations*→Biology→Microbiology

Another student took only Path I and yet another chose both Paths A and E but not I. Although these students were not the majority in the group, their approach seemed to suggest that some users would probably behave similarly in navigating hierarchically. It may be plausible to think that the average user, lacking any training in information organization, will more likely choose a form facet before a topical one than would LIS students.

### *5.2. Precision vs. Recall*

When the two search methods were compared, navigating through Yahoo! directories yielded high precision but low recall and keyword searching resulted in low precision but high recall. Applying the former method, Path A, for example, led to 10 sites with only one of them being a broken link; the other 9 sites all met the search criteria. However, many sites of other relevant organizations in microbiology were missed because they had not been added by Yahoo's human organizers. On the other hand, more than 140,000 sites were retrieved as a result of a search by the keywords "microbiology organizations." Many more relevant sites were included in the keyword search result set, but a majority of the sites retrieved were irrelevant.

A particular feature in Yahoo! directories contributing to a higher precision rate was the embedded hierarchical structure. In navigating the structure, the user was led from a broad class to its narrower concepts to achieve a desired level of specificity. The labor-intensive process of establishing hierarchical relationships among concepts in the context of each discipline (or area of interest) was meant to gather together only relevant sites under each concept, thus maximizing the precision rate. This advantage of directory navigation was articulated by one third of the participating students (8).

An interesting question to ask is: "Is precision or recall a preference in information seeking?" Twenty out of the 24 participants in the study mentioned either precision or recall in their discussions of the two searches. This high percentage was not surprising because the participants were LIS students who had already had at least some exposure to theories of information retrieval. As shown in Table 8, among the 11 students who cited high precision as one of the advantages of hierarchical navigation, only 3 clearly preferred this search method, 1 preferred keyword searching, 4 stated that it depended on the user's need or situation, and 3 mentioned no preference. Among the 11 students who mentioned low recall as a disadvantage of hierarchical navigation, 1 preferred this type of search, 1 preferred keyword searching, and 9 either said that it depended on the user's need or did not indicate their preference. Twelve students cited high recall as an advantage of keyword searching, of whom 2 preferred hierarchical navigation, 2 preferred keyword searching, and 8 either said that it depended on the user's need or did not indicate a preference. The fact that keyword

| Advantage or Disadvantage | | Preferred Hierarchy | Preferred Keyword | Depending on Need | Preference Not Stated |
|---|---|---|---|---|---|
| Hierarchical navigation | High precision (n= 11) | 3 | 1 | 4 | 3 |
| | Low recall (n=11) | 1 | 1 | 5 | 4 |
| Keyword search | High recall (n=12) | 2 | 2 | 5 | 3 |
| | Low precision (n=10) | 1 | 0 | 4 | 5 |

*Table 8. Preferences of search methods.*

searching resulted in a low precision rate was mentioned as a disadvantage of keyword searching by 10 students, of whom 1 preferred hierarchical navigation, none preferred keyword searching, and 9 either said that it depended on the user's need or did not indicate a preference. Further, among the 3 students who preferred keyword searching, 2 mentioned high recall as an advantage of keyword searching but none saw low precision as a disadvantage of this type of search. It seemed that neither precision nor recall appeared to be an overriding concern of the participants despite their awareness of precision and recall issues.

### *5.3. The "Time" Factor*

A clear majority of the students (20) mentioned "time" (see Table 9). In comparing the two search methods, 13 students thought that hierarchical navigation was quick, easy, and not time-consuming. Some of them, rightly, pointed out that the task involved in this assignment required only a few easy steps and many real-life inquiries might lead to more complicated and prolonged exploration of numerous paths in each search; in other words, this type of search could be time-consuming. Three other participants, however, said that following Yahoo! directories in this assignment took more steps and, thus, was more time-consuming. As for keyword searching, 2 students said that being time-saving was one of its advantages; but 16 others said that it was more time-consuming. The keyword search was time-consuming, according to the students, because it resulted in too many relevant as well as irrelevant hits and too much repetition with many sites listed numerous times. In addition, the large number of retrieved sites and the low precision rate made it necessary to take more care in examining the appropriateness of each site.

| The "Time" Factor | Number of Students |
|---|---|
| Mentioned time | 20 |
| Hierarchical navigation was quick, easy, and not time-consuming. | 13 |
| Hierarchical navigation was more time-consuming. | 3 |
| Keyword search was time-saving. | 2 |
| Keyword search was more time-consuming. | 16 |

*Table 9. The "time" factor mentioned by the students.*

Among the 6 students who preferred hierarchical navigation, 5 mentioned its time-saving advantage (as compared to only 3 who mentioned its high precision) and 5 said that keyword searching was time-consuming (as compared to only 1 who said that keyword searching resulted in a low precision rate). Two out of the 3 who preferred keyword searching cited time-saving as one of its advantages and one of those two students also said that hierarchical navigation took more steps and was, thus, more time-consuming. It appears that among those who indicated their preferences between the two types of searching, "time" may have been a more significant factor than "precision/recall."

The idea of what is time-saving was somewhat ambiguous. One of the 3 students who said that hierarchical navigation took more steps also acknowledged the fact that keyword searching in this case was, overall, more time-consuming because more time was spent weeding out many irrelevant items. After reviewing these three students' descriptions, it became clear that they were only concerned with the steps taken in searching. In other words, hierarchical navigation took more steps and was more time-consuming in the searching stage; but keyword searching required more time spent on evaluating the appropriateness of the retrieved sites individually, many of which did not meet the search criteria. In this way, the precision and recall of search results were clearly linked to time taken.

*5.4. The "Knowledge/Skill" Factor*

The degree of difficulty, as seen by the participating students, seemed to be associated with how much knowledge and/or skill the user brought to each search (Table 10). On the one hand, subject knowledge was a particular concern: 3 students thought hierarchical navigation was a useful search for subject experts; 8 students said that subject knowledge was necessary for conducting a successful hierarchical search; and 1 student maintained that keyword searching was the best if the user knew little about the subject. Search skills possessed by the user were also mentioned as a significant factor: 2 students indicated that the user did not need to generate search terms or know complex search techniques in navigating the directories; 4 others said that it was necessary for the user to understand the concept of hierarchical arrangements in order to conduct such a search; 2 stated keyword searching to be user-friendly; and, 9 students said that the user needed to know either useful keywords or advanced search techniques or both to conduct an effective keyword search. Since the assignment was uncomplicated, the students' perception of the difficulties involved might not reflect the complexities in many real-life searches.

| The "Knowledge/Skill" Factor | Number of Students |
|---|---|
| Hierarchical navigation was useful for subject experts. | 3 |
| Subject knowledge was necessary for successful hierarchical search. | 8 |
| Keyword search was best if the searcher was unfamiliar with the subject. | 1 |
| Hierarchical navigation required no knowledge of useful terms or complex search techniques. | 2 |
| Users would need to know hierarchical arrangements in navigating directories. | 4 |
| Keyword search was user-friendly. | 2 |
| Keyword search required knowledge of useful terms and complex search techniques | 9 |

*Table 10. The "knowledge/skill" factor mentioned by the students.*

Of the 24 students, 17 discussed subject knowledge and search skills. Among the 6 who preferred hierarchical navigation, 1 considered it easier for experts because it would require both subject knowledge and more experience in searching; another mentioned the need for subject knowledge; the third only listed disadvantages of keyword searching in terms of the need for both advanced search techniques and useful keywords; and the rest did not mention this issue. Among the 3 students who preferred keyword searching, only 1 mentioned the need to know useful keywords; the other two did not touch on the issue of subject knowledge or search skills.

As shown in Table 11, 8 of the 24 students thought that the need for subject knowledge was one of the disadvantages of hierarchical navigation. Among the 8, 2 preferred hierarchical navigation (neither talked about the difficulties in keyword searching), none preferred keyword searching, and 6 either said that it depended on the user's need or did not indicate a preference. Four students thought that the need to understand hierarchical arrangements was one of the disadvantages of hierarchical navigation. Among them, 1 preferred hierarchical navigation, another preferred keyword searching, and 2 either said that it depended on the user's need or did not indicate a preference. Six of the 24 participants thought that the need to know useful keywords was a disadvantage of keyword searching. Among the 6, 1 preferred hierarchical navigation, 1 preferred keyword searching, and 4 either said that it depended on the user's need or did not indicate a preference. Five students thought that the need to know advanced search techniques was a disadvantage of keyword searching. Among the 5, 1 preferred hierarchical navigation, none did keyword search, and 4 either

| Disadvantage | | Preferred Hierarchy | Preferred Keyword | Depending on Need | Preference Not Stated |
|---|---|---|---|---|---|
| Hierarchical navigation | Need subject knowledge (n=8) | 2 | 0 | 3 | 3 |
| | Need to know hierarchical arrangements (n=4) | 1 | 1 | 1 | 1 |
| Keyword search | Need to know useful keywords (n=6) | 1 | 1 | 2 | 2 |
| | Need to know advanced search techniques (n=5) | 1 | 0 | 2 | 2 |

*Table 11. The "knowledge/skill" factor and student preferences*

said that it depended on the user's need or did not indicate a preference. This pattern did not give the impression that knowledge/skills played a significant role in the students' preferences even though they considered it a factor worth noting.

## 6. Conclusion

What answers might emerge from this exploratory study? Is further research merited and, if so, what directions might it fruitfully take?

The first question we asked is: Do searchers use hierarchical directories successfully? The students in this study were generally effective in navigating the hierarchy of the directories. In this instance, it was a hierarchy likely to be familiar to many users. The hierarchy follows the conventional notion common in bibliographic classifications of classification by discipline with form as the last facet. Whether or not this capability would be present in other groups of searchers is open to question. Yahoo! does not require searchers to use traditional citation order, although that is the citation order displayed at the head of the results screen. Searchers can find the same results by using any progression that comes to mind as long as they select the same categories. So "Science" to "Organizations" to "Biology" to "Microbiology" will retrieve the same screen as "Science" to "Biology" to "Microbiology" to "Organizations." Since the former, unconventional citation order was used by three students in this study, it suggests that exploration of alternative citation orders for searching might be worthwhile in other contexts. Bibliographic classification has limited the flexibility of citation order, presumably because of its use in determining the physical location of items. However, making individual facets searchable in some manner (perhaps a PRECIS-like arrangement) might be worthwhile if further research suggests that conventional citation order is not intuitive.

Do searchers prefer hierarchical navigation or keyword searching? LIS students, schooled in the need to listen to users, leaned toward the answer: it depends. However, few preferred keyword searching categorically while nearly as many preferred hierarchical navigation as suggested that "it depends." While the numbers in this study are insufficient to draw conclusions, they do suggest that further research is merited. Quite a few students were new to the Yahoo! directories, which are placed well down the busy Yahoo! homepage. Google has gone even further to take directories off of the homepage (a searcher must click on "more" search options to find a set of icons that includes "directories" represented rather cryptically by a graphic of an open book). Library catalogs and indexing and abstracting databases often make classification searching equally obscure, typically, under "call number" search or "numeric" search in a drop-down menu if available at all. (Of more use in library catalogs is the option to click on a call number in an individual bibliographic record and retrieve a list of other items at that call number. Sometimes, the searcher is dropped into an online shelflist through which they can scroll. However, this approach still does not give the browsability of lists of categories arranged hierarchically.) If options for hierarchical navigation were made more readily accessible, would searchers prefer them, at least in certain contexts, to keyword searching? Might users employ hybridized techniques that draw on hierarchical navigation to assist keyword searching and vice versa? These are questions worth further investigation through system and interface development.

What factors potentially influence searchers' success and preferences? The results of this study in relation to such factors are complex. LIS students were able to express their perceptions in terms such as precision and recall and considered issues such as search expertise, knowledge of the discipline, and user-friendliness. However, no consensus was reached in terms of how these factors would or should influence the decision to use hierarchical navigation or keyword searching. They did generally agree that the hierarchical navigation produced higher precision and keyword searching produced higher recall. They did not agree on which approach is more user-friendly, more or less time-saving or time-consuming, or more appropriate for experts or novices. In discussing these factors, the students considered their own experience with the two searches and also tried to anticipate the experiences of other searchers. Sorting out these various factors will require further research, probably of a qualitative nature, with data collection more targeted toward the variables raised here than this pedagogical exercise allowed.

What characteristics of classification are germane to searchers' performance? Related to the question of the advantages and disadvantages of hierarchical navigation are two interesting implications regarding the nature of classification. First is the advantage of showing relationships between topics. While relationships can also be shown in an alphabetically arranged controlled vocabulary such as a thesaurus

(not available in Yahoo!), hierarchical or classification-like structures make relationships more visible, arranging related topics in proximity to each other for browsing. The second implication for classification is that several students referred to the structure of the hierarchy as a limitation, requiring knowledge of the particular subject hierarchy of hierarchical searching and limiting what might be searched. This limitation needs testing, perhaps by comparing multiple conventional hierarchies with each other or with more creative structures such as might allow searchers to determine citation order (e.g. elaborated "sort by" options).

In sum, the evidence provided by this study closes no doors to future research. It points to hierarchical navigation as a viable alternative to the ubiquitous keyword searching. It calls for clarification of the roles of expertise and time in hierarchical navigation. It leaves us with an "it depends" conclusion about the desirable balance between hierarchical navigation and keyword searching, with the preference for precision or recall as one continuing factor affecting that balance. Two types of research are suggested by these conclusions: continuing studies, both quantitative and qualitative, of users' performance and preferences in hierarchical navigation; and, the development and testing of experimental systems and interfaces that offer easier and more effective hierarchical navigation than is currently available. The former should include more complex user-defined search topics, other types of classified databases, larger samples, different categories of users, and different cultural contexts so as to move beyond the limitations of the present study. Studies of users' preferences in existing systems can inform development and testing of experimental systems and interfaces. Only with both research streams can we answer the question: Is classificatory structure a powerful tool for the 21st century?

## References

Ainsbury, Bob. 2002. Cataloging's comeback: Classifying and organizing corporate documents. *Online* (May/April): 27-31.

Borgman, Christine, Sandra G. Hirsch, Virginia A. Walter, and Andrea L. Gallagher. 1995. Children's searching behavior on browsing and keyword catalogs: The Science Library Catalog Project. *Journal of the American Society for Information Science* 46: 663-84.

Callery, Anne. 1996. Yahoo! Cataloging the web. *Untangling the Web*. Available at: http://www.library.ucsb.edu/untangle/callery.html Accessed: October 6, 2004.

Chan, Lois Mai, Xia Lin, and Marcia Zeng. 1999. Structural and multilingual approaches to subject access. *65th IFLA Council and General Conference, Bangkok, Thailand, August 20-August 28, 1999*. Available at: http://www.ifla.org/IV/ifla65/papers/012-117e.htm Accessed October 8, 2004.

Cleverdon, Cyril W. 1962. *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Cranfield, Eng.: College of Aeronautics.

Cleverdon, Cyril, and Michael Keen. 1966. *Factors determining the performance of indexing systems. Vol. 2: test results*. Cranfieldfield, Eng.: College of Aeronautics.

Cousins, Shirley Anne. 1992. In their own words: An examination of catalogue users' subject queries. *Journal of information science* 18: 329-41.

Cutter, Charles A. 1904. *Rules for a dictionary catalog*. 4th ed. Washington: Government Printing Office.

Dodd, David G. 1996. Grass-roots cataloging and classification: Food for thought from World Wide Web subject-oriented hierarchical lists. *Library resources & technical services* 40: 275-86.

Drabenstott, Karen Markey, and Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential.* San Diego: Academic Press.

Ellis, David, and Ana Vasconcelos. 1999. Ranganathan and the net: Using facet analysis to search and organise the World Wide Web. *Aslib proceedings* 51: 3-10.

Ellis, David, and Ana Vasconcelos. 2000. The relevance of facet analysis for World Wide Web subject organization and searching. *Journal of internet cataloging* 2n3/4: 97-114.

Hildreth, Charles R. 1995. Online catalog design models: Are we moving in the right direction? Part 5. Browsing and exploring: A new paradigm for IR/OPAC system design. Available at: http://www.ou.edu/faculty/H/Charles.R.Hildreth/clr-five.html Accessed January 30, 2005.

Hudon, Michèle. 2003. Expanding audiences for education-related information and resources: Classificatory structures on the World Wide Web. *Cataloging & classification quarterly* 37n1/2: 81-96.

Koch, Traugott, Michael Day, et al. 1997. Specification for resource description methods. Part 3. The role of classification schemes in Internet resource description and discovery. http://www.ub.lu.se/desire/radar/reports/D3.2.3/class_v.10.html January 30, 2005.

Larson, Ray R. 1991. The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science* 42: 197-215.

Mai, Jens-Erik. 2004. Classification of the web: Challenges and inquiries. *Knowledge organization* 31: 92-97.

Markey, Karen. 1986. Class number searching in an experimental online catalog. *International classification* no. 3: 142-50.

*McGraw-Hill Encyclopedia of Science & Technology Online*. 6th ed. 2003. New York: McGraw-Hill.

McIlwaine, I.C. 2003. The *UDC* and the World Wide Web. In I.C. McIlwaine, ed. *Subject retrieval in a networked environment: Proceedings of the IFLA Satellite Meeting Held in Dublin, OH 14-16 August 2001.* München: K.G. Saur, pp. 170-76.

Micco, Mary. 1996. Subject authority control in the world of the internet. Part two. *Libres* 6n3. Available at: http://libres.curtin.edu.au/libre6n3/micco_2.htm Accessed January 17, 2005.

Notess, Greg R. 1997. WEB sites: Directories. *Database magazine* 20n1: 61-64.

Pack, Thomas. 1995. WEB search engines: LOOKSMART, EBLAST, YAHOO! Inc., SNAP. *Database magazine* 22n1: 51-54.

Said, Mirza Pahlevi, and Hiroyuki Kitagawa. 2005. Conveying taxonomy context for topic-focussed web search. *Journal of the American Society for Information Science & Technology* 56: 173-88.

Steinberg, Steve G. 1996. Seek and ye shall find (maybe). *Wired, issue 4.05*. Available at: http://www.wired.com/wired/archive/4.05/indexweb_pr.html Accessed January 20, 2005.

Sullivan, Danny. *Search engine watch*. Available at: http://searchenginewatch.com/ Accessed January 30, 2005.

Sullivan, Danny. July 14, 2004. Nielsen NetRatings Search Engine Rating. http://searchenginewatch.com/reports/article.php/2156451 Accessed August 5, 2004.

Svenonius, Elaine. 1986. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science* 37: 331-40.

Van de Walt, Marthinus S. 1997. The role of classification in information retrieval on the Internet: some aspects of browsing lists in search engines. In *Knowledge organization for information retrieval: Proceedings of the Sixth International Study Conference on Classification Research*. The Hague: International Federation for Information and Documentation (FID), pp. 32-35.

Van der Walt, Martin. 1998. The structure of classification schemes used in Internet search engines. In W. Mustafa el Hadi, J. Maniez and S.A. Pollitt, eds. *Structures and relations in knowledge organization: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France.* Würzburg, Germany: Ergon Verlag, pp. 379-87.

Warner, Julian. 2000. Can classification yield an evaluative principle for information retrieval? In R. Marcella and Arthur Maltby, eds. *The future of classification*. Aldershot, England: Gower, pp. 33-41.

Weinberg, Bella Hass. 1999. Improved internet access: Guidance from research on indexing and classification. *Bulletin of the American Society for Information Science* 25n2:. 26-29.

Wheatley, Alan. 2000. Subject trees on the internet: A new role for bibliographic classification? *Journal of Internet cataloging* 2n3/4: 115-41.

White, Martin. 2001. Architecture, search, integration: Classification is the common denominator. *EContent* (September): *54-55.*

Williamson, Nancy J. 1997. Knowledge structures and the Internet. In *Knowledge organization for information retrieval: Proceedings of the Sixth International Study Conference on Classification Research*. The Hague: International Federation for Information and Documentation (FID), pp. 23-27.

Yee, Martha M. 1998. *Guidelines for OPAC displays*. Prepared for the IFLA Task Force on Guidelines for OPAC Displays. Available at http://www.ifla.org/ifla/VII/s13/guide/opac-d.pdf Accessed March 29, 2005.

**Appendix: Assignment Description**

A user asks you to help her find five national or international scholarly/professional organizations in the field of microbiology. You decide to search on the Internet by using Yahoo! (Do not use any other search engines) in two different ways: **one by navigating through the categories** (Yahoo directories) and **the other by typing in the keywords** (in this

order, please). Do not mix these two methods in each search.

Compare and contrast the two types of searches (classification vs. subject terms). Write a two-page (double-spaced) summary of your findings, including a discussion of the merits and problems in each.

A hint: Start your category search with "Science."