

AN EMPIRICAL STUDY OF THE IMPACT OF SUBJECT HEADINGS
ON INFORMATION PERFORMANCE
- INVESTIGATION OF NURSING DOMAIN USERS

by
Ok nam Park

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2003

Approved by:

Advisor

Ok Nam Park. An Empirical Study of the Impact of Subject Headings on Information Performance - Investigation of Nursing Domain Users. A Master's paper for the M.S. in I.S. degree. April, 2003. 60 pages. Advisor: Claudia Gollop

This paper is designed to understand and identify a user's search, in a particular database, when using keyword search and subject headings search, while at the same time emphasizing real life relevance.

This study was conducted using Medical Subject Headings (MeSH) in the Medline databases for 7 nursing domain users at the University of North Carolina at Chapel Hill. Search performance using keyword, subject headings and combined search methods was assessed from a user's perspective, based on transaction logs, relevance assessments and interviews. Any mismatch between the user's keyword search statement and subject headings was analyzed to discover the relationship with precision, recall and number of retrieved records. The relationships among variables were compared.

Headings:

Subject Headings

Keyword

Search Performance

Medline

Medical Subject Headings (MeSH)

Nursing Domain Users

Table of Contents

1	Introduction _____	1
2	Literature Review _____	3
	2.1. Difficulties in Searching Information from the End User's Perspective _____	3
	2.2. Characteristics of Keyword and Subject Headings _____	4
	2.3. History of the Debate over Subject Headings versus Keywords _____	6
	2.3.1. Era One: Controlled vocabulary _____	7
	2.3.2. Era Two: the keyword search versus Subject Heading search Controversy _____	7
	2.3.3. Era Three – Case Studies _____	8
	2.3.4. Era Four – New advances in user-based systems _____	13
	2.3.4.1. OPAC Searching _____	14
	2.3.4.2. CD-ROM or Online Database Searching _____	16
	2.3.4.3. Medical Subject Headings (MeSH) _____	18
	2.4. Limitations of Prior Research on Keyword and Subject Heading Searches _____	20
	2.5. Information Retrieval Performance Measures _____	21
3.	Objectives _____	22
4.	Methods _____	22
	4.1. System and Subject Headings – Medline and MeSH _____	22
	4.2. Subject Selection _____	24
	4.3. Data Collection Tools _____	24
	4.4. Data Collection Procedures _____	25
5.	Data Analysis _____	27
6.	Discussion _____	28
	6.1. Background Information on Users' computer use _____	28
	6.2. Perception of Searching Strategies _____	30
	6.3. Search Performance – Recall and Precision _____	32
	6.4. Matching keyword terms in MeSH and the relationship with Search Performance _____	35
	6.5. Satisfaction with Search Strategies _____	40
	6.6. Post-Preferences and Perception of Search Strategies _____	43
7.	Conclusion _____	44
	References _____	47
	Appendix A: Transaction Log Form for Keyword Search _____	52
	Appendix B: Transaction Log Form for Subject Heading Search _____	53
	Appendix C: Transaction Log Form for Combining Search _____	54
	Appendix D: Interview Questionnaires _____	55

List of Figures

Figure 1:	Years of Computer Experience _____	29
Figure 2:	Years of Database Searching Experience _____	30
Figure 3:	Frequency of Use of Databases _____	30
Figure 4:	The Pre-Perception of Searching Strategies _____	32
Figure 5:	The Pre-Pattern of Searching Strategies _____	32
Figure 6:	Search Statements _____	33
Figure 7:	Precision _____	34
Figure 8:	Recall _____	34
Figure 9:	Number of retrieved documents and relevant documents retrieved _____	35
Figure 10:	The Value of Precision and Recall _____	35
Figure 11:	Mapping MeSH terms using Combined Search _____	38
Figure 12:	Mapping MeSH terms in Subject Headings Searching _____	39
Figure 13:	Unique citations _____	40
Figure 14:	Satisfactions with Citations Retrieved _____	42
Figure 15:	Satisfaction with Citation Retrieved by subjects _____	42
Figure 16:	Satisfaction with the Search Process _____	43
Figure 17:	Satisfaction with the Citations Retrieved, by Subject _____	43

1. Introduction

The bigger and more usable databases with many resources are, the more difficult it is to find material in them. Although many search strategies such as searching by abstract, author, citation, keyword, subject headings, etc. have been suggested, it is especially hard for a non-expert in information retrieval to search appropriate resources effectively using these proposed search strategies. Effective search strategies and database systems should be developed from the perspective of the end-user and his search. To do so, understanding a user's search strategies and search performance is critical. Numerous studies of the two main search strategies---keyword search and subject headings search--have been conducted in the last several decades and there have been an ongoing debates concerning the most effective search strategy. One of the search strategies commonly used by end-users, i.e. non-information search professionals, is a keyword search. However, the limitations of a keyword search are well known: ambiguity, big retrieval sets, large numbers of irrelevant records retrieved, etc. (Dubois, 1987; Rowley, 1994; Sloan, 1999). Moreover, subject headings search has been proposed as a complementary way to conduct information retrieval. However, there are still limitations to subject headings searching, such as the high cost, frequency of human errors, inadequacy of coverage, etc. (Sclafani, 1999; Chan, 2001). Each strategy has its own benefits and limitations. Previous studies have had some design limitations in the areas of applying subject headings, applying real user's information needs, etc. In addition, retrieval

performance is affected by numerous variables, thus the most effective search strategy is not identical in every situation.

This study investigated keyword searching and one of the existing subject headings searches (Medical Subject Headings (Mesh)) using the Medline databases for nursing domain users' real world searches. In the current study, search performance using keyword, subject headings and combined search was assessed from a user's perspective, based on transaction logs, relevance assessments and interviews. Any mismatch between the user's keyword search statement and subject headings was analyzed to discover the relationship with precision, recall and number of retrieved records. The relationships among variables were compared. The study aims to provide more effective search strategies from the perspective of end users based on the empirical study. The results of this study will contribute to better query interface design for electronic databases. It will serve as a preliminary study for proposing new end user domains and additional subject headings. The main research questions are:

- How much information users in nursing domain know about subject headings and their use in their real search?
- In real world searches, how many keyword terms used by end users match the subject headings? Is there a relationship with search performance?
- When an end user's search information includes multiple concepts including both keywords and subject headings, is search performance more effective than when only keywords or subject headings are employed?
- When is an end user most satisfied? Why?

2. Literature Review

2. 1. Difficulties in Searching Information from the End User's Perspective

As database searching by end users becomes more common and more information is available (Wildemuth, 1995), it is becoming more difficult for an end-user to find the appropriate resources efficiently because of the growing number of resources. Especially in the large database environment, it is harder for an end user to conduct searches effectively. Even known-item searches become slower, and possibly more difficult as well (Tibbo, 1994). Tillotson has shown in a study of student searches (1995) that the percentage of searches with relevant citations using a smaller database is higher than the corresponding percentage with a larger database. Therefore, the comprehensive and effective searching of a database becomes increasingly more important (Andrews 1991; Wildmuth 1995). Turner (2002) has pointed out that users don't know what they don't know about searching, even though many students have confidence in the search information gathered with their computers since most have experience in searching the World Wide Web using Yahoo or Google. Wildemuth (1995) has observed that librarians often point out that users fail to take advantage of many opportunities to use new search strategies, including subject headings search, because comprehensive searching may require knowledge of data structure, subject headings, etc. In this context, Andrews (1991) has suggested using subject headings searches for a more effective search. She analyzed subject headings for searching information of transcultural health

and nursing in different main databases such as CINAHL, MEDLINE, Psychological, etc. to help users find key terms from MeSH (Medical Subject Headings). Heer (2001) also has suggested several strategies, such as using subject headings, citations, abstracts, etc. However, a study by Connaway (1997) has shown that users, particularly undergraduate students, confuse keyword searching with subject headings searching. She also observed that as the database grows larger and more heterogeneous, it becomes more difficult for users to conduct subject or keyword searches effectively. Wildemuth (1995) also has pointed out the ambiguity of natural language and the usefulness of subject headings when searching through a large database.

2.2.Characteristics of Keyword and Subject Headings

When searching, using the appropriate vocabulary is important for effective and efficient information retrieval. The search expressions employed in information retrieval could be expressed in the form of a keyword search and a subject headings search (Muddamalle, 1994). Subject Headings, which are controlled or indexed by a person, usually a professional indexer, are indexing languages that represent the subjects of a resource. Usually there is a list of terms, namely subject headings, or a thesaurus. Professional indexers or catalogers assigned terms to a document. Typically, the searcher refers to the list of terms as part of his or her search strategy (Rowley, 1994). One of the purposes of subject headings is to compensate for the limitations of keyword searching during information retrieval. Subject headings are designed to enhance the precision and recall of a search. Subject headings terms are sometimes called index terms, controlled vocabulary, or descriptors (Svenonius, 1986). Keywords, which are sometimes called

free-text terms, natural language, or ordinary language, are part of the language of the document itself. Keyword searching is a way to search without translating a user's vocabulary into the vocabulary of the system (Svenonius, 1986). In real searches, keywords can be any terms appearing in the document, such as terms in the title, the abstract or the full text of the document (Rowley, 1994). In keyword searching, every word has equal retrieval value and a user can search resources from the full information content (Dubois, 1987).

Both the keyword search and subject headings search have their own advantages and disadvantages for information retrieval. Keyword search is simple to employ, involves cost-efficient management of the database, and has a low probability of human indexing errors (Dubois, 1987). However, keyword search is not effective in retrieving all relevant resources. Sometimes the keywords that a user inputs are so ambiguous that keyword search may generate too large a retrieval set and a number of irrelevant resources may be retrieved. Even though a user can recognize which are the relevant resources (given enough time), he may be overwhelmed and therefore miss some of the most useful resources (Rowley, 1994; Sclafani, 1999). A subject headings search is often contrasted with a keyword search (Schauder & Gove, 1999). Sclafani (1999) indicated that this is the main advantage of subject headings search. The subject headings search presents smaller and more specific retrieval sets and provides cross-references that can greatly enhance a search. It also might solve many semantic problems, such as ambiguity, that are brought from a user's query because subject headings are standard search terms. Chan (2001) has pointed out additional benefits of subject headings, such as control for synonyms and the ease of doing generic searches. Subject headings approach also has

some limitations. The main ones are its lack of flexibility and the fact that a user sometimes requires training to use subject headings. In addition, different catalogers may use different subject headings for the same resources. Therefore, there is a chance that a given subject headings will be incorrectly, inadequately or inconsistently employed. In addition, subject headings do not always have the desired specificity for searching because of the lack of subject headings that can represent the contents of certain resources. Finally, the maintaining and updating of subject headings is expensive in both money and time. Especially for rapidly changing areas of the world, it is hard to keep abreast of all the new terms that are introduced as subject headings (Dubois, 1987; McCarthy, 1986; Muddamalle, 1998).

2.3. History of the Debate over Subject Headings versus Keywords

Svenonius (1986) and Muddamalle (1998) divide the history of the debate over subject headings versus keyword search into three eras: Era One - the dominance of subject headings (or controlled vocabulary); Era Two - the keyword (or free-text) versus subject headings (or controlled vocabulary) controversy; and Era Three - case studies and experimental studies comparing the relative utility of keyword searches versus subject headings searches. Rowley (1994) later added an Era Four, characterized by new advances in user-based systems. In the present history, studies and tests are reviewed within the framework of these four eras; the review focuses especially on research conducted in Era Four.

2.3.1. Era One: Controlled vocabulary

Era One started in the nineteenth century, when main subject headings depended on classification schemes that were used to order bibliographic records in classified or classed catalogs. However, as library collections grew, several problems became evident: 1) users encountered so much difficulty finding a specific title using this classification scheme that precision was poor, and 2) the vocabularies used in classified catalogs were not easy for every user to use (Rowley, 1994 ; Svenonius, 1986). The use of uncontrolled vocabulary was suggested as a solution, that is, keywords from title terms. The debate then shifted to the effectiveness of free keyword searching. Cutter (as cited in Svenonius, 1986) argued for the use of title term indexing for two reasons.

(1) The title itself does not always capture all of the true subjects included in the resources. (2) The resources on the same subject would be separated into two or more different subjects if the phraseology of their titles were different (Rowley, 1994; Svenonius, 1986).

2.3.2. Era Two: the keyword search versus Subject Headings search controversy.

Era Two was the period of controversy between keyword search and subject headings search, which began with the advent of computer-based information retrieval systems. The Era begins in 1959, when keyword in context indexing (KWIC) was introduced; KWIC allowed information scientists to recognize that indexing could be achieved with the aid of computers. Era two (between 1959 and 1970) was marked by experiments and controversy (Rowley, 1994; Svenonius, 1986). One major experiment during this era became known as the Cranfield experiment (Swanson, 1965). This experiment compared the effectiveness of different indexing languages during retrieval

and tested the effectiveness of vocabulary control. The results indicated that a minimally controlled indexing language (e.g. synonyms and word-endings control) performed as well or better than subject headings indexing language. Some critics claimed, however, that most of these experiments did not reflect reality (Rowley, 1994; Svenonius, 1986). Most were conducted in laboratory rather than natural settings (Swanson, D.R (1965) and most experiments were dependent on only precision and recall. Researchers during this era tried to show that keyword searches produced results equivalent to subject headings searches.

2.3.3. Era Three – Case Studies

Era Three began in the middle of the 1970s and was characterized by case studies that compared keyword search with subject headings search search. During this period, controversy was typical. Hersey et al. (1971) conducted experiments based on general subjects such as population biology, pesticides, etc. to compare recall and precision with subject headings search and keyword search for general topics. Their experiments demonstrated that there was better recall and more relevant values with subject headings searches. This study showed that subject headings searches could be quite helpful. However, a study by Bhattacharyya reached a different conclusion (1974). He tested retrieval performance of keyword search in many different subject areas, focusing on the basic sciences (chemistry, botany, geology and physics). The purpose of this experiment was to show the advantages of keyword searching and keyword indexing with no or only minimal control, and was based on Cranfield's experiment. During the course of these experiments, Bhattacharyya looked at the state of terminology in these fields and reported that: (1) no formal control exists in the basic sciences; (2) many codes do not follow the

standard guidelines; and (3) scientists commonly use non-standard names/terminology. For these reasons, he concluded that subject headings were not updated often enough to keep current with new or quickly changing languages in rapidly evolving subject fields. In such fields, keyword search is more effective. Following these studies, there was a trend toward suggesting the use of both types of search.

Henzler (1978) evaluated keyword search, especially by title words, and subject headings search using the CANCECNET database within broad topics. His experiment showed that 35% of all title words did not match the descriptions in the subject headings and 50% of descriptions in subject headings assigned to resources could not find acceptable keyword representation. In addition, a search by subject headings generated higher recall and precision than keyword search for general topics. However, Henzler also pointed out that the keyword search is not useful in the case of new or innovative topics. He concluded that keyword and subject headings search should be combined for ideal searching, and should be made generally available in information retrieval.

Markey et al. and Katzer et al. also compared keyword search and subject headings search effectiveness. Markey et al. (1980) assessed keyword search and subject headings search based on six general subjects, e.g. sports, women, incentives, etc., using the ERIC database. This study showed that subject headings searches resulted in greater precision, but keyword searches provided better recall. Markey conducted another experiment in 1982 using 165 keyword search statements in the ERIC database. She discovered that the concepts expressed by keywords could also be stated by ERIC descriptors. Katzer et al. (1982) also compared performance measures – recall and

precision - and measures of overlap using the INSPEC database in terms of seven document representations. The seven representations were:

TT – free-text terms from titles

AA – free-text terms from abstracts

DD – descriptor terms (subject headings)

II – identifier terms chosen by indexer from the document

TA – free-text terms from titles and abstracts (combining TT and AA)

DI – indexer selected terms (combining DD and II)

ST – stemmed free-text terms from titles and abstracts

These investigators concluded that 70% of all documents could be retrieved from only three representations of documents – TT, AA and II. Overlap of representations was found to be low. Finally, Katzer noted that there was a relationship between relevance and overlap. The most relevant documents had more overlap.

Tenopir (1985) conducted similar experiments using a full-text database, the Harvard Business Review Online Database. She used 36 queries and reported that full text searches using keywords yielded better recall, while subject headings searches produced poorer precision. A combined search strategy resulted in significantly improved precision, up to 78.57% better. Tenopir also concluded that combining keyword and subject headings search is optimal.

Calkin (1980) took a somewhat different approach, comparing subject headings searches with keyword searches in terms of time and cost statistics. In her experiments, each search requester, who all was engineers, gave his search request to professional indexers. The professional indexers then constructed a search strategy using keyword

terms and subject headings and found resources using COMPENDES and ENVIROLINE databases. Professional indexers were paid \$30 per hour. The requester evaluated the search output. From this output, she concluded that keyword searching didn't retrieve all the relevant information, while subject headings searches required more time and money. She suggested that only retrievals using a combination of the two strategies could produce the best retrieval.

Svenonius (1986) attempted to summarize the keyword and subject headings search controversy and literature up to 1986. She organized the history of this debate into the three eras referred to earlier. She mentioned that many researchers such as Markey had demonstrated that keyword search produced higher recall and lower precision. She suggested that two questions still needed to be addressed: why searching by subject headings generated better precision and why subject headings search improves recall as well as precision. She suggested that future research should consider the following questions:

- (1) What is the ability of users to verbalize search requests?
- (2) What is the relevance or aboutness to an article of different kinds of terms?
- (3) What is the semantic structure of a typical text in a given discipline?

McCarthy (1986) and Dubios (1987) both attempted to summarize the advantages and disadvantages of keyword search and subject headings search.

McCarthy (1986) enumerated several weaknesses of the latter.

- (1) Subject catalogers of books dealing with the same topic may choose different terms.

- (2) The level of specificity may be too high, and divide a book from others on closely related topics.
- (3) Not enough headings are assigned to bring out the contents of the book.
- (4) Subdivisions may have been applied at different levels of specificity.
- (5) Cross references are made too sparingly

Dubios (1987) suggested that subject headings search and keyword search should no longer be viewed as mutually exclusive choices. Each has advantages and disadvantages in certain contexts and using both strategies is optimal. He enumerated the advantages and disadvantages of each in Table 1 below.

Advantages of Keyword Searches

Low cost
Simplified searching
Full information content searchable
Every word has equal retrieval value
No human indexing errors
No delay in incorporating new terms

Disadvantages of Keyword Searches

Greater burden on searcher
Information implicitly but not overtly included in text may be missed
Absence of specific-to-generic linkage
Vocabulary of discipline must be known

Advantages of Subject Headings Searches

Solve many semantic problems
Permit generic relationships to be identified
Map areas of knowledge

Disadvantages of Subject Headings Searches

High cost
Possible inadequacies of coverage
Human error
Possible out-of-date vocabulary
Difficulty of systematically incorporating all relevant relationships between terms

Table 1. Dubios (1987)

In this area, most studies depended on case studies, not large-scale experiments like the Cranfield experiments.

2.3.4. Era Four – New advances in user-based systems

Era Four was added with the development of end-user-based systems such as hypertext and GUI, including online public access catalogues and databases on CD-ROM by Rowley (1994). This era began in the late 1980s and includes the present. In this era,

research has been conducted on different types of systems and has identified the key issues raised by those systems.

2.3.4.1. OPAC Searching

Klugman (1989), Tibbo (1994), Tillotson (1995), Curl (1995) and Connaway (1997) considered the difficulties of subject headings searching and analyzed the reasons for these difficulties using an Online Catalog. Klugman (1989) listed some of the reasons for the failure of patrons' subject headings searches of the online catalog at U.C.L.A. When new terminology emerges, catalogers do not usually go back and change earlier subject headings. Therefore, subject headings are not always current with new trends or changes. Second, catalogers are not consistent in assigning subject headings to documents; there are even cases in which the same cataloger has assigned different subject headings to the same topic documents. Users are sometimes required to match search terms with inconsistent and ambiguous subject headings and thus have problems finding matches. Tibbo (1994) has also commented on the difficulty of finding relevant information and relevant subject headings in large databases. In her study, a graduate assistant searched a number of materials by subject headings at the local OPAC at the University of North Carolina, Chapel Hill. She pointed out some of the difficulties encountered in finding documents using subject headings, and the trade-offs involved in using a large number of subject headings per record. A large number of subject headings per record increase the chance of retrieving the articles. However, the size of the database will grow and this will generate a large retrieved set. It is inherently difficult to balance precision and recall in a large database. Tillotson (1995) compared keyword and subject

headings search using the Library of Congress subject headings in OPAC interfaces using librarians and library science students as his subjects. His results were different from others'; keyword searching sometimes generated too many useless references, or a user could not retrieve all the relevant material using keyword search. He reported that 50 percent of subject headings searches resulted in no hits at all. He suggested that if a user could enter the exact keyword instead of subject headings, this would improve the results. However, he pointed out that both keyword search and subject headings search should be available, since the study did not confirm that a user was happier with the results when a user used keyword searching. He also stated that OPAC provides both options, but it explains how to use subject headings search less well than other search options such as keyword search. If OPAC offered better explanations and help with subject headings searches, it might produce better results. Curl (1995) also has indicated that library user feel frustrated by inadequate subject access to the online catalog. He cited the same problems with subject headings that McCarthy mentioned (1986). Because catalogers have paid little attention to the serviceability of the subject headings, patrons continue to have difficulty. He suggested that subject headings are applied incorrectly and inadequately as well as inconsistently within a catalog. In his view a "cataloger has the responsibility to "produce an outstanding, dynamic, information source for our patrons". The online catalog should be designed with a view toward improving keyword and subject headings access. Connaway (1997) also studied users' opinions of the online catalog at the University of Wisconsin, Madison, by conducting focus group interviews with undergraduates, graduate students and faculty. This study also showed that users,

especially undergraduate students, confused keyword search and subject headings search. However, they tended to utilize subject headings search strategies more often.

2.3.4.2. CD-ROM or Online Database Searching

Rowley (1990) compared subject headings searching with keyword searching in a computer-based environment. His study used precision and recall to evaluate search effectiveness using the ERIC thesaurus and subject headings from the Sears List of Subject Headings. The subjects/users were 40 second-year students in a bachelor's degree program in Library and Information Studies in the UK. In this study, keyword searches were characterized by poorer recall and better precision. There was little difference between keyword searching and subject headings searching in mean recall or mean precision. Throughout the study, he found that there was notable variability in performance between searchers. That is, some searchers have high recall and low precision, while other searchers tend toward low recall and high precision. In addition, some searchers have better skills in choosing search terms and using search strategies. Therefore, he suggested that information retrieval systems and database producers need to investigate the relative effect of keyword searching and subject headings search on the retrieval performance of their specific users.

Fidel (1991 & 1992) investigated the reasons why users choose to emphasize keywords or subject headings during their searches. She analyzed 281 searches and observed 5 searches from a total 47 professional searchers from many bibliographic databases. First, she studied the keywords ratio for a searcher. The keywords ratio of medical searchers is 34%, that of social sciences and humanities searchers is 39%, that of no-subject-specialty searchers is 57% and the ratio of science and technology searchers is

76%. Second, she investigated the rate of subject headings neglect. The rate of medical searchers is 0%, that of social sciences and humanities searchers is 13%, that of non-subject-specialty searchers is 29% and that of science and technology searchers is 32%. Third, she analyzed the reason for neglecting subject headings. The reasons were: did not trust the subject headings or the indexing (28%), the term was not in the subject headings (23%), had to perform a multi-database search (25%), and finally, had no access to a relevant thesaurus (24%). She suggested that both keyword and subject headings are necessary for good searching and that it is important to offer simple, flexible, and low-cost access to the subject headings.

Muddamalle (1998) compared keyword search and subject heading search of the SOILSC Database, which was designed for scientists and engineers working in the field of soil mechanics, and focused on precision and recall. The retrieval tests showed equal performance but the combination of both strategies generated a 5% improvement in the result. He concluded that studying the two strategies separately does not have meaning anymore, because a combination of both strategies is the “ideal” strategy.

In 1998 (Morgan, 1998), Seven Ovid databases such as Ageline, AIDSLINE, CancerLit, CINAHL, MEDLINE, PscyINFO, and Sociological Abstracts were searched. “Family caregivers” was selected as the query term for test. Subject headings searches retrieved more unique information than keyword searches. In the subject headings searching, only 33% of the citation was duplicated. Medline had the highest number of unique citations. However, during keyword searching, only 33% of the citations were unique.

Scalfani (1999) also looked at searching for keyword and subject headings, especially Library of Congress subject headings, to find ten current health and human services related topics. He stated that subject headings searches had the advantage of offering a smaller and more specific retrieval set, as well as cross-references. However, LC subject headings are very numerous and complicated. He also mentioned that keyword searching sometimes generated so many citations that it might overwhelm some users and causes them to miss important material. He suggested that a user should start with a keyword search, look at bibliographic subject headings, and then search with subject headings. He also recommended that expert assistance would be another way to maximize the effectiveness of the search.

2.3.4.3. Medical Subject Headings (MeSH)

Lingle (1992) found that applying MeSH was frustrating and complicated for end users. Users, even users who are computer literate, do not always understand why there is differentiation or why they obtain different choices. Therefore, he also pointed out that there is a need to train users in the use of MeSH and end-user software needs to become more sophisticated. However, Richiwine reported that there had been improvement in the search performance of MeSH. Richiwine, et al. (1993) studied the Medical Subject Headings (MeSH) terms in an online catalog. Her testing tried to retrieve bibliographic records using MeSH terms. She concluded that employing relevant MeSH terms increased the recall rate to 88%, even though the MeSH terms assigned did not successfully retrieve all the relevant books.

Hallett (1998), Hersh, et al. (1998), Scalfani (1999), Milstead (1999), Qin (2000), French, J.C. et al. (2001), Sievert (2001) and Gault (2002) each studied the Medical

Subject Headings (MeSH) in Medline. (2002). Hallett et al. (1998) tested the unique system-defined subject headings search features of the Medical Subject Headings (MeSH) in Medline in two systems, Dialog and Ovid. Preston Medical Library users at the University of Tennessee Medical Center conducted this test using nine search topics. This study showed that MeSH could increase the precision of search results if it were consulted before searching, even though the two systems generated asymmetrical search results. Hersh, et al. (1998) evaluated Medline using MeSH by physicians and medical students. He found that it was beneficial for experienced clinicians or librarian searchers to use MeSH headings, while features such as explosions and subheadings were not as useful. This experiment also showed that the recall of experienced clinicians and librarians was 48%-49%, and the precision of librarians (58%) was better than that of experienced clinicians (49%). However, novice clinicians had lower recall (27%) and precision (38%) than experienced clinicians and librarians. Hersh et al. therefore suggested that guidelines or instructions should be provided to novice users who use MeSH in Medline.

In addition to these studies that demonstrate the usefulness of MeSH, Qin (2000), French, et al. (2001) and Gault (2002) have added that it is important to develop efficient retrieval systems that provide ease of access and use of MeSH. Qin (2000) has observed that the most successful document-indexing strategy is one that maintains independent free-text and MeSH indexes that are used in combination during retrieval. French et al. (2001) added that the techniques to map user queries into the subject headings, especially MeSH, in Medline are a good way to improve retrieval performance. He tested multiple collections based on the OHSUMED, an interactive retrieval evaluation and test

collection. This experiment showed that document retrieval using MeSH had better precision. Gault, et al. (2002) noted that users usually do not know about the MeSH terms assigned to the citations. They generally depend on the system to map their query terms into the appropriate subject headings. Guault et al. (2002) also pointed out that users usually enter non-MeSH terms for their searches. In this study, a number of patrons (e.g. faculty, students, and staff from the colleges of nursing, medicine, pharmacy, public health, dentistry, and health and human development) at the Health Sciences Library at the University of Illinois at Chicago participated in the search testing. They tried to find citations using Medline in different interface systems. The test showed that 432 of 466 patron query terms matched the appropriate MeSH terms. In the Ovid interface, 66% of the terms were mapped to the first term of the MeSH mapped list. 96% of terms were corresponded to the first screen of the MeSH mapped list. 93% of the terms were mapped into the first to fifth terms of the first MeSH mapped screen and 4% of the terms were mapped into the sixth to tenth terms of the first MeSH mapped screen. This investigator also reported that keyword searching generated more irrelevant citations than MeSH searching. However, even if the test resulted in quite successful matches, novice searchers were not comfortable about finding the appropriate MeSH term. Guault et al. (2002) concluded that the ability of the system to map users' query terms to appropriate subject headings such as MeSH is important in balancing precision and recall.

2.4. Limitations of Prior Research on Keyword and Subject Headings Searches

Most previous studies have included flawed experiments. First, subject headings retrieval studies used only general topics such as sports, women, etc. New or upcoming topics should have been included. Second, studies have been conducted based on simple

search statements. Users' actual search statements and their searches should be examined, because in a real life search environment, users often use more complicated search statements and their queries may include multiple concepts. Third, the evaluation criteria for search performance only included recall and precision, not satisfaction. That is, previous studies have not considered: different search statements from dissimilar users in diverse disciplines, a range of search performance criteria, and different subject headings in different databases. Retrieval performance is affected by so many variables that effective search strategies are not identical in every situation.

2.5.Information Retrieval Performance Measures

There have been ongoing debates for many years about what successful information retrieval is and how to measure it. Most studies have used recall and precision as retrieval performance measures, at least since Cranfield proposed the use of recall and precision as retrieval performance measures (Hildreth, 2001). Recall is the proportion of relevant material retrieved. Precision is the proportion of retrieved material that is relevant. Recall is hard to calculate because it requires knowing all relevant material in a given database for each query (Su, 1992). Cooper issued user's satisfaction with search results data while developing a user-oriented approach (Hildreth, 2001). Since then, there have been many experiments to show relationships between search results (Recall and Precision) and satisfaction. Su (1992) conducted experiments using 20 measures based on relevance, utility, and user satisfaction with information retrieval performance to identify the best evaluation measures. Her study argued that recall is more important to the user than

precision. However, Hildreth (2001) reported finding that there is no relationship between satisfaction and search results.

3. Objectives

This study is designed to understand and identify a user's search when using keyword search and subject headings search, while at the same time emphasizing real life relevance. The latter includes the real information needs of the user, a specific user domain, a specific database and specific subject headings.

The purpose of the current study is three-fold.

- Analyze the match between the terms that a user uses in keyword search and in subject headings search (MeSH- Medical Subject Headings) when searching Medline.
- Compare the search performance of keyword search, subject headings search, and combined search in terms of precision, recall (number of retrieved resources) and satisfaction.
- Identify any relationships between the above variables of search performance.

4. Methods

4.1. System and Subject Headings – Medline and MeSH

Medline is produced by the National Library of Medicine in Bethesda, Maryland (Burnham, 1993). It has become one of the most powerful tools for searching the literature of health-related fields. It includes citations from Index Medicus, the Index to Dental Literature and the International Nursing Index, etc. (Renford, 1989). Medline

began coverage in 1966; it indexes more than 3,600 journals and selected monographs of congresses and symposia (<http://eresources.lib.unc.edu/eid/description.php?EDID=91>).

One of the most notable characteristics of Medline is the Medical Subject Headings (MeSH), which was developed by the National Library of Medicine. MeSH is a controlled vocabulary to categorize health-related literature; it provides hierarchical structure linked to medical subject headings (Renford, 1989). MeSH currently includes 17,000 terms. The National Library of Medicine revises MeSH annually and issues its revisions late in the year (normally in November) to reflect changes in the biomedical literature and the health sciences community (Backus, 1987; Renford, 1989). Every Medline citation entered into the database needs to be indexed by MeSH. First, the National Library Medicine indexers select some MeSH terms from the list of MeSH terms that best describe the citation (Renford, 1989). Approximately eight to twelve MeSH terms with three main subject headings are assigned to each citation to be indexed (Burnham, 1993).

Medline in Ovid provides an automatic mapping function of a user's terms into subject headings to aid a user's search with MeSH terms (Jasco, 1994). When a user checks the box "Map To Subject Headings" and enters the terms, the system automatically tries to map the user's terms into related subject headings. Then the system displays mapped subject headings and other alternative terms. If a user selects any of the terms displayed, the system executes the search as requested. If there is no exact match, the system tries to retrieve documents from the query against the text of the database (Milstead, 1999).

4.2. Subject Selection

A convenience sample of subjects was recruited based on researcher's affiliation with the School of Nursing at University of North Carolina, Chapel Hill (UNC-CH). Subjects consisted of seven people, which includes three doctoral students and four postdoctoral fellows in the School of Nursing at UNC-CH. The informed consent of all subjects was obtained for data collection, data analyzing and data evaluation.

4.3. Data Collection Tools

Various data collection tools such as transaction logs, coding for relevance assessments, interviews, audiotape recordings and MeSH for subject headings were used for this study. Data Collection Tools include:

1. *Transaction logs.* A transaction log was applied to keep track of all processes and data for each search session. A subject noted each search statement, query terms, subject headings used for subject headings searches and combining searches, all retrieving steps, number of retrieved records, and number of relevant records in a transaction log form (Appendix A). Each transaction log was numbered. For a subject's first search session, the keyword search transaction log was numbered 1.1, the subject headings search transaction log was numbered 1.2, and the combined search transaction log was numbered 1.3.
2. *Coding for Relevance Assessment.* A two-point scale was used for relevance assessment: Relevant (R) and Not Relevant (NR). This assessment-coding scheme has been borrowed from another study (Saracevic et al., 1988).

3. *Interview*: An interview with each subject was conducted as soon as all search sessions for each subject were ended. All interviews were tape recorded and noted under permission from a subject. In this process, simple demographic questions, the pre-perception and the use of keyword and/or subject headings search before search sessions, preferences after the search sessions were asked. In addition, the subject's satisfaction with the search session and the citations retrieved was queried using a five point Likert Scale (Hildreth, 2001). Open questions were used to in the follow-up to identify the reasons for the evaluation (Appendix D).

4.4. Data Collection Procedures

All searches were done using Medline in Ovid Interface; it has system-defined subject headings and search features and uses MeSH for subject headings and for a combined search.

The basic procedure for data collection follows.

1. Seven subjects were included in the sample. The study was explained to each subject and the subject's permission and agreement were sought on the IRB consent form.
2. Each subject formulated one search statement for her experiment based on real personal information needs in the nursing domain. Each search statement should include at least two concepts. Formulating the search statement was depended completely on the subject.

3. Before beginning a search session with a subject, a sheet of instruction for keyword searching and subject headings searching in Medline, as well as requirements (e.g., transaction log, etc) for the experiments were provided.
4. When a subject started a search session, three transaction log forms were provided: one for keyword search, one for subject headings search, and one for combined search. Each subject started the keyword searching session first, then a subject headings searching was conducted, and finally a combined searching was be done last. The keyword searching allowed a subject to use Boolean operators such as AND, OR and NOT. The subject headings searching permitted a subject to use the “Map Terms to Subject Heading” function on the main search screen to identify the appropriate subject headings from MeSH. A subject began each search session based on his or her search statement. Each subject was fully responsible for each of his search sessions. While conducting a search, all data was documented on a transaction log form.
5. Each subject was to print out all retrieved records. She assessed the relevance of each record for the first thirty highest ranked retrieved records.
6. While retrieving records, each subject kept track of the search process on a transaction log and relevant assessment. After finishing one search session, each subject went back to perform another search session. The process was same for the keyword searching, subject headings searching, and combining searching.
7. After finishing the search session, each subject was interviewed. All interviews were tape recorded and transcribed and noted under permission from a

subject. Open-ended questions were used to probe participant's search process, her interaction with searching and using keyword and subject headings, her background and satisfaction of each search, reasons for satisfactions.

5. Data Analysis

After gathering data from the subjects based on the transaction logs, relevance evaluations and interviews, the data was analyzed. Two types of analysis were conducted, a quantitative analysis using statistical methods and a content analysis. The analysis of the transaction logs and relevance assessments was quantitative. Quantitative analysis of the transaction logs was used for calculating precision and recall. To analyze search performance, a framework for search performance evaluation was developed by modifying previous theoretical models such as Rowley's (1990), McJunkin's (1995) and Hildreth's (2001). In a framework for evaluation, precision, recall, number of retrieved records and degree of satisfaction were applied based on the transaction logs and relevance assessments. In addition, based on a list of citations retrieved by each subject, the number of unique citations and number of unique relevant citations were compared.

To calculate Recall (R) and Precision (P), the method Rowley (1990) defined was employed for each search type.

Precision = (Number of Relevant Documents Retrieved/ Number of Documents Retrieved) x 100%

Recall = (Number of Relevant Documents retrieved / Total Number of Relevant Documents in the system) x 100%

Calculation by this method used the following parameters:

- (1) *Number of documents retrieved.* This was calculated by totaling the number of documents retrieved for a given search type.
- (2) *Number of relevant documents retrieved.* This was calculated by counting all of the relevant documents retrieved for each search type.
- (3) *Total number of relevant documents.* This was calculated by counting all of the relevant documents that could have potentially been retrieved in response to the various questions.

After developing the search performance framework, the relationships among satisfaction, precision and recall were analyzed. In addition, the relationships between MeSH match terms and search performance were probed as well.

Second, content analysis was used to analyze the verbal data from the interviews. In this section, a subject's experience with computers, database searching, retrieving information with Medline and using search strategies, as well as the subject's search process and detailed reasons for degree of satisfaction were all analyzed.

The final analysis was completed by comparing the results from both types of analysis.

6. Discussion

6.1. Background Information on Users' computer use

In the first part of the subject interviews, I collected data on background information such as years of computer experience, years of database searching experience, frequently used databases and the frequency of use of those databases. First, regarding the years of computer experience as seen in Figure 1 below, one respondent had nine

years, two respondents had ten years, three respondents had eleven years and one respondent had twelve years of experience with computers. On average, respondents had 10.6 years of computer experience. For years of database searching experience, as seen in a Figure 2 below, one subject had four years experience, one subject had five years, three subjects had seven years, four subjects had eight years, and one subject had ten years of experience. On average, these subjects had 6.9 years of experience with database searching. Usually they answered that they had started to search databases with CD-ROMs before the World Wide Web was launched. Among the frequently used databases, every respondent had often used Medline and some of them habitually use CINAHL and PubMed. For frequency of database searching, as seen in Figure 3 below, two people responded that they search databases more than weekly, three people search weekly, one searches biweekly and one searches monthly. Participants in this study not only have extensive experience with computer use and database searching, but also regularly search databases such as Medline, CINAHL and PubMed.

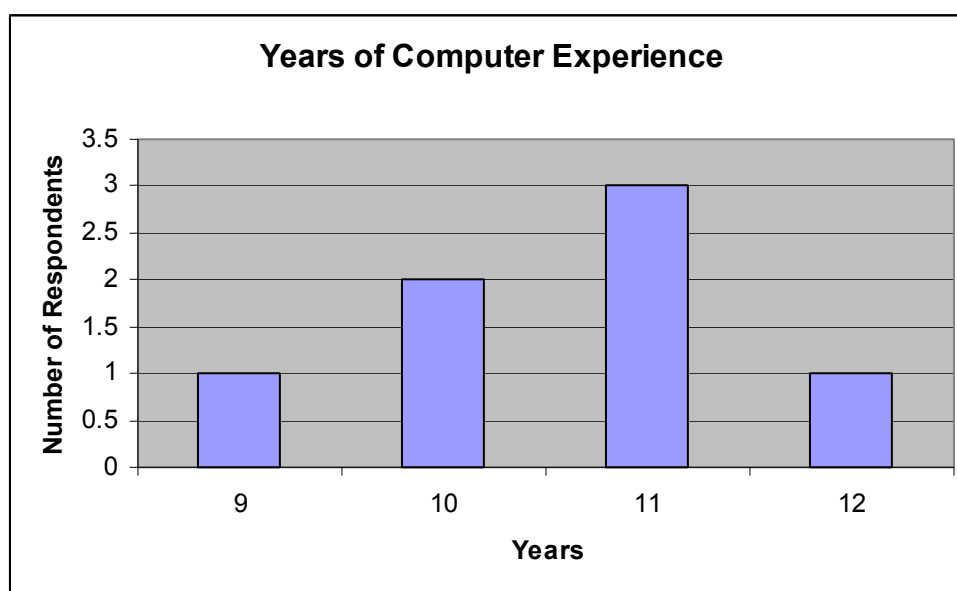


Figure 1. Years of Computer Experience

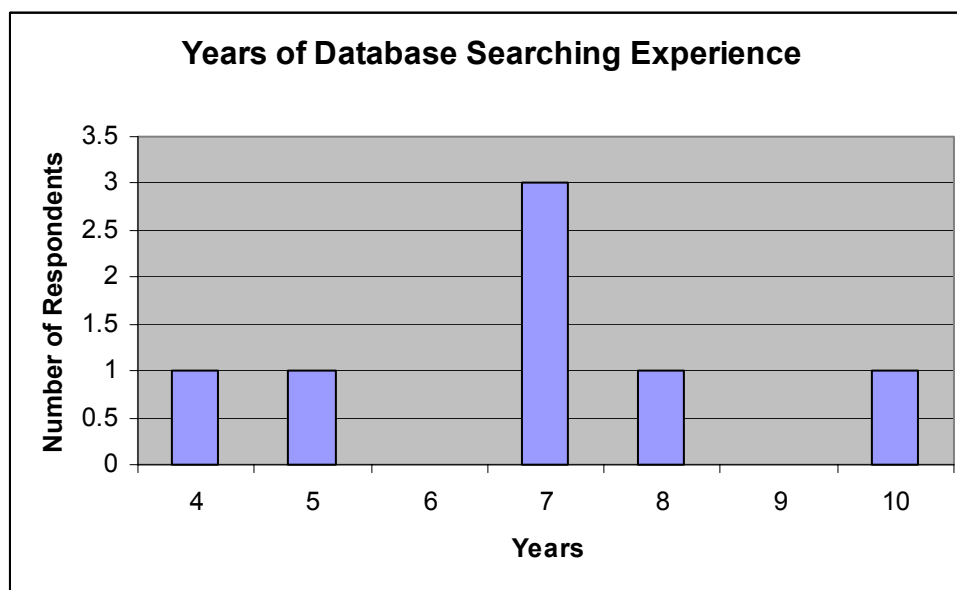


Figure 2. Years of Database Searching Experience

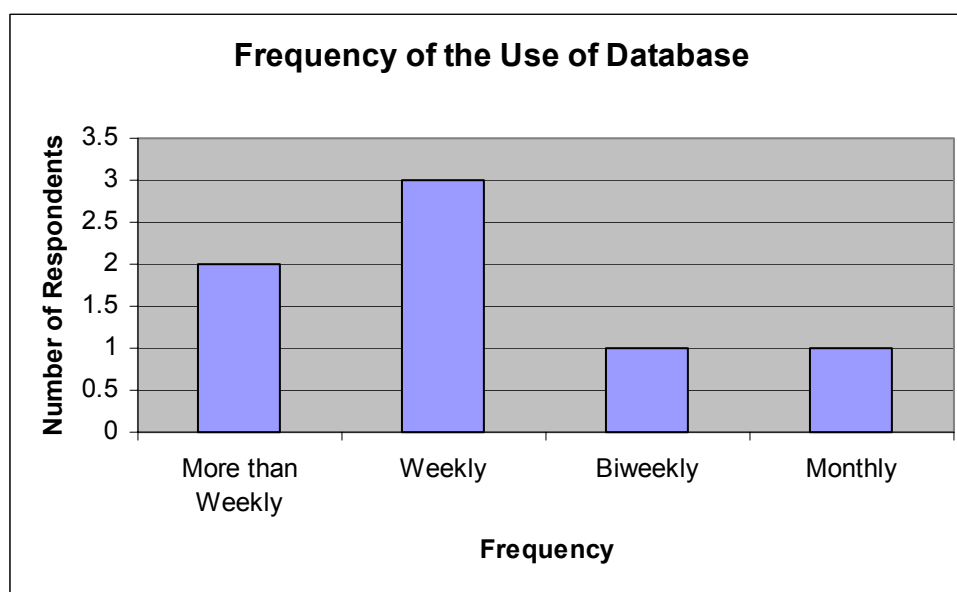


Figure 3. Frequency of Use of Databases

6.2. Perception of Searching Strategies

In the interview, the users were asked how much they in the nursing domain, knew about subject headings and their use in the students' research search was also queried. First, as seen in Figure 4 below, six respondents answered that they had heard of

both search strategies (keyword and subject headings searches) and one subject responded that she had only heard about keyword searching. However, every respondent who replied that he had heard of the two search strategies added that he had not figured out how subject headings worked. One person said that she had just guessed that it might refer to the subjects of a citation. As seen in Figure 5 below, three respondents had only used keyword search and four participants answered that they had used both search methods. However, four respondents who had used both strategies answered that they only began to use subject headings because some databases such as Medline and CINAHL in Ovid checked subject headings mapping as a default and automatically retrieved mapped subject headings. Three of the respondents added that this function was confusing, because the system sometimes generated a list of mapped subject headings and sometimes did not. The results show that they did not understand keyword and subject headings terminology well, and had never used subject headings intentionally, even though most of these respondents have extensive experience with computers and database strategies, and they are familiar with the terminology of keywords and/or subject headings and have used both strategies in their searches.

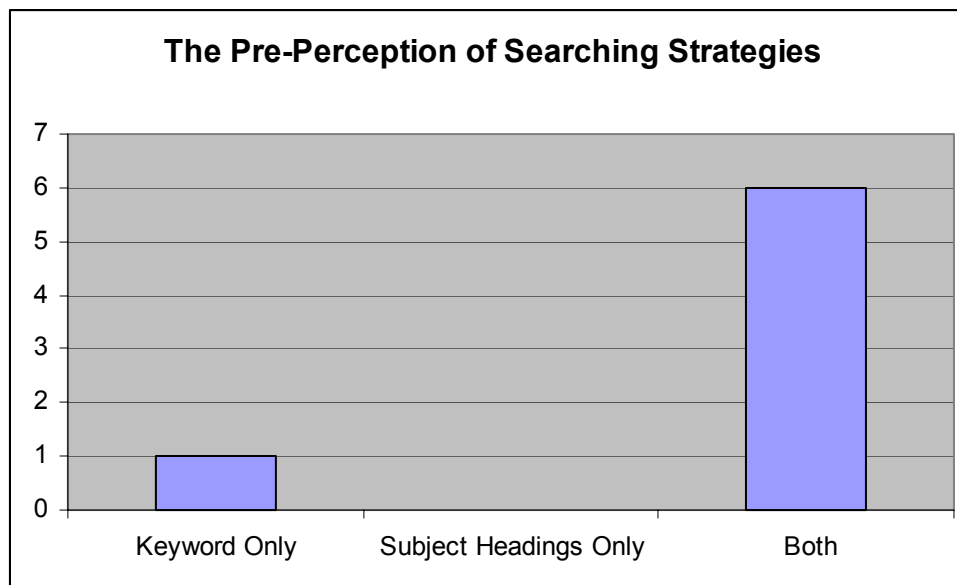


Figure 4. The Pre-Perception of Searching Strategies

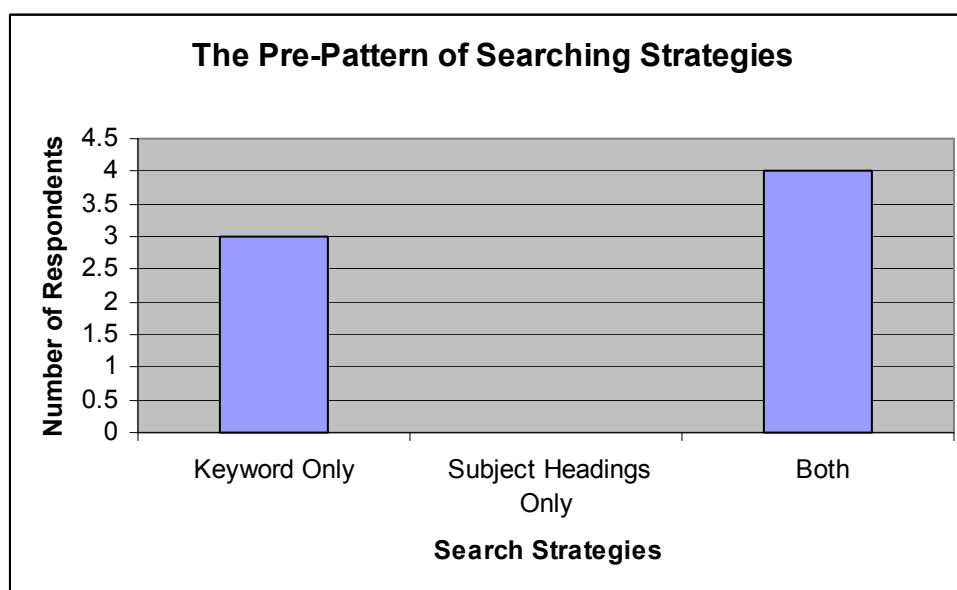


Figure 5. The Pre-Pattern of Searching Strategies

6.3. Search Performance – Recall and Precision

All search experiments were conducted based on users' real search statements related to the nursing domain. The search statements employed by seven subjects can be seen in Figure 6 below. In the search experiments from these seven participants, precision and recall were calculated based on subjects' relevance assessment.

Participant Number.	Search Statements
Participant 1	Stigma AND infectious disease
Participant 2	Smoking cessation AND transtheoretical model
Participant 3	Nursing AND case management AND strategies
Participant 4	Nutrition AND Adolescent
Participant 5	Psychiatric AND seclusion AND restraint
Participant 6	End-of-life AND decision AND factor
Participant 7	Nurse AND outcomes AND staffing

Figure 6. Search Statements

Subject headings generated the highest precision (P) in three of the seven cases, as seen Figure 7 below. In two cases, combined searches generated the greatest precision. Only one case showed the highest precision with keyword searching and one case failed to retrieve the citation. This result corroborates the results of previous studies such as those of Markey et al. (1980) and French et al. (2001)'s, both of which showed that subject headings searches were more precise.

As seen in Figure 8 below, comparing recall with keywords versus subject headings showed that six of the seven search experiments generated better recall with keyword searches. And only one of the seven search experiments showed better recall with subject headings searches. Also, as seen in Figure 9, the number of retrieved articles is highest with keyword searching in four cases out of seven, with combined searches in only two cases, and with subject headings search only once. This result supported the results previously reported by Markey et al. (1980) and French et al. (2001), namely that keyword searches generated better recall than subject headings searches.

However, when I compared recall from keyword, subject headings and combined searches, as seen in Figure 9, subject headings and combined searches yielded different results. Combined searching resulted in better recall in five of the seven search

experiments. One keyword search had better recall and one subject headings search had better recall. The value of precision and recall, which is simply the sum of recall and precision, is a simple measure of the overall performance of the different types of searching since it reflects performance in terms of both precision and recall (Rowley, 1990). The results for this value demonstrated that combining searching is better on precision and recall in three of the seven cases, as seen in Figure 10. In two cases, subject headings searches generated better values. In the last two cases, keyword searching produced better values. These data confirm the results of Henzler, Calkin (1980), Tenopir (1985) and Muddamalle (1998), all of who argued that a combination strategy is ideal for searching and results in the best retrieval.

	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6	Subject7
Keyword:	28.6%	70%	100%	30%	60%	100%	76.7%
Subject	36.7%	100%	0%	20%	36.7%	10%	100%
Headings:							
Combined:	21.4%	73.3%	100%	33.3%	63.3%	40%	93.3%

Figure 7. Precision

	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6	Subject7
Keyword:	12.5%	39.6%	57.7%	45%	37.5%	31.8%	41.4%
Subject	68.75%	18.9%	0%	5%	22.9%	13.6%	8.9%
Headings:							
Combined:	18.75%	41.5%	41.5%	50%	39.6%	54.5%	50%

Figure 8. Recall

		Subject1	Subject2	Subject3	Subject4	Subject5	Subject6	Subject7
Keyword:	Total Retrieved document	7	61	44	6833	89	7	79
	Relevant among first 30 hits	2	21	30	9	18	7	23
Subject Headings:	Total Retrieved document	17	10	Zero	5	30	136	5
	Relevant Among first 30 hits	11	10	Zero	1	11	3	5
Combined:	Total Retrieved document	14	55	22	6781	151	323	42
	Relevant Among first 30 hits	3	22	22	10	19	12	28

Figure 9. Number of retrieved documents and relevant documents retrieved

	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6	Subject7
Keyword:	41.1	109.6	157.7	75	97.5	131.8	118.1
Subject Headings:	105.45	118.9	N/A	25	59.6	23.6	108.9
Combined:	40.15	114.8	141.5	83.3	102.9	94.5	143.3

Figure 10. The Value of Precision and Recall

6.4. Matching keyword terms in MeSH and the relationship with Search

Performance

First the researcher analyzed the matching of keyword terms in MeSH in terms of subject headings and combination searches. Second, I looked at the relationship between matching keywords terms and MeSH terms. However, our data did not reveal any relationship between matching in MeSH and search performance measures such as precision and recall. However, there was a relationship between matching keyword terms

into MeSH terms and number of unique citations and number of unique relevant citations retrieved.

When they used combination searches, most users tried to map one query term among their search terms to MeSH terms. Four searches among all the combined searches succeeded in mapping a keyword term exactly into MeSH terms (participants 1, 2, 3 and 4). In the case of searches of participant 5, 6 and 7, the system did not show the exact mapped terms and the subjects tried to map terms taken from a list of MeSH terms suggested by Medline, as seen in Figure 11. In the combined searches, as seen in Figure 11 and Figure 13, searches 5, 6 and 7 did not match exact terms and/or the two selected MeSH terms generated more unique citations and unique relevant citations than did searches 1, 2, 3 and 4. Search 5 produced 17 unique citations among the first 30 hits and 11 unique relevant citations among the first 30 hits. Search 6 had 28 unique citations and 10 unique relevant citations and Search 7 included 10 unique citations and 9 unique relevant citations. Combined searching of search 1 and 2 included all of the citations from keyword searching; keyword searching of search 3 contains all of the citations from combined searching.

During subject headings searching, participants tried to map their query terms one by one into MeSH terms as seen in Figure 12 below. In this search, every search statement contained at least two concepts. The mapping process required two or more steps to map each term into MeSH terms. Subject headings searches generated more unique citations, unique relevant citations and fewer duplicate citations. For searches 1 and 6, only 1 citation out of the first 30 hits duplicated the citations from the keyword search and subject headings search; the remaining 29 citations were unique and did not

duplicate those from the keyword or subject headings searches. Only searches 5 and 7 had 3 and 4 duplicate citations, respectively. Ten of the 11 relevant citations retrieved for search 1, 10 of 10 for search 2, 1 of 1 for search 4, 9 of 11 for search 5, 2 of 3 for search 6, 1 of 5 for search 7 were unique relevant citations retrieved using subject headings. These data show that subject headings and keyword searching generated different retrieval sets.

These activities also reveal that if users' search terms are mapped well, the number of unique citations is small. For example, searches 5 and 6 only mapped one term exactly, "restraint" for search 5 and "decision" for search 6. These searches included many unique citations: 27 unique citations for search 5 and 29 citations for search 6. However, this does not mean that these searches included more unique relevant citations. For search 6, only two citations are uniquely relevant among the 29 unique citations out of first 30 hits. However, I can see that if the system mapped the user's query terms into MeSH terms more exactly, it retrieved fewer unique citations. In the case of search 7, it matched two terms among three query terms into MeSH terms more exactly and it only included one unique citation and one unique relevant citation. For search 5, it matched two terms out of two query terms into MeSH terms, and only generated 5 unique citations and 1 unique relevant citation.

In general, subject headings searches generated more unique citations and sometimes produced a larger number of unique relevant citations.

Participant Number	Keyword terms	Mapped MeSH terms
1	Stigma	Stigma
2	Smoking cessation	Smoking cessation
3	Case management	Case management
4	Nutrition	Nutrition
5	Seclusion	Attitude of health personnel /or patient advocacy /or social isolation/ or restraint, psychical /or patient isolation / Or mentally ill persons/
6	End-of-life	Decision making
7	Outcomes	Outcome assessment” (health care) /or “outcome and process” (health care)

Figure 11. Mapping MeSH terms using Combined Search

Participant Number	Keyword Terms	Mapped MeSH terms
1	Stigma	Stigma
	Infectious disease	Communication disease
2	Smoking cessation	Smoking cessation
	Transtheoretical model	Asian American
3	Case management	Case management
	Nursing	Nursing
	Strategies	Health Planning
4	Nutrition	Nutrition
	Adolescent	Adolescent nutrition /or child nutrition
5	Seclusion	Attitude of health personnel /or patient advocacy /or social isolation /or restraint, psychical /or patient isolation /or mentally ill persons/ Hospital, psychiatric /or psychiatric aids /or psychiatric department, hospital /or psychiatric nursing
	Psychiatric	
	Restraint	Restraint, physical
6	End-of-life	Decision making
	Decision	Decision making
	Factor	Factor analysis, statistical /or f factor
7	Outcomes	Outcome assessment” (health care) /or “outcome and process” (health care)
	Nurses	Nurses
	Staffing	Personal staffing and scheduling

Figure 12. Mapping MeSH terms in Subject Headings Searching

Search Type		Subject Number						
		1	2	3	4	5	6	7
Keyword:	Number of Unique citations among first 30 hits	0	0	8	9	17	5	7
	Number of Unique citations among relevant	0	0	8	1	10	5	4
Subject Headings:	Number of Unique citations among first 30 hits	29	10	Not applicable	5	27	29	1
	Number of Unique citations among relevant	10	10	Not applicable	1	9	2	1
Combining:	Number of Unique citations among first 30 hits	7	1	0	9	17	28	10
	Number of Unique citations among relevant	1	1	0	2	11	10	9

Figure 13. Unique citations

6.5. Satisfaction with Search Strategies

During our post-interviews with the participants, I asked about satisfaction with the citations retrieved and with the whole search process.

The satisfaction with the three search strategies is seen in Figure 14; the satisfaction with the citations retrieved by keyword searches and combined searches have the same mean values, while the mean value of citations from subject headings searches is lower, at 3.6. The reason given by some respondents for their satisfaction with keyword searching was that the relevance was high. However, other respondents reported that the relevance was quite low. Among the responses to the satisfaction question after combined searching were: it mapped the wrong terms, it limited the searching too much, and was easy to lose, its relevance was high, it generated irrelevant citations, and it was

good for narrowing down the possibilities even when it did not retrieve relevant articles. The reasons given for satisfaction/dissatisfaction with the citations retrieved through subject headings searches included: too many irrelevant citations, it did not map, its relevance was low, and its relevance was high. The study showed that many responses to the satisfaction question were related to the relevance of the citations. Many researchers have studied the relationship between satisfaction and search results. Su (1992) has observed that satisfaction is more closely related to recall than precision. But Hildreth (2001) argued that there is no relationship among satisfaction, precision and recall. However, most people reported that their satisfaction with citations was related to their relevance. For example, as seen in Figure 15, search 1, recall and precision in subject headings searches are higher than with other searches and gave higher satisfaction values; for search 4 and 6, recall and precision in subject headings are lower than in other types and gave lower satisfaction values. It appears that there is a relationship between satisfaction and search results. However, it is hard to tell whether precision or recall is more closely related to satisfaction in this study.

With regard to satisfaction with the search process, keyword search ranks highest and subject headings search is lowest, as seen in Figures 16 and 17. Subjects answered that keyword searching is faster and easier to use, but it is harder to decide on appropriate query terms and to think of other related terms. On the other hand, people responded that subject headings search is relatively easy, fast, and showed multiple related words. However, it requires more complicated steps to find mapped subject headings and to look at a list of MeSH terms. The question for combination searching elicits the reasons given for both subject headings searching and keyword search. This study shows that users are

more familiar with the keyword search process than with the other processes, even though the others result in better precision and recall.

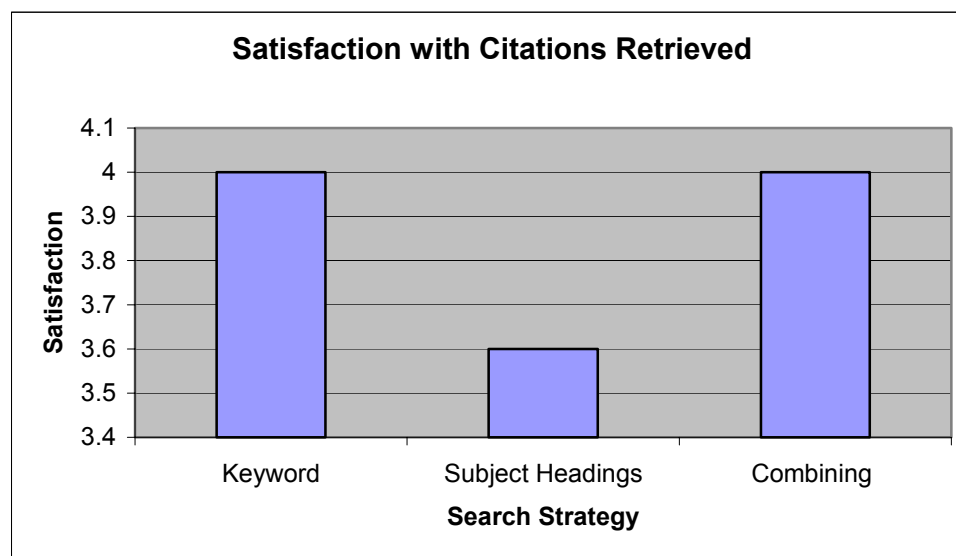


Figure 14. Satisfaction with Citations Retrieved

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
Keyword	3	4	5	3	4	4	5
Combining	3	4	5	4	4	3	5
Subject Headings	4	4	2	3	4	3	5

Figure 15. Satisfaction with Citation Retrieved by subjects

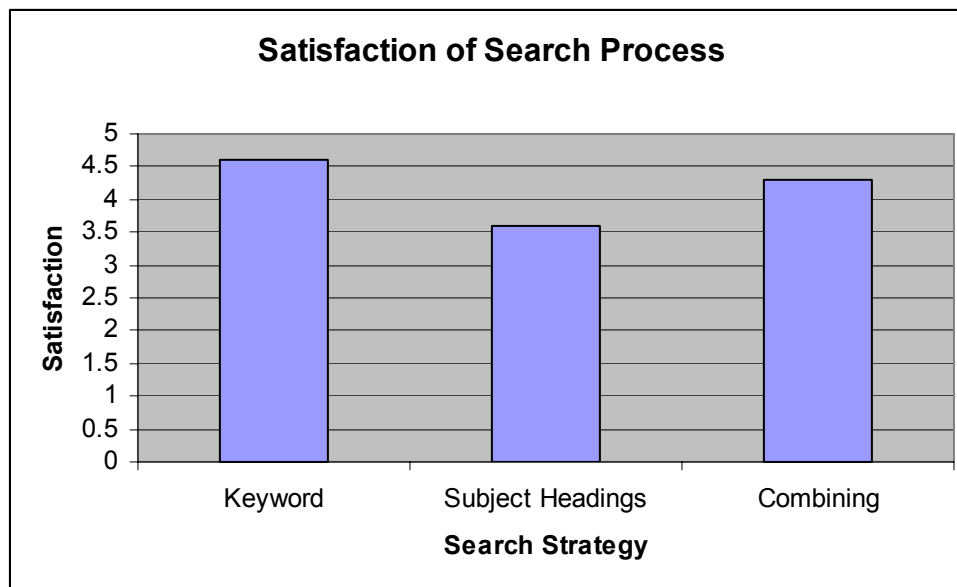


Figure 16. Satisfaction with the Search Process

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
Keyword	4	5	4	5	5	5	4
Combining	5	4	4	5	5	3	4
Subject Headings	4	4	3	3	4	3	4

Figure 17. Satisfaction with the Citations Retrieved, by Subject

6.6. Post-Preferences and Perception of Search Strategies

This study also examined the subjects' preferences for different search strategies after participating in the search experiments, and the usefulness of the search strategies. One respondent preferred subject headings search because she thought that it was neither too limited nor too broad. Some subjects preferred keyword searching because it is more

specific and easy to use and subject headings searches can generate too few citations. However, subjects also pointed out that subject headings searches can be useful in some cases and/or subject headings searches are needed when keyword searches are not sufficient. One subject preferred combination searching because it gave her more satisfactory citations and the best results. One respondent told us that she needed to use both keyword and subject headings strategies, since each generated different and unique citations and each had its own strengths: keyword searching uses only one related term while subject headings searches permit the user to select multiple related words. One respondent mentioned that subject headings and keyword searches have different purposes, and that keyword searching only retrieved the citations that include the terms a user wants, while subject headings searches generated other citations.

7. Conclusion

The purpose of this study of Medline use by nursing students was to learn more about the search parameters of three different search strategies - keyword search, subject headings search and combined search - whose merits have been debated for a long time.

First the researcher surveyed the nursing subjects' pre-perceptions of search strategies and search patterns in the nursing domain. Although both keyword and subject headings search strategies have been available to these subjects, they did not understand what these terms mean and they have usually favored only keyword searches.

Second, when the researcher looked at search performance, it was found that subject headings searches resulted in greater precision than either keyword searches or

combined searches, while combined searches were associated with better recall than the other two strategies and keyword searches generated better recall than keyword searches only. Combination strategies yielded better values for both recall and precision.

Therefore, combining the two strategies resulted in the best search performance in terms of balancing recall and precision.

Third, when matching keyword terms into MeSH terms and the relationships with citations, most of our participants could find mapped MeSH terms for their queries when they combined key word searching with subject headings searching except for one case. However, only some of the keyword terms could be matched exactly with MeSH terms. Searching by exactly matched MeSH terms generated fewer unique citations; if users selected multiple MeSH terms or imperfectly matched MeSH terms, they retrieved more unique citations. In addition, subject headings searches resulted in more unique citations and unique relevant citations than did keyword or combination searches.

Fourth, the study showed that user satisfaction with the citations retrieved is related to search result parameters such as relevance, recall, and precision. In most cases, the satisfaction reported with the poorest precision and poorest recall was low. And the satisfaction with the search process is related to ease of use, the number of steps needed for searching and the amount of time required. Keyword searches were associated with higher satisfaction than the other strategies, while subject headings searches produced the least satisfaction.

I also looked at the post-preferences of subjects for the three search strategies. Some of the nursing students preferred keyword searching because it retrieved the citations matched to the terms that they had entered and was easier and simpler to use.

However, it was found that some of the respondents mentioned that keyword searching itself is not enough and that they felt that combination strategies are needed. In addition, some of respondents pointed out that the two main strategies, keywords and subject headings, are different and that they each retrieved different citations and have different purposes.

The author has learned that there is a general recognition among the subjects that subject headings, keyword and combination searches have different merits, but that keyword searching should usually be used in conjunction with subject headings search. The researcher looked at the real information needs of nursing students using Medline, in contrast to previous studies that have focused on broad concepts and used professional catalogers or searchers as their research subjects. However, this study employed only 7 subjects and 7 search statements and used only one information search system, Medline in Ovid. It, therefore, has obvious limitations, if I wish to generalize the results to other domains and groups of users. The debate thus endures. Nevertheless, this study plays a useful role as a model for similar studies with different users or different systems. The research also serves as a basis for developing information retrieval systems and database indexing systems designed to support users who want to use multiple search strategies effectively and efficiently.

References

- Andrews, M.M. (1999). How to search for information on transcultural nursing and health subjects: internet and CD-ROM Resources. *Journal of Transcultural Nursing*, 10(1), 69-74.
- Backus, J.E.B., Davidson, S., & Rada, R. (1987). Searching for Patterns in the MeSH. Vocabulary. *Bulletin of Medical Library Association*, 75(3), 221-227.
- Burnham, J. (1993). Comparison of CINAHL, EMBASE, and Medline databases for the nurse researcher. *Medical Reference Services Quarterly*, 12(3), 45-57.
- Calkins, M.L. (1980, June). Free text or controlled vocabulary? – A case history step-by-step analysis... plus other aspects of search strategy. *Database*, 53- 59.
- Chan, L.Mai. (2001). Subject Vocabulary for Web Resources. *CASLIN*, from <http://www.caslin.cz:7777/caslin01/subjectvoc.html>.
- Connaway, L.S., Johnson, D.W., & Searing, S.E. (1997, September). Online Catalogs from the Users' perspective: the Use of Focus Group Interviews. *College & Research Libraries*, 403-419.
- Curl, M.W.(1995). Enhancing subject and keyword access to periodical abstracts and indexes: possibilities and problems. *Cataloging & Classification Quarterly*, 20(4), 45-55.

- Dubios, C.P.R. (1987). Free text vs. Controlled vocabulary; a reassessment. *Online Review*, 11(4), 243-253.
- Fidel, R. (1991). Searchers' selection of search keys: II. Controlled vocabulary of free-text searching. *Journal of the American society for information science*, 42(7), 501-514.
- Fidel, R. (1992, Winter). Who needs controlled vocabulary. *Special Libraries Association*, 1-9.
- French, J.C., & Powell, A.L. (2001). Exploiting a controlled vocabulary to improve collection selection and retrieval effectiveness. *CIMK*.
- Gault, L.V., Shultz, M., & Davies, K. (2002). Variations in Medical Subject Headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of matting lists. *Journal of Medical Library Association*, 90(2), 173-180.
- Hallet, K.S., & Todd, A.L. (1998). Separate but equal? A system comparison study of Medline's controlled vocabulary MeSH. *Bulletin of Medical Library Association*, 86(4), 491-495.
- Heer, D.L.B. (2001). Clinical Medicine Coverage in PASCAL Biomed and Medline. *Health information and Library Journal*, 19, 110-116.
- Henzler, R.G. (1978). Free or Controlled vocabularies. *International Classification*, 5, 21-26.
- Hersey D.F., Foster, W.R., Stalder, E.W., & Carlson, W.T. (1971, September). Free text word retrieval and scientist indexing: Performance Profiles and Costs, *Journal of Documentation*, 27, 167-183.

Hildreth, C.R. (2001). Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study. *Information Research*, 6(2), Available: <http://informationr.net/ir/6-2/paper101.html>.

Jasco, P. (1994, February). OVID/MEDLINE. *database*, 31-38.

Katzer, L., McGill, M.J., Tessier, J.A., Frakes, W., & DasGupta, P. (1982). A study of the overlap among document representations. *Information Technology: Research and Development*, 2, 261-274.

Klugman, S. (1989). Failures in subject retrieval. *Cataloging and Classification Quarterly*, 10 (1/2), 9-35.

Lingle, V.A., & Delozier, E.P. (1992). Medline and MeSH: challenges for end users. *Medical Reference Services Quarterly*, 11(3), 29-46.

Markery, K., Atherton, P., & Newton, C. (1980). An analysis of controlled vocabulary and free text search statements in online searches. *Online Review*, 4(3), 225-236.

McCarthy, C. (1986). The reliability factor in subject access. *College & Research Libraries*, 47(1), 48-56.

McJunkin, M.C. (1995). Precision and recall in title keyword searches. *Information Technology and Libraries*, 14(3), 161-171.

Milestead, J.L. (1999). Cross-file searching: How vendors help – and don't help – improve compatibility. *Searcher*, 7 (5), 44-55.

Morgan, S. (1998, Spring). More than Medline: Strategies for Better Searching. *Health Science Libraries* [Online]. Available: <http://oldsite.library.upenn.edu/services/publications/pub-biomedical/newss98.pdf>.

- Muddamalle, M.R. (1998). Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics. *Journal of The American Society for Information Science*, 49(10), 881-887.
- Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: an investigation in the antibiotic resistance literature. *Journal of the American society for information science*, 51(3), 166-180.
- Renford, B.L. (1989). MeSH and the quality of searching, *DLA Bulletin*, 9(3), 3-5.
- Richwine, P.W. (1993). A study of MeSH and UMLS for subject searching in an online catalog. *Bulletin of Medical Library Association*, 81(2), 229-233.
- Rowley, J. (1994). The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of Information Science*, 20(2), 108-110.
- Saracevic, T., & Kantor, P. (1988). A study of Information seeking and retrieving II. users, question, and effectiveness. *Journal of the American Society for Information Science*, 39(5), 177-196.
- Schauder, C., & Gove, K. (1999). Metadata: Controlled vocabulary and Metadata. *Connections* [Online]. Available: <http://www.curriculum.edu.au/scis/connect/cnetw99/30meta.htm>.
- Sclafani, F. (1999). Guest Essay: Controlled subject heading searching versus keyword searching. *Technicalities*, 19(9), 7-15.
- Su, L.T. (1992). Evaluation measures for interactive information retrieval. *Information Processing & Management*, 28(4), 503-516.

- Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. *Journal of the American society of information science*, 37(5), 331-340.
- Swanson, D.R. (1965). Evidence Underlying the Cranfield Results. *Library Quarterly*, 35,1-20.
- Tenopir, C. (1985). Full Text database retrieval performance. *Online Review*, 9(2), 149-164.
- Tillotson, J. (1995, May). Is keyword searching the answer?. *College and Research Libraries*, 199-206.
- Tibbo, H.R. (1994, Spring). The epic struggle: subject retrieval from large bibliographic databases. *American Archivist*, 57, 310-326.
- Turner, N.B. (2002). befuddled or bemused: Testing students' use of the online catalog. *College and Undergraduate Libraries*, 9(1), 71-79.
- Wildemuth, B.M., & Moore, M.E. (1995). End-user search behaviors and their relationship to search effectiveness. *Bulletin of Medical Library Association*, 83(3), 382-387.

Appendix A: Transaction Log Form for Keyword Search

Form Number: 1-1 (Keyword Search)

Transaction Log Form (Please Fill Out This Form)

Date:

Search Statement	
Query Terms you <u>used</u>	
Retrieving Steps. (Please fill out all steps in as much detail as possible.)	1. 2. 3. 4. 5. 6. 7. 8. 9. 10.
Number of Retrieved Records	
Number of Relevant Records	
Total Search Time:	
Notes:	

Appendix B: Transaction Log Form for Subject Headings Search

Form Number: 1-2 (Subject Headings Search)

Transaction Log Form (Please Fill Out This Form)

Date:

Search Statement	
Query Terms you <u>used</u>	
Retrieving Steps. (Please fill out all steps in as much detail as possible.)	1. 2. 3. 4. 5. 6. 7. 8. 9. 10.
Number of Retrieved Records	
Number of Relevant Records	
Total Search Time:	
Notes:	

Appendix C: Transaction Log Form for Combining Search

Form Number: 1-3 (Combining Search)

Transaction Log Form (Please Fill Out This Form)

Date:

Search Statement	
Query Terms you <u>used</u>	
Retrieving Steps. (Please fill out all steps in as much detail as possible.)	1. 2. 3. 4. 5. 6. 7. 8. 9. 10.
Number of Retrieved Records	
Number of Relevant Records	
Total Search Time:	
Notes:	

Appendix D: Interview Questionnaires

Form Number: 1

Interview Questionnaires

Date:

1. University Status

a) Undergraduate Student b) Master Student c) Doctoral Student

d) Post- Doctoral Student e) Faculty f) Other _____

2. Years of Computer Experience _____

3. Years of Database Searching Experience _____

4. What kinds of database you usually use? _____

5. How often do you use Database for searching? _____

6. Before this search experiment, did you understand or feel familiar with keyword or subject headings terminology?

7. Before this search experiment, have you ever used keyword and/or subject headings for your search? Yes _____ No _____

If so, which (please circle)?

a) Keyword only b) Subject Headings only c) Both

8. After this search experiment, what is your preference for search?

a) Keyword only b) Subject Headings only c) Combining both

9. Compare the three search sessions when you answer the following questions:

A. Keyword search

(1) How satisfied were you with finding the information you sought?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(2) Please list the reasons why you gave the answer you did on Question (1).

(3) How satisfied were you with search process?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(4) Please list the reasons why you gave the answer you did on Question (1).

B. Subject Headings Search

(1) How satisfied were you with finding the information you sought?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(2) Please list the reasons why you gave the answer you did on Question (1).

(3) How satisfied were you with search process?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(4) Please list the reasons why you gave the answer you did on Question (1).

C. Combining Search

(1) How satisfied were you with finding the information you sought?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(2) Please list the reasons why you gave the answer you did on Question (1).

(3) How satisfied were you with search process?

Very Unsatisfied-> 1 2 3 4 5 <-Very Satisfied

(4) Please list the reasons why you gave the answer you did on Question (1).