Subject Headings and Subject Search: A Comparative Study[1]

Jingchi Guo
East China Normal University
China
jcguo@library.ecnu.edu.cn

Jie Huang
University of Oklahoma
United States
lilyh@ou.edu

Abstract: *Subject headings, a controlled vocabulary, in bibliographic records are considered as an important instrument to help increase the precision of information search results. Yet, many studies in the past showed that subject search is the most difficult searching method for library users. In 2007, librarians at the University of Oklahoma Libraries in the United States did a study on users' subject searching behavior. The findings showed that among the OPAC searching methods examined (i.e., title, author, subject and keyword), subject search was used the least and caused the most difficulty. This paper investigates how college students conduct subject search with Chinese authorized subject headings. The data collected in the library instruction classes at the East China Normal University are compared with that from the University of Oklahoma Libraries. The reasons for failed searches are analyzed. Through comparison and analysis, the authors found that the reasons causing the difficulty in subject search are similar at both universities in spite of the fact that subject headings from different authority databases were used. The main reasons for users' difficulty include the complexity of authorized subject headings, the lack of information literacy training, and the unfamiliarity with the controlled vocabulary and structures used in bibliographic records. Some feasible suggestions for improving library users' ability to utilize the controlled vocabulary and for increasing the success rate of subject search from the perspectives of both reference and cataloging are provided.*

I. Introduction

Since the Online Public Access Catalog (OPAC) emerged in the 1970s, it has been the most important tool for library users to find materials and resources held at the library to

[1] This article is based on a conference presentation entitled "Improve the Efficiency of Subject Searching: A Comparative Study", which appeared in *City Life and Library Service: The Proceedings of the Fifth Shanghai (Hangzhou) International Library Forum* (pp. 236-247). Shanghai: Shanghai Scientific and Technological Literature Publishing House, 2010.

meet their information needs. It has been libraries' desire to make OPAC easy to use and to provide users with as many access points as possible. Among all access points, subject headings, such as Library of Congress Subject Headings (LCSH), are widely used in the library community as a tool for creating and providing access points to traditional library collections as well as digital resources (Yi & Chan, 2010).

Subject headings as a controlled vocabulary assigned by librarians are considered as a key instrument that can increase the precision of information search results. Theoretically, when users do not know the exact titles, authors, or series, the best way to find the information they are looking for is to use "subject search". However, the findings in many studies showed that library users do not understand the structure of authorized subject headings and therefore have difficulty identifying the correct subject terms to complete a comprehensive subject search (Drabenstott, Simcox, & Fenton, 1999; Olson & Boll, 2001). It often makes users feel frustrated and dissatisfied with either too many results or none when they use this type of search. Some users do not even know the existence of subject terms (Antell & Huang, 2008).

As the Web search engines, such as Google, Yahoo, Baidu[2], and Wikipedia, get more and more popular nowadays, their searching methods have a bigger impact on users' searching behavior. Keyword search has become a popular search means among many users, especially with the younger generation such as college students. Users' expectation of OPAC has also changed greatly (Yu & Young, 2004; Carstens & Buchanan, 2004).

In 2007, librarians at the University of Oklahoma (OU) Libraries in the United States did a study on users' subject searching behavior. The findings showed that among the searching methods examined, namely title, author, subject and keyword search in OPAC, subject search was used the least and caused the most difficulty (Antell & Huang, 2008).

This paper investigates how students do subject search with Chinese authorized subject headings at East China Normal University (ECNU) with the data collected from the

[2] Baidu, Inc. (Chinese: 百度; pinyin: Bǎidù, NASDAQ: BIDU), simply known as Baidu and incorporated on January 18, 2000, is a Chinese web services company headquartered in the Baidu Campus in Haidian District, Beijing, People's Republic of China. -- Wikipedia

library instruction classes. The results are compared with that from the OU Libraries. Through studying the subject terms used by college students and their searching behavior, the authors hope to identify the problematic areas of subject search and make some feasible suggestions on their improvement.

II. Literature Review

Many librarians know the difficulties that users encounter when they use the subject search method to look for information. To make this valuable searching method more user-friendly, both reference and cataloging librarians in North America have done many studies on this topic from various perspectives.

The discussion and debate over the usefulness of the controlled subject vocabularies started after some research revealed that many library users, especially the younger generation, prefer to use keyword search over subject search. “The increasing reliance on keyword or natural language searching has also prompted information professionals to question if controlled vocabularies such as LCSH are still necessary in the online environment” (Steele, 2010, p. 6).

In 1995, this issue was addressed by the Association for Library Collections and Technical Services in its program called “Crisis in Subject Cataloging and Retrieval”. It pointed out that there was an administrative push to cut back or eliminate subject cataloging because of the availability of keyword searching (Steinhagen, 1996).

The debate has escalated “in the intervening thirteen years, during which the availability of keyword searching resources has increased exponentially due to the ubiquitousness of the Internet and its many freely available search tools” (Antell & Huang, 2008, p. 69).

Some advocated that the OPAC “be completely redesigned into a Web search engine that looks and functions like Google” (Russell & Huang, 2009, p. 77). But most of reference librarians and cataloging librarians still highly value the controlled vocabularies and subject search. It was claimed that authorized subject headings remain viable for the specificity it brings to subject search (Mann, 2003; Beall, 2006). Many librarians believe

that subject headings provide greater precision in subject search (Carstens & Buchanan, 2004). “The appropriate use of subject searching is a powerful tool that enables users to extract relevant information from the OPAC” (Antell & Huang, 2008, p. 73).

In the meantime, a number of librarians are convinced that subject headings do have their limitations and need to be improved so that they are more user-friendly and easier to use. It has been suggested that efforts “be made to render the semantics and syntax of LCSH more consistent and predictable,” and that their structures “be made more rigorously hierarchical” (Yi & Chan, 2010, p. 686).

Guidance on subject search needs to be further strengthened through the library’s information instruction class (Antell & Huang, 2008).

Based on the data collected from surveys, interviews, and transaction logs, etc. in the study of subject seearch over the years, many good suggestions and ideas, such as allowing users to supply descriptive tags to catalog records, utilizing a new type of external knowledge source *folksonomy*, incorporating less-controlled vocabulary through broader usage of cross-references, updating and adding new subject terms more quickly, providing “just in time” instruction and assistance, have been introduced (Antell & Huang, 2008; Rolla, 2009; Russell & Huang, 2009; Yi & Chan, 2010). Librarians in America are sparing no efforts in making OPAC behave like a search engine that will retrieve precise, accurate, and quality information with easier and quicker searches.

In China, librarians have carried out similar studies on OPAC. Some of them discuss the issues related to traditional subject search, such as subject headings are not exactly reflecting or corresponding to the user’s search terms in the natural language. Thus, they propose to optimize subject search by combining the merits of both controlled subject vocabulary and natural language so as to improve the traditional subject search (Xu, 2006).

Some people also see subject headings as one of the factors that affect the quality of OPAC. The common issues with subject headings are that the terms assigned are too general or too specific, inadequate, imprecise, incomplete, wrong in order, lacking in

meaning, etc. They propose to standardize subject headings according to the rules in the *Chinese Classified Thesaurus (CCT)*, to assign sufficient subject headings, and to select subject headings scientifically (Zhang, 2004).

In examining OPAC as a system, some studies compare OPAC systems in China and abroad and some compare the systems used in the top 30 academic libraries in China. The studies found that most OPAC systems have subject searching method. All the six foreign OPAC systems studied have subject keyword search, and five of them have implemented subject standardization.

Some libraries in China have even adopted the tagging technology in their OPAC. Indeed, tagging helps users find popular materials. However, many words tagged on the catalog are not the controlled vocabulary and therefore may decrease the precision of search results.

Based on their investigations, Song & Zhan (2005) and Si & Chen (2009) have proposed various ways to improve the browsing functions of OPAC, such as matching subject terms, supplying keywords and call numbers automatically by the system, and displaying the "subject tree" of sub-fields in OPAC.

Nevertheless, there are not many studies on OPAC users' searching behavior or in-depth studies on OPAC subject search in the Chinese literature.

III. Research Methodology

By means of a questionnaire and interviews, this study analyzes the data obtained from a general education selective course at ECNU in 2009. It explored the ways the students used Chinese subject terms for their subject search. The findings are then compared with that by the librarians at the OU Libraries. It also analyzed the cases of search failures and their causes.

This selective course, designed to cultivate students' ability to search for information, was open to both undergraduates and graduates at the master's and doctoral levels at ECNU. It was divided into two main sections: one for social sciences, humanities and arts

majors and the other for natural sciences majors. The participants of the study were the social sciences and humanities majors at the graduate level during the fall semester of 2009.

The composition of the students is as follows. The 94 students in three classes were mainly from music (12.8%), education (25.5%), and economics and finance (21.3%). Most were master's students, with only 6.4% as doctoral students. The majority consisted of those who had done their undergraduate studies elsewhere, with only 7.5% of them had received their bachelor's degree from ECNU. Those who had graduated from other universities almost had no previous training in information literacy education. Only two former ECNU undergraduates had taken a course in this area previously. But even they acknowledged their lack of experience using the Internet and databases.

Those students use email, chat room (MSN or QQ[3]), and search engines (Google or Baidu) on a daily basis. When searching for information on OPAC, they use natural language with no knowledge about the recall and precision ratio.

The questionnaire was assigned as the homework for the course, and it was done twice. Before it was assigned for the first time, the requirements of OPAC subject search had been explained briefly. But the results of the homework handed in were not valid due to the fact that most students lacked knowledge about subject search. Most of them used keyword search and title search instead. Those who did use subject search used natural language and gave up their search after a few failed attempts. Thus, before it was redone, further explanation of OPAC subject search was made, how to choose subject terms demonstrated, and students' hands-on practice search arranged, in class. The students' questions were answered on the spot, but some students were too shy to ask questions and, for that reason, some questions were not resolved in time.

Each student was required to choose 3 topics for search, which came from their major coursework or research projects with their professors, or from their personal interests.

[3] Tencent QQ, generally referred to as QQ, is the most popular free instant messaging computer program in Mainland China. As of September 30, 2010, the active QQ users accounts for QQ IM amounted to 636.6 million, possibly making it the world's largest online community. -- Wikipedia

With each topic they could make 5 search attempts to find related information. If unsatisfied with the results, they could change the subject terms, recording each topic and subject term they searched, the number of results of each search, and satisfaction level of each search. After 5 unsuccessful attempts on a search topic, the student should answer the following questions on the questionnaire: How would you proceed with the next step in search? Would you expect to find the "correct" subject term and continue to use OPAC? Would you consider giving up on OPAC and searching somewhere else? If yes, where?

IV. Findings and Analysis

A total of 94 students participated in the research via questionnaire. Each student was to choose 3 topics for search, with a maximum of 5 search attempts for each topic. Therefore, the maximum number of possible search attempts was 1,410 (94 x 3 x 5), but the actual number of search attempts was 1,205, because some students had not made 5 attempts on all 3 topic searches.

The search results and the answers to the questionnaire indicate:

1. Relatively low frequency in OPAC use

None of the students used OPAC on a daily basis. Only 50.0% used it 2-5 times a week; 27.7% used it 2-6 times a month; 18.1% used it once a month; and 4.2% never used it.

The reasons for the lack of OPAC use among the students are: First of all, 79.8% of the participants were first-year graduate students who came from other universities and were not familiar with the ECNU library or its website and OPAC. Their use of the library consisted of studying inside the library and checking out some items found on the shelves. The other students (8.5% of second-year and 4.3% of third-year graduate students, and 7.5% attended ECNU as undergraduates) were relatively more familiar with the library's OPAC and electronic resources. Those students preferred to use the electronic resources, but the OPAC contained mainly hard-copy materials with only a few electronic resources.

2. OPAC subject search as the least chosen search method

The most commonly-used methods for searching OPAC were keyword search (53.8%), title search (29.0%), and author search (13.8%). Among the least used methods are

subject search (43.1%), ISBN/ISSN search (24.2%), and call number search (22.3%). (See Figure 1)

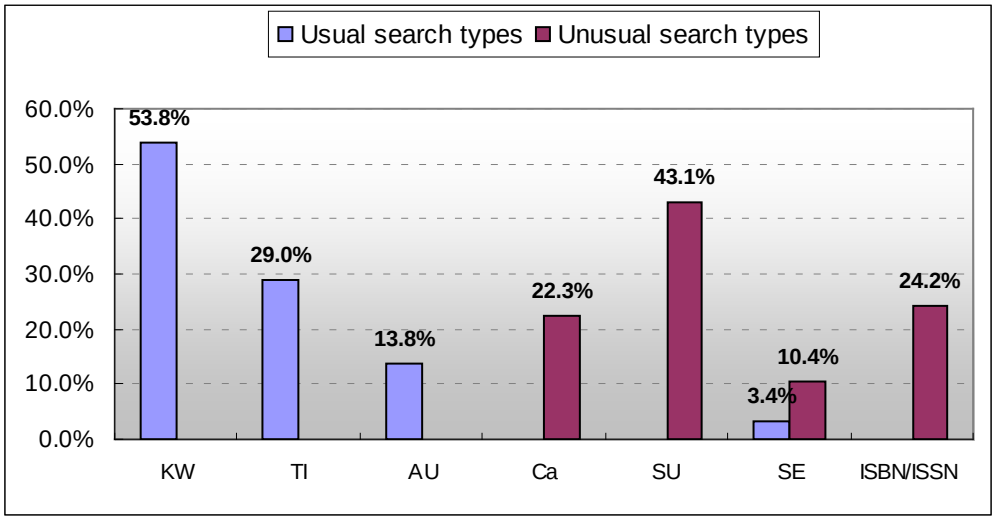

Figure 1. Frequency of each search type (ECNU)

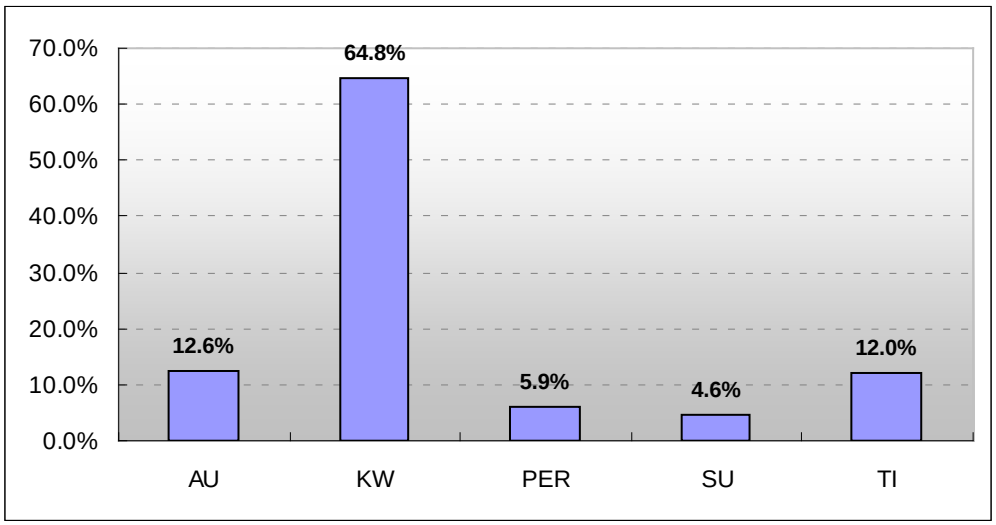

Figure 2. Frequency of each search type (OU)

Comparing these findings with that from the analysis of OPAC transaction logs at the OU Libraries (See Figure 2) (Antell & Huang, 2008), the following similarity are manifest: Search methods that using information contained in the original content, such as keyword, title, and author, were preferred over searches via the methods which contain the information provided by librarians, such as subjects and call numbers. This phenomenon gives rise to the question: Should OPAC searches be adjusted according to users' habits

or should users be instructed to search in a more professional way?

3. OPAC abandoned if search results were unsatisfactory

If the search results were not satisfactory, only 35.1% of the participants would continue using OPAC until they found a correct search term while 64.9% opted to give up on OPAC, believing that it did not contain the information sought.

Among those who found search results unsatisfactory, 37.0% would switch to Google search, 23.9% to online databases, 13.1% would go to the library to browse the items on the shelves, 16.1% would go and ask their professors, and 10.0% would seek for help from librarians. (See Figure 3)

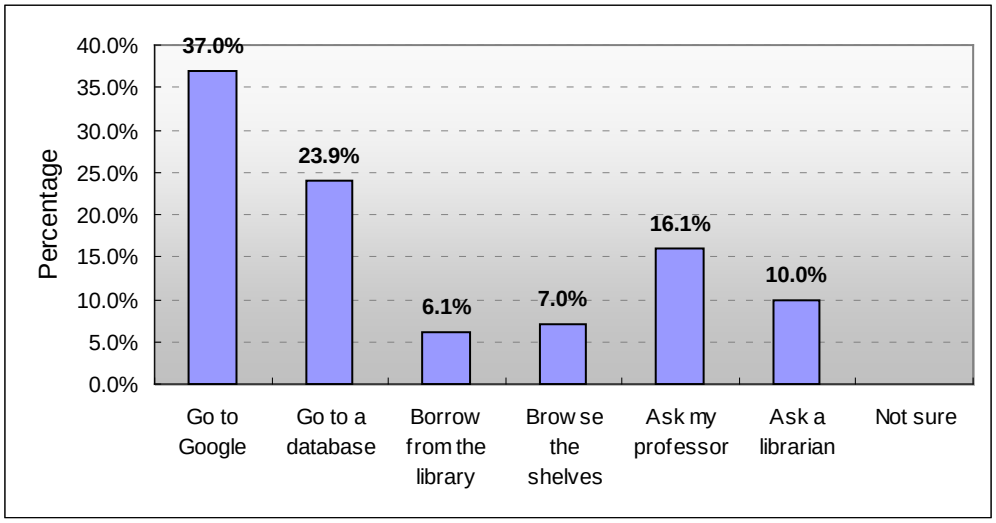

Figure 3. Next step after an unsuccessful topic search (ECNU)

A similar tendency was found in comparison with the interviews at the OU Libraries (Antell & Huang, 2008). (See Figure 4)

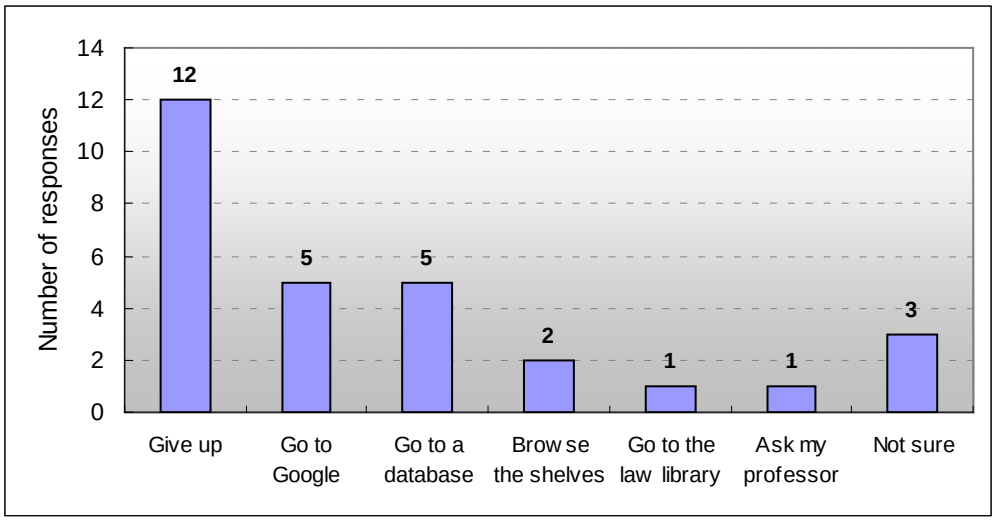

Figure 4. Next step after an unsuccessful topic search (OU)

As can be seen from these results, the library does not lack resources or services. What lacks are the means to make its resources and services known to users, or to make them available to users in a timely and suitable manner.

As pointed out by OCLC in its document entitled "How academic librarians can influence students' web-based information choices", most of college students choose search engines as their primary online resources because they are simple, easy and convenient to use (OCLC, 2010).

4. Comparison between subject and keyword searches

The results between subject and keyword searches were compared. For subject searches, 45.7% had zero results, 49.1% had 1-499 results, and 5.1% had 500 or more results. For keyword searches, 17.5% had zero results, 59.8% had 1-499 results, and 22.7% had 500 or more results. (See Figure 5)

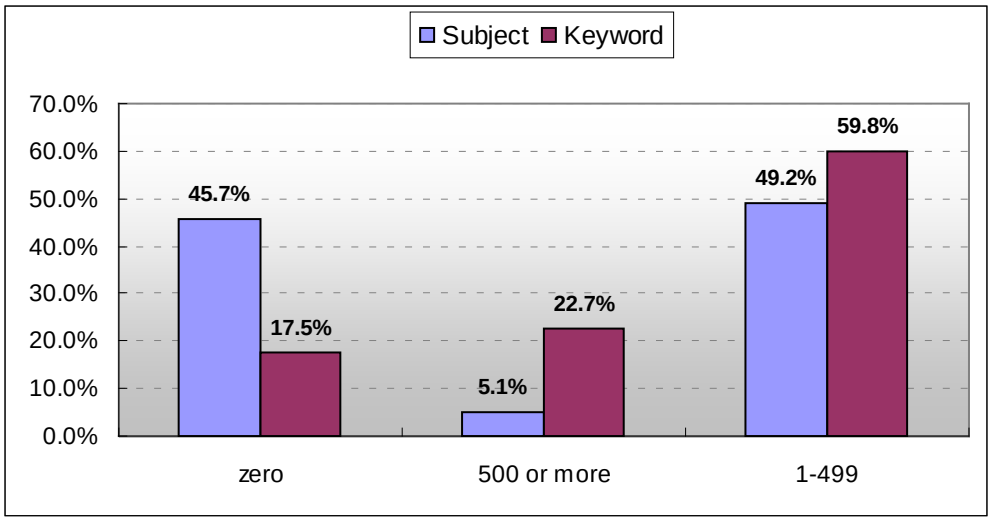

Figure 5. Result yielded from different types of search (ECNU)

A similarity was found with the OU Libraries' OPAC transaction log analysis (Antell & Huang, 2008). (See Figure 6)

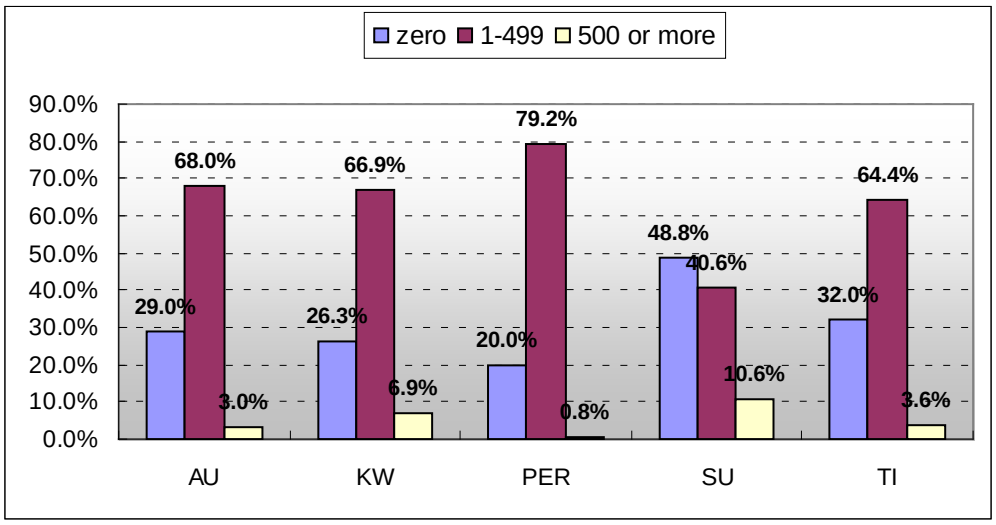

Figure 6. Result yielded from different types of search (OU)

Subject search led to much more results with either zero or over 500 hits than other types of search. The quantitative results show why students seldom chose subject search: They were very frustrated when half of their subject searches led to zero results, owing to their lack of knowledge about standardized or controlled subject terms.

ECNU differed from OU when the search results were more than 500. The reason for the difference is that at ECNU, subject search is limited to standardized subject terms

whereas at OU, it was allowed to use both standardized subject terms and keywords in subject terms.

5. How to adjust search terms when search results are zero?

When search results are zero, 61.6% would change search terms, 24.8% would take OPAC's suggestions, and 13.6% would go back to the search terms already used. (See Figure 7)

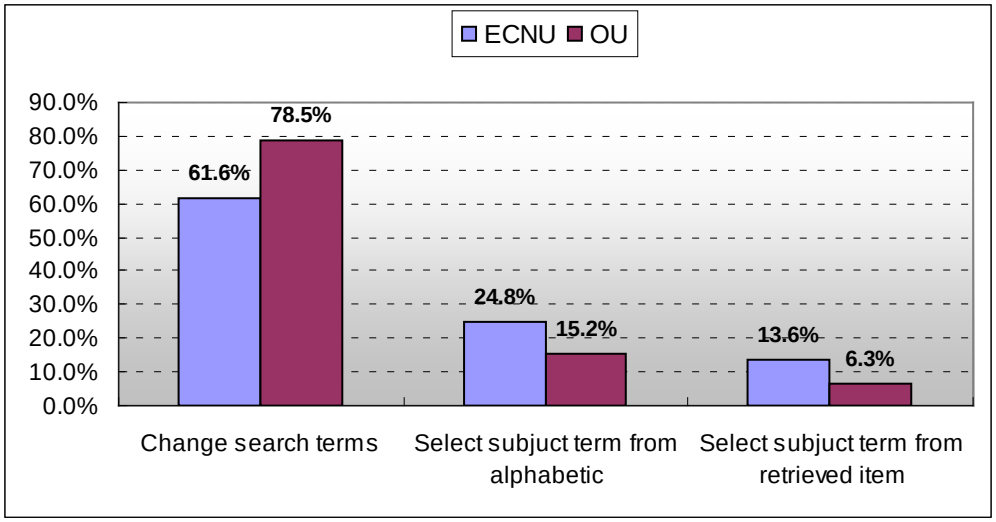

Figure 7. How to adjust search terms when search results are zero (ECNU/OU)

The interviews conducted at OU showed the similar result (see Figure 7) (Antell & Huang, 2008). Zero results made users at a loss as to what to do. They would most likely change the search terms. With zero results, the OPACs at both universities would provide suggested subject headings for the topic searched. More users (24.8%) from ECNU took the suggested headings for their continued search.

V. Further Discussion

The satisfaction level was relatively low for OPAC subject search. None of the respondents thought it was "excellent," or "very good." Only 12.8% thought it was "good," 28.7% thought it was "average," and 58.8% thought it was "not good."[4] The main reasons lie in the choice of subject terms, as listed below.

[4] "Excellent" means 90% of search results are relevant to the topic, "very good" means 75%. "good" means 60%, "average" means 45%, and "not good" means 30%.

1. Using very general subject terms for subject search would lead to an extremely large number of results. The user would have to spend too much time going over the titles on hundreds of pages. For example, the input of the term “international trade” would result in 949 entries displayed on 461 pages on ECNU’s OPAC, and the input of “real estate” would result in 1030 entries displayed on 624 pages.

2. If a user was used to the natural language for search engines such as Google, and input a long phrase or even a sentence, it is very likely that he or she will get very few or even zero results. For instance, using “parents’ affection and healthy children” as a search term would yield many relevant results in Google but zero results in the OPAC.

3. If a user was used to the search terms used in certain databases and input very specific search terms, the search results on OPAC would likely be very few or even zero. For instance, “ink divided into five colors” is a special term of Chinese painting in fine arts. Because it is too specific, the results shown on ECNU’s OPAC would be zero. The search query or formula “biology AND course evaluation AND (US OR UK OR Hong Kong OR Taiwan OR Australia OR New Zealand)” could retrieve many relevant results in a database, but would produce zero results from the OPAC.

4. Without knowledge about basic search rules, a user would likely fail to find the results in need. For example, if no results were found by inputting “Mozart sonata,” then followed by “Mozart idyllic sonata,” it indicates that the user did not know that an increase in the number of search terms entered would reduce rather than increase the number of search results.

5. Newly created terms may be too new to serve as search terms in OPAC. For instance, “low-carbon economy” may not have been collected into the subject heading list and therefore would not yield any search results.

VI. Conclusion and Suggestions

The data and analysis from this study indicate that library users, especially college

students, have a different searching behavior. They are more used to the search method adopted by Internet search engines. In their search for information, natural language is used much more often than controlled vocabulary provided by cataloging librarians. If libraries want to continue to play a vital role in information provision, especially in today's digital world, more efforts need to be made to improve OPAC's functionality, provide more user-friendly subject headings, and meet users' higher expectations. Subject search should be a key component of the improvement.

1. Subject terms need to be standardized. Subject headings need to be assigned to records carefully. If the subject headings assigned to records are too general or too broad, users will get too many hits. Consequently, the relevance and precision of results will be low. On the other hand, if the headings are too specific or too narrow, users may not find all relevant materials. That will negatively affect the recall ratio. So each record should contain both broader subject headings and more specific subject headings to balance the recall and precision of search results.

2. In the ECNU and OU library systems, some old print materials either do not have any subject headings or the subject headings are out-of-date due to lack of manpower and sufficient resources to add or update in the past. With more advanced technologies, some tasks traditionally carried out by librarians can be accomplished by the library system. That will free librarians from some simple repetitive daily work and enable them to add subject access points to those old records. The enhancement of old records will increase the recall of subject search.

3. To make subject search easier for users who do not have any knowledge about the controlled vocabulary, incorporating the natural language into authorized subject headings, as some librarians (Beall, 2006; Muddamalle, 1998) suggested, is a good practice that libraries should adopt. "Keyword search" should be implemented within the subject search in the libraries to help increase the recall of subject search.

4. Libraries should use new technologies and features available to them, such as Web 2.0, tagging, and word cloud, to make OPAC search more user-friendly. Such new

technologies and features allow librarians to interact with users online and provide users with opportunities to participate in the construction of OPAC. Librarians can use some terms, especially new scientific terminology, tagged on the OPAC by users, to supplement the controlled vocabulary since it takes time for such terms to be incorporated into the authorized subject headings list. A user-friendly library catalog search tool, called "AquaBrowser Library," offers a revolutionary search method for users. After a search term is entered on OPAC, a Word Cloud will appear, displaying related terms, alternate spellings, and even translations of that search term. It also shows a comprehensive list of options and suggestions to help users refine and pin-point their searches. This search technology is especially helpful for those users who do not know the authorized subject terms well but want to find information through subject search. Many libraries in the U.S. and U.K. have already started using this tool.

5. The comparison between the two universities discussed above shows that it is hard to find students with information literacy in its true sense. Reference librarians need to find more efficient ways to train users. Appropriate use of subject search is an important way of getting accurate relevant search results from OPAC. To guide students to correct use of subject headings is an important duty of reference librarians. Association of College & Research Libraries (ACRL) published *Information Literacy Competency Standards for Higher Education* in 2000. It emphasizes the use of controlled vocabulary, "The information literate student . . . selects controlled vocabulary specific to the discipline or information retrieval source." (ACRL, 2000, p. 11). At present time, the Library Society in China (LSC) has not formulated a national criteria for information search capability. However, *The Index System for Information Search Capability for the Universities in Beijing* was published in 2005, which states that students with information search capability should "know the basics of controlled search terms and how to use them"; "be able to make comprehensive use of natural language and controlled language and its vocabulary list, and to decide on search terms (e.g., subject terms, keywords, synonyms, and related terminology)"; and "know the similarities and differences between searches made on the Internet

search engines and those made in the information search systems provided by libraries." Besides teaching freshmen how to use the library through general training, some special seminars with contents varying to some extent according to the majors also need to be offered. The core contents should include the subject search to introduce the structure of controlled vocabularies, some commonly used subject headings in the field, and how to use the related subject headings provided by the library system to adjust the search terms, etc.

Since subject search is the most difficult search method for many library users, cataloging librarians and reference librarians need to corporate and work together to make the subject headings easy to use and to teach users how to master this method better. The authors believe that subject search as a valuable search method will significantly increase the precision of search results when it is used properly.

References

Antell, K., & Huang, J. (2008). Subject searching success: Transaction logs, patron perception, and implications for library instruction. *Reference & User Services Quarterly, 48*(1), 68-76.

Association of College and Research Libraries (ACRL). (2000). *Information literacy competency standards for higher education*. Retrieved April 20, 2010 from: http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf

Beall, J. (2006). Ethnic groups and Library of Congress Subject Headings. *Colorado Libraries, 32*(4), 37-44.

Carstens, T., & Buchanan, H. (2004). The future of the catalog: A user-friendly academic search engine. *Technical Services Quarterly, 22*(2), 37-47.

Drabenstott, K. M., Simcox, S., & Fenton, E. G. (1999). End-user understanding of subject headings in library catalogs. *Library Resources & Technical Services, 43*(3), 140-160.

Mann, T. (2003). Why LC subject headings are more important than ever. *American Libraries, 34*(9), 52-54.

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.

OCLC. (2002). *How academic librarian can influence students' web-based information choices*. Retrieved April 20, 2010, from: http://www5.oclc.org/downloads/community/informationhabits.pdf

Olson, H. A., & Boll, J. J. (2001). *Subject analysis in online catalogs* (2nd ed). Englewood, CO: Libraries Unlimited.

Rolla, P. J. (2009). User tags versus subject headings: Can user-supplied data improve subject access to library collections? *Library Resources & Technical Services*, *53*(3), 174-184.

Russell, S. E., & Huang, J. (2009). Increasing information retrieval through subject access improvements. In C. M. Garcia & T. A. Flores (Eds.), *Academic Librarianship in the 21st Century* (pp. 75-90). Hauppauge, NY, Nova Publishers.

Si, L., & Chen, X. (2009). The investigation and analysis in OPAC system of 30 university libraries in China. *New Technology of Library and Information Servic*e, 7/8, 28-32. (in Chinese)

Song, D., & Zhan, M. (2005). A comparative research on OPACs inland and overseas. *Document, Information & Knowledge*, 3, 44-46. (in Chinese)

Steele, G. (2010). The wisdom of the cataloguers: Using collective intelligence to assign Library of Congress Subject Headings. *Catalogue & Index*, 161, 6-9.

Steinhagen, E. N. (1996). Cataloging news. *Cataloging & Classification Quarterly, 22*(2), 89-97.

The Index System for Information Search Capability for the Universities in Beijing. (2005). Retrieved April 20, 2010, from: http://edu.lib.tsinghua.edu.cn/InformationLiteracy/index.htm

Xu, Z. (2006). Subject retrieval system optimization technique. *Modern Information*, 10, 149-150, 175. (in Chinese)

Yi**,** K., & Chan, L. M. (2010). Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. *Journal of the American Society for Information Science and Technology*, *61*(4), 677-687.

Yu, H., & Young, M. (2004). The impact of web search engines on subject searching in OPAC. *Information Technology and Libraries, 23*(4), 168-180.

Zhang, M. (2004). The analysis of some factors on the quality of library OPAC system. *New Century Library*, 2, 42-44. (in Chinese)

Submitted to CLIEJ on 5 May 2011.

Guo, Jingchi, & Huang, Jie. (2011). Subject headings and subject search: A comparative study. *Chinese Librarianship: an International Electronic Journal*, 31. URL: http://www.iclc.us/cliej/cl31GH.pdf