Wei-Hsin Su. A Visual Enhancement for Metadata Generation Tools: A Semi-Automatic Approach via KWIC and Highlighting. A Master's Paper for the M.S. in I.S degree. April, 2008. 48 pages. Advisor: Jane Greenberg

This paper reports on a study that examined a visual enhancement for NC Health Info, an online health information portal for NC residents. The research goal was to improve the Health Topic assignment with a semi-automatic approach via KWIC and highlighting. The study had three components: a contextual inquiry investigating improvable areas; a prototype developed according to the contextual inquiry findings; and a comparative user study evaluating the effects of the proposed approach on the assignment of Health Topics and users' perceptions of two systems. The experiment results proved that the prototype significantly reduced the cataloging time and may potentially improve metadata quality. Additionally, measured users' perceptions of the proposed system were positive. This approach is expected not only to improve NC Health Info services but further enhance metadata generation tools in the future.

Headings:

Highlighting

KWIC

Metadata Generation Tools

Metadata Quality

Semi-Automatic Metadata Generation

Subject Metadata

A VISUAL ENHANCEMENT FOR METADATA GENERATION TOOLS: A SEMI-AUTOMATIC APPROACH VIA KWIC AND HIGHLIGHTING

by
Wei-Hsin Su

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 2008

Approved by

_____

Jane Greenberg

Table of Contents

1 Introduction 3
2 Literature Review 6
2.1 Metadata for Web Documents 6
2.2 Automatic Metadata Generation 7
2.3 KWIC 10
2.4 Highlighting 10
3 Research Methods 13
3.1 Contextual Inquiry 13
3.1.1 Documentation Analysis 13
3.1.2 Database Analysis 14
3.1.3 Usability Evaluation 15
3.2 Prototype Design 17
3.3 Experiment Design 23
3.3.1 Participants 23
3.3.2 Resource Sample 23
3.3.3 Environment 24
3.3.4 Measurements 24
3.3.5 Procedure 24
3.4 Follow-ups 25
3.4.1 Reexamination 25
3.4.2 Data Analysis 25
4 Experiment Results 26
4.1 Summary of Participant Data 26
4.2 Performances 27
4.2.1 Cataloging Time 27
4.2.2 Score 28
4.3 Perceptions 31
5 Discussions, Limitations, and Recommendations 32
5.1 Discussions 32
5.1.1 Performances 32
5.1.2 Perceptions 34
5.2 Limitations 35
5.3 Recommendations 36
6 Conclusions 36
Acknowledgements 37
Bibliography 38
Appendices 44
Appendix A. Principles and Key Points of Contextual Design Employed in the Study 44
Appendix B. Pre-test Questionnaire 45
Appendix C. Post-test Questionnaire 46

List of Tables

Table 1. Examples of Applications Using Color Highlighting 12
Table 2. A Resource Assigned with Different Local Terms and the Same Health Topic 15
Table 3. Statistics of Cataloging Times (min) 27
Table 4. Statistics of Scores 28
Table 5. Statistics of Revised Scores 30
Table 6. Statistics of Perceptions 32
Table 7. Perceptions of Each Cataloger 32
Table 8. Quotations from Open-Ended Discussions 35

List of Figures

Figure 1. The Front Page of NC Health Info Web Site 4
Figure 2. The Traditional KWIC Index 10
Figure 3. Search Terms Highlighted in the Search Result 11
Figure 4. Search Terms Highlighted in the Google Cached Page 11
Figure 5. Simplified ER Diagram 15
Figure 6. The Distribution of Numbers of Health Topics Assigned to a Resource 15
Figure 7. The Current Interface 17
Figure 8. The Reduced Visible Areas of Parallel Windows 17
Figure 9. The Conceptual Draft of the Proposed Approach 19
Figure 10. The Initial Draft of the Current Interface Prototype 20
Figure 11. The Initial Draft of the Proposed Interface Prototype 20
Figure 12. The Final Version of the Current Interface Prototype 22
Figure 13. The Final Version of Proposed Interface Prototype 22
Figure 14. The Calculation of the Score 24
Figure 15. Average Cataloging Times for Each Resource 27
Figure 16. Cataloging Times for Each Resource by Each Cataloger (compared) 28
Figure 17. Cataloging Time for Each Resource by Each Cataloger (individually) 28
Figure 18. Scores of Each Resource Received by Each Cataloger (compared) 29
Figure 19. Scores of Each Resource Received by Each Cataloger (individually) 29
Figure 20. Revised Scores of Each Resource Received by Each Cataloger (compared) 31
Figure 21. Revised Scores of Each Resource Received by Each Cataloger (individually) 31

1 Introduction

With the rapid increase in Internet resources and a growing awareness of health management, the World Wide Web has become one of the major sources of health information serving the public. A series of online health information seeking and usage behavior studies conducted by the Pew Internet & American Life Project from 2000 to 2007 had reported this significant growth (Fox & Rainie, 2000, 2002; Fox & Fallows, 2003; Fox, 2005, 2006; Madden & Fox, 2006). In 2006, approximately 80% U.S. Internet users which counted up to a total number of 113 million adults sought medical and health information online (Fox, 2006). It was found that many patients, their family and friends would rather go to the Internet for health information needs because it is easily accessible and the best for certain issues (Ferguson & Kelly, 1999; Fox & Rainie, 2005).

Medical information can be a matter of life and death. One of the biggest issues with online health information is its lack of quality control. Not only is there the risk of Web sites containing false or insufficient information, but patients may not be able to select relevant resources from a huge volume of results returned by search engines and assess the contents accurately (Impicciatore, et al. 1997; Eysenbach & Diepgen, 1998; McClung, Murray & Heitlinger, 1998; Shon & Musen, 1999; Berland, 2001). Fox (2006) reported that most health information seekers start at a search engine. Although most of them feel assured of information they found, very few check the information source and the updated time. Furthermore, a quarter of them are frustrated or confused by the difficulty to find relevant information (Fox, 2006). The National Library of Medicine (NLM) has recognized the necessity of providing access to trustworthy online health service information as a gatekeeper, especially at the local level (Hilligoss & Silbajoris, 2004). NC Health Info (http://www.nchealthinfo.org)(Figure 1) was thus developed by Health Science Library (HSL) at University of North Carolina-Chapel Hill (UNC-CH) to address this urgent need.

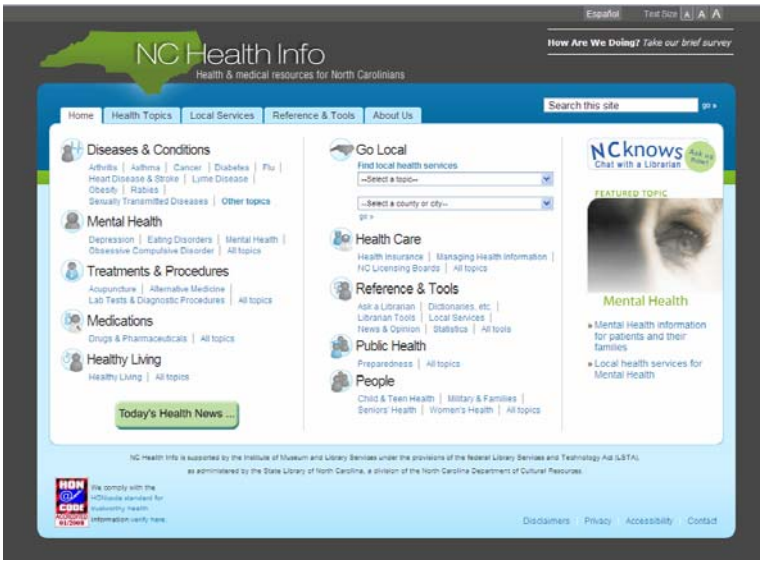


Figure 1. The Front Page of NC Health Info Web Site

Launched in 1999, NC Health Info is the first NLM funded Go Local site and served as the leading example to 24 other Go Local projects. It offers NC residents a central access portal to quality online information about local health services and the diseases these services address. Specifically, NC Health Info does this by finding, selecting, describing, and presenting these authoritative resources. One significant responsibility of NC Health Info catalogers is to describe the resources accurately and appropriately for the discovery purpose. A key focus is the assignment of adequate Local Terms and Health Topics for each resource from local controlled vocabularies. In NC Health Info, Local Term controlled vocabulary developed by HSL medical librarians consists of terms describing service types of resources, such as “Cancer Clinics”, “Pain Clinics”, and “Weight Management Programs”, etc., where as Health Topic controlled vocabulary is a subset of Medline Plus health topics used to depict specifically the health services provided, for instance, “Lung Cancer”, “Back Pain”, “Eating Disorders”, and so forth. Topic pairings are combinations of Local Terms and Health Topics developed by NC Health Info catalogers to expedite assignment process. For example, a pairing of “Cancer Clinics – Lung Cancer” indicates the common association.

Health Topics complement Local Terms by providing information about the diseases and conditions resources addressed. Nevertheless, the assignment process of Health Topics is time-consuming and requires human intelligence. More than 400 Health

Topics were in use according to the data of 2007, among which 32 were topic parents grouping related topics together to facilitate the cataloging. Furthermore, close to 1,200 synonyms of Health Topics and 9,618 topic pairings of Local Terms and Health Topics have been developed to broaden the topical coverage and expedite the process, respectively. There is no limitation on the number of Heatlh Topics assigned to a resource; rather the catalogers are required to generate extensive but specific topic metadata. A previous study analyzing annotations about problematic metadata between catalogers during work revealed that most annotations focused on topical scope. The results implied that catalogers sought consensus on Health Topics assignment through iterative annotations (Blake, et al., 2005). The time and effort it takes to assign Health Topics to a resource also depends on the complexity of Web sites. For instance, a Web site with only one page is very easy-cataloging compared to one with a complicated hierarchy of multiple pages. For the latter it usually takes catalogers a while to identify important pages and make sure that no relevant concept is missed. Furthermore, in order to assure the quality of information presented to consumers, resources have to be reviewed once a year and the metadata have to be amended to reflect the changes of the services indicated in the resource content. This work is time demanding for catalogers, and limits the time they can devote to discovering and cataloging new resources (Hilligoss & Silbajoris, 2004). The complicated Web sites and dynamic content make it difficult to handle Health Topic issues. Manual assignment of Health Topics was therefore identified as a crucial problem to be improved in this study.

By the end of year 2007, NC Health Info contained approximately 6,200 Web sites/pages, which was three times larger than the size during its first year. The exceedingly large expansion of the existing collection and new resources have pushed NC Health Info to seek a solution to improve the cataloging process, strike a balance between metadata generation and maintenance, and further provide high quality services. The study aims to improve the Health Topics assignment by proposing a visual enhancement for metadata generation tools, which is a semi-automatic approach via KWIC and highlighting. This approach is expected not only to improve NC Health Info services but further enhance metadata generation tools in the future.

2 Literature Review

2.1 Metadata for Web Documents

The World Wide Web has been expanding at an unimaginable pace ever since 2000, with new Web pages being generated every moment in time. It was estimated last year that there were around 30 billion Web pages existing in the cyberspace, which was almost three times of year 2005 (Pandia, 2007). While the Web presents a growing volume of valuable and ready information, it also presents more and more inaccurate poorly rated information. In other words, Internet resources are easily accessible with a simple click, but large amounts of irrelevant pages retrieved by search engines just because they contain the submitted search terms hinder users from the information they desire. The characteristics of uncontrollability and instability of online resources call for well management (Vellucci, 1997; Koehler, 1999).

Librarians and information specialists have been devoted to providing better access to Internet resources meeting specific user needs by selecting “good-quality” resources in particular domains and organizing them in a manner similar to traditional library catalogs. Although the cataloging principles for print materials can apply to the Internet, a number of limitations are found, due to the nature of Web documents such as various formats, ephemeral existence, and mixed contents, and make it difficult to describe Internet resources (Flannert, 1995), even for professionals. Koehler in his study indicated that there were two kinds of metamorphosis of Web resources, namely persistence (the existence) and change (the content), by which catalogers were greatly challenged (1999).

Metadata has been generally used to describe and organize information resources. Within this context, subject metadata has been shown to be crucially helpful to resource discovery and to provide access (Spink, et al., 2001). However, it is one of the biggest challenges because a single Web page may contain information of any kinds of subjects, a Web site may consist of complicated hierarchical Web pages, and the content is not static (Ellis & Vasconcelos, 1999). Previous studies noted that content changes including the breadth and depth of information a Web site or Web page contains occur more frequently than structural changes (Vellucci, 1997; Koehler, 1999). The cataloging and maintenance process is often time-consuming and requires human judgment (Weihs,

1996; Anderson & Perez-Carballo, 2001; Ciravegna & Wilks, 2003; Greenberg, 2004). Catalogers usually have to browse back and forth looking up to down within Web sites and Web pages to decide proper subject metadata, which makes this task more difficult (Banerjee, 1997).

2.2 Automatic Metadata Generation

Many studies and projects had been conducted exploring automatic metadata generation for Web resources because manually cataloging is a very time-consuming, labor-intensive, and high-cost work, and may result in inconsistency. There are two kinds of automatic metadata generation applications depending on the generation methods, that is, metadata harvesting and extraction (Greenberg, 2004).

Harvesting tools gather metadata from META tags created either manually by the authors or automatically/semi-automatically by software between <header> and </header> in HTML source codes (Greenberg, 2004). DC-dot (http://www.ukoln.ac.uk/cgi-bin/dcdot.pl) is an example of this kind of application outputting metadata following Dublin Core standards (UKOLN; Greenberg, 2004). Nevertheless, research has shown that Web pages rarely contained metadata, not to mention those adopting Dublin Core guidelines (Lawrence & Giles, 1999; Vinyard, 2001). Even META tags used are likely not relevant to the content of Web pages due to the abuse or the misuse of metadata (Marchiori, 1998; Lawrence & Giles, 1999). It is risky and unrealistic to rely solely on this kind of harvesting tools.

Instead of collecting existing metadata from META tags, metadata extraction software employ text/data mining algorithms to capture key concepts from the content. Related issues in this area include automatic web classification and automatic web indexing. Several Singapore researchers raised the issues in Web data mining in 1999 (Madria, et al.). In the same year, Garofalakis and others also explored the Web data mining. Gietz in his report introduced the importance of automatic classification and significant systems at that time (2001). In 2005, Sukakanya and Porkaew proposed a framework to automatically classify e-Business Web content.

Large-scale efforts have been devoted to develop sophisticated mining algorithms to facilitate web content management. Natural language processing (NLP) techniques

such as rules-based approaches, statistical analysis, Bayesian probability, neural network, support vector machine (SVM), semantic and linguistic clustering are prevailing methods to support knowledge extraction (KE) for Web documents. In 1999, Ardö and Lyngby in the DESIRE project applied heuristics and weighting schemes to improve the classification process in which texts of Internet documents were extracted and matched with the Ei-thesaurus. Further study was conducted later in the DESIRE II project cooperating with OCLC to automatically classify Web services (Koch & Vizine-Goetz, 1999). GERHARD is another project using linguistic techniques to classify Web pages based on the Universal Decimal Classification schema (Muller, et al., 1999), while Jenkins and his colleges grounded on another classification scheme – Dewey Decimal Classification (1998). Loia and Luongo conducted their study to automatically categorize Web pages using a genetic-based fussy clustering methodology (2001). Taiwanese researchers proposed a semantic approach extracting classification knowledge of Web documents with mining term associations (Lin, et al., 1998). In the same year, Mase tested to categorize Web pages for IR system automatically according to a Knowledge Base. A positive example based learning method using SVM was applied to classify Web pages in 2002 (Yu & Han). Calvo and two Korean researchers used a Naïve Bayes algorithm to manage Web content (2004). Later, Cho and Richards further posed a BayesTH-MCRDR algorithm combining the naïve Bayesian algorithm using Threshold and the MCRDR algorithm (2004). Golub classified Web pages based on a controlled vocabulary by counting the term frequencies and assigning weight (2006).

Ontology is another popular approach that can be integrated with NLP methods for automatic on-line resources management. In 2003, Korean researchers presented two studies: one study examined an automatic Web page classifier using adaptive ontology (Noh, et al.), and the other study used an ontology which expresses terminology information and vocabulary in the Web content in a hierarchical structure (Song, et al., 2006). A latter study also classified Web documents by extracting concepts from ontologies (Litvak, et al., 2007).

Besides analyzing the semantic content with NLP techniques, other methods approach automatic Internet document classification and indexing based on the features of the Net such as HTML structure and hyperlinks. Previous studies found that important

concepts were discovered by analyzing the structure of Web documents such as HTML tags and links (Attardi, et al, 1999). Golub and Ardö also put great emphasis on the importance of HTML structural elements to automated subject classification (2005). Yi in 2007 further explored how hyperlinked external resources contribute to subject discovery.

With the development of automatic Web document processing, scholars have focused on improving automatic metadata generation and evaluation. Liddy, et al. conducted research in this area applying the NLP techniques (2002). Another effort is the SAmgI (Simple Automatic Metadata Generation Interface) project, which designed and implemented an innovative application attempting to reach the goal of automatic metadata generation (Meire, Duval, & Ochoa, 2007). Kris in his unpublished doctoral thesis depicted a general framework for automatic metadata generation (2007).

Contrasting with focused automatic algorithms, very few studies approached metadata generation from an overall system functionality perspective. Greenberg, et al. (2002) examined iterative design aspects for metadata creation tools. The AMeGA (Automatic Metadata Generation Application) project provided a thorough overview on the functionalities required for automatic metadata generation applications (Greenberg, Spurgin & Crystal, 2006).

Although machines are capable of generating acceptable metadata, Greenberg concluded that humans are still better adequate to produce metadata for domain needs, especially the judgment-demanding subject metadata (2003, 2004). Furthermore, scholars claimed that automatic metadata generation would not be successfully adopted without human being's trust (Irvin, 2003). The best way to solve this dilemma seems to be combining the power of machines and valuable human intelligence to improve the quantity and quality of Web metadata (Vellucci, 1997; Hirokawa, Itoh & Miyahara, 2003; Greenberg, 2003, 2004). It was proposed that subject metadata generation of Web documents could be improved by automatically mapping manually generated metadata to a thesaurus (Greenberg, 2003). In fact, the combination of human and machines is possible to be effectively and efficiently achieved with the concepts of Keyword-in-Context (KWIC) and highlighting.

2.3 KWIC

Adapted from Crestadoro's Keywords in Titles indexing system by Luhn in 1959, KWIC, standing for Keyword-in-Context, is a technique to produce a special word index. The method applied to information retrieval is based on the belief that a title is the compact abstract of a document, within which each keyword of a title serves as an entry point and is displayed sandwiched with its surrounding words, namely the context (Sedano, 1964; Baldonado, 1998; Käki, 2006). It has been proved implicitly with our daily experiences and is supported by previous research that context is of crucial importance in order to catch the connotations, avoid false negative errors, and make accurate relevance judgment (Marchionini, 1995; Hussam, et al., 1998).

A traditional KWIC index is aligned with keywords in the middle for rapid browsing (Sedano, 1964; Baldonado, 1998; Käki, 2006) as illustrated in Figure 2, an example of how this paper's main title "A Visual Enhancement for Metadata Generation Tools" would appear in a KWIC index. In other words, position is used in KWIC to direct users' visual attention and keep them from distraction, which provides similar functions as the color of a highlighter. However, dynamically rearranging text position is disturbing and impractical when users' are viewing search results or within documents. The concept of KWIC is later generally adopted highlighting keywords via various font formats such as **bold**, *italic*, and color.

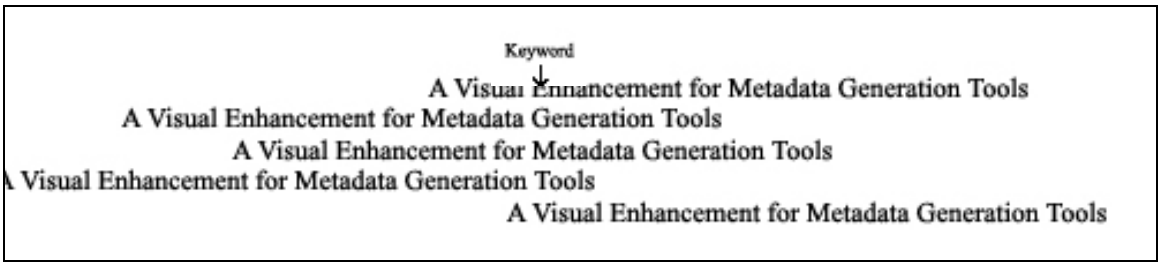


Figure 2. The Traditional KWIC Index

2.4 Highlighting

Highlighting seems to be such a natural behavior of human being. A mother makes circles on the calendar to remind her of family's birthdays; students mark the important sentences on their textbooks using highlighter with different colors; analysts

underscore the crucial statistic data. People highlight everyday everywhere for various purposes: put emphasis, direct attention, allow for easy comparison and skimming later. In the digital environment, highlighting keywords in context using color is seemed as the *de facto* standard of visual enhancement for browsing and searching, and is widely used by IR systems, including general search engines, databases, and within-document search functions such as document editing/viewing tools, especially electronic books (Brown, T.J., 1991; Baldonado, 1998; Wu & Yuan, 2003; Chi, et al., 2004). The following figures and table give examples of applications providing the highlighting function/option to facilitate navigation and relevance judgment (Figure 3 & 4; Table 1).

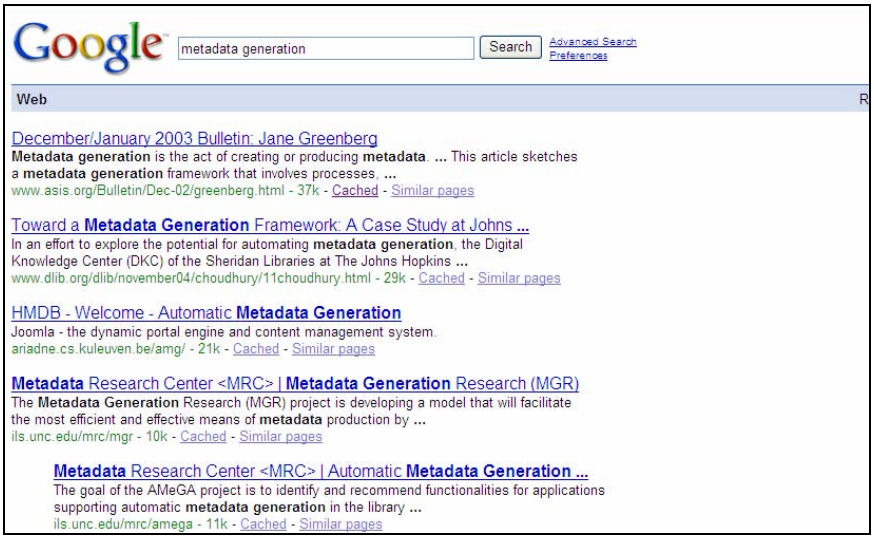


Figure 3. Search Terms Highlighted in the Search Result

Figure 4. Search Terms Highlighted in the Google Cached Page

Category	Application
Search Engines	Google
	Yahoo!
	AOL
Databases	ISI Web of Science
	ScienceDirect
	The Philosopher's Index
Within-Document Search	Google Book Search
	SuperBook
	ScentIndex of 3book
	Adobe Reader
	Notepad++

Table 1. Examples of Applications Using Color Highlighting

Users often feel lost when performing search tasks or online reading wondering which and where are worthy further examination (Byrd, 1999). Studies in this area often take advantage of highlighting to assist users in quickly identifying potentially relevant keywords in context to visualize the locations or the occurrence distributions of these relevant elements by histogram bar, scrollbar, and pie chart (Hearst, 1995; Feldman, Dagan & Hirsh, 1998; Byrd, 1999; Harper, Coulthard & Sun, 2002). Research has found that this kind of information visualization can greatly improve users' browsing and searching performance by providing visual cues indicating potentially relevant passages within documents. This, in turn, can reduce the cognitive load and expedites fact finding and relevance judgment, particularly for a specific goal (Hearst, 1995; Hussam, 1998; Byrd, 1999; Harper, Coulthard & Sun, 2002; Wu & Yuan, 2003; Chi, et al., 2004, Harper, et al., 2004; 2005; Deller, 2007). Furthermore, highlighting technique is also utilized supporting automatic/semi-automatic annotation generation and evaluation for semantic web, which is very similar to web metadata generation (Erdmann, et al., 2000; Uren, et al., 2005).

The effects of the design of metadata generation tools still need to be investigated further. The studies reviewed suggest that a combination of human intelligence and machine operation may offer the best solution to topical metadata generation challenges, such as those noted in NC Health Info project. More specifically, research literature indicates that the application of KWIC and highlighting may expedite fact finding and improve relevance judgment process. These conclusions have led to the research topic of this study, which is that a visual enhancement for metadata generation tools may improve

the Health Topic metadata assignment process of NC Health Info with a semi-automatic approach via KWIC and highlighting.

3 Research Methods

This study had three main components: a contextual inquiry investigating improvable areas; a prototype developed according to the contextual inquiry findings; and a comparative user study evaluating the effects of the proposed approach on the assignment of Health Topics and users' perceptions of two systems. Follow-ups include result reexamination and data analysis. Detail procedures and findings before the experiment are described below.

3.1 Contextual Inquiry

The contextual inquiry followed the principles proposed by Beyer and Holtzblatt in 1998 regarding four areas: context, partnership, interpretation, and focus. See Appendix A for key points extracted from their work *Contextual Design* and employed in this research.

The investigator first met with the director to obtain an overview understanding of the project mission, framework, and its urgent need, which is to improve current metadata generation and maintenance performance for on-line health information resources, quantitatively and qualitatively. Meanwhile, she also studied important documentation to gather implicit data. Furthermore, the content and schema of the back-end database were investigated to understand the structure of the collection and the database itself. Last but not least, the current interface was evaluated based on usability principles.

3.1.1 Documentation Analysis

Many essential facts were found from the documentations: the cataloging training manual indicated clearly that the assignment of Health Topics should be extensive and specific (NC Health Info, 2003); the antecedent usability study of NC Health Info noted the effects of the presentation methods of Health Topics and the working windows (Ellington, 2004); Blake, et al.s' research analyzed annotations about problematic metadata between catalogers during work and revealed that most annotations were about

topical scope and implied that through iterative annotations catalogers sought consensus on Health Topic assignment (2005); the introductory study described the efforts to modify the Health Topic controlled vocabulary and to manage the growing collection (Hilligoss & Silbajoris, 2004). All of these reconfirmed the importance of Health Topic assignment and its acute need to be improved. The investigator also met with the previous usability study conductor to explore important issues of NC Health Info interface design and usability test, testing environment, record methods, etc.

3.1.2 Database Analysis

The schema and content of the NC Health Info database were analyzed to understand the structure of the collection and the relationship between the input-interface and the back-end database. From the ER diagram it was found that the RESOURCES table is connected with LOCAL_TERMS and HEALTH_TOPICS and tables via RESOURCES_INDEXING and INDEX_MAPPING tables (Figure 5). To prevent influencing the current collection and database, a stand-alone read-only backup of December 27, 2007 was studied instead. At that time, there were 6,206 resources, 416 Health Topics among which 32 were topic parents grouping the related topics, 1,126 synonyms, and 9,654 indices (pairings of Local Terms and Health Topics). The employment of indices indicated that careful data cleansing must be performed before calculating number of Health Topics assigned to each resource since a resource might be assigned with different Local Terms combined with the same Health Topic. Below is an example of this situation (Table 2). It was found that numbers of Health Topics assigned to a resource ranged from 1 to 94, the average number was 8.85, and most of resources (90.04%) were assign 1 to 20 Health Topics (Figure 6). These data was very helpful to the sample design later.

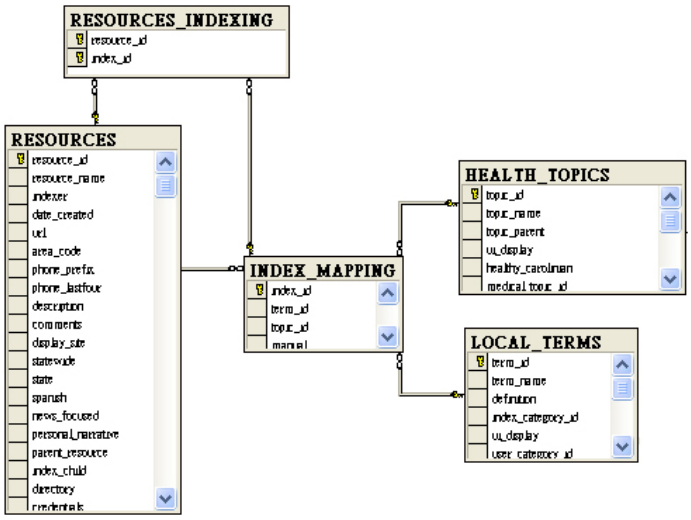


Figure 5. Simplified ER Diagram

resource_id	resource_name	index_id	term_id	term_name	topic_id	topic_name
7	Structure House	930	19	Exercise/Fitness Programs	277	Exercise and Related Topics
7	Structure House	931	108	Personal Trainers	277	Exercise and Related Topics

Table 2. A Resource Assigned with Different Local Terms and the Same Health Topic

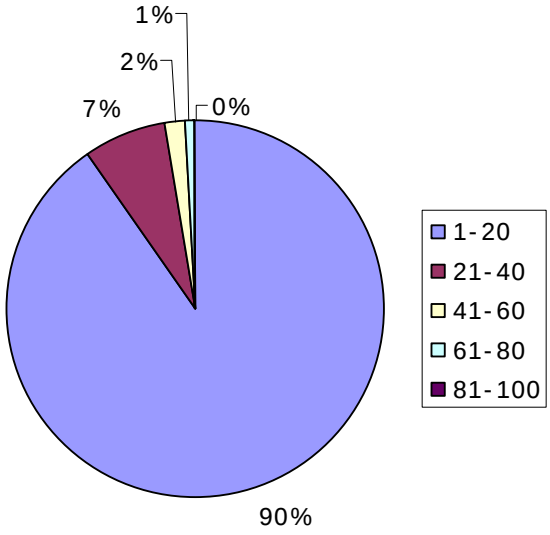


Figure 6. The Distribution of Numbers of Health Topics Assigned to a Resource

3.1.3 Usability Evaluation

System usability can greatly impact on the task performance. Thus, a simple usability evaluation of current interface according to GNOME's human interface guidelines 2.0 (2004) and Nielsen's usability heuristics (2005) was conducted to probe

into any potentially improvable area via field observation and informal interview. Catalogers were observed when performing cataloging tasks in the real work environment. During this period, the investigator asked questions such as “How did you feel when learning this system?”, “Why did you select this Health Topic for this resource?”, “Do you have any idea that might improve the system based on your daily experience?”. These questions were further discussed in the informal interview later. The investigator also used the system in person to discover the difficulties catalogers might encounter. The current interface is a three-column window (Figure 7). The typical Health Topic assignment process for a resource is simply described as follows: first, click on one of drop down menus in the left column and select a Local Term; the Health Topic Groups (topic parents) belonging to that local term would then be showed in the middle column; the Health Topics of selected Health Topic Group would appear in the right column; finally an appropriate Health Topic for the resource is assigned to the resource. This process would iterate several times until all needed Health Topics are selected. Several usability problems were detected. Although topic mapping and topic pairings were developed to expedite this process, it is still unintuitive and asks for lots of clicks (five clicks for most of topic pairings). Besides, catalogers have to work in separate windows (the system window and the resource window). It seems that during the back and forth actions, short-term memory overload had an impact on their ability to resume the examination of the resource after working on the system window, because the visual attention was broken and the focal point was lost. One cataloger works with double screens and the other two resize their windows aligning them parallel to each other so that they could work on two windows at the same time. However, this might affect the performance since the visible areas would be greatly reduced (Figure 8). Last, it might take catalogers for a while to find the locations of important concepts when the resource is a complicated Web site with unorganized content or unfriendly navigation.

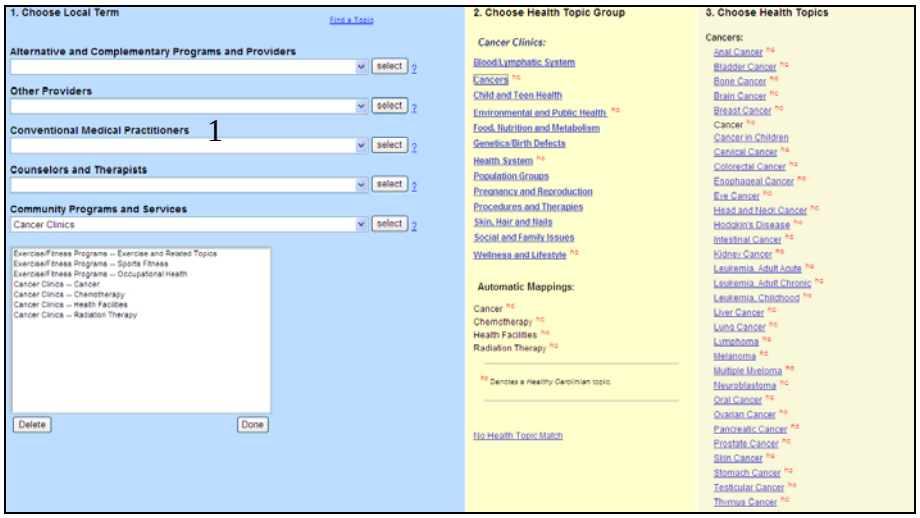


Figure 7. The Current Interface

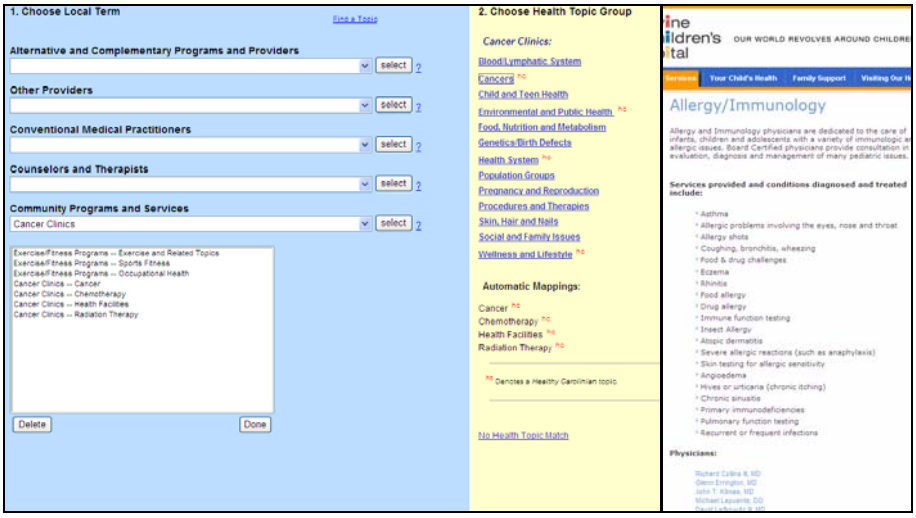


Figure 8. The Reduced Visible Areas of Parallel Windows

3.2 Prototype Design

From the extensive literature review and previous inquiry, it was found that the combination of human knowledge and machine operation may improve topic metadata generation by visualizing the locations and distributions of possible relevant keywords with KWIC and highlighting, According to these findings, a visual enhancement for metadata generation tools which may enhance Health Topic assignment process in NC Health Info with a semi-automatic approach via KWIC and highlighting is posed. In the beginning, a controlled vocabulary of the variations of Health Topics would be developed by sophisticated algorithms and human experts together. Web sites would be

automatically retrieved by URL, archived to local disks at 3 levels, and then be processed. During the automatic processing phase, source codes between HTML (Hypertext Markup Language) tags <body> and </body> of each Web page would be scanned line to line to look for possible relevant keywords. Special CSS codes (Cascading Style Sheets) would be inserted before and after Health Topics and their variations to make the highlight effects. The importance of a Health Topic and a Web page within a Web site would be counted. The calculation method is as follows. The occurrence frequencies of Health Topics including their variations in a Web page would be recorded individually. Let *k* be the number of distinct variations of a Health Topic occurring in a Web page, *j* be the number of distinct Health Topics occurring in a the same page, and *i* be the number of pages containing a certain Health Topic. For shorthand purposes, Health Topic term frequency is called HTTF here. The Page HTTF of a page *x* could be calculated as follows:

$$PageHTTF_x = \sum_{y=1}^{j} \sum_{z=1}^{k} HTTF_{xyz} \qquad (1)$$

Also, the Total HTTF of a Health Topic y within a Web site could be calculated as follows:

$$TotalHTTF_y = \sum_{x=1}^{i} \sum_{z=1}^{k} HTTF_{xyz} \qquad (2)$$

For example, if“obesity” and “overweight” occur 3 times and 4 times in a page, then the HTTF of Health Topic “Obesity” in this page would be 7; Total HTTF of “Obesity” would be the sum of the HTTFs of its all variations occurring in this Web site; the Page TTF would be the sum of the HTTFs of all variations of all Health Topics occurring in that page.

While Total HTTF could be seen as the comparative importance of a Health Topic to a resource, Page HTTF indicates the comparative importance of a page to the whole Web site. Only pages with positive Page HTTF would be considered worthy of examination and be presented to catalogers ranked by Page HTTF, and possible relevant Health Topics would be displayed and sorted by their Total HTTF as suggested Health

Topics because both GNOME usability project (2004) and Nielsen (2005) suggested that users' memory load should be minimized by removing irrelevant objects and making options visible. Whether one Health Topic should be assigned to a resource relies on human experts to make the final decision because term frequency does not necessarily mean relevancy especially when serving special purpose. For example, a Health Topic "Juvenile Rheumatoid Arthritis" may only occur once within the whole Web site, but still it is one of diseases the resource addresses. However, the automatic tasks above mentioned were manipulated by the researcher manually in this study due to research scale and time limitation. A sophisticated automatic algorithm/system is expected be introduced and work with the proposed method together in the future.

Conceptual drafts were used to make these ideas more concrete and communicate with users during the initial design process (Figure 9-11). Considering the time limitation and the risk of influencing the current collection, two prototypes of current system with only Health Topic assignment function and proposed interface were built with PHP and MySQL instead of altering the real system in use. Several modifications were made for the final versions of both interfaces. Pause button is available so that the experiment result would not be affected if the participant would like to stop for a while for any reason.

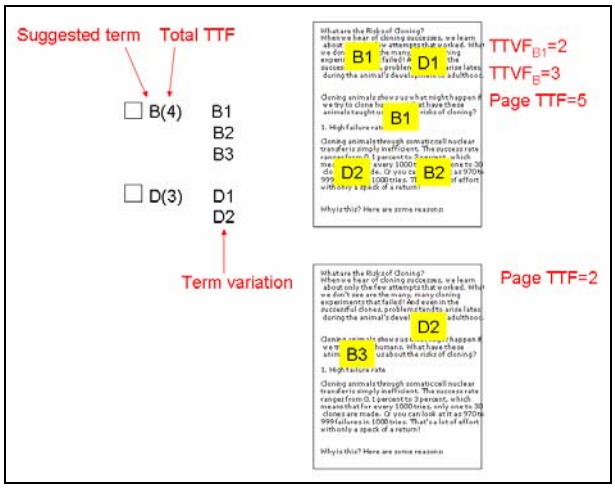


Figure 9. The Conceptual Draft of the Proposed Approach

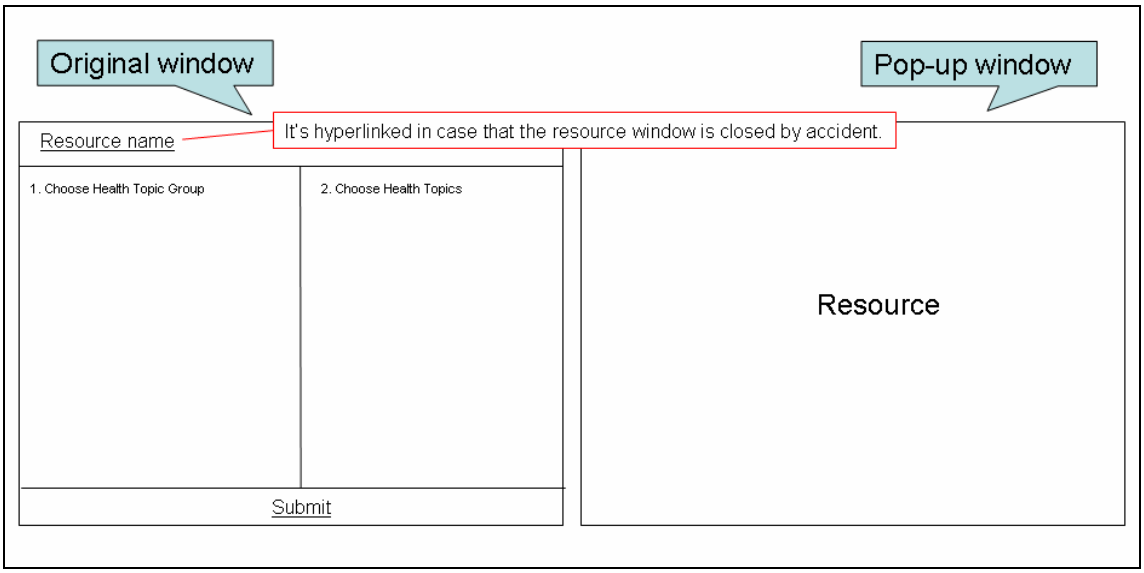


Figure 10. The Initial Draft of the Current Interface Prototype

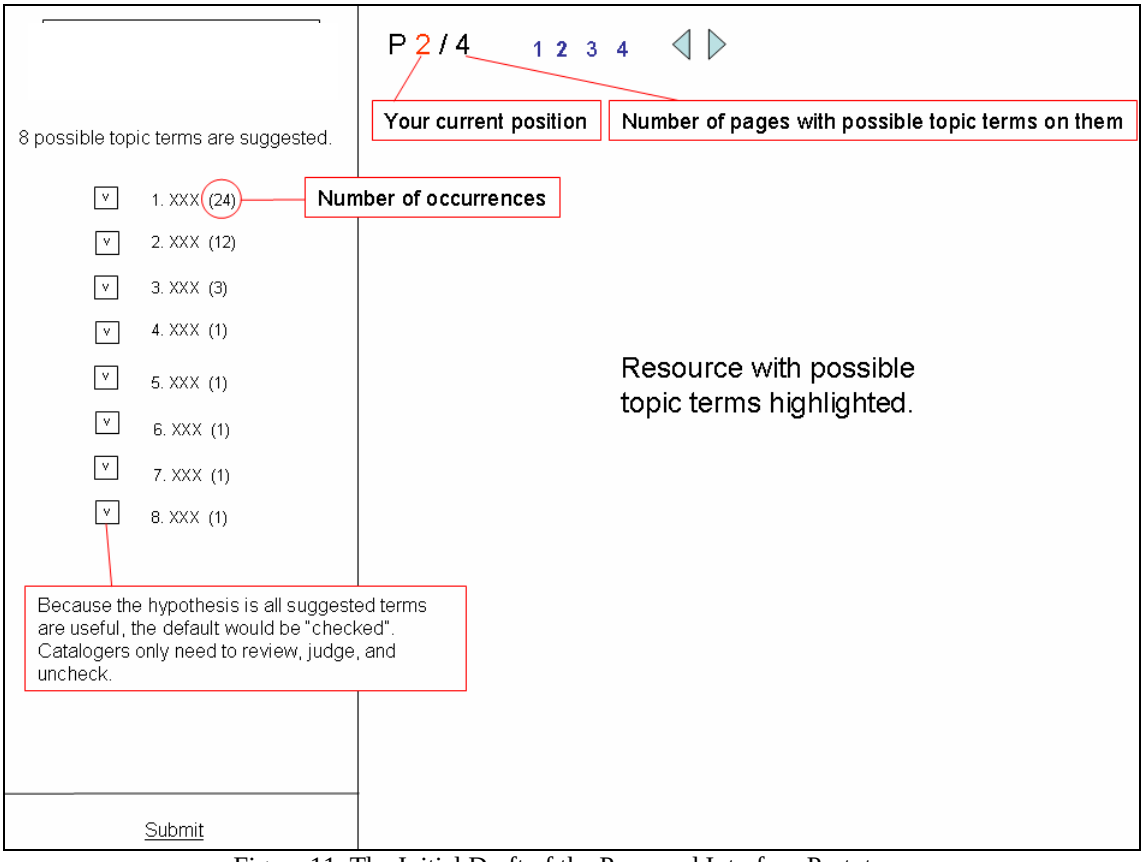


Figure 11. The Initial Draft of the Proposed Interface Prototype

The current interface prototype is a four-column window (Figure 12). The top column indicates the resource name and provides buttons for reopening the resource window and pause. The three columns below are very similar to the current interface physically and functionally. Only the participants start with selecting Health Topic Groups instead of Local Terms.

The proposed interface prototype is a two column window (Figure 13). The left one is the working panel and the right one displaying the resource Web page. The number buttons under the resource name indicates the number of important pages identified within this Web site in important order and the page currently examined showed with the number only. This feature aids with navigation too. The suggested Health Topics and their HTTFs and distributions are listed below on the left-hand side. However, an important adjustment of the conceptual drafts here is that only Health Topics occuring in the page currently examined would be displayed. This is again according to GNOME (2004) guidelines and Nielsen's (2005) suggestion to minimize users' memory load. Besides examining a Web page with all possible relevant Health Topics and their variations highlighted, users could click on a Health Topic and see it highlighted in this page only. Also, once a Health Topic is identified relevant by a user and selected, the radio button in front of it would be disabled so that users do not have to pay attention to it again, even when reviewing other pages. This approach was selected because it can reduce noise and the short term memory load, expedite the relevance judgment, and decrease the chance that important concepts would be missed.

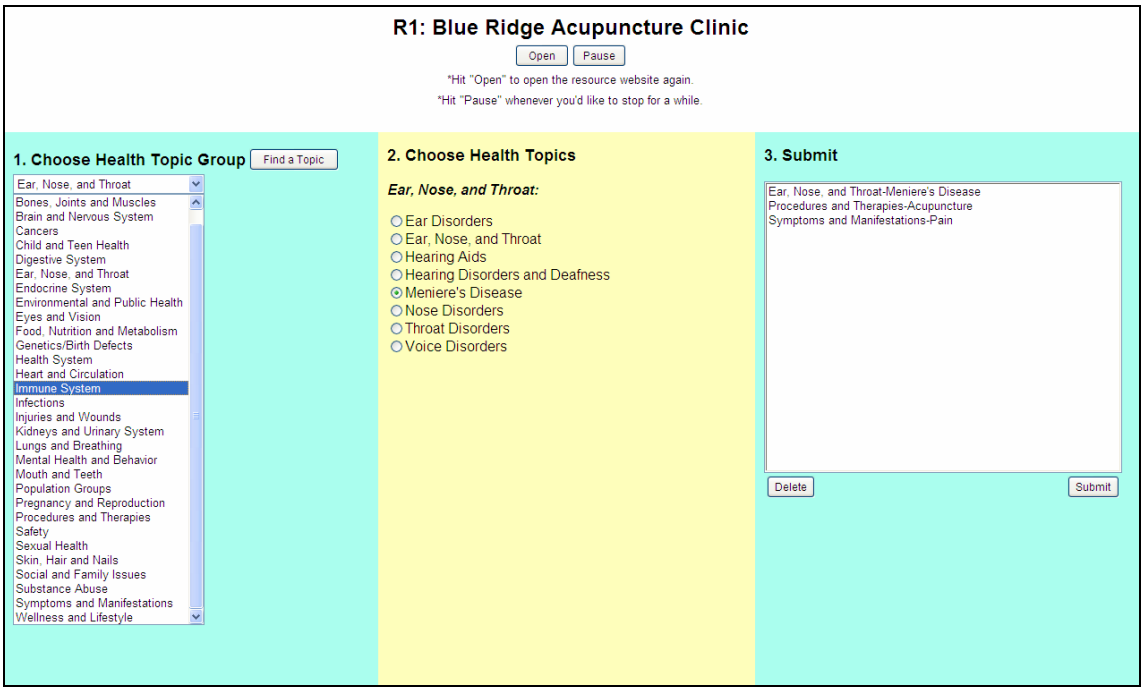


Figure 12. The Final Version of the Current Interface Prototype

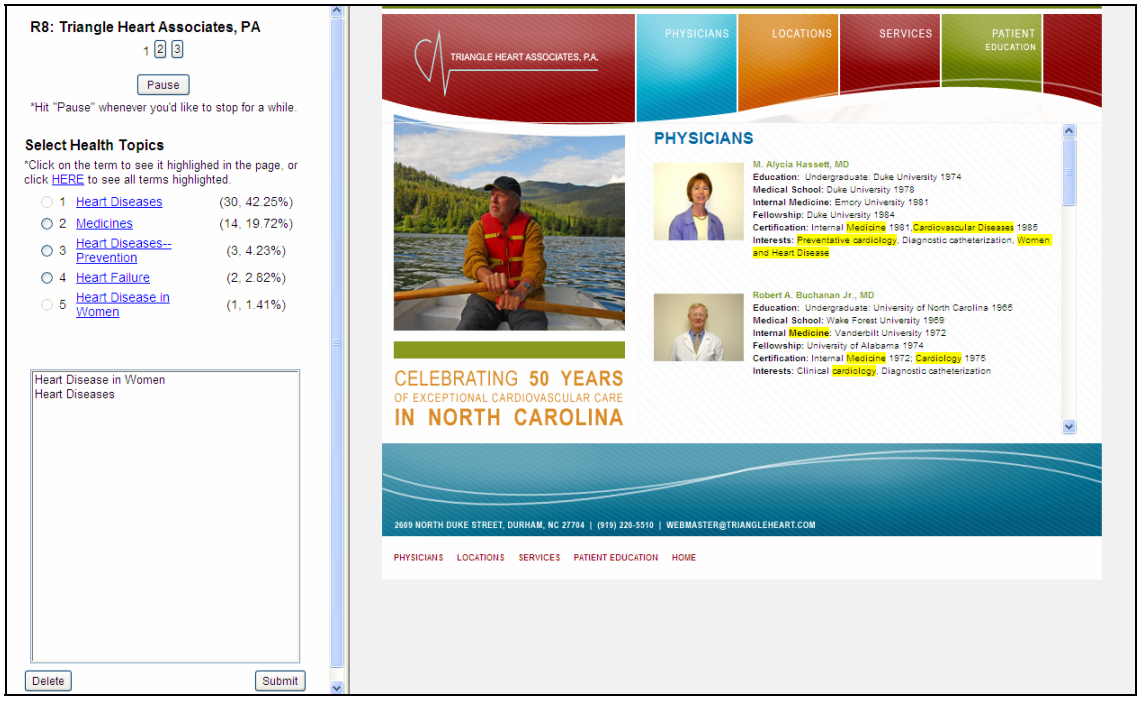


Figure 13. The Final Version of Proposed Interface Prototype

3.3 Experiment Design

Based on the previous findings, five hypotheses regarding the effects of the proposed visual enhancement for the Health Topic assignment in NC Health Info are posed and the experiment was conducted to verify them. H_{1a} and H_{1b} are about the effects on the performance, qualitative and quantitative, while $H_{1c\text{-}d}$ are hypotheses on users' perceptions. Measurements used for the study are cataloging time, accuracy rate which is later translated into score, perceived ease of learning (PEOL), perceived ease of use (PEOU), and perceived usefulness (PU).

H_{1a}: Using the proposed interface shortens cataloging time.
H_{1b}: Using the propose interface increases scores.
H_{1c}: Proposed interface is easier to learn.
H_{1d}: Proposed interface is easier to use.
H_{1e}: Proposed interface is more useful.

3.3.1 Participants

The study included four participants, two HSL librarians and two graduate students at School of Information and Library Science (SILS) at UNC-CH. Each participant has had previous experience using NC Health Info system because knowledge of specific cataloging procedure and principles are required. HSL librarians' experiences with NC Health Info are 7.5 years and 2.5 years, and students' experiences are 1 year and 0.5 year.

3.3.2 Resource Sample

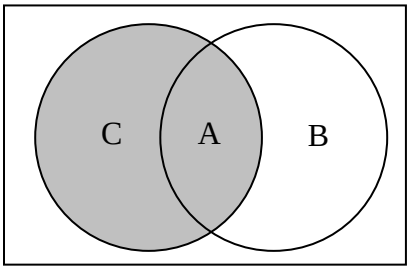
To facilitate the research, a stratified random sample of 12 resources was selected. Each of the 12 resources had been previously assigned with no more than 20 Health Topics from the NC Health Info Health Topic controlled vocabulary. The sample size and the limitation of resources being assigned no more than 20 Topic Terms previously were a practical consideration to keep the experiment duration in approximately an hour.

3.3.3 Environment

The experiment was conducted in a reserved study room at Health Science Library. To make the testing environment similar to the real working setting, keyboard, mouse, and 17" LCD were provided.

3.3.4 Measurements

The cataloging performance was measured quantitatively by cataloging time and qualitatively by accuracy rates that was translated to scores later. Cataloging times of each resource by each cataloger were logged automatically by the machine. As for the accuracy rate, because catalogers might make false positive and false negative errors which occur when they assign irrelevant Health Topics or miss relevant Health Topics, both conditions should be taken into consideration carefully. In the figure below (Figure 14), the circle in full grey is cataloger's answer set, and the circle with part of grey area is the correct answer set. Let A be the correct part of cataloger's answer, B be cataloger's false negative error, which means these Health Topics should be selected but not, and C be cataloger's false positive error, which means these Health Topics should not be selected. Thus the number of distinct Health Topics within the whole set would be the sum of A, B, and C. The accuracy rate would hence be A divided by N. Perceptions of ease of learning, ease of use, and usefulness were measured with 5-point Likert scales.



N = A+B+C;
Accuracy rate = A/N.

Figure 14. The Calculation of the Score

3.3.5 Procedure

The experiment was an unobtrusive user study consisting of five parts, including the pre-test questionnaire (Appendix B), the first cataloging session using the current interface prototype for six resources, the second cataloging session using the proposed

interface prototype for another six resources, the post-test questionnaire (Appendix C), and the open ended discussion.

After the participant signed the consent form, the investigator introduced the study and explained what the participant would be asked to do during the experiment. Two interfaces were demonstrated using a demo account. After making sure that the participant had no questions toward the experiment, the investigator left the room and then came back to discuss several open-ended questions with the participant after the cataloging sessions and questionnaires were completed. The pre-test and post-test questionnaires were built with a credible on-line survey software Qualtrics provided by the Odum Institute at UNC-CH to maintain survey quality and protect participants' privacy.

3.4 Follow-ups

3.4.1 Reexamination

After the experiment, the data including questionnaire results and cataloging task performances was gathered and analyzed. The inconsistencies were found when comparing Health Topics selected by different catalogers and Health Topics previously assigned to resources. The inconsistent Health Topics were collected in a spread sheet and sent back to the participants to be reexamined because they were the most and the only adequate professionals for this task. To avoid the potential bias, no identifiable participant and resource information was included, and catalogers performed the reexamination individually without interference. Finally, only Health Topics identified relevant more than three times were integrated into the answer set.

3.4.2 Data Analysis

Data collected was organized and analyzed using Microsoft Office Excel, Access and SPSS. Accuracy rate here was multiplied by 100 to be transferred into an easy-comparable score. Cataloging times, scores, and perceptions were calculated, compared, and tested with Wilcoxon Signed-Ranks test (Harper, et al. 2004; Käki, 2006). Detail analysis is described in the next section.

4 Experiment Results

This section reports quantitative and qualitative results on users' (catalogers) performance and their perceptions toward the two metadata systems. The first six resources were cataloged using current interface, which they use in their daily work, and the other six were cataloged with the proposed interface. System performance measured quantitatively by the cataloging time, and qualitatively by the accuracy of Health Topic assignment. Besides cataloging performance, users' perceptions of two systems including perception of ease of learning (PEOL), perception of ease of use (PEOU), and perception of usefulness (PU) were investigated from a usability point of view. Participants were also asked about if the proposed interface would be helpful to improve and maintain metadata generation work. Participants' data is first summarized in 5.1, and cataloging performances are presented in 5.2, and then perceptions in 5.3. Hypotheses are tested respectively with the non-parametric Wilcoxon Signed-Ranks test (Harper, et al. 2004; Käki, 2006). Tables and figures are displayed to give a clear description.

4.1 Summary of Participant Data

Participants are four females aged from 26 to 50. They were recruited to take part in the study because all of them have NC Health Info cataloging experience. Two of them are current HSL librarians with Master's degree in Library Science, and they had each previously been NC Health Info catalogers. The other two are graduate students at School of Information and Library Science, UNC-CH and they are current catalogers of NC Health Info. The pre-test questionnaire captured data on their previous experience with NC Health Info. HSL librarians' experiences are 7.5 years and 2.5 years, and students' experiences are 1 year and 0.5 year. To protect participants from disclosure, four participants are separated into two groups A and B according to their experience with the current system and sorted randomly. Group A are more experienced ex-catalogers and group B are less experienced current catalogers. In the rest of this section and next sections, participants are referred to as A1, A2, B1, and B2.

4.2 Performances

4.2.1 Cataloging Time

H_{1a}: Using the proposed interface shortens cataloging time.

The cataloging times for each resource by each cataloger were logged automatically in the database. The average cataloging time for a resource using current interface is 4 minutes and 49 seconds, while the average cataloging time using proposed interface is 2 minutes and 33 seconds. H_{1a} is thus proved to be true from an overall viewpoint and is significant at the level $p = 0.028$. It is also accepted in individual case except participant A2.

	Current Interface		Proposed Interface		*p*-value
	Mean	S.D.	Mean	S.D.	
A1	6.56	1.11	3.25	1.25	.028
A2	4.18	1.32	2.68	0.92	.075
B1	4.29	1.02	1.99	0.61	.028
B2	4.23	1.04	2.27	0.75	.046
AVG	4.82	1.47	2.55	0.98	.028

Table 3. Statistics of Cataloging Times (min)

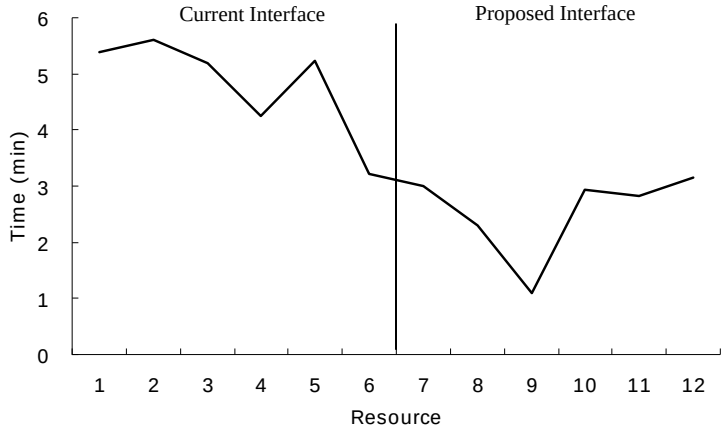


Figure 15. Average Cataloging Times for Each Resource

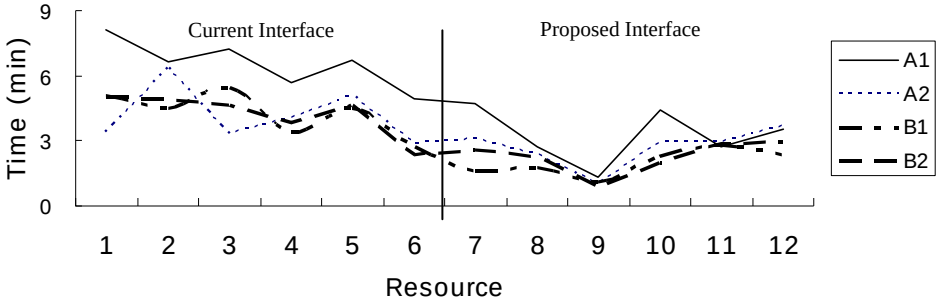


Figure 16. Cataloging Times for Each Resource by Each Cataloger (compared)

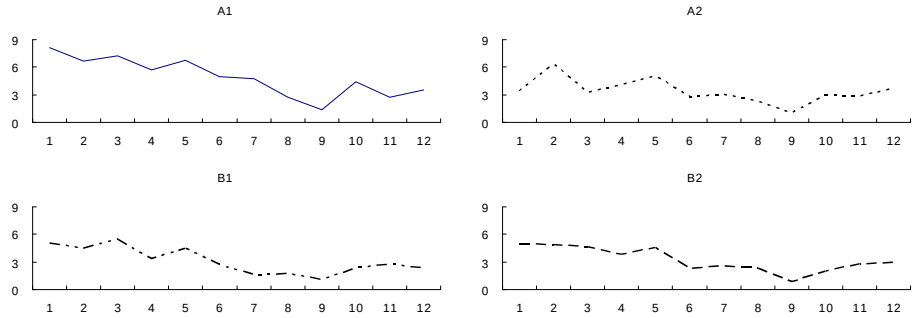


Figure 17. Cataloging Time for Each Resource by Each Cataloger (individually)

4.2.2 Score

H_{1b}: Using the propose interface increases scores.

The average scores of two interfaces are 47 and 54, respectively. Although only one participant's (A2) score significantly increased, the improvement may be implied by the increased scores of participant B1 and the average.

	Current Interface		Proposed Interface		*p*-value
	Mean	S.D.	Mean	S.D.	
A1	49.65	18.37	46.43	17.39	.600
A2	49.24	9.47	85.39	12.55	.028
B1	38.39	19.68	47.86	29.07	.116
B2	48.95	13.89	38.18	16.95	.753
AVG	46.56	15.56	54.47	26.32	.463

Table 4. Statistics of Scores

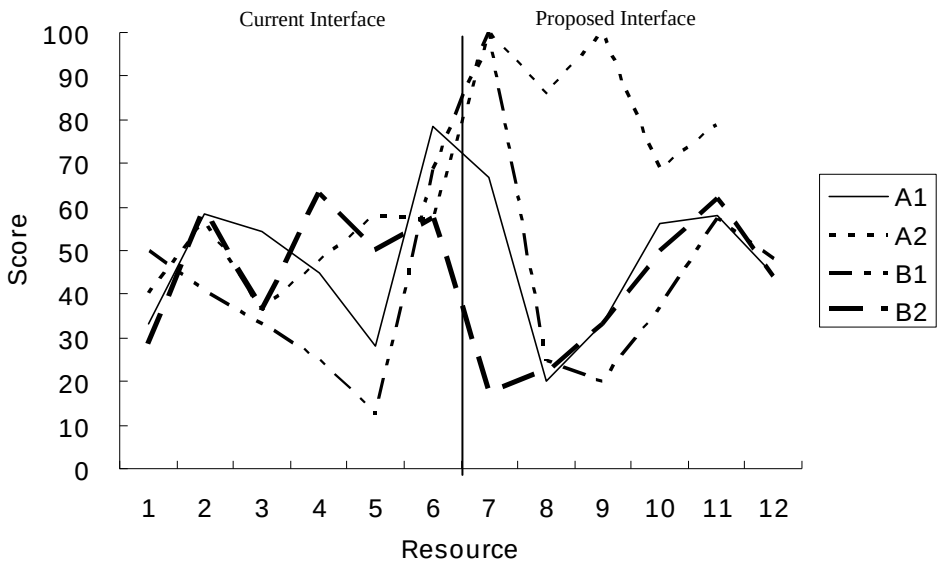


Figure 18. Scores of Each Resource Received by Each Cataloger (compared)

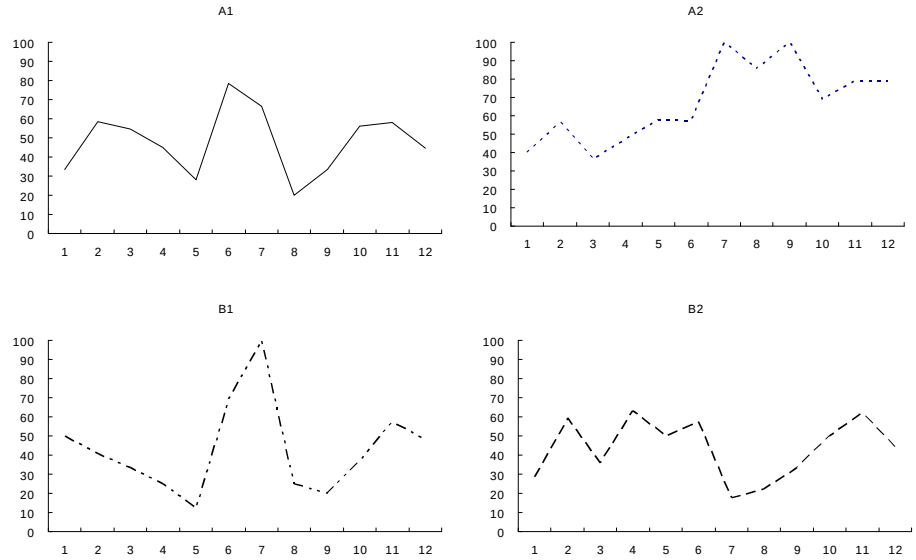


Figure 19. Scores of Each Resource Received by Each Cataloger (individually)

During the reexamination of the results and data analysis period, the limited improvement was found to result partly from the incomplete suggested Health Topics. First, the proposed systems suggests potentially relevant Health Topics identified by calculating the occurrences of Health Topic variations detected in the Web pages. This detection task was executed manually in the study due to the insufficient time to develop

and implement a sophisticated algorithm. A few variations were missed by accident. Second, some missing terms are actually implicit concepts such as “Heart Disease in Women” of Resource 10 (R10) and “Financial Assistance” of Resource 11 (R11). While heart disease was identified as important keyword of R10 due to its high occurrence, there was no obvious hint that it addresses women’s heart disease except phrases like “all types of circulation problems” and “all types of chest surgery”. “Financial Assistance” of R11 was in fact derived from “If you have no insurance, Medicaid, or Medicare, you may apply for our sliding fee discount for services.”. These kinds of semantic differences thus affect the results.

To eliminate influences of human errors and semantic differences and have a closer examination into if KWIC and highlighting are helpful to relevant judgment, the result was analyzed again after the inconsistencies were removed and presented below (Table 5; Figure 20-21). Although the overall average scores were still not statistically significant different, the improvement is more evident from a reasonable perspective.

	Current Interface		Proposed Interface		*p*-value
	Mean	S.D.	Mean	S.D.	
A1	49.65	18.37	54.65	21.17	.463
A2	49.24	9.47	98.61	3.40	.028
B1	38.39	19.68	54.40	28.99	.116
B2	48.95	13.89	45.89	22.86	.345
AVG	46.56	15.56	63.38	29.00	.116

Table 5. Statistics of Revised Scores

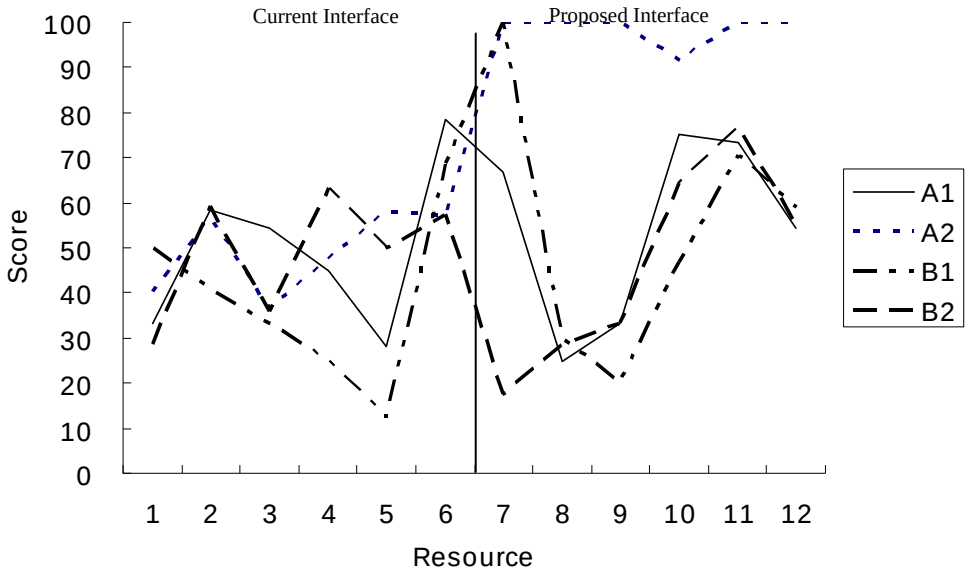


Figure 20. Revised Scores of Each Resource Received by Each Cataloger (compared)

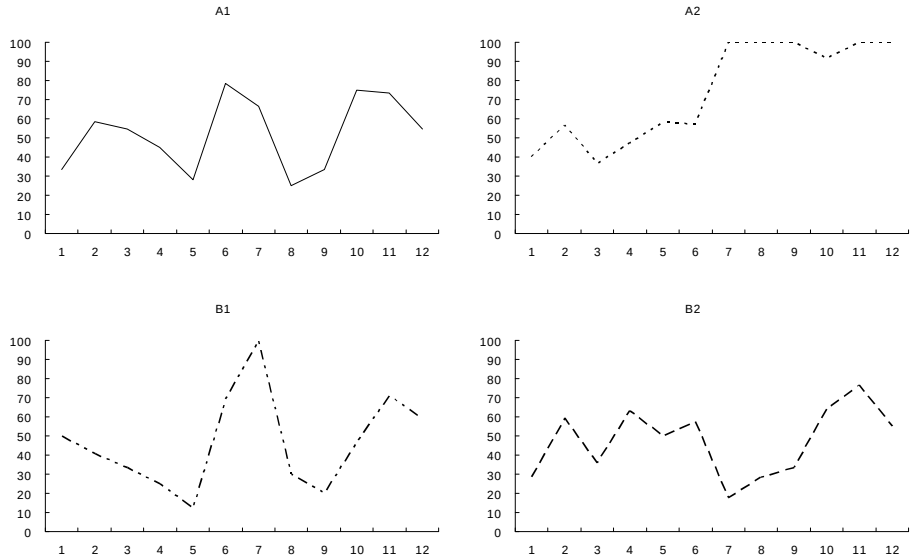


Figure 21. Revised Scores of Each Resource Received by Each Cataloger (individually)

4.3 Perceptions

H_{1c}: Proposed interface is easier to learn.
H_{1d}: Proposed interface is easier to use.
H_{1e}: Proposed interface is more useful.

Three perceptions toward two systems measured on a 5-point Likert scale are examined together here as 0 being poor and 4 being excellent. In general, the ratings for the proposed interface were more positive than the ratings gathered for the current one,

although the difference between them was not statistically significant. All users highly rated the proposed system, participant B1 in particular. It is noticeable that participant B2 gave the same rate to two interfaces regarding the ease of learning. At the end of post-questionnaire, participants were asked about if they considered that the proposed system would improve the cataloging work and help maintain metadata quality at NC Health Info. All of them agreed with it except one participant was neutral to the quality maintenance part.

	Current Interface		Proposed Interface		p-value
	Mean	S.D.	Mean	S.D.	
PEOL	2.50	0.58	3.00	0.82	.157
PEOU	2.25	0.96	3.25	0.50	.157
PU	2.50	1.00	3.25	0.50	.317

Table 6. Statistics of Perceptions

	PEOL		PEOU		PU	
	Current Interface	Proposed Interface	Current Interface	Proposed Interface	Current Interface	Proposed Interface
A1	3	3	3	3	3	3
A2	2	3	1	3	3	3
B1	3	4	2	4	1	4
B2	2	2	3	3	3	3
AVG	2.50	3.00	2.25	3.25	2.50	3.25

Table 7. Perceptions of Each Cataloger

5 Discussions, Limitations, and Recommendations

In this section, the experiment results presented earlier are further discussed from a larger perspective of the field of library and information science and metadata generation tool development. Limitations of the study are noted and suggestions to future research are proposed.

5.1 Discussions

5.1.1 Performances

From the experiment results, it was found that the cataloging time was significantly reduced with the proposed visual enhancement. The results also indicated that the system has the potential to improve metadata quality, although the increase of measured scores did not achieve a significance level of 0.05. The absence of significance

in metadata quality improvement may be explained by several reasons. First, catalogers are very familiar with the current system but it was their first time using the proposed system. Trust and job-related outcome expectation have been seen as important constructs of users' acceptance and usage behavior of a technology (Venkatesh, et al., 2003; Cody-Allen & Kishore, 2006; Lee & Lei, 2007). The uncertainty about the new system and Health Topics it suggested, and the potential future consequence may all have an impact on their performances. To reduce the distrust and anxiety about the possible outcome brought by the new system, a follow-up study should be conducted after users become familiar with it, and fully understand how it works and what would probably happen to their work if the visual enhancement is really implemented. Second, catalogers at NC Health Info have, in the past, sought a consensus on topic metadata via iterative annotations between each other (Blake, et al, 2005). The experiment forced them to make the decision at once on their own, which may influence their decision making. Third, topic pairings which are common and established combinations of Local Terms and Health Topics developed by catalogers are frequently used in NC Health Info to facilitate the cataloging process. The current study only focused on the assignment of Health Topics due to a pressing need to reduce the time required for generation and maintenance of this type of metadata. Although the lack of Local Terms was one of reasons for the lack of insignificance in quality improvement, it noted the importance to employ the associations of Local Terms and Health Topics in the next step. Last, the suggested Health Topics were not complete due to accidental human errors and difficulties handling semantic issues, such as a general topic "Pain" versus a specific topic "Back Pain", and implicit concepts unidentifiable with the current approach. In addition, users were limited to select only from Health Topics suggested when using the proposed system. The above mentioned conditions impeded them from some relevant Health Topics which were not recognized by the system. The identification and highlighting of potentially relevant Health Topics were executed manually in this study due to time limitation and in order to determine a proof-of-concept. A sophisticated automatic algorithm or system should be introduced and work as the back-end support of the proposed interface. Furthermore, users should be allowed to assign Health Topics which are not suggested by the system and to recommend new pairings of related Local Terms and Health Topics. Several

functions of the current system such as topic definitions look-up and Health Topics in the same group browse should also be integrated into the proposed system for their reference.

Although the quality of Health Topic assignment was not enhanced significantly in the study, it seems that there is a potential to achieve the goal once the noted limitations are addressed. Moreover, important facts are found in the results. During the experiment period, it was proved that the important concepts relevant to the collection usually gather on several and certain types of pages, such as services and programs, and via presenting only the important pages and highlighting all the potentially relevant Health Topics could be very helpful to improve the assignment performance quantitatively. Considering the cataloging task could be iterative and tedious, a second saved is a second earned. This time saving could eventually contribute to a cost savings and better service. Although some common Health Topics such as “Medicine” and “Pain” are often suggested and may contribute “noise” to relevance judgments when only occurrence frequencies were considered in the study, participants responded that these kinds of common topics only required a glance in order to make the decision with the visual enhancement. They further indicated that important concepts would not be missed even if related Health Topics occur only once or twice because all possibly relevant Health Topics are suggested. Although human catalogers tend to make false negative or false positive errors, they are able to identify implicit concepts related to the collection, such as the examples of “Heart Disease in Women” and “Financial Assistance” in the last section. Human errors may be alleviated via training, and it would be more cost effective and quality assured than to develop and employ a fully automatic system.

5.1.2 Perceptions

The proposed visual enhancement was also reviewed from user’s point of view. The results demonstrated notable changes with three perception measurements PEOL, PEOU, and PU, although they did not reach the significant level. Again, the lack of significance in perception improvement may be attributed to the fact that catalogers are very familiar with the current system and they have no difficulty manipulating it now, while the proposed system is totally new to them. Nevertheless, the average perceptions of three measurements toward the proposed system were rated higher, and all participants

were very positive about it. From their responses in the post-test questionnaire and the quotations in the table below (Table 8), it is confirmed that the visual enhancement proposed in the study is helpful to reduce catalogers' cognitive and work load when performing Health Topic assignment and the implementation of the proposed interface is welcome.

"It is really exciting!"
"It was very helpful to see how they (the terms) were used (in context)."
"They (the suggested terms) are already there for me."
"I don't have to think hard or spend too much time."
"There's no need to go back and forth."
"It is the least effort."

Table 8. Quotations from Open-Ended Discussions

5.2 Limitations

A number of limitations stemming from practical research constraints have been mentioned above, such as uncertainty and anxiety of the possible outcome resulted from the proposed system, the difference between the experiment and the real world, the lack of pairings with Local Terms and a supportive automatic algorithm/system, and the missing useful functions originally provided by the current system. Additional limitations noted here are as follows: Firstly, only four catalogers participated in the study due to knowledge and experience requirement, and only twelve resources selected from the same level of difficulty assigned with no more than twenty Health Topics were cataloged. More catalogers should be involved and resources at different levels of difficulty should be tested. A longitudinal study should be able to more realistically reflect the effects of the proposed visual enhancement for metadata generation. Secondly, although catalogers were not required explicitly to participate and they all gave positive responses to the study and the system, the work environment setting is mandatory itself. Willingness to volunteer is also one of major constructs to users' intention and usage behavior of a technology (Sharp, 2007). Thirdly, the questions used to measure users perceptions were mostly adopted from the previous NC Health Info usability study (Ellington, 2004). Although these questions were simple and straightforward, they may not have allowed the research to fully study users' perceptions toward the two systems, given that they

were developed for a previous study. To closely approach users' perceptions, multi-dimensional questions should be developed and tested instead. The additional limitations noted here also stem from practical research constraints and could be addressed in future research.

5.3 Recommendations

This study proposes a series of recommendations based on the research, including the experiment results. The recommendations consider the visual enhancement for metadata generation tools and automatic algorithms/systems to be implemented as the back-end support are proposed. First, catalogers' identification and relevance judgment of Health Topics should be further studied to improve the automatic algorithms. The occurrence frequency of Health Topic alone is not sufficient to extract the implicit concepts specifically needed by the collection. Secondly, the effect of presenting the distribution of occurrence frequencies remains unknown since some participants like it and some neglected it. There is a need to clarify what information would assist in relevance judgment and how it should be presented in addition to the application of KWIC and highlighting. Third, even general Health Topics only take users a few seconds to decide whether they are worth further examination, it may be helpful to give different weights according to comparative generalization of Health Topics. For example, "Back Pain" should receive higher weight than "Pain". Lastly, special care should be taken to implement the automatic algorithm/system to support the proposed interface, especially while calculating occurence frequencies of Health Topics and executing the highlighting. For instance, if "Back Pain" is detected and counted, the last half of it which is "Pain" should not be identified as the general term "Pain". Also, be careful not to break the original HTML structure when making highlighting effects. Only contents within <body> and </body> and also outside of HTML tags should be processed. It is hoped that some of these suggestions could be pursued in the future.

6 Conclusions

In conclusion, the proposed visual enhancement via KWIC and highlighting combining human intelligence and automated technology was able to reduce the

assignment time for Health Topics in NC Health Info, The experiment results suggested a potential for metadata quality improvement. Users' perceptions toward the new system were positive, indicating that the implementation of the semi-automatic approach is welcome. Future studies should be conducted to address the noted limitations and move this research further, in order to improve metadata generation within NC Health Info and other Go Local initiatives.

Acknowledgements

The author would like to thank Dr. Jane Greenberg, and NC Health Info director, Christie Silbajoris for their unceasing encouragement and constructive suggestions; and Beth Ellington, Linda Johnsen, and Megan Rudolph for their participation, time, and valuable feedback.

Bibliography

Anderson, J.D., & Perez-Carballo, J. (2001). The nature of indexing: how humans and machines analyze messages and texts for retrieval. part I: research, and the nature of human indexing. *Information Processing and Management, 37*(2), 231-254.

Ardö, A., & Koch, T. (1999). Automatic classification applied to the full-text Internet documents in a robot-generated subject index. *Online Information 99: 23rd International Online Information Meeting Proceedings*, 239-246.

Attardi, G., Gulli, A., & Sebastiani, F. (1999). Automatic Web page categorization by link and context analysis. *Proceedings of THAI'99, European Symposium on Telematics, Hypermedia and Artificial Intelligence*, 105-119.

Baldonado, M.Q.W., & Winograd, T. (1998). Hi-Cites: dynamically created citations with active highlighting. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 408-415.

Banerjee, K. (1997). Describing remote electronic documents in the online catalog: current issues. *Cataloging & Classification Quarterly, 25*(1), 5-20.

Blake, C., et al. (2005). Cataloging on-line health information: a content analysis of the NC Health Info portal. *Proceedings of the Annual American Medical Informatics Association Conference (AMIA 2005)*, 56-60.

Brown, T.J. (1991). Visual display highlighting and information extraction. *Proceedings of the Human Factors Soceity 35th Annual Meeting*, 1427-1431.

Byrd, D. (1999). A scrollbar-based visualization for document navigation. *Proceedings of the Fourth ACM Conference on Digital Libraries*, 122-129.

Calvo, R.A., Lee, J.M., & Li, X. (2004). Managing content with automatic document classification. *Journal of Digital Information, 5*(2). Retrieved November 2, 2007, from http://jodi.tamu.edu/Articles/v05/i02/Calvo/

Chi, E.H., et al. (2004). eBooks with indexes that reorganize conceptually. *CHI Extended Abstracts 2004*, 1223-1226.

Cho, W.C., & Richards, D. (2004). BayesTH-MCRDR algorithm for automatic classification of Web document. *Lecture Notes in Computer Science, 3339*, 344-356.

Ciravegna, F., & Wilks, Y. (2003). Designing adaptive information extraction for the Semantic Web in Amilcare. In S. Handschuh & S. Staab (Eds.), *Annotation for the Semantic Web*. IOS Press, Amsterdam, The Netherland

Cody-Allen, E., & Kishore, R. (2006). An extension of the UTAUT model with e-quality, trust, and satisfaction constructs. *Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research*, 82-89.

Ellington, B. (2004). *An evaluation of cataloging input interface tools for NC Health Info System*. Unpublished manuscript.

Ellis, D., & Vasconcelos, A. (1999). Ranganathan and the Net：using facet analysis to search and organize the World Wide Web. *Aslib proceedings, 51*(1), 3-10.

Erdmann, M., et al. (2000). From manual to semi-automatic semantic annotation: about ontology-based text annotation tools. *Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, 79-86.

Feldman, R., Dagan, I. &, Hirsh, H. (1998). Mining text using keyword distributions. *Journal of Intelligent Information Systems, 10*, 281-300.

Flannert, M.R. (1995). Cataloging Internet resources. *Bulletin of the Medical Library Association, 83*(2), 211-215.

Fox, S. (2005). Health Information online. *Pew Internet & American Life Project*. Retrieved April 3, 2008, from http://www.pewinternet.org/pdfs/PIP_Healthtopics_May05.pdf

Fox, S. (2006). Online health search 2006. *Pew Internet & American Life Project*. Retrieved April 3, 2008, from http://www.pewinternet.org/pdfs/pip_online_health_2006.pdf

Fox, S., & Fallows, D. (2003). Internet health resources. *Pew Internet & American Life Project*. Retrieved April 3, 2008, from http://www.pewinternet.org/pdfs/PIP_Health_Report_July_2003.pdf

Fox, S., & Rainie, L. (2000). The online health care revolution. *Pew Internet & American Life Project*. Retrieved April 3, 2008, from http://www.pewinternet.org/pdfs/PIP_Health_Report.pdf

Garofalakis, M.N., et al. (1999). Data mining and the Web: past, present and future. *Proceedings of the 2nd International Workshop on Web Information and Data Management*, 43-47.

Gietz, P. (2001). *Report on automatic classification systems*. Retrieved November 2, 2007, from http://www.daasi.de/reports/Report-automatic-classification.html

Golub, K., & Ardö, A. (2005). Importance of HTML structural elements and metadata in automated subject classification. *Proceedings of ECDL 2005 - the 9th European Conference on Research and Advanced Technology for Digital Libraries*, 368-378.

Greenberg, J. (2003). Metadata and the World Wide Web. In M.S. Drake (Ed.), *Encyclopedia of Library and Information Science* (2nd ed., pp.1876-1888). New York: Marcel Dekker, Inc.

Greenberg, J. (2004). Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. *Journal of Internet cataloging, 6*(4), 59-82.

Greenberg, J., et al. (2002). *Iterative design of metadata creation tools for resource authors*. Retrieved November 3, 2007, from http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf

Greenberg, J., Spurgin, K., & Crystal, A. (2006). Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. *International Journal of Metadata, Semantic and Ontologies, 1*(1), 3-20.

Harper, D.J., Coulthard, S., & Sun, Y. (2002). A language modeling approach to relevance profiling for document browsing. *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, 76-83.

Hearst, M.A. (1995). TileBars: visualization of term distribution information in full text information access. *Proceedings of ACM Human Factors in Computing Systems Conference (CHI '95)*, 59-66.

Hilligoss, B., & Silbajoris, C. (2004). MedlinePlus Goes Local in NC: the development and implementation of NC Health Info. *Journal of Consumer Health on the Internet, 8*(4), 2004.

Hirokawa, S., Itoh, E., & Miyahara, T. (2003). Semi-automatic construction of metadata from a series of web documents. *Proceedings of Australian conference on artificial intelligence, AI-03, Lecture Notes in Computer Science, 2903*, 942-953.

Hussam, A., et al. (1998). Semantic highlighting. *CHI 98: Conference Summary on Human Factors in Computing Systems*, 191-192.

Irvin, K.M. (2003). *Comparing information retrieval effectiveness of different metadata generation methods*. Unpublished master's thesis, University of North Carolina-Chapel Hill.

Jenkins, C., et al. (1998). Automatic classification of Web resources using Java and Dewey Decimal Classification. *Computer Networks & ISDN Systems, 30*, 646-648.

Käki, M. (2006). fKWIC: frequency-based keyword in context index for filtering web search results. *Journal of the American Society for Information Science and Technology, 57*(12), 1606-1615.

Koch, T., & Vizine-Goetz, D. (1999). Automatic classification and content navigation support for Web services: DESIRE II cooperates with OCLC. *OCLC*. Retrieved October 29, 2007, from http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003489

Koehler, W.C. (1999). Classifying Web sites and Web pages: the use of metrics and URL characteristics as markers. *Journal of Librarianship and Information Science, 31*(1), 21-31.

Kris, C. (2007.) *A dynamic learning object life cycle and its implications for automatic metadata generation*. Unpublished doctoral dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.

Lawrence, S., & Giles, C. (1999). Accessibility of information on the Web. *Nature, 400*, 107-109.

Lee, C.B.P., & Lei, U.L.E. (2007). Adoption of e-government services in Macao. *Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance*, 217-220.

Liddy, E.D., et al. (2002). *Automatic metadata generation & evaluation*. Retrieved November 5, 2007, from http://web.syr.edu/~diekemar/ruimte/Papers/p401liddy.pdf

Lin, S.H., et al. (1998). Extracting classification knowledge of Internet documents with mining term associations: a semantic approach. *Research and Development in Information Retrieval, SIGIR*, 241-249.

Litvak, M., Last, M., & Kisilevich, S. (2007). Classification of Web documents using concept extraction from ontologies. *Lecture Notes in Computer Science, 4476*, 287-292.

Loia, V., & Luongo, P. (2001). An evolutionary approach to automatic Web page categorization and updating. *Lecture Notes in Artificial Intelligence, International Conference on Web Intelligence(WI)*, 477-478.

Madria, S.K., et al. (1999). Research issues in Web data mining. *Proceeding of Data Warehousing and Knowledge Discovery, First International Conference, DaWaK1999*, 303-312.

Marchionini, G. (1995). *Information Seeking in Electronic Environment*. Cambridge: Cambridge University Press.

Marchiori, M. (1998). The limits of Web metadata, and beyond. *Computer Networks and ISDN Systems, 30*, 1-9.

Mase, H. (1998). *Experiments on automatic web page categorization for IR system*. Technical report, Stanford University, California. Retrieved October 15, 2007, from http://citeseer.ist.psu.edu/cache/papers/cs/1719/http:zSzzSzwww-db.stanford.eduzSzpubzSzpaperszSzhisao.pdf/mase98experiments.pdf

Meire, M., Ochoa, X., & Duval, E. (2007). SAmgI: Automatic Metadata Generation v2.0. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, 1195-1204).

Muller, et al. (1999). Automatic classification of the World-Wide Web using the Universal Decimal Classification. *Proceedings of the 23rd International Online Information Meeting (Online99)*, 231-237.

NC Health Info. (2003). NC Health Info documentation: NC Health Info cataloging training manual. Unpublished manuscript.

Nielsen, J. (2005). Ten usability heuristics. *useit.com*. Retrieved March 3, 2008, from http://www.useit.com/papers/heuristic/heuristic_list.html

Noh, S., et al. (2003). Classifying Web pages using adaptive ontology. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2144-2149.

Pandia. (2007, February 25). The size of the World Wide Web. *Pandia Search Engine News*. Retrieved March 3, 2008, from http://www.pandia.com/sew/383-web-size.html.

Pew Internet & American Life Project. (2008). Retrieved March 15, 2008, from http://www.pewinternet.org/PPF/c/5/topics.asp

Sedano, J.M. (1964). *Keyword-in-context (KWIC) indexing: background, statistical evaluation, pros and cons, applications*. Unpublished doctoral thesis, Pittsburgh University. Retrieved March 28, 2008, from http://stinet.dtic.mil/cgibin/GetTRDoc?AD=AD443912&Location=U2&doc=GetTRDoc.pdf

Sharp, J.H. (2007). Development, extension, and application: a review of the technology acceptance model. *Information Systems Education Journal, 5*(9), 1-11.

Song, M.H. (2006). Ontology-based automatic classification of Web documents. *Lecture Notes in Computer Science, 4114*, 690-700.

Spink, A., et al. (2001). Searching the Web: the public and their queries. *Journal of the American Society for Information Science, 53*(2), 226-234.

Sukakanya, U., & Porkaew, K. (2005). A framework for automatic classification of e-business Web content. *International Journal of the Computer, the Internet and Management (IJCIM), 13*(Suppl. 3), 13.1-13.8.

The GNOME Usability Project. (2004). *GNOME Human Interface Guidelines 2.0.* Retrieved February 1, 2008, from http://library.gnome.org/devel/hig-book/stable/

UKOLN. (N.A.). *DC-dot Dublin Core metadata editor*. Retrieved March 3, 2008, from http://www.ukoln.ac.uk/cgi-bin/dcdot.pl

Uren, V., et al. (2005). Browsing for information by highlighting automatically generated annotations: a user study and evaluation. *Proceedings of the 3rd International Conference on Knowledge Capture (K-CAP '05)*, 75-82.

Vellucci, S.L. (1997). Options for organizing electronic resources: the coexistence of Metadata. *Bulletin of the American Society for Information Science, 24*(1), 14-17.

Venkatesh, et al. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Vinyard, P.E. (2001). *An analysis of embedded metadata usage on the World Wide Web*. Unpublished master's thesis, University of North Carolina-Chapel Hill.

Weihs, J. (1996). Solving the Internet cataloging nightmare. *Technicalities, 16*(4), 4-6.

Yi, K. (2007). A pilot study of enhancing subject discovery of textual Web resources. *Proceedings of the 35th Conference of the Canadian Association for Information Science (CAIS)*. Retrieved October 25, 2007, from http://www.cais-acsi.ca/proceedings/2007/yi_2007.pdf

Yu, H., & Han, J. (2002). PEBL: positive example based learning for Web page classification using SVM. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery in Databases (KDD02)*, 239-248.

Appendices

Appendix A. Principles and Key Points of Contextual Design Employed in the Study

Principle	Key points
Context	"Go where the work is to get the best data." "Avoid summary data by watching the work unfold." "Avoid abstractions by returning to real artifacts and event."
Partnership	"Keep the customer concrete by exploring ongoing work." "Help customers articulate their work experience." "Alternate between watching and probing." "Find the work issues behind design ideas." "Let the customers shape your understanding of the work." "It's a goal to be nosy." "Partnership creates a sense of shared quest."
Interpretation	"Determine what customer words and actions mean together." "Design ideas are the end product of a chain of reasoning." "Design is built upon interpretation of facts – so the interpretation had better be right." "Sharing interpretations with customers won't bias the data." "Sharing interpretations teaches customers to see structure in the work." "Customers fine-tune interpretations."
Focus	"Clear focus steers the conversation." "Focus reveals detail." "Commit to challenging you assumptions, not validating them."

(Beyer & Holtzblatt, 1998)

Appendix B. Pre-test Questionnaire

Please read all the choices and circle only one that answers the question.

1. What is your job?

Health Science Library librarian at UNC-CH, for _____ year(s)
Graduate student at SILS, UNC-CH

2. What is your highest level of education completed?

Bachelors degree Masters degree Doctoral degree Other
In _____ major (if applicable).

3. Have you had any previous cataloging experience? Yes No
(Stop here if no to this question.)

4. Please describe your level of cataloging experience on a scale of 0 to 4, with 0 being "New" and 4 being "Very experienced".

0 1 2 3 4

5. Have you had any previous web-based cataloging experience? Yes No
(Stop here if no to this question.)

6. Please describe your level of web-based cataloging experience on a scale of 0 to 4, with 0 being "New" and 4 being "Very experienced".

0 1 2 3 4

7. Have you previously cataloged resources for NC Health Info? Yes No
(Stop here if no to this question.)

8. How long have you worked on NC Health Info?

Less than one month
1-6 months
6 month to 1 year
1-2 years
2-3 years
Over 3 years

9. On a scale of 0-4 with 0 being poor, 2 being neutral and 4 being excellent, please rate your impression about the following elements of the current NC Health Info cataloging system when assigning Health Topics:

	Poor				Excellent
Ease of learning	0	1	2	3	4
Ease of use	0	1	2	3	4
Usefulness	0	1	2	3	4

Appendix C. Post-test Questionnaire

Please read all the choices and circle only one that answers the question.

1. On a scale of 0-4 with 0 being poor, 2 being neutral and 4 being excellent, please rate your impression about the following elements of the *proposed* NC Health Info cataloging system when assigning Health Topics:

	Poor				Excellent
Ease of learning	0	1	2	3	4
Ease of use	0	1	2	3	4
Usefulness	0	1	2	3	4

2. Do you think the proposed interface will improve the cataloging work of NC Health Info?

Strongly disagree	0	1	2	3	4	Strongly agree

3. Do you think the proposed interface will help maintain the metadata quality of NC Health Info?

Strongly disagree	0	1	2	3	4	Strongly agree

Thank you for your participation!