Review Article　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Designing a Subject-based Information Retrieval system in Digital Libraries

Biglu MH[1] and Ghavami M[2]*

[1]Department of Basic Sciences, Tabriz University of Medical Science, Tabriz, Iran
[2]Radiology Department, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract

Objective: Information visualization is the study of interactive depictions of abstract and data to strengthen the human cognition. Designing an appropriate information visualization system may be very useful technique for scholars, who intent to get scientific information in digital libraries. The objective of current study was to map and visualize the key-information of dissertations in academic libraries. To achieve the aim an information retrieval system was designed to present the interactive graphic view of dissertations' subjects in academic.

Methods: We designed an information retrieval system by information visualization toolkits that presented the related subjects of dissertations in academic libraries. In addition, the satisfaction-levels of library-users was analyzed by administrating a standard questionnaire (QUIS Questionnaire).

Results: The study indicated that the designed IR system helped to provide a user-friendly environment through displaying subjective relations of dissertations, overwhelming variety of colors in displaying information. Fast and easy access to the cover-to-cover information of dissertations and user-interaction facilities were the advantages of designed IR. Analysis of data furthermore indicated, the user satisfaction from the system was from medium to high grade.

Conclusion: Designing the IR-system revealed an excessive influence on users' satisfaction; therefore, proposing such systems for employing in academic libraries is very suitable and its implementation is necessary.

Keywords: Information visualization; Information Retrieval system; Digital libraries

Introduction

Information visualization is the study of interactive depictions of abstract data to strengthen the human cognition. Designing an appropriate information visualization system can be very useful technique for scholars, who intent to get scientific information in digital libraries. Many users in digital libraries usually begin their searches by an unclear vision of what they want to accomplish. If the information retrieval systems can clarify the users' information-needs, and retrieve the proper information, we can announce that a correct search has been performed.

To achieve this purpose, an information retrieval system should provide the translation of user's keywords into controlled vocabulary terms, based on the subject headings of related thesaurus. There is no doubt that choosing the accurate and consistent keywords would provide the possibilities to enhance the content representation and retrieval of related literature.

Using the subject heading of controlled vocabulary is the most important step of retrieving the desired information. It makes to reduce the amount of irrelevant results, that come up; therefore, selection the appropriate keywords play a major role in information retrieving processes.

Other things that may affect the users' satisfaction in the use of information retrieval systems are the way of information-displaying, and the simplicity of use of the system.

The goal of this current paper is to depict the graphically relations between dissertations in order to provide effective subject-based, information retrieval within digital libraries. Keeping in mind that physical access to the underlying digital collection is not an option, efficient subject-based information retrieval is crucial for a successful digital library [1].

A proof of concept is also provided in this study to demonstrate the applicability of the proposed research to current and yet-to-come visual, interactive, and subject-based, information retrieval systems within digital libraries [1].

The Role of Information Visualization in Retrieval Process

Since the last two decades, several information specialists have paid attention to Human Computer Interaction (HCI) and the user interface design of information retrieval systems, as well as the digital libraries [2].

The study of Daniel Hienert et al. showed that graphics and interaction techniques can be used intuitively and the questions can be answered very fast [3]. One important technique for designing such user-friendly interfaces is information visualization [4].

Information visualization utilizes computer graphics and interaction to assist humans in solving problems [5].

"Information retrieval visualization" refers to the information resources, the user questions, information retrieval model, the retrieval process and the retrieval results of invisible interior semantic relationship convert into graphics or images, and displayed in a two-dimensional, three-dimensional or multidimensional visual space,

these intuitive image of spatial information to help users understand the search results, get the retrieval direction to guide the retrieval process, in order to progress the efficiency and performance of information retrieval [6].

Retrieval process visualization can provide information feedback, allow the retrieval process reversible, support retrieval strategy control, and a user's interaction [6].

Information Retrieval Visualization Within Digital Libraries

Information visualization is a set of technologies that use visual computing to strengthen the human cognition with abstract information. However, the major aim of information visualization is to increase the cognitive presentation of information, it should help the human to vast the insights [7]. Examples of websites that have used information visualization technique in their IR systems are: Library of Congress [8]. The National Center for Biomedical Ontology (BioPortal) [9], Microsoft Academic Search [10], visual thesaurus [11], etc. Medical Subject Headings (MeSH) also visualized [12,13] and Has been used for different purposes.

Proposed Approach

In academic libraries, Thesis' sector is special importance, Quick retrieval possibility and awareness of subject relations between thesis and dissertations, is important for students, who want to choose a topic for their dissertation, as well as for researchers, who want to know the results of other researches on a particular topic.

In most Academic libraries, the only way to retrieve thesis information is searching manually or electronically by librarians; On the other hand, not only it doesn't use a particular Subject classification scheme for thesis indexing, but also It doesn't use of controlled vocabularies in their keywords selection, hence retrieving the exact subject of the dissertation is not beyond the realms of possibility. In this study, the information visualization principle was used to retrieve the visualized information of dissertations based on subject areas in digital libraries. We created subject relations between themes based on the subject headings.

In this current proposed system, users can find theses, keywords, supervisors as well as their interesting topics, visually. Users can manage the search-strategy simultaneously between keywords and subjects.

Case Study

The proposed approach has been applied to the digital library of theses and dissertations in the management and medical informatics Faculty of Tabriz University of medical science. This IR system provides information via displaying subject relations between various titles of thesis. To design this system, we used the perforce visualization tool, and it is appropriate using by web browsers.

In the section 6 (comparative evaluation) we made a comparative assessment between an ordinary used system (MAKTUB) in the Tabriz University of Medical Sciences and our proposed system. The evaluation showed the major advantage of our designed IR system clearly.

Comparative Evaluation

In the current IR system of digital library of Tabriz University of Medical Sciences (MAKTUB) (Figure 1) users are able to search in all the libraries of the university or do in a particular college-library.

By typing the keyword (evaluation) in the search box, Results are provided as a list of resources includes the undergraduate, masters and doctoral theses that the term "evaluation" presents in the part of their bibliographic information. User can find bibliographic details by choosing one of the titles.

Due to lack of subject headings and Controlled vocabulary in resource indexing, search results for two words ("evaluation" & "assessment") retrieves different results. And the search results are not subject-based.

In this system subject relations between these are not visible to users.[1]

In proposed visual IR system

In our visual information retrieval system, the disadvantages of the current system have been fixed.

The user chooses management and medical informatics Faculty at first (Figure 2). On the first page of software, topics and subjects of all theses and their subject-based correlations will be showed graphically. The information is presented in two separated languages, English and Persian. The user can choose his/her arbitrary language and continue searching (Figure 3).

[1] In the present application (MAKTUB), the dissertations information is only in Persian, and English language search is not possible, so the search operation has not been show by pictures.
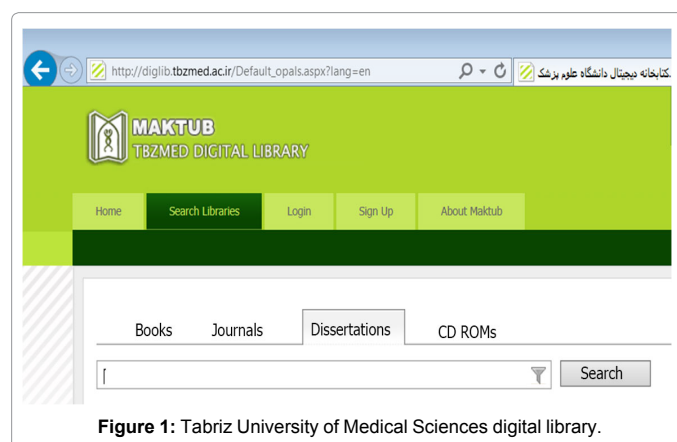

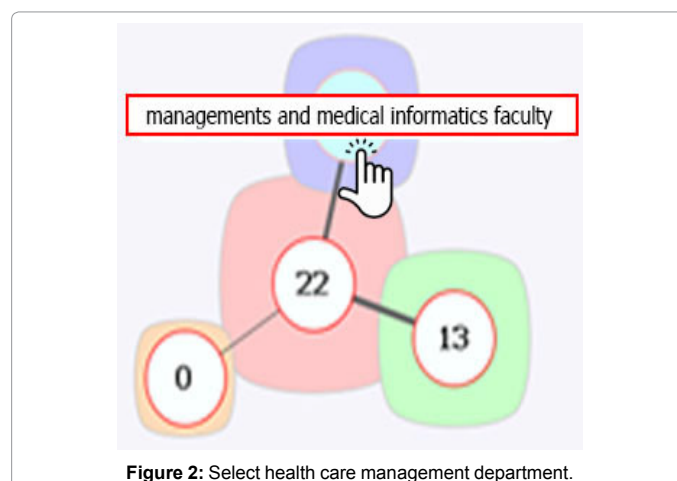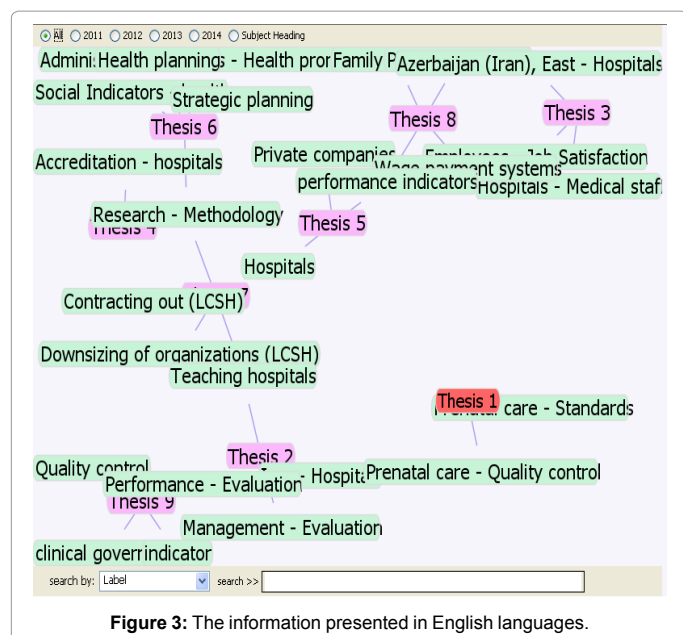**Figure 1:** Tabriz University of Medical Sciences digital library.

Figure 2: Select health care management department.

Figure 3: The information presented in English languages.

Figure 4: The system begins to search letter by letter and eliminates the unmatched results (In English).

As typing a word, the system begins to search letter by letter and eliminates the unmatched results and achieves the most relative results finally (Figure 4).

Discussion of Retrieval Information in Proposed Visual IR System

Issues devoted to thesis have controlled by using the subject headings.

The subject headings used to show the relationship between theses to assign the precise subject matter of the thesis.

Keywords, thesis and masters related to them are shown in the software's interface by moving the cursor on the topics (Figure 5).

The system has the ability to search keywords and subject headings assigned to thesis simultaneously, so that searching for two words with same meaning (e.g., evaluation–assessment) similar results retrieves. (Figures 6 and 7).

Providing graphical information, possibility of user interaction with the system and displaying the subject relations between the thesis

using a graph, has caused to easily diagnose the subject relations with the user, so the user can assemble information to meet all his/her needs about special subjects.
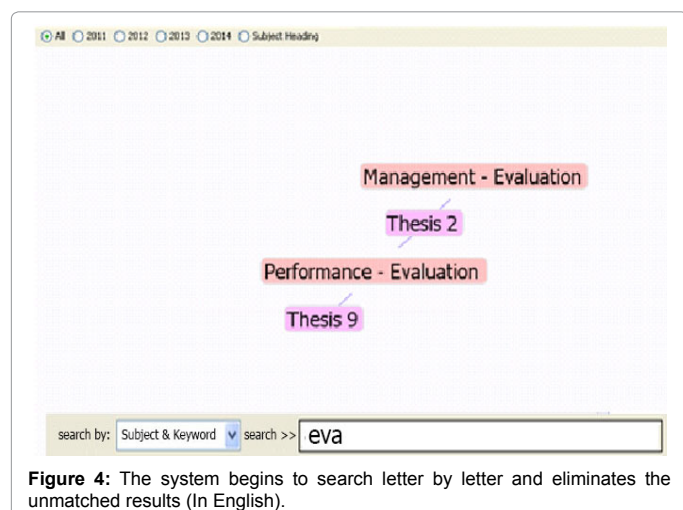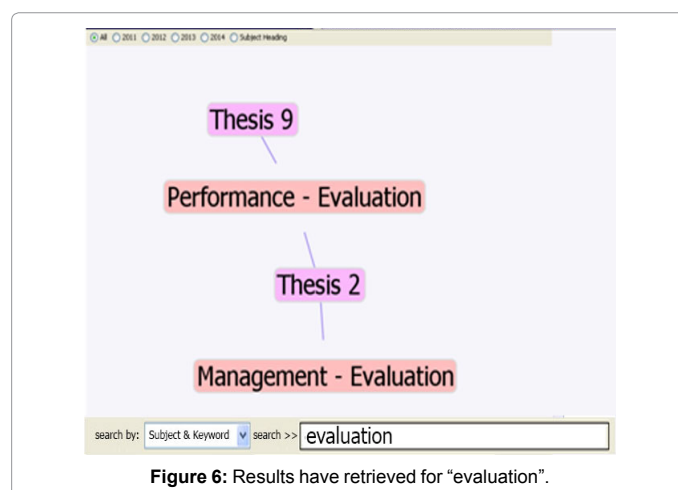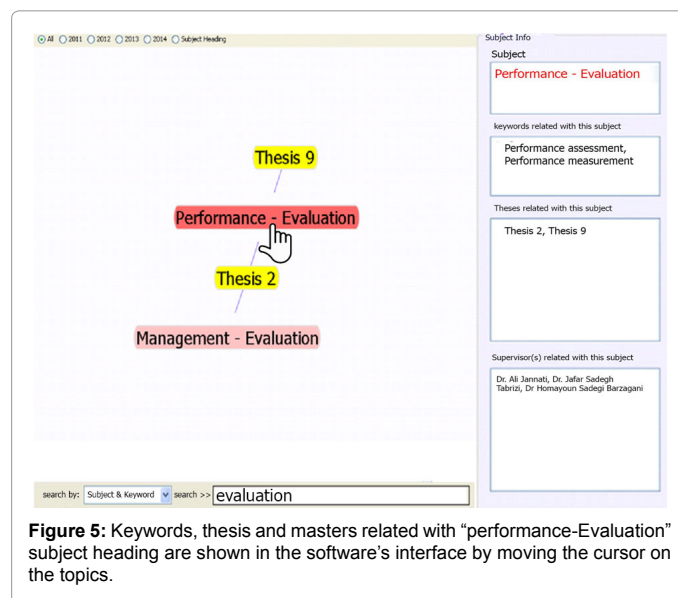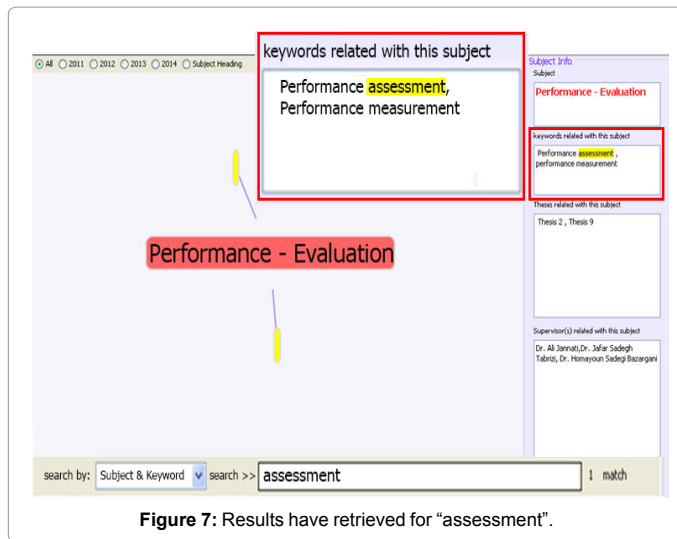
Conclusions, Future Work

In this article, an effort is to graphically exploit the relations between dissertations in order to provide more effective subject-based, information retrieval within digital libraries. Keeping in mind that physical access to the underlying digital collection is not an option, efficient subject-based information retrieval is crucial for the efficacy of digital library.

The proposed methodology could be applied to other information resources within digital libraries, thus providing a universal tool for user-centered, interactive, subject-based information retrieval.

During the process, users enhance their cognitive learning, since they are able to discover which subject headings correspond to their information needs.

Future work is oriented towards the exploration of other semantically rich information structures produced by librarians. The

Figure 5: Keywords, thesis and masters related with "performance-Evaluation" subject heading are shown in the software's interface by moving the cursor on the topics.

Figure 6: Results have retrieved for "evaluation".

Figure 7: Results have retrieved for "assessment".

ultimate goal of this research is to contribute to the development of next-generation digital libraries capable of exploiting the wealth of knowledge accumulated within traditional libraries [14].

References

1. Gerken J, Heilig M, Jetter HC, Rexhausen S, Demarmels M, et al. (2009) Lessons learned from the design and evaluation of visual information-seeking systems. Int J Digit Libr 10: 49-66.

2. Sastry HG, Reddy LC (2009) User Interface Design Principles for Digital Libraries. International Journal of Web Applications 1: 86-91.

3. Hienert D, Sawitzki F, Schaer P, Mayr P (2012) Integrating interactive visualizations in the search process of digital libraries and IR systems. Advances in Information Retrieval: Springer pp: 447-450.

4. Wan G (2006) Visualizations for digital libraries. Information technology and libraries 25: 88-94.

5. Ware C (2012) Information visualization: perception for design. Elsevier.

6. Liao Z, Gao M, Yan F (2012) Application Study of Information Visualization in Digital Library. National Conference on Information Technology and Computer Science (CITCS 2012). Atlantis Press Lunzhou, China pp: 161-164.

7. Card SK, Mackinlay JD, Shneiderman B (1999) Readings in information visualization: using vision to think. Morgan Kaufmann.

8. (2013) Library of Congress Linked Data Service: Authorities and Vocabularies.

9. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, et al. (2011) BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Research 39: W541-W545.

10. He J, Dai X, Zhao X (2007) PLAN: a web platform for automating high-throughput BLAST searches and for managing and mining results. BMC Bioinformatics 8: 53.

11. Visual thesaurus: Thinkmap Inc. 1998-2013.

12. Sheth N, Cai Q (2003) Visualizing MeSH dataset using radial tree layout.

13. Du H, Yoo T (2006) Indexing Similarity Visualization over the Medial Subject Headings (MeSH).

14. Biglu MH (2008) Scientometric study of patent literature in MEDLINE & SCI (Doctoral dissertation, Humboldt-Universität zu Berlin, Philosophische Fakultät I).