

สัญญาเลขที่ PDF/25/2540

รายงานวิจัยฉบับสมบูรณ์

โครงการ The Development of a Visualization Information System for Structuring and Retrieval of
Research Documents

อัมภาพร หุขประยูร
จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว.ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

This research is dedicated to the mentor, Prof. Dr. Korfhage, for his long support to Visualization in Information Retrieval Interface research.

บทคัดย่อ

Application of visualization to document collection was at an early stage. There were several proposed visualization tools for document collections. A visualization to information retrieval system, GUIDO (Graphical User Interface for Document Organization), was preliminary proven of its usability for users in a given set of information retrieval tasks. This research attempted to discover the usability of GUIDO on specific purpose in conducting document search for research development. A usability test on a set of representatives of researchers was set up. The result confirmed the same level of usability of the visualization tool. However, the learnability of the tool was rated lower than the usability of the tool.

เนื้อหาทางวิจัย

บทนำ

Researchers are overwhelmed by piles of documents. Traditionally, researchers store the documents related to their interest in hardcopy forms and retrieve them by recalling which documents contain the information they want. As a consequence, researchers will have a potential of memory overloading to remember the information related to their researches which result in time-consuming retrieval or inaccurate citation.

The researches can use the library as an alternative. Most library retrieval tools are implemented using bibliographic search, which includes searches by title, authors, and a limiting set of keyword indexes. However, researchers also used three other important, but somewhat more complicated, search strategies in addition to bibliographic search. These strategies include analytical search which is a search by a combination of keyword indexes, search by analogy which is a search by a model document, and browsing which allows the searcher to look around the already arranged document in the collection. Traditional library retrieval tools have limitation in serving these search strategies.

Scientific visualization has been documented as playing an important role in computational science and engineering, such as molecular modeling, medical imaging, environmental control, meteorology, gas and fluid dynamics. Focuses of these scientific visualizations are on a three-dimensional modeling of physical objects. Visualization can also be applied to the more abstract concepts of discrete values can be computed and mapped onto a pixel colors. Therefore, applications of visualization to a collection of documents are useful in that it allows the users to see the explicit similarity relationship among the documents in similar way to the library stack. The resulting visualization tools also enhance the searchers with the problematic search strategies, analytical search, search by analogy and especially browsing. The marriage of visualization in an information retrieval system yields an interface named visual information retrieval interface (VIRI).

This research has the following objectives.

1. To confirm the result of the preliminary study which concluded that certain similarity measures on vector space model result in significantly better retrieval performance than others.
2. To study the usability of the GUIDO interface.
3. To study the learnability of the GUIDO interface.

ระเบียบวิธีวิจัย

This research involve three steps: visualization tool on information retrieval system development, document collection preparation, and usability testing on subjects

1. Visualization tool refinement.

As the setting of environment is in Thailand, in order to test the usability of the GUIDO (Graphical User Interface for Document Organization), the tool must be developed to accept Thai characters as well as English. The retrieval engine was vector space model based on Euclidean similarity. The tool allowed users to input the indexed keywords for each document manually. Then, it constructs a vector as a representation for each document in the collection.

2. Document collection Preparation

Document Collections used in this research are excerpts from senior projects, which 4th year students in the Statistics undergraduate programs must complete. Normally, when completing their senior projects, students were asked to submit a full report. In order to use them as the document collection in this study, students were asked to add an executive summary of their reports. Then each senior project executive summary was read and indexed manually by two readers and agreed on a set of index of that summary. This manually indexing was chosen to guarantee the consistency of the index keywords in the developed system. The document collection in this research was student senior projects completed in the past three academic years: 1997 1998, and 1999. There were about 300 documents in the collection. A vector for each document was calculated based in the keywords input as indexes.

3. Usability and learnability testing on the visualization tool.

Usability and leanability testing on the developed GUIDO interface was done with the third year undergraduate students Statistics Program. These students were required to complete their senior projects individually in their fourth year. To start their projects, students headed towards finding their topics and literature reviews. Although there was a diverse resource for their literature reviews, one resource in common was the senior projects in the past years. These students could be representatives of Thai researchers in the searching on academic documents.

There were two issues to be tested on: confirmation of the GUIDO interface usability on targeted Thai subjects and learnability of the GUIDO interface – the ease with which new users

can begin effective interaction and achieve maximal performance. Seventy 3rd year students of academic 1999 were randomly selected as subjects and asked to participate in the experiment at the kick-off time of their senior projects.

The subjects were given an oral explanation on how to use the GUIDO interface, then they were asked to perform search for any topics of their interest for half an hour or until they were satisfied with the search. The search satisfaction can be either getting the document with respect to their intended topics or finding that there was no such document in the system relevant to their topics. Then, they were asked to answer questionnaires. There were 3 parts of questions in a questionnaire: background information including what experience a subject had with an information retrieval prior to the study, visualization tool satisfaction, and learnability of the tool. The questions in the middle part of the questionnaire were modified from the prior study and translated into Thai. The questions in the last part of the questionnaire adopted the concept from Kristoffersen (2000) resulting in 5 questions: predictability, synthesizability, consistency, generalizability, and familiarity.

Predictability is an evaluation on the user's ability to judge what system response is going to be as a response to the next user action. Synthesizability is a measure on the user's ability to understand which actions have lead to the current state, and what the system did to get there. Consistency is a measure on the similar functionality a system should offer from comparable situations. Generalizability is the state where existing knowledge can be successfully applied. Lastly, familiarity is the extent to which the user experiences a real world parallel to the system.

บทวิจารณ์

1. Visualization tool refinement.

The first attempt was to refine the user interface of the original program in order to make it easier to use. The attempt to upgrade the program in the early stage was hard. The student who worked on this as her senior project did not understand the concept of Information Storage and Retrieval or the concept of visualization. She spent a lot of time reading and discussing the concepts as well as studying the development tools before starting to code.

While the upgraded version was not ready and delayed, an experimental design was conducted on the original version for subjects drawn from 4th year students in the department. However, they had trouble with English content and the result was inconclusive.

The student graduated; the upgraded version (version 1.1) was still not ready for usability testing. The new replacement continued to catch up what the first student did but, again, the developer lacked the

understanding of concept of Information Storage and Retrieval and Visualization because this concept was not taught in the curriculum.

However, the developer managed to change the look of the user interface. At version 2.0, the visualization of document organization was tested with English content. However, the pilot subjects had failed the test due to the English user interface and document content. Changing to Thai collection required modifying the input module of the program. The executive summary of the senior projects completed in the past two academic years: 1997 and 1998 were input into the system as a test collection. A pilot test was done again on 20 fourth year students. The feedback from the pilot subjects were about too small set of document collection in the system. Half of the students complained that they could not find any relevant document, but the visualization part was interesting. Moreover, attempt to explain the subject the vector space model as opposed to Boolean model was hard, time-consuming, and not effective. However, the explanation on how to interpret the display was understandable. The first version of questionnaire on usability and learnability, as translated into Thai, was also tested in this pilot test. Feedbacks from the pilot test helped improve the language used in the questionnaire.

The revised GUIDO visualization tool was completed with a document collection with manual indexing, vector representation calculation, and the visualization for query and search result display part. A set of the 1999 senior project executive summaries were added into the document collection. All documents were indexed manually with at least 3 to 6 keyword for each document.

2. Usability testing on the visualization tool.

The background information collected from the subjects showed that third year students starting to work on their first independent study had little experience on information retrieval system, even the one in the library. They had little understanding on both bibliographic search and Boolean search. Out of 70 subjects, only 14 of them had experienced library search.

The data collected from the questionnaire showed that the subjective rating on GUIDO from the targeted Thai subjects was not different from the prior study. In 7-point Likert scale, the table below show the average evaluation of the subjects in this study.

Subject's rating on the usability of the GUIDO interface

Usability Questions	Average
Ease of use	6.22
Reference points creation	6.20
Appropriate retrieval threshold calculation selection	6.78
Relevant document selection	5.98
Interface manipulation	5.98

Usability Questions	Average
Interpretation of GUIDO display	5.53
Overall Satisfactory	5.82

The data collected from the third part of the questionnaire emphasizing on learnability of the tool on 7-point Likert scales is shown the following table.

Subject's rating on learnability of the GUIDO interface

Learnability Questions	Average
Predictability	5.20
Synthesizability	5.80
Consistency	4.20
Generalizability	4.88
Familiarity	3.23

Although the subjects rated GUIDO positively for the usability, they tended to rate the learnability of the tool less than the usability. Moreover, the familiarity of the tool was rated negatively.

The visualization for an information retrieval system might be a new kind of user interface to the subjects in this study, the 3rd year students seeking topics for their senior projects. Although the behavior these students can be a replica of those of researchers, the students were novices in both the information retrieval system experience and search experience. The subjects might not have good enough mental model on how to search for a research topic.

อ้างอิง

Aalbersberg 1995, <http://www.philips.com>

Benford, S.D., D. Snowdon, C. Greenhalgh, R. Ingram, I. Knox, & C.Brown 1995. VRVIBE: A virtual environment for co-operative information retrieval. Eurographics '95, August 30 - September 1, Maastricht, The Netherlands, 349-360.

Borgman, C.L. 1996. The user's mental model of an information retrieval system: an experiment on a prototype online catalog. International Journal of Man-Machine Studies, Vol.25, no1: 47-64.

Cutting, Douglass R., Karger, David R., & Pedersen, Jan O. (1993). Constant interaction-time Scatter/Gather browsing of very large document collections.. In Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 126-134). Pittsburgh, Pennsylvania: ACM.

- Dix, A., Finlay, J., Abowd, G., and Beale, R. (1997). Human-computer interaction. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Hearst, M.A. (1995). TileBars: Visualization of term distribution information in full text information access. In Proceedings of CHI '95: Human Factors in Computing Systems. Denver, Colorado: ACM.
- Hearst, M.A. 1995. TileBars: Visualization of term distribution information in full text information access. CHI '95 Proc., 213-220.
- Hemmje, Matthias, Kunkel, C., & Willett, A. (1994). LyberWorld – A visualization user interface supporting fulltext retrieval. In Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 249-258). Dublin, Ireland: ACM.
- Hemmje, M., C. Kunkel, & A. Willett. 1994. LyberWorld – a visualization user interface supporting fulltext retrieval. SIGIR '94, 249-259.
- Kennedy, J.B., K.J. Mitchell, & P.J. Barclay. 1996. A framework for information visualisation. SIGMOD Record 25(4): 30-34.
- Kim, H. 1996. Ph.D. dissertation, School of Information Sciences, University of Pittsburgh.
- Kim, H. & R.R. Korfhage. 1994. BIRD: Browsing interface for the retrieval of documents. Proc. 1994 IEEE Symposium on Visual Languages, St. Louis, 176-177.
- Kristoffersen, S. 2000. Learnability and Robustness of User Interfaces: Towards a Formal Analysis of Usability Design Principles.
- Lin, Xia (1992). Visualization for the document space. In Proceedings of Visualization '92 (pp. 274-281). Boston, Massachusetts: IEEE.
- Korfhage, R.R. 1997. Information Storage and Retrieval. New York: John Wiley & Sons.
- Korfhage R.R., D.S. Dubin, & E.M. Housman. 1998. What good is visualization? – three experiments. Proc. SPIE Conference on Visual Data Exploration and Analysis V, San Jose, California. 3298: 196-207.
- Koshman, S. 1996. Usability testing of a prototype visualization-based information retrieval system. Ph.D. dissertation, School of Library and Information Science, University of Pittsburgh.
- Lin, X. 1997. Map displays for information retrieval. JASIS 48(1): 40-54.
- Marchionini, G. 1995. Information Seeking in Electronic Environments. New York: Cambridge University Press.
- Morse, E. & M. Lewis. 1997. Why information visualizations sometimes fail. Proc. IEEE International Conference on Systems Man and Cybernetics. Orlando, Florida, October 12-15, 1997.
- Morse, E., M. Lewis, R.R. Korfhage, & K. Olsen. 1998. Evaluation of text, numeric and graphical presentations for information retrieval interfaces: User preferences and task performance measures.

- Proc. IEEE International Conference on Systems Man and Cybernetics, San Diego, California, October 11-14, 1998.
- Nuchprayoon, A. 1996. PhD. dissertation, School of Library and Information Science, University of Pittsburgh.
- Nuchprayoon, A & Korfhage, R. R. 1994. GUIDO: A visual tool for retrieving documents. In Proceedings of 1994 IEEE Symposium on Visual Languages (pp. 64-71). St. Louis, Missouri: IEEE Computer Society.
- Olsen, Kai A., Korfhage, Robert R., Sochats, Kenneth M., Spring, Michael B., & Williams, James G. (1993). Visualization of a document collection: The VIBE system. *Information Processing & Management* 29, 1, 69-81.
- Olsen, K.A., R.R. Korfhage, M.B.Spring, K.M. Sochats, & J.G. Williams. 1993. Visualization of a document collection: The VIBE system. *Information Processing & Management* 29(1): 69-81.
- Pejtersen, A. M. (1991). Icons for representation of domain knowledge in interface. In Proceedings of 1st International ISKO Conference: Tools for Knowledge Organization and the Human Interface, Vol. 2 (pp. 175-193).
- Pirolli, P., P. Schank, M. Hearst, & C. Diehl. 1996. Scatter/Gather browsing communicates the topoi structure of a very large text collection. *CHI '96*, 213-220.
- Rose, Daniel Eric, Mander, Richard, Oren, Tim, Ponclen, Dulce B., Salomon, Gitta, & Wong, Yin Yin (1993). Content awareness in a file system interface: Implementing the 'pile' metaphor for organizing information. In Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 260-269). Pittsburgh, Pennsylvania: ACM.
- Salton, Gerard & Allen, James (1994). Text retrieval using the vector processing model. In Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval (pp. 9-22). Las Vegas, Nevada: University of Nevada Las Vegas.
- Spoerri, Anselm (1993). InfoCrystal: A visual tool for information retrieval. In Proceedings of Visualization '93 (pp. 150-157). San Jose, California: IEEE Computer Society.
- Williamson, Christopher & Shneiderman, Ben (1992). The Dynamic HomeFinder: Evaluating dynamic queries in a real-estate information exploration system. In Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 338-346). Copenhagen, Denmark: ACM.
- Trilk, J. & F. Schutz. 1998. Visualization: A really generic approach or the art of mapping data to graphical objects. *Proc. SPIE Conference on Visual Data Exploration and Analysis V*, San Jose, California. 3298: 72-79.

- Wise, J.A., J.J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, & V. Crow. 1995. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. Proc. Information Visualization, October 20-21, 1995, 51-58.
- Zhou, M.X. & S.K. Feiner. 1998. Visual task characterization for automated visual discourse synthesis. CHI '98, 392-399.

Output จากโครงการวิจัย

- Nuchprayoon, A and Korfhage, K.K. (1997). Visualization Information Retrieval Interfaces: A Survey. Proceeding
- Nuchprayoon, A and Korfhage, K.K. (1997). GUIDO: visualizing document retrieval. In Proceeding of 1997 Symposium on Visual Languages. September 1997.