Using Concept Maps as a Cross-Language Resource Discovery Tool for Large Documents in Digital Libraries

Ryan Richardson Edward A. Fox
Virginia Tech
Digital Library Research Laboratory
Blacksburg, Virginia 24061 USA
+1-540-231-3615
ryanr, fox@vt.edu

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering, search process.*

General Terms

Algorithms, Experimentation, Human Factors.

Keywords

Concept maps, crosslingual information retrieval, automatic summarization.

ABSTRACT

Project Gutenburg, the Million Book Project, the Networked Digital Library of Theses and Dissertations, Amazon's book search service, and the recently announced collaboration of Google and leading libraries, all aim to make available large numbers of book-length objects, in a variety of languages. Traditional approaches to discovering a suitable book for a particular purpose have generally relied on catalog records, sometimes enhanced with abstracts. Full-text searching – popular, e.g., with legal and government documents – and passage retrieval techniques, suitable for encyclopedias and reference works, have not been adequately tested with large collections of large objects.

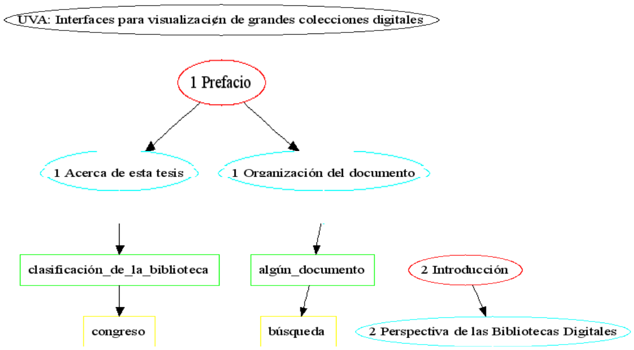
UVA: Interfaces para visualizaci¢n de grandes colecciones digitales
1 Prefacio
1 Acerca de esta tesis
1 Organización del documento
clasificación_de_la_biblioteca
algún_documento
2 Introducción
congreso
búsqueda
2 Perspectiva de las Bibliotecas Digitales

Figure 1: Automatically generated

We demonstrate an alternative approach, integrating visualization, text mining, and simple machine translation techniques. We show how electronic theses and dissertations (ETDs), in both English and Spanish, can be summarized as concept maps, according to three different design layouts: whole document maps, chapter-level maps, and table of contents-based maps. Our initial experiments are in the computer science and industrial systems engineering fields.

JCDL'05, June 7–11, 2005, Denver, Colorado, USA.
ACM 1-58113-876-8/05/0006.

Concept maps (see Figure 1) are "graphical representations of knowledge that are comprised of concepts and the relationships between them" [4]. Concept maps allow students to acquire knowledge more quickly than usual [2]. However, concept maps have yet to be extensively researched as aids to knowledge discovery. Our demonstration will show how this can work, including for multi-lingual collections. We argue that 1) users can more quickly grasp the key ideas of a work from a concept map than from an abstract alone; 2) it is easier to automatically generate concept maps than good abstracts; and 3) it is easier to automatically translate concept maps than abstracts.

We have tested several methods of finding relations in the texts, such as association rules [1], t-scores, Pearson's χ^2, Dice's coefficient, and mutual information. Using part-of-speech information from MontyTagger [3], we can find noun phrases which appear as nodes in the maps. Other parts-of-speech, usually verbs or prepositions, are used as the links between the nodes. Our demonstration highlights these techniques.

ACKNOWLEDGMENTS

This work was partially supported by NSF grants DUE-0121679 and IIS-0080748, and by SAIC as part of the Research and Development Experimental Collaboration—Applied Research Center. We also would like to thank Marcos Gonçalves, Rao Shen, and Jesús Trespalacios who assisted with user experiments.

REFERENCES

[1] Agrawal, R. and Srikant, R. *Fast Algorithms for Mining Association Rules in Large Databases*. Presented at 20th International Conference on Very Large Databases (VLDB'94), Santiago, Chile, Sept. 1994.

[2] Chmielewski, T. L. and Dansereau, D. F. *Enhancing the recall of text: Knowledge mapping training promotes implicit transfer*. Journal of Educational Psychology, vol. 90(3), pp. 407-413, Sept. 1998.

[3] Liu, H. "MontyTagger", 1.2 ed. Cambridge, Mass, 2003. *http://web.media.mit.edu/~hugo/montytagger/*.

[4] Novak, J. D. and Gowin, D. B. *Learning How To Learn*. Cambridge, UK, Cambridge University Press, 1984.