F. Sudweeks, H. Hrachovec and C. Ess (eds). Proceedings Cultural Attitudes Towards Communication and Technology 2010, Murdoch University, Australia, 45-52.

DESIGNING A SEMANTICALLY RICH VISUAL INTERFACE FOR CULTURAL DIGITAL LIBRARIES USING THE UNESCO MULTILINGUAL THESAURUS

ALI SHIRI[1], STAN RUECKER[2], CARLOS FIORENTINO[3], AMY STAFFORD[4], MATTHEW BOUCHARD[2] AND MARK BIEBER[5]
[1]*School of Library and Information Studies*
[2]*Humanities Computing Program and Department of English and Film*
[3]*Department of Art and Design*
[4]*Department of English and Film Studies*
[5]*Department of Computing Science*
University of Alberta, Canada

Abstract. This paper reports on the design of a visual user interface for the UNESCO digital portal. The interface makes use of the UNESCO multilingual thesaurus to provide visualized views of terms and their relationships and the way in which spaces associated with the thesaurus, the query and the results can be integrated into a single user interface.

1. Introduction

The purpose of this project was to develop and study thesaurus-enhanced multilingual visual interfaces for cultural digital libraries. The specific aim was to design a series of visual interfaces utilizing the UNESCO thesaurus, a multilingual thesaurus in English, French and Spanish. The UNESCO thesaurus covers such areas as education, science, culture, social and human sciences, information and communication, and politics, law and economics; making it a perfect candidate for developing a multilingual visual interface for the humanities and social sciences. Our experimental interface is intended to support humanities and social science scholars in effectively and efficiently exploring the information space of digital libraries. This project built on a series of small projects started in 2006 that studied the feasibility of designing thesaurus-based visual interfaces (Shiri et al., 2006a, 2006b, 2007; Ruecker et al., 2007a; Stafford et al., 2008). Our previous studies have focused on bilingual interfaces using a general purpose thesaurus. The goal of the current project is to extend this work by developing and evaluating visual interfaces for multilingual cultural digital libraries containing multimedia digital information in a variety of formats, such as text, images and audiovisual.

2. Previous Research

Jorna and Davis (2001) point out the importance of tools to support multilingual information retrieval. They note that in order to facilitate cross-cultural communication in an increasingly global information society multilingual thesauri can play a significant role. Thesauri have played an important role in modern information storage and retrieval systems. While initial proposals to utilize thesauri focused on their ability to ensure consistent analysis of documents during input to information retrieval systems, they have increasingly become vital as aids to effective retrieval (Shiri, 2000). As Aitchison et al. (1997) have stated, the role of the thesaurus is changing, but it is likely to remain an important retrieval tool. This refocusing of the use of thesauri within information retrieval systems means that it is imperative that professionals are cognizant of the potential of thesauri as essential components of the largest information retrieval environment, namely the World Wide Web (Shiri, 2000).

Digital libraries are multifaceted and complex information structures that offer a wide range and variety of information bearing objects. They vary in their content, subject matter, cultural characteristics, language etc. Arms (2000) notes that "a digital library is only as good as the interface it provides to its users." The variety of digital objects and materials in a digital library poses challenges to the design of usable and easy to understand user interfaces. Visual interfaces to digital libraries have recently found widespread attention. This development is mainly due to the fact that information visualization techniques allow for rich representation of information bearing objects within digital libraries. Borner and Chen (2001) suggest that visual interfaces for digital libraries shift users' mental load from slow reading to faster perceptual processes such as visual pattern recognition. Zaphiris, et al. (2004) explore the application of information visualization in digital libraries and identify three key tasks in digital libraries, namely searching, browsing and navigation to which information visualization can make contribution.

Over the last decade, a number of digital libraries and online initiatives have incorporated knowledge organization systems such as thesauri and classification systems into their user interfaces to provide support for query formulation, collection browsing and other search tasks (Hodge, 2000; Hudon and Hjartarson, 2002;Shiri and Molberg, 2005). A few prototype interfaces have utilized graphical as well as two- or three-dimensional category hierarchies using the MeSH Thesaurus. TraverseNet (McMath et al., 1989), MeSHBrowse (Korn& Shneiderman, 1995), Cat-a-cone (Hearst &Karadi, 1997), Visual MeSH (Lin, 1999), and the Integrated Thesaurus-Results Browser (Sutcliffe et al., 2000) are among the prototype thesaurus-enhanced interfaces. There are also some studies that have found that thesaurus-enhanced search interfaces can support users' query formulation and expansion (Beaulieu, 1997; Shiri, 2006c). Jorna and Davis (2001) note that in order to facilitate cross-cultural communication in an increasingly global information society multilingual thesauri can play a significant role.

A number of reports cite the general usefulness of the UNESCO Thesaurus. For example, Williamson (2007) highlights the UNESCO Thesaurus as one of many online knowledge organization systems in her analysis of the development of online finding aids. She notes that thesauri are powerful navigational aids for Internet users because of the hierarchical and relational nature typical to these information organization systems.

With regard to UNESCO in particular, Williamson notes that the Thesaurus is "simple and effective to use" and has clear instructions, but that the two presentation modes – as a hierarchy or as an alphabetical list – may pose browsing problems because the user cannot view or browse both modes at once. With regard to the usefulness of UNESCO in relation to specific projects or resources, an oft-cited example comes from Garrod (2000), who describes the relatively early adoption of the Thesaurus for a digital archive of the UK's National Digital Archive of Datasets (UK NDAD) as the finding aid of choice for the datasets. Garrod (2002) writes that, in the context of the UK's National Archives Network, numerous archival projects have adopted the Thesaurus because of its broad subject reach, availability in electronic format, adherence to both British and ISO standards for thesauri, and the fact that UNESCO itself is willing to share the Thesaurus for non-profit use.

Our goal to present the design of a visual user interface developed to support users of the UNESCO digital libraries in searching, browsing, navigating and exploring the content. One of the core components of the interface will be the UNESCO multilingual thesaurus, which will assist humanities scholars to formulate, and expand their queries in a semantically rich visual environment where users search terms can be enhanced through interaction with the thesaurus.

3. Methodology

3.1. THEORETICAL FRAMEWORK

The theoretical framework for the design of the interface will be based on two key concepts. The first is the idea of rich-prospect interfaces, in which individual representations of every item in a collection are combined with emergent tools (Ruecker and Chow, 2003). Using this conceptual framework, Ruecker et al. have subsequently developed a number of metadata enhanced visual interfaces to support users' information search and exploration activities (Ruecker et al., 2006, 2007b). The second set of principles draws on the design ideas for thesaurus-based search interfaces suggested by Shiri et al. (2002), including:

- Providing hierarchical and alphabetical lists to support different strategies.
- Allowing flexible ways of choosing terms.
- Facilitating movement between a descriptor and its hierarchical structure.
- Catering for the selection of alternative Boolean operators.
- Providing a *term pool* option for saving the descriptors.
- Integrating thesaurus and retrieved documents displays.
- Making thesaurus options available in all stages of the search process.

4. User Interface Design

In the proposed interface, the aim was to provide the user with the following spaces within the interface:

- Query space: for formulating search statements
- Thesaurus space: for browsing and navigating the thesaurus
- Document space: for viewing document representations

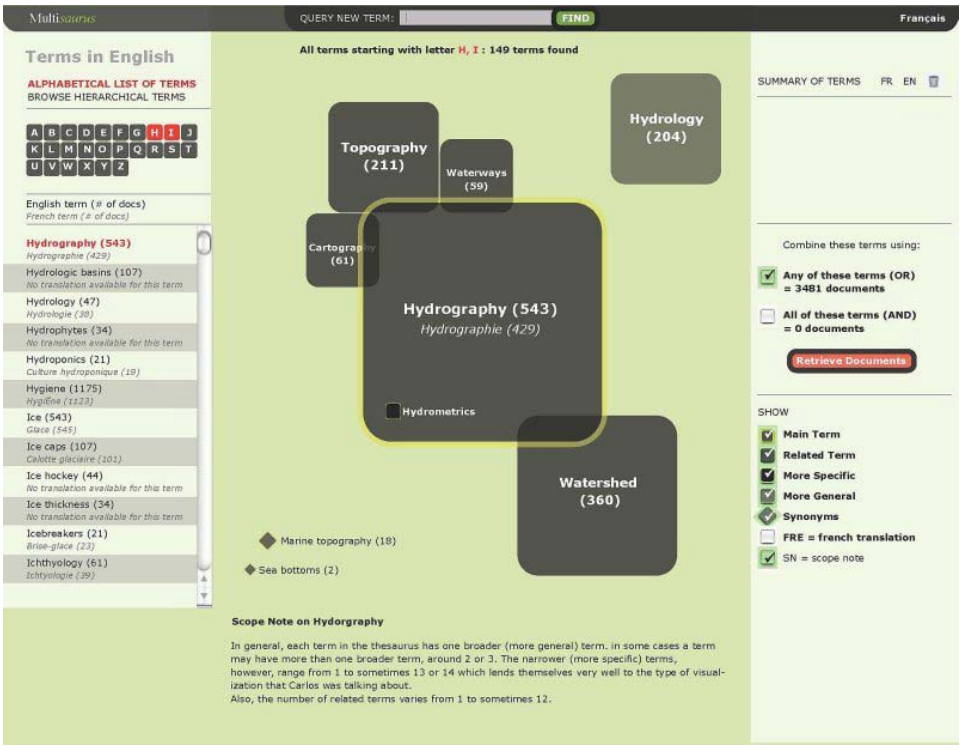

Figure 1. The Thesaurus and Query spaces

The Query space is located across the top and on the right side of the screen while the Thesaurus space is located on the left and in the centre. Users can search for a single term in the thesaurus by entering it in the query box at the top of the page and clicking the Find button. If the term exists in the thesaurus it will appear in the centre of the screen with a number in parentheses beside it, which indicates the number of documents in the collection that include the selected term. Users can also browse all the terms in the thesaurus using the panel on the left, which can be sorted either alphabetically or hierarchically by category. Again, each term has a number beside it in parentheses indicating how many documents in the collection contain the term. When a term in the list is clicked, it will appear in the centre of the screen.

When a term is selected by either method it is represented by a square in the central Thesaurus space. By utilizing the checkboxes in the bottom of the right-hand panel, users can choose to view the thesaurus terms that are related, narrower (more specific), broader (more general), and preferred or non-preferred (synonyms) compared with the selected term. These associated terms are also represented in the Thesaurus

space by squares or diamonds and their relationship to the selected term is represented by their relative proximity and opacity.

In figure 1, for example, Hydrometrics is a more specific term than the selected term, Hydrography. Therefore, the square for Hydrometrics is much darker and completely overlaps the square for Hydrography. Hydrology, on the other hand, is a more general term than Hydrography, so it is further away and more transparent. Furthermore, the size of each shape represents the number of documents in the collection that contain the represented term. Thus, the square for Hydrology, which appears 204 times in the document collection, is smaller than the square for Hydrography, which appears 543 times.

Users can also use the checkboxes in the right-hand panel to show the terms in more than one language at once and to view scope notes for selected terms (Figure 2). When users decide to add a term to their query, they do so by clicking on its square in the centre of the screen, at which time it is added to the Summary of Terms list, or term pool, at the top of the right-hand panel. Users can add as many terms as they like, delete them at any time, choose to keep them in only one language rather than multiple languages, and combine them using the Boolean operators below the list. When they have finished formulating their query they click Retrieve Documents to view the results.

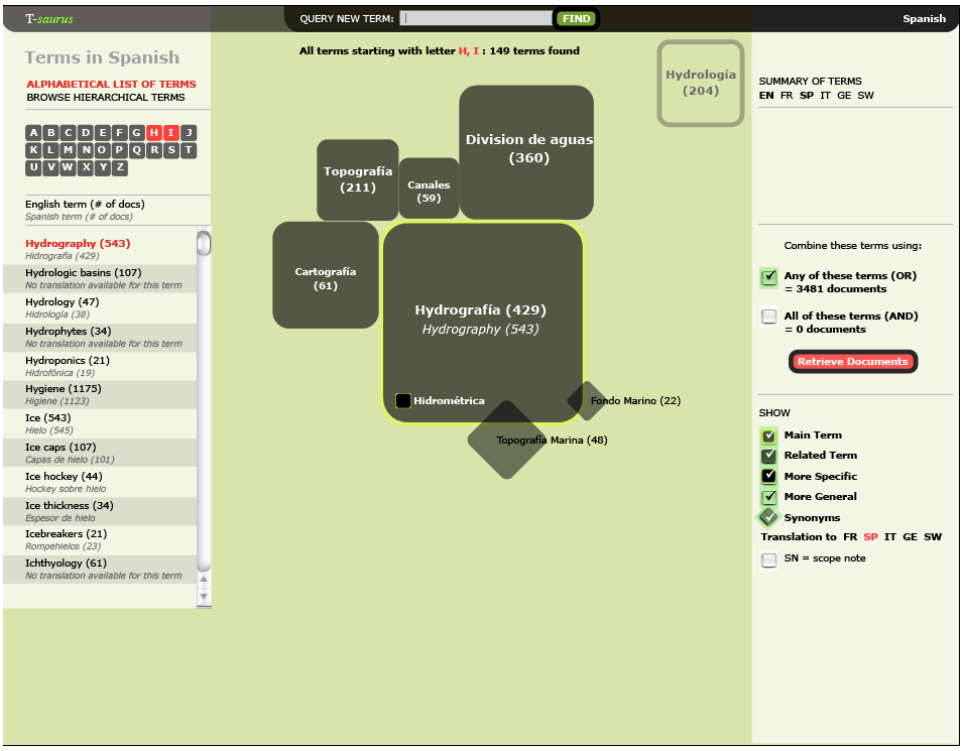

Figure 2. The user is working in Spanish.

The Document space now replaces both the Thesaurus space in the centre and part of the Query space on the bottom right-hand side of the screen. In the centre of the screen the document results are represented visually by red squares. Again, a document's proximity to the selected term is significant as it represents the document's degree of relevance. The document results are also displayed in the right panel as a sortable list of titles. Both the title in the list and the representative square can be clicked to open a PDF version of the document.

5. Conclusion

The main idea behind the above user interfaces is to make use of the power of semantics in thesauri and current visualization techniques to facilitate searching and browsing in digital information environments.

There are a number of novel features and functionalities that were incorporated into the these interfaces. The first novel aspect of these interfaces lies in their approach to visualize semantic relationships held in standard thesauri, namely broader, narrower, related and synonymous terms. The second novel aspect is the use of such visual cues as location, size, colour, font type and the use of space on the interface along with visualization techniques such as word clouds and the notion of terms as visual objects. In addition, these interfaces are developed based on the idea of combining three different spaces into one single user interface, namely, thesaurus space, query space and document space. The next step is to develop the operational prototype for these interfaces and to conduct user-centred evaluation studies to establish their usability, learnability and usefulness.

Acknowledgements

This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) Strategic Research Grant.

References

Aitchison, J., Gilchrist, A. and Bawden, D. (1997). *Thesaurus Construction and Use: A Practical Manual*, 3rd ed., Aslib, London.

Anvik, Karl, Ali Shiri, Ximena Rossello, and Stan Ruecker. “A Tabular Thesaurus Browser to Support Multilingual Queries.” Paper presented at the Society for Digital Humanities (SDH/SEMI) conference. York University, Toronto. May 29-31, 2006.

Arms, William Y. (2000). *Digital Libraries*. M.I.T. Press.

Beaulieu, M. (1997). Experiments of interfaces to support query expansion. *Journal of Documentation*, 53 (1), 8-19.

Bertin, J.(2000/2001). Matrix Theory of Graphics. *Information Design Journal*. 10(1), 5-19.

Börner, K. and Chen, C. Visual Interfaces to Digital Libraries, *Lecture Notes in Computer Science* 2539, Berlin Heidelberg: Springer-Verlag, 2002.

Borner, K.; Chen, C. (2001). Visual Interfaces to Digital Libraries– The *First International Workshop at the First ACM/IEEE Joint Conference on Digital Libraries*, 2001. Available at: http://www.sigir.org/forum/S2001/DLVis.pdf

Garrod, P. (2000). Use of the UNESCO Thesaurus for archival subject indexing at UK NDAD. *Journal of the Society of Archivists.* 21(1), 37-54. Retrieved July 16, 2009, from http://www.ukat.org.uk/downloads/case03.pdf.

Hearst, M.A. &Karadi, C. (1997). Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In: *SIGIR '97:Proceedings of the 20th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, Philadelphia, PA, 27–31 July1997 (ACM, New York, 1997).

Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*, CLIR, Washington, Available at: http://www.clir.org/pubs/abstract/pub91abst.html

Hudon, M., & Hjartarson, F. (2002). Governments meet people: developing metathesauri in the framework of "government online" initiatives. In *Advancing knowledge : Expanding horizons for information science : Proceedings of the 30th Annual Conference of the Canadian Association for Information Science, 30 May - 1 June 2002, Toronto, Canada*, p.46-60. Edited by Lynne C. Howarth, Christopher Cronin, and Anna T. Slawek. Toronto, ON: CAIS, 2002.

Jorna, K. & Davies, S. (2001). Multilingual thesauri for the modern world - no ideal solution? *Journal of Documentation*. 57(2), 284-295.

Korn, F. & Shneiderman, B. (1995). *Navigating terminology hierarchies to access a digital library of medical images*, University of Maryland Technical Report HCIL-TR-94- 03, 1995.

Lin, X. (1999). Visual MeSH. In: M. Hearst, F. Gey and R. Tong (eds), *SIGIR'99: Proceedings of 22nd Annual InternationalACM/SIGIR Conference on Research andDevelopment in Information Retrieval*, Berkeley, CA, 15–19 August 1999 (ACM, New York, 1999).

McMath, C.F.; Tamaru, C.S. &Rada, R. (1989). A graphical thesaurus-based information retrieval system, *InternationalJournal of Man–Machine Studies.* 31, 121–147.

Milstead, J. (1998). Thesauri in a full-text world. In Cochrane, P.A. and Johnson, E. (Eds), Visualizing Subject Access for 21st Century Information Resources, Proceedings of the 1997 Annual Clinic on Library Applications of Data Processing, Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign, IL, pp. 28-38.

Ruecker, Stan and Rosan Chow. (2003). The significance of prospect in interfaces to health-related web sites for the elderly. *Proceedings from Include 2003*. Helen Hamlyn Research Institute, Royal College of Art, London England March 25-8, 2003. CD format. pp. 273-7.

Ruecker, S., Lewcio, M., Plouffe, M., Wynne, M.(2006). I never forget a face, a rich-prospect image browser for conferences. Paper presented at the Society for Digital Humanities (SDH/SEMI) conference. York University, Toronto. May 29-31

Ruecker, S., Given, L., Simpson, H., Sadler, E., Ruskin, A. (2007a). Design of a Rich-Prospect Browsing Interface for Seniors, A Qualitative Study of Image Similarity Clustering. *Visible Language*. *41*(1), 4-22.

Ruecker, Stan, Stéfan Sinclair, and Milena Radzikowska.(2007b). Confidence, Visual Research and the Aesthetic Function. *Partnership: the Canadian Journal of Library and Information Practice and Research*. *2*(1).

Shen, R., Srinivas, N., Vemuri, S. N., Fan, W., Torres, R. S. , Fox, E. A. (2006). Exploring digital libraries: integrating browsing, searching, and visualization. In *Proceedings of the 6th*

ACM/IEEE-CS Joint Conference on Digital libraries JCDL 2006, Chapel Hill, NC, USA, June 11-15, 2006, 1-10.

Shiri, A. A.; Revie, C. (2000).Thesauri on The Web: Current Developments and Trends. *Online Information Review*, 24(4): pp.273-279.

Shiri, A. A.; Revie, C.; Chowdhury, G. (2002). Thesaurus-Enhanced Search Interfaces. Journal of Information Science. 28(2): pp.111-122.

Shiri, A.; Molberg, Keri (2005). Interfaces to Knowledge Organization Systems in Canadian Digital Library Collections. *Online Information Review*. 29(6), 604-620.

Shiri, A.;Ruecker, S.; Anvik, K.; Rossello, X. (2006a). Thesaurus-enhanced Visual Interfaces for Multilingual Information Retrieval. Proceedings of the American Society for Information Science and Technology (ASIS&T) Annual Conference, Austin, Texas, U.S., November 3 - 9, 2006.

Shiri, A.;Ruecker, S.; Anvik, K.; Rossello, X. (2006b). Rich Prospect Visual Interfaces for Multilingual Query Enhancement Using Thesauri. *Proceedings of theCanadian Symposium of Text Analysis (CASTA) Annual Conference*, Fredericton, New Brunswick, Canada, October 11-14, 2006.

Shiri, A. &Revie, C. (2006c). Query Expansion behaviour within a thesaurus-enhanced search environment: A user-centred evaluation. *Journal of the American Society for Information Science and Technology*, 57 (4), 462-478.

Shiri, A.; Ruecker, S.; Rossello, X.; Bouchard, M.; Mehta, P. (2007). Development of A Thesaurus-enhanced Visual Interface for Multilingual Digital Libraries. *Proceedings of the 35th Annual Conference of the Canadian Association for Information Science, Information Sharing in a Fragmented World: Crossing Boundaries*, K. Dalkir and C. Arsenault (eds.), Montreal, May 10-12, 2007.

Shneiderman, B., Williamson, C. and Ahlberg, C. (1992). Dynamic Queries: Database Searching by Direct Manipulation. *Proceedings of the SIGCHI conference on human factors in computing systems.* pp. 669-670. http://doi.acm.org/10.1145/142750.143082. Accessed March 10, 2006.

Stafford, A., Shiri, A., Ruecker, S., Bouchard, M., Mehta, P., Anvik, K., Rossello, X. (2008). Searchling: User-Centered Evaluation of a Visual Thesaurus-Enhanced Interface for Bilingual Digital Libraries. In: *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries: ECDL 2008*, Aarhus, Denmark, September 14-17, 2008.

Sutcliffe, A.G. , Ennis, M. & Hu, J. (2000). Evaluating the effectiveness of visual user interfaces for information retrieval, *International Journal of Human–ComputerStudies* 53(5), 741–763.

Williamson, N. J. (2007). Knowledge structures and the Internet: progress and prospects. *Cataloging & Classification Quarterly*, 44(3), 329-342.

Zaphiris, P., Gill, K., Ma, T. H. , Wilson, S., Petrie, H. (2004). Exploring the use of information visualization for digital libraries. *New Review of Information Networking*, 10 (1), May 2004, 51 – 69.