Exploratory Search Interfaces for the UNESCO Multilingual Digital Library: Combining Visualization and Semantics

Stan Ruecker, Ali Shiri, Carlos Fiorentino, Matt Bouchard, Amy Stafford, & Mark Bieber

2011

Journal of the Chicago Colloquium on Digital Humanties and Computer Science
Volume 1 Number 3 (2011)

Exploratory Search Interfaces for the UNESCO Multilingual Digital Library: Combining Visualization and Semantics

Stan Ruecker, Humanities Computing Program, University of Alberta
Ali Shiri, School of Library and Information Studies, University of Alberta
Carlos Fiorentino, Department of Art & Design, University of Alberta
Matt Bouchard, Humanities Computing Program, University of Alberta
Amy Stafford, Department of English and Film Studies, University of Alberta
Mark Bieber, Department of Computing Science, University of Alberta

Abstract

The objective of this paper is to report on the design of semantically–rich and dynamic visual user interfaces to support exploratory interaction with the UNESCO digital library materials. The UNESCO multilingual thesaurus, which supports English, French, and Spanish, has been utilized to provide a semantic and multilingual environment where users can browse the thesaurus using word buckets, formulate queries, explore the conceptual space of their query terms and view results in a dynamic and highly interactive environment. The interface design focuses on integrating searching and browsing and multilingual elements in a visual environment where terms and retrieved documents are designed as visual objects. To facilitate users' interaction with terms, thesaural relationships are shown using such elements as colour, size, location, and distance.

Introduction

Highly interactive and dynamic user interfaces for exploratory browsing and searching of digital information collections has been the focus of some recent studies. White and Kules[1] note that in exploratory search, users generally combine querying and browsing strategies to foster learning and investigation. Marchionini[2] points out that to engage people more fully in the search process and put them in continuous control, researchers are devising highly interactive user interfaces. In this paper we report the design of an exploratory user interface that takes advantage of dynamic views supported by the multilingual UNESCO thesaurus. The key features of the interface are a) combining searching and browsing, b) supporting dynamic exploration of the conceptual structure of a thesaurus, c) using word buckets to provide high level overviews of the terms and the collection, d) utilizing a novel technique to implicitly show thesaural relationships using colour, size and distance, and e) visualized view of the results along with thesaurus terms. The user interface supports understanding concepts and their relationships, examining and comparing results and reformulating queries, as well as serendipitous browsing.

Note: For larger, higher quality versions of the figures reproduced here, please refer to the *Supplementary Data* section accompanying this article online at http://jdhcs.uchicago.edu

[1] R. W. White, B. Kules, S. M. Drucker, and M. C. Schraefel, "Supporting Exploratory Search," Introduction, Special Issue. *Communications of the ACM* 49, no. 4 (2006): 36-39.

[2] G. Marchionini, "Exploratory search: from finding to understanding," *Communications of the ACM* 49, no. 4 (2006): 41-46.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Theoretical Framework

The theoretical framework for the design of the interface is based on three 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.[3] The second set of principles is the design ideas for thesaurus-based search interfaces suggested by Shiri et al.[4] including:

- Providing hierarchical and alphabetical lists to support different strategies.
- Allowing flexible ways of choosing terms.
- Facilitating moving 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.

In this project we made use of size, colour, and distance to implicitly show thesaural relationships to reduce users' cognitive load. Traditionally many user interfaces that incorporated standard relationship types such as broader, narrower, related, or synonymous terms were criticized for being too overwhelming for the user who is not familiar with an information retrieval thesaurus. To alleviate this problem, we adopted a new approach which takes advantage of word clouds as well as distance between terms, size, and colour.

User Interface Design

The interface design provides users with three different spaces, namely thesaurus space, query space, and the retrieved documents space. This design approach emphasizes the notion of exploratory search spaces where users have the opportunity to explore dynamically updated search results and reformulate queries and interact with semantically related terms. The multilingual thesaurus-enhanced visual interface supports users in formulating sophisticated search queries in multiple languages. One of the key and novel features of the design is to integrate searching and browsing in a visual environment where exploratory and directed searching can be effectively accommodated. The semantic relationships in the UNESCO multilingual thesaurus and various visualization techniques form the basis for the design of the exploratory user interfaces. The interface makes use of visual objects as well as such features as size, colour, location, zoom in and zoom out features, and the word cloud idea to distinguish between various types of thesaurus terms and their relationships.

The first version of these interfaces shows a core of visual elements consisting of a set of "buckets" organized in the center of the screen. Figure 1 shows the first version of the interface. The number of buckets represents the number of terms found by the query. The size of the buckets represents the number of matches for that particular term, while proximity and opacity represent scope and accuracy of the term in relation to pre-established hierarchies for the query: main terms, related

---

[3] Stan Ruecker and Rosan Chow, "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): 273-7.

[4] A. Shiri, C. Revie, and G. Chowdhury. "Thesaurus-enhanced search interfaces," *Journal of Information Science* 28, no. 2 (2002): 111-122.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

terms, more specific and more general terms, and synonyms. Within this interface users can easily switch between searching and browsing and can shift their attention from browsing the alphabetical list to exploring the conceptual map of a particular term. They can also use the language control feature on the right-hand side to switch to a different language. The interface will then dynamically show the terms in the selected language. The column on the left shows the alphabetical list, the centre of the interface visualizes the concept space of the term 'hydrography' along with its narrower, broader, and related terms, and the right column shows the language and query formulation options.

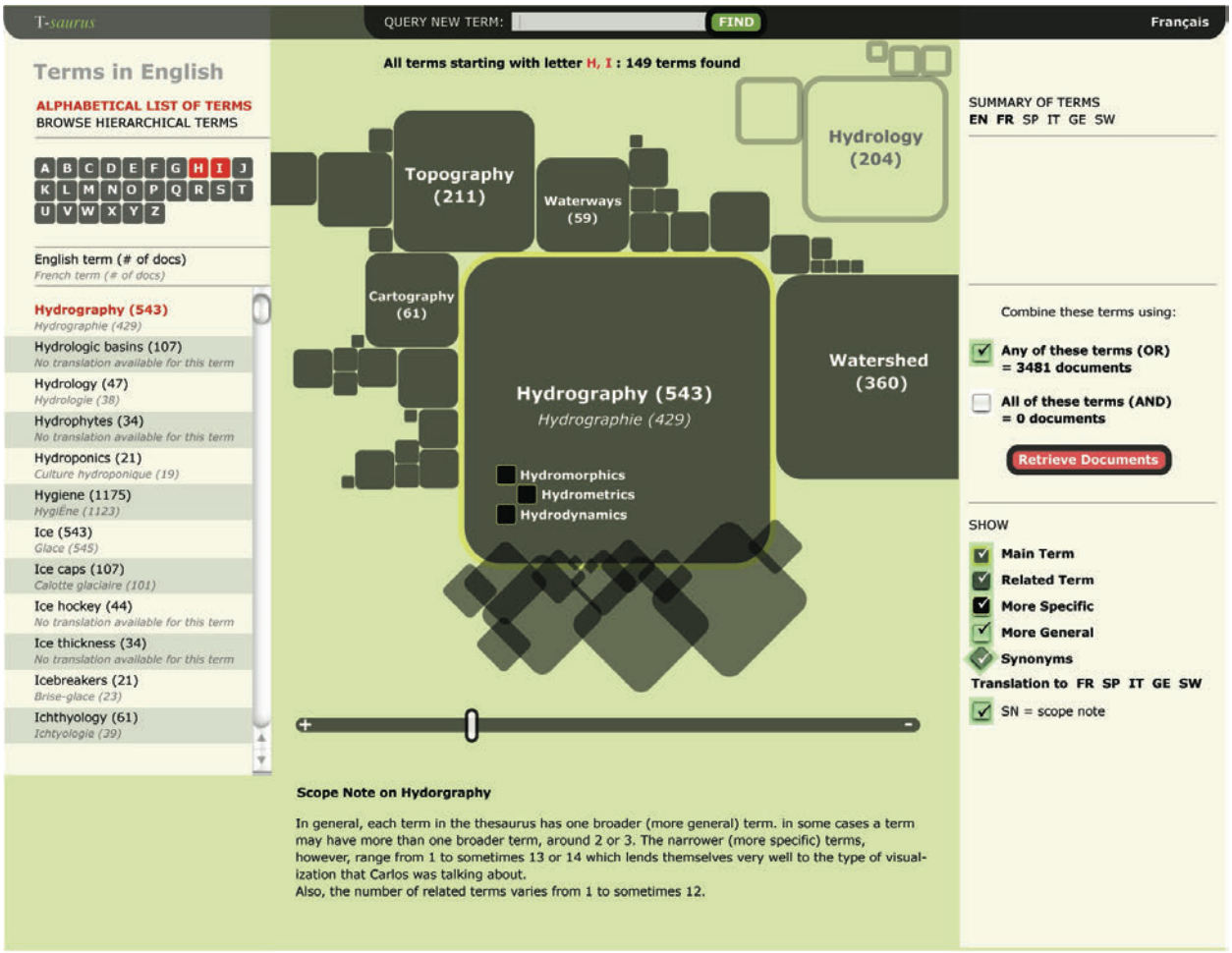

Figure 1. Query formulation interface.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

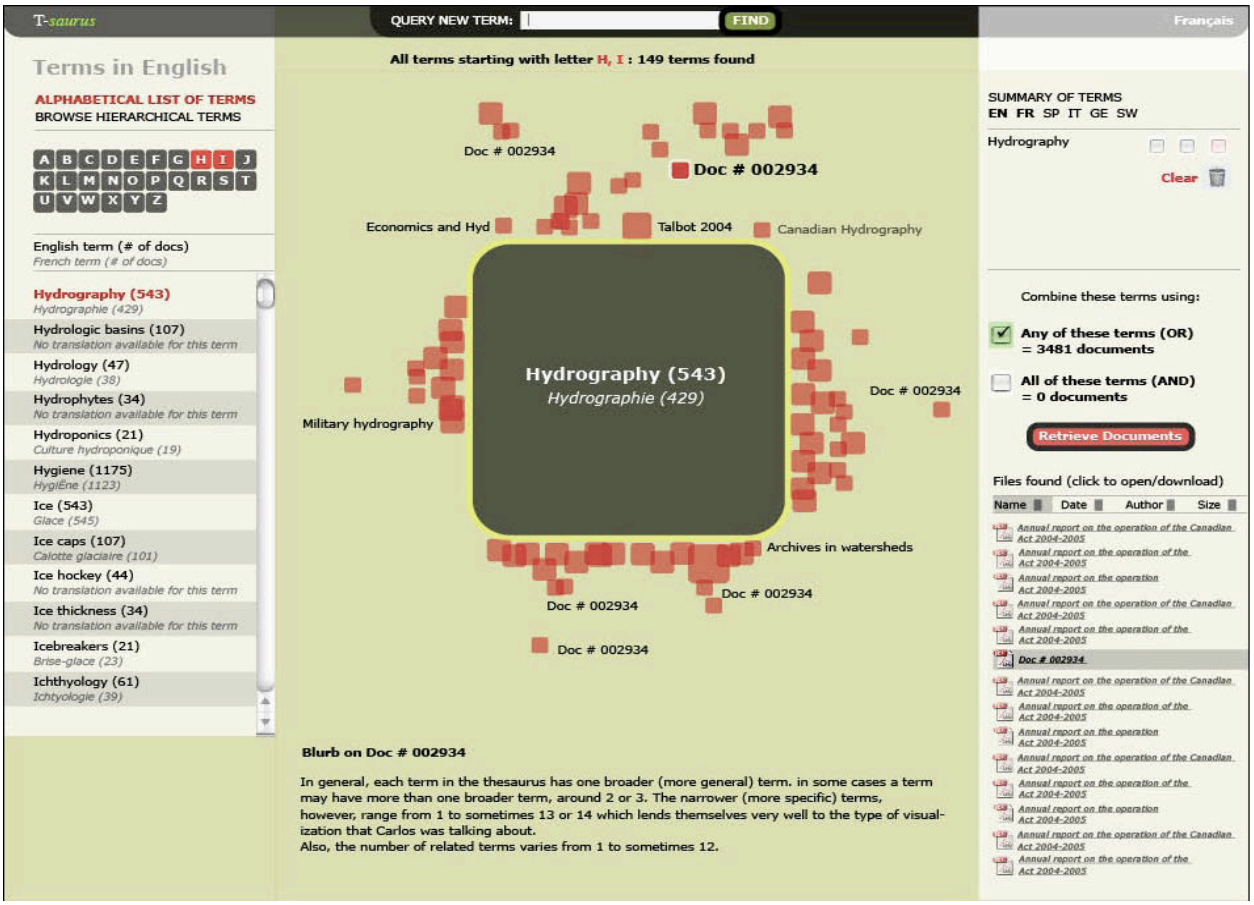

Figure 2. Results display interface.

Figure 2 shows red boxes in the middle around the green box, representing the results retrieved for the chosen term. The green box in the middle shows the thesaurus term and its French equivalent, as well as the number of documents indexed using that term. The column on the right-hand side functions as a control mechanism for sorting results based on such metadata elements as date, author, and size. The user can click on a thesaurus term shown on the left or type in a term in the query box to switch to a different term and its associated terms.

Conclusion and Future Work

We reported the design of semantically rich visual user interfaces to support exploratory and dynamic interaction with multilingual digital libraries. The main idea behind the above user interfaces is to utilize the power of semantics in thesauri as well as current visualization techniques to facilitate and combine searching and browsing to support users' interaction with digital information collections. There are a number of novel features and functionalities that were incorporated into the interfaces. The first novel aspect lies in our approach to visualizing semantic relationships held in standard thesauri; namely broader, narrower, related, and synonymous terms based on such elements as colour, font size, and distance. To the best of our knowledge there is no previous study that has made use of word buckets and thesaural relationships to develop interactive and exploratory user interfaces for multilingual collections. The second novel aspect of these interfaces is related to the combined use of alphabetical listing, word clouds, and 'buckets' to support various information-

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

seeking strategies, and to provide an overview of the conceptual space along with query formulation support features. In order to provide the user with more control over the exploration process, a slider feature has been designed to support users' exploration of the conceptual space of the thesaurus through zooming in and out. The next step is to use exploratory search tasks to conduct a series of comparative user-centered evaluation studies to establish the usability, learnability, and usefulness of these interfaces.

Bibliography

Marchionini, G. "Exploratory search: from finding to understanding," *Communications of the ACM* 49, no. 4 (2006): 41-46.

Ruecker, Stan and Rosan Chow. "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): 273-7.

Shiri, A, C. Revie, and G. Chowdhury. "Thesaurus-enhanced search interfaces," *Journal of Information Science* 28, no. 2 (2002): 111-122.

White, R. W., B. Kules, S. M. Drucker, and M. C. Schraefel. "Supporting Exploratory Search," Introduction, Special Issue. *Communications of the ACM* 49, no. 4 (2006): 36-39.

Source URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago