Interactive Visualization for Multilingual Search
by Stan Ruecker, Ali Shiri and Carlos Fiorentino

Over the past 10 years we have been carrying out a series of experiments in rich-prospect browsing, where some meaningful representation of every item in a collection is combined with tools for manipulating the display [1]. Example projects include the Mandala Browser for the visual construction of complex Boolean queries of XML documents [2], the Texttiles browser for visual grouping of items in RSS feeds and the structured surfaces interfaces for the Just in Time Research (JiTR) project [3]. What these interfaces each provide is an environment for people to explore digital materials where some combination of grouping and searching supports both information access and information exploration.

In addition, we have been exploring the use of interactive visualization in the design of user interfaces that leverage knowledge organization systems as multilingual thesauri. The goal is to support a variety of information seeking tasks, namely searching, browsing, navigation and exploration. More specifically, depending on the level of sophistication that can be offered in the interface, visual user interfaces have the potential to support such interactive tasks as query formulation, modification or expansion. In pursuit of this goal, we have developed two visual user interfaces, called Searchling and T-Saurus that use the Government of Canada Core Subject Thesaurus, a bilingual thesaurus in English and French, and the UNESCO multilingual thesaurus in English, French and Spanish [7].

Theoretical Framework and Design of Searchling and T-Saurus
The design of the Searchling and T-Saurus user interfaces was based on two key elements – the first was the idea of rich-prospect interfaces; the second was the set of principles for design ideas for thesaurus-based search interfaces suggested by Shiri, Revie and Chowdhury in 2002 [4], including the following:

- 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.

The two interfaces provide the user with the following three spaces within a single screen:
query space: for formulating search statements
thesaurus space: for browsing and navigating the thesaurus
document space: for viewing document representations.

We created two versions of the Searching interface: one for the Government of Canada Core Subject Thesaurus [5] and the other for the UNESCO multilingual thesaurus [6]. Our comparative evaluation of T-Saurus and Searching made use of the UNESCO thesaurus. The functional prototypes of these user interfaces are available at http://thesaurusbrowser.info.

Searching User Interface

The Government of Canada Core Subject Thesaurus is a well-structured and well-established thesaurus, which is currently being used by a number of government agencies and information centers in Canada for indexing and information representation [7]. The thesaurus is bilingual, which allows for multilingual features to be designed based on the terminology of the thesaurus.

The goal of the Searching interface is to make the thesaurus information readily accessible to the user during the process of query formulation or reformulation (Figure 1).

A tabular view allows quick navigation through the five kinds of data, namely broader terms, narrower terms, related terms, and preferred and non-preferred (synonymous) terms, to help inform the user where a given term falls within the language of the thesaurus. In addition, a side panel presents a list of the highest-level facets of the thesaurus, which gives a user unfamiliar with the system a possible list of starting points. A language switch provides a means of checking for corresponding terms in another language. These terms are always visible as microtext satellites of the query terms. Their persistent presence in the thesaurus table both reminds the user that more than one language is available and also provides a quick means of switching back and forth between languages.

This function is also served by an explicit language selection choice, made with a radio button in the panel to the right of the main table. The Searching user interface is similar to faceted search user interfaces, except that it provides various thesaurus-based browsing and search functionalities in addition to high-level facets.

In using the system, people can add as many terms as they like to the Selected Terms list and can delete them at any time or choose to keep them in only one language rather than both. When users have finished formulating their query by selecting terms, choosing languages and combining them using the Boolean operators below the Selected Terms list, they click “Retrieve Documents” to get a list of relevant documents in the collection.

left, the Query Formulation space on the right and the top of the Results space at the bottom.

Searching User Evaluation

The evaluation method for Searching drew upon information search behavior and usability techniques. Fifteen researchers were recruited from
the Department of Modern Languages at the University of Alberta. Eight of the 15 participants were bilingual: five were fluent in English and French, two in English and Spanish and one in English and Italian. Four of the five English/French speakers currently conduct research in both languages.

User evaluation methods consisted of interviews, usability and affordance strength questionnaires and usability testing software. Participants completed four search tasks that required interaction with the thesaurus, conducting searches and viewing results. In other words, they were provided with opportunities to interact with the three different spaces within the user interface.

The bilingual researchers in the study agreed that Searchling’s ability to facilitate searches simultaneously in both languages is very useful, with the qualification that this feature is most useful when it allows them to collect a larger quantity of information, not just the same information repeated in both languages. In other words, they are not interested in using the interface as a translation tool, as they are fluent in both languages and can search equally well in either themselves. They are interested, however, in being able to find all the documents that a collection may hold on a specific topic at once, no matter the language of those documents. In practical terms, different collections will vary in terms of coverage in each language.

The most promising finding for Searchling is that it has the potential to solve the greatest problem that people encounter with other search tools, namely formulating queries using terms from a thesaurus. Some users said they would find Searchling most useful at the beginning of a research project on an unfamiliar topic, because they could start by browsing through general categories for relevant terms and use the Thesaurus to help them narrow or broaden their search (and consequently see how many documents the collection has to offer them on each topic). Most users also appreciated that they could keep as many items as they liked in the Selected Terms list and that they could keep them there without adding them to the search for documents by unchecking the language boxes beside each term.

**T-Saurus User Interface**

The T-Saurus user interface design (Figure 2) is based on the idea of representing the information as a set of visual elements rather than a series of text lists. This approach allows users to interact dynamically and intuitively with the information as objects, optimize the process of retrieving information and obtain results more quickly. The interface shows a core of visual elements consisting of a set of “buckets” organized in the center of the screen. 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 terms, more specific, more general and synonyms.

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 center 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 to the selected term. These associated terms are also represented in the Thesaurus space by squares, and their relationship to the selected term is represented by their relative proximity and opacity.

People can use the checkboxes in the right-hand panel to show the terms in more than one language at once and also to view scope notes for selected terms. 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 3). The red dots in the middle around the green box represent the results retrieved for the chosen term. The green box in the middle shows the thesaurus term, its French and Spanish equivalents and the number of documents indexed using that term.

When a person mouses over a red dot, the title of the document is immediately shown. In addition to the visual representation of the retrieved documents, the interface shows titles of retrieved documents on the bottom right-hand side of the interface.

**T-Saurus User Evaluation**

A comparative evaluation [10] was carried out to examine users’ attitudes, impression and thoughts about both the Searchling and T-Saurus user interfaces. As part of this study participants were asked to categorize themselves as either visual learners or linear thinkers. This decision was made to examine which user interface would be evaluated positively by the two categories of users. Twenty-five participants were recruited for this study by purposive, maximum variation snowball sampling. Although the participant pool included students and faculty members across various departments, multilingual volunteers – particularly those from the Department of Modern Languages and Cultural Studies – were specifically targeted throughout the recruitment process. The resulting participant pool was diverse, comprising professors and graduate and undergraduate students. This study used a wide range of data gathering tools, including a pre-test, post-test and usability questionnaires; interviews; audio, video and screen capture; the think-aloud technique; and direct observation. Three search tasks were designed to allow participants to interact with the interfaces, thesauri and query formulation mechanisms.

The results show that the visualization in both interfaces was found to be comprehensible to users. The Searchling interface was found to be more favorable and easier to use in terms of multilingual features, thesaurus and search functions, as well as users’ motivation to use such an interface for research purposes. Though T-Saurus was preferred by fewer users than Searchling, the most promising finding for T-Saurus is that it has the potential not only to support browsing, searching and query formulation, but also to transform these processes. It was found that linear thinkers prefer Searchling, whereas visual learners like T-Saurus. Searchling is a linear, sequential and visual interface that uses faceted structure as its default interface, and the thesaurus terms such as more general, more specific and related terms are shown automatically as soon as the user selects a term. T-Saurus, on the other hand, provides users with a more interactive and dynamic visualization interface where users need to interact with and choose the individual thesaurus term relationships to be shown.
Conclusion

Our research into the design, development and evaluation of a wide range of experimental visual browsing interfaces, including the two detailed above, has demonstrated that users’ cognitive and interactive preferences and skills may influence how they evaluate visualization user interfaces and environments. In particular, visualizing such textual information as thesauri can provide alternative ways of interacting with text and formulating queries. Within these two studies we showed that domain-specific knowledge organization systems, such as thesauri in the humanities and social sciences, can be effectively reused and repurposed to support information access and retrieval in semantically rich user interfaces that provide seamless support for searching, browsing and results displays. The approach taken in these projects can be extended to other kinds of thesauri as well as to other controlled vocabularies. There are many different domain-specific thesauri available in the areas of humanities and social sciences that can serve as sources of term selection, collection visualization and understanding, and query formulation.

Acknowledgements. The authors wish to acknowledge the Social Sciences and Humanities Research Council of Canada (SSHRC) for providing the funding support and our research assistants, Matt Bouchard, Mark Bieber, Ximena Rossello, Amy Stafford, Karl Anvik, Lindsay Doll and Paras Mehta, for their contributions to the above research projects.