

# CAIS Poster: SINCERITY: A new bilingual search engine for image retrieval in a bilingual context

Elaine Ménard (McGill University) and Jonathan Dorey (McGill University)

**Abstract:** This poster summarizes the results of the first and second evaluation phases of a search engine (SINCERITY) dedicated to image retrieval in a bilingual context (French and English). SINCERITY constitutes a definite benefit for image searchers unfamiliar with more than one language, by giving them user-friendly access to visual resources.

**Résumé :** Cette affiche présente les résultats de la première et deuxième phase d'évaluation d'un moteur de recherche (SINCERITY) dédié au repérage d'images en contexte bilingue (français et anglais). SINCERITY est avantageux pour les chercheurs d'images qui ne connaissent pas plus d'une langue, en leur donnant un accès convivial aux ressources visuelles.

This poster presents a research project that aims to develop an interface model for image retrieval in a bilingual (French and English) context, that is, when the query language differs from the indexing language. Some search engines already offer increasingly sophisticated search mechanisms and features; however, much remains to be done to ensure universal access to non-textual documents.

In the first phase of the study, a best practices review was completed in order to acquire knowledge of the existing image search features and functionalities and to assess how they could be integrated in the development of a bilingual search interface. An examination of 159 resources that offer image retrieval was carried out. This investigation provided valuable suggestions and guidelines for the eventual interface design (Ménard and Smithglass 2014).

Using the features and functionalities identified in the first phase of the study, a survey was developed and administered to two groups of participants: 20 English-speaking respondents and 20 French-speaking respondents. The questionnaire comprised 27 closed questions and 8 open-ended questions. Complementing the survey data, the research team did semi-structured interviews with 20 participants who were asked to identify and discuss the functionalities they usually employ to search for images. Descriptive statistics were used on the collected data, while the content of the open-ended questions and the interviews was analyzed and coded by two evaluators (Ménard and Khashman 2014).

The data analysis suggested several implications for the interface design. For example, it revealed the importance to keep the search interface as simple as possible. It also exposed that image searchers still have difficulty formulating an appropriate query to retrieve images. Consequently, we hypothesized that including a taxonomic structure to begin the search process would be a feature that improves the usability of the interface. Furthermore, the collected data underlined the lack of interest of the respondents in most content-based image retrieval (CBIR) methods (e.g., colours, textures, shapes). This explains why these techniques are not included in the new search interface. Finally, a majority of respondents expressed interest in retrieving images associated with words from different languages. As a result the interface model includes a mechanism that facilitates the retrieval of images indexed in two different languages (English and French).

Combined with the exploration of the best practices for image retrieval, the results obtained in the second phase of the project provided the foundation for the development of the search interface (SINCERITY – Search **IN**terfa**CE** for the **R**etrieval of Images indexed with a **T**axonom**Y**). The interface was developed using open source technologies: Apache Solr for indexing, Ajax Solr for the browser-based interface, and jsTree for the taxonomy navigation (Neugebauer and Ménard in press).

SINCERITY allows initiating the queries with keywords, but also with a browsing structure called TIIARA (**T**axonomy **I**mage **I**ndexing **A**nd **R**etrieval), a bilingual taxonomy designed for the indexing of

ordinary digital images. Taxonomies have recently emerged as powerful tools that provide helpful information for use by indexers or information specialists who are describing visual resources, from works of art to archival materials. TIARA already constitutes a successful tool that provides access to ordinary images with seven main categories. The number of top-level categories, as well as the depth of the taxonomic structure, was kept to a minimum in order to avoid frustration and increase the degree of satisfaction of the eventual users. TIARA presents great potential in providing image searchers with intuitive access points and constitutes an aid for cross-language retrieval, that is, when image searchers are not familiar with images indexed in English, which is still the dominant language of the Web.

A user testing was conducted to measure efficiency and effectiveness and to ensure that the final product is clear, comprehensive and consistent. This involves an exhaustive simulation of the retrieval process using SINCERITY to complete typical image retrieval tasks with a sample of 30 images and 60 participants. It is followed by a questionnaire in order to measure the degree of user satisfaction.

This poster will summarize the results of the first and second phases of the project and will present the preliminary results of the user testing. Once fully developed, SINCERITY is intended to be an innovative tool for image searchers who are looking for ordinary images. The main contribution of this project lies in bridging a gap for unilingual image searchers. The bilingual search interface will constitute a definite benefit for image searchers unfamiliar with more than one language, by giving them user-friendly access to visual resources.

## **References**

- Ménard, E. and N. Khashman. 2014. "Image retrieval behaviours: The real users are leading the way." *Library Hi Tech* 32 (1): 50-68.
- Ménard, E. and M. Smithglass. 2014. "Digital image access: An exploration of the best practices of online resources." *Library Hi Tech* 32 (1): 98-119.
- Neugebauer, T. and E. Ménard. In press. "SINCERITY: the making of a search engine for ordinary images." *OCLC Systems and Services* 31 (3).