

**Elaine Ménard**  
**Université de Montréal, Montréal, Québec, Canada**

## **Indexing and retrieving images in a multilingual world**

Extended abstract for the position paper submitted to the North American Symposium on Knowledge Organization (NASKO), Toronto, Ontario, 2007. Full paper to appear in *Knowledge Organization* vol. 34 no. 2 (2007).

**Extended abstract:** The Internet constitutes a vast universe of knowledge and human culture, allowing the dissemination of ideas and information without borders. The Web also became an important media for the diffusion of multilingual resources. However, linguistic differences still form a major obstacle to scientific, cultural, and educational exchange. With the ever increasing size of the Web and the availability of more and more documents in various languages, this problem becomes all the more pervasive. Besides this linguistic diversity, a multitude of databases and collections now contain documents in various formats, which may also adversely affect the retrieval process.

This paper presents the context, the problem statement, and the experiment carried out of a research project aiming to verify the existing relations between two different indexing approaches: (1) traditional image indexing recommending the use of controlled vocabularies or (2) free image indexing using uncontrolled vocabulary, and their respective performance for image retrieval, in a multilingual context. The use of controlled vocabularies or uncontrolled vocabularies raises a certain number of difficulties for the indexing process. These difficulties will necessarily entail consequences at the time of image retrieval. Indexing with controlled or uncontrolled vocabularies is a question extensively discussed in the literature. However, it is clear that many searchers recognize the advantages of either form of vocabulary according to circumstances (Arsenault, 2006). It appears that the many difficulties associated with free indexing using uncontrolled vocabularies can only be understood via a comparative analysis with controlled vocabulary indexing (Macgregor & McCulloch, 2006).

This research compares image retrieval within two contexts: a monolingual context where the language of the query is the same as the indexing language; and a multilingual context where the language of the query is different from the indexing language. This research will indicate if one of these indexing approaches surpasses the other, in terms of effectiveness, efficiency, and satisfaction of the image searchers. For this research, three data collection methods are used: (1) the analysis of the vocabularies used for image indexing in order to examine the multiplicity of term types applied to images (generic description, identification, and interpretation) and the degree of indexing difficulty due to the subject and the nature of the image; (2) the simulation of the retrieval process with a subset of images indexed according to each indexing approach studied, and finally, (3) the administration of a questionnaire to gather information on searcher satisfaction during and after the retrieval process. The quantification of the retrieval performance

Ménard, E. (2007). Indexing and retrieving images in a multilingual world. (extended abstract). *Proceedings of the North American Symposium on Knowledge Organization*. Vol. 1. Available : <http://dlist.sir.arizona.edu/1913>

of each indexing approach is based on the usability measures recommended by the standard ISO 9241-11, i.e. effectiveness, efficiency, and satisfaction of the user (AFNOR, 1998).

The need to retrieve a particular image from a collection is shared by several user communities including teachers, artists, journalists, scientists, historians, filmmakers and librarians, all over the world. Image collections also have many areas of application: commercial, scientific, educational, and cultural. Until recently, image collections were difficult to access due to limitations in dissemination and duplication procedures. This research underlines the pressing necessity to optimize the methods used for image processing, in order to facilitate the images' retrieval and their dissemination in multilingual environments. The results of this study will offer preliminary information to deepen our understanding of the influence of the vocabulary used in image indexing. In turn, these results can be used to enhance access to digital collections of visual material in multilingual environments.

## References

- Arsenault, C. (2006). L'utilisation des langages documentaires pour la recherche d'information. *Documentation et bibliothèques*, 52 (2): 139-148.
- AFNOR. (1998). *Exigences ergonomiques pour travail de bureau avec terminaux à écrans de visualisation (TEV) - partie 11 : lignes directrices relatives à l'utilisabilité*. Genève: Organisation internationale de normalisation.
- Macgregor, G., & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55 (5): 291-300.