

Ordinary Image Retrieval in a multilingual context: a comparison of two indexing vocabularies

Elaine Ménard, School of Information Studies, McGill University
elaine.menard@mcgill.ca

Abstract

Purpose

This paper examines image retrieval within two different 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 study also compares two different approaches for the indexing of ordinary images representing common objects: traditional image indexing with the use of a controlled vocabulary and free image indexing using uncontrolled vocabulary.

Design/methodology/approach

This research uses three data collection methods. An analysis of the indexing terms was employed in order to examine the multiplicity of term types assigned to images. A simulation of the retrieval process involving a set of 30 images was performed with 60 participants. The quantification of the retrieval performance of each indexing approach was based on the usability measures, that is, effectiveness, efficiency and satisfaction of the user. Finally, a questionnaire was used to gather information on searcher satisfaction during and after the retrieval process.

Findings

The results of this research are twofold. The analysis of indexing terms associated with all the 3,950 images provides a comprehensive description of the characteristics of the four non-combined indexing forms used for this study. Also, the retrieval simulation results offers information about the relative performance of the six indexing forms (combined and non-combined) in terms of their effectiveness, efficiency (temporal and human) and the image searcher's satisfaction.

Originality/value

The findings of this study suggest that in the near future, the information systems could benefit from allowing an increased coexistence of controlled vocabularies and uncontrolled vocabularies resulting from collaborative image tagging, for example, and giving the users the possibility to dynamically participate in the image indexing process, in a more user-centred way.

Introduction

For a few years, the diffusion of images has increased, mainly because of the development of digital technologies and the unprecedented growth of the Web. Confronted by this profusion of images, individuals now speculate on how to retrieve images successfully. Although it has recently become easier to retrieve images with, among other things, the use of specifically designed search engines, successful image retrieval still presents a significant degree of difficulty. Indeed, the complexity in retrieving images with precision is clearly apparent to anyone who has ever searched for images on the Web. Unlike text-based documents, images do not have full-text data that can be used for retrieval purposes. As such, the selection of appropriate indexing terms to describe the image, whether in a monolingual or a multilingual retrieval context, is crucial for their retrieval. Among the many types of images available, ordinary images (i.e., images of common objects such as a bicycle, a hammer, a coffee cup, etc.) occupy an important part of Web users' searches. Developing effective means to access these

images is becoming increasingly necessary as more individuals and organizations are interested in ordinary images they find on the Web.

Two problems tend to arise when searching for images on the Web using a textual query. First, the image searcher needs to transpose or transfer the image's visual representation into a textual description (Jansen, 2008). This operation is hazardous due to the tendency of having multiple interpretations of the visual content of the images; this leads to ambiguity and error. The difficulties relating to the use of word-based systems to retrieve images are numerous and are presently without an effective solution to "translate" images into words. Consequently, the image searcher will not necessarily describe the same image with the same concepts or words. The second critical stage of image retrieval concerns the great linguistic diversity that exists on the Web. Yet given this multiplicity, one would expect a similar variety in the languages of texts associated with the images. When using a textual query, the image searcher faces a double challenge when trying to retrieve images: the query must correspond to the text associated with the images and the language of the query must also match the language of the text associated with the images.

In this study, two linguistic contexts are compared: 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 also examines two different approaches for the indexing of ordinary images representing common objects. The approaches are the traditional image indexing with the use of a controlled vocabulary and the free image indexing using uncontrolled vocabulary.

Current State of Image Indexing and Retrieval

In general, two types of systems are used for the indexing and retrieval of images: systems based on the content or the physical characteristics of the image (Content-Based Image Retrieval – CBIR) and systems focused on the context or description of the image (Concept-Based Image Retrieval). With CBIR systems, images are associated to the values of certain parameters such as colour, texture and form. When searching for images with CBIR systems, users can initiate the retrieval process using a sketch or a similar image and thus avoid the constraints of formulating the query in words (Boudry and Agostini, 2004). However CBIR systems are complex to develop and use. For now, methods of indexing and retrieval using the physical characteristics of the image are still in the experimental stage (Kherfi, Ziou and Bernardi, 2004). Therefore, indexing using the textual description of the image is still the most widely used technique.

In general, metadata schemas used for image indexing suggest the use of controlled vocabularies for some elements used for the description of the images. Among the various types of controlled vocabularies that can be used for indexing are classification schemes, subject headings lists and thesauri. A classification scheme is a documentary language built using linguistic signs or symbols. This language is based on the structuring of topics into classes. The classes and the relations between them are usually represented by indices. Traditionally, classification schemes are most often used in processing monograph resources. However, these schemes may also be used for indexing documents such as those based on digital imaging. Among the major document classification schemes are the Library of Congress Classification (LCC), the Universal Decimal Classification (UDC) and the Dewey Decimal Classification (DDC). In addition to the general schemes, other classifications specifically dedicated to visual material have also emerged. Among them is the Iconclass system, an international classification system that can be used to describe visual material.

A subject heading list is “a controlled vocabulary of terms in natural language that are designed for both precoordination and postcoordination” (Chu, 2005). Among the popular general subject heading lists, we have the Library of Congress Subject Headings (LCSH), the Medical Subject Headings (MeSH), the Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU), the Canadian Subject Headings (CSH) and the Répertoire de Vedettes-Matière de l'Université Laval (RVM). Although there are many lists presently available, subject headings lists are not extensively used for image indexing (Jørgensen, 2003).

Thesauri are controlled vocabularies created with terms extracted from the natural language and designed specifically for post-coordination searches. Thesauri help reduce problems caused by natural language such as polysemy and synonymy (Hudon, 2006). Thesauri can be monolingual or multilingual with several being presently used to describe digital images. For example, the Art and Architecture Thesaurus (AAT) is a controlled vocabulary for describing and indexing cultural heritage. The AAT offers traditional thesaural relations (equivalence, hierarchical and associative) but also the semantic relations based on the logical relationships between the concepts, activities and objects. Another example of a commonly used thesaurus for indexing visual material is the Thesaurus for Graphic Materials (TGM). This thesaurus was originally developed for the Library of Congress to catalogue and identify any visual material (prints, photographs, drawings, moving images, etc.) whether it was part of a book, a manuscript or a visual collection.

Finally, other documentary languages may be used for indexing images, including authority lists and visual dictionaries (Jørgensen, 2003). The major authority lists used for image indexing include the Union List of Artist Names (ULAN), the Authority File of the Library of Congress (LC), the controlled vocabularies of the American Library Association (ALA) and the Multilingual Glossary for Art Librarians of the International Federation of Library Associations (IFLA). With regard to visual dictionaries, even if they may not be formally considered as controlled vocabularies due to their lack of hierarchical structure or equivalence, they can still be used to describe ordinary images. However, visual dictionaries offer two significant benefits for image indexing. First, unlike traditional dictionaries, which require that the user be first familiar with the words to be used, the illustrations contained in visual dictionaries directly provide keywords for an idea or an image. Second, this type of dictionary is often offered in a bilingual or multilingual format, which is rarely the case of most usual controlled vocabularies.

In general, controlled vocabularies facilitate the indexing process and offer many advantages for retrieval, browsing and interoperability. Nevertheless, controlled vocabularies also present some weaknesses, the main one being that they often represent concepts in an artificial manner. Indeed, the indexing terms offered by controlled vocabularies generally have few linkages with the terms commonly used by the image searcher (Furnas et al., 1987). Another disadvantage of controlled vocabularies is that they quickly become obsolete and neologisms take a long time to enter controlled vocabularies. Finally, not only is the use of these vocabularies a complex task for most indexers, the most commonly used controlled vocabularies for indexing images are available only in English. Therefore, an indexer with little knowledge of English who wishes to use these vocabularies will face a significant language problem unless an effective translation mechanism is provided.

Consequently, image indexing with uncontrolled vocabulary constitutes an alternative to the traditional indexing with controlled vocabularies. Uncontrolled vocabularies are words or keywords freely chosen by either the searcher or the indexer. They are not validated by an authoritative list of terms (Arsenault, 2006). For example, we encounter this kind of free indexing in photo sharing systems we encounter on the Web, such as Flickr. These systems allow massive

image storage and diffusion on the Web. The users upload their own images and index them using their own terms. The indexing terms found in such systems can be drawn from a single language or a combination of several languages, bringing a distinct advantage over controlled vocabularies usually written in one language. However, using uncontrolled vocabularies for image indexing also presents several disadvantages. For example, certain ambiguities often arise when the same keywords are employed by many individuals, but in different contexts. Moreover, this lack of vocabulary control results in a variety of keywords often pointing to the same concepts.

The use of controlled vocabulary or uncontrolled vocabulary leads to a number of difficulties at the time of indexing that impact directly on the image retrieval. Literature relating to image indexing highlights the many advantages and disadvantages of both indexing approaches, with controlled and uncontrolled vocabularies, as well as the combination of these two approaches. Numerous studies have examined the advantages of both indexing approaches but none reach a clear consensus on which approach is superior. Two specific indexing tendencies can be outlined: the uncontrolled indexing approach through the concept of collaborative indexing and the value of combining the two types of vocabularies for image indexing (Matusiak, 2006; Macgregor and McCulloch, 2006; Enser et al., 2007).

In a multilingual retrieval context, we can conjecture what indexing approach, with controlled or uncontrolled vocabulary, is preferable for ordinary image indexing in order to facilitate their retrieval in a multilingual context. This crucial question has been little studied until now. The image indexing process remains a sensitive issue that deserves special attention in order to achieve effective and efficient image retrieval in a multilingual context, as well as to obtain a superior degree of image searcher satisfaction. The purpose of the present research was to verify the differences between the characteristics of these two indexing approaches and their respective performance for retrieval of ordinary images representing common objects. Two specific objectives were established: to identify the characteristics of each of the two image indexing approaches and to expose the differences in terms of effectiveness, efficiency and searcher satisfaction when retrieving ordinary images indexed with these approaches. This research examined if either of these indexing approaches outperformed the other for image retrieval in a multilingual context.

Methodology

This research compared image retrieval within two contexts: a monolingual context, that is, where the language of the query (French) was the same as the indexing language (French), and a multilingual context, that is, where the language of the query (French) was different from the indexing language (English). As a first step, a database containing ordinary images extracted from a commercial online catalogue representing common objects was constructed. A total of 3,950 images were selected and then indexed according to four indexing forms: with French controlled vocabulary, with French uncontrolled vocabulary, with English controlled vocabulary and with English uncontrolled vocabulary. For the indexing process using the controlled vocabulary, the indexer was asked to use the *Nouveau dictionnaire visuel multilingue* (Corbeil and Archambault, 2003). Even if this dictionary cannot be formally considered an authentic controlled vocabulary, it was chosen for three reasons. First, this dictionary contains words and images of common objects similar to the type of images included in the database. Second, it provides a form of term standardization allowing maximum control on regionalisms, archaisms and anglicisms. Finally, this dictionary offers both French and English terms (in addition to Spanish, German and Italian terms), which facilitated the indexing process while maintaining a

form of interlinguistic consistency. The *Nouveau dictionnaire visuel multilingue* was used for both French and English controlled vocabulary indexing. The process of indexing proceeded as follows: the indexer was shown the image to be indexed and then located the best corresponding image in the dictionary. The French and English terms associated with the image of the dictionary were then used as indexing terms. Between one and five controlled indexing terms could be assigned to each image. For the indexing process with uncontrolled vocabularies (French and English), the indexers also saw the images and then used their own words as indexing terms. No limit was imposed on the number of indexing terms to assign to one image.

Then, the 16,920 terms assigned to the images during the indexing process were examined in order to identify the key characteristics of the four indexing forms being studied. An analysis grid was developed and used for each of the four indexing forms. This grid consisted of three groups of attributes (terminological, perceptual and structural) and highlighted the similarities and differences between each one. The terminological attributes were related to lexicographical aspects of the indexing terms. The perceptual attributes included the attributes associated with the physical, functional and visually identifiable aspects of the images. Finally the structural attributes specifically identified the relationships between words.

In a second step, a retrieval simulation of 30 images randomly extracted from the database was performed. During this experiment, 60 participants were presented each of the 30 images, in the same order of presentation. Each participant was randomly associated with one of the six indexing forms established for this study, that is, one of the four forms mentioned, in addition to the combination of French controlled and uncontrolled vocabularies and the combination of English controlled and uncontrolled vocabularies. Four retrieval performance measures were used for this research: the effectiveness of image retrieval as measured by the success rate calculated with the total number of retrieved images; the temporal efficiency measured by time, in seconds, as determined by the time needed to retrieve an image; human efficiency as measured by human effort represented by the number of queries formulated for each image retrieved; and the satisfaction of the image searcher after each one of the retrieval task.

Finally, to complement the data collected during this simulation, a questionnaire to gather the impressions of the participants with regard to the retrieval process and performance was developed. The questionnaire was submitted to participants in two stages: after each retrieval task, the image searcher was asked to evaluate his degree of satisfaction with this particular task; and a questionnaire containing closed- and open-ended questions was given to the participants once all the retrieval tasks were complete.

Results

The results of this research are twofold. The analysis of indexing terms associated with all the 3,950 images provided a comprehensive description of the characteristics of the four non-combined indexing forms used for this study. Also, the retrieval simulation results offered information about the relative performance of the six indexing forms (combined and non-combined) in terms of their effectiveness and efficiency (temporal and human), and the image searcher satisfaction.

Analysis of Indexing Terms

The analysis of indexing terms used to describe images representing common objects reveals that both the controlled and uncontrolled vocabulary indexing approaches differ from the terminological, perceptual and structural perspectives.

For the terminological analysis, five types of indexing terms have been considered: single terms (one word/one single concept, e.g., cup), multiple terms (several words/one single concept, e.g. tennis racket), complex terms (several words/several concepts, e.g., headphones with radio and antenna), abbreviations (e.g., GPS) and neologisms (e.g., home theatre). Two indexing levels have been considered: the terms of a generic nature (e.g., fence) and the terms of a specific nature (e.g., snow fence). Our analysis revealed that the indexing approach with controlled vocabulary diverges from the indexing approach with uncontrolled vocabulary by a larger allocation, on average, of indexing terms to a single image. Furthermore, indexing using the controlled vocabulary offers proportions of single/multiple terms and specific/generic terms which are quite similar. For the perceptual attributes, the indexing approach with controlled vocabulary offers less variety than the indexing approach with uncontrolled vocabulary in the choice of attributes assigned to images. With regard to the structural relations, the indexing approach with controlled vocabulary does not generally use these relations to any great extent. For its part, the indexing approach with uncontrolled vocabulary uses more specific and complex terms.

On the perceptual point of view, four groups of attributes were analyzed. The first group consisted of the physical attributes of the object depicted on the image: the terms referring to the form (e.g., square plate), dimensions (e.g., big), colour (e.g., blue cup), texture (e.g., knitted hat), equipment (e.g., wood hanger) and the parts of the body (e.g., foot bath). The second group is related to functional attributes of the objects: the function (e.g., cooking utensils), the result/product (e.g., pleasure garden), activity/action (e.g., blinking light) and location (e.g., garden table). The third group consists of the attributes of objects identifiers represented: the trademark (e.g., MotoMaster), the product name (e.g., Fridgidaire), the sponsor (e.g., Nascar T-shirt) and location (e.g., marble from Italy). A final group of perceptual attributes identified indexing terms with a temporal or chronological element (e.g., Christmas plate), numerals (e.g., ink cartridge HP 78), other types of adjectives not covered by the categories already mentioned (e.g., electronic) and attributes related to people (e.g., child seat). As for the perceptual attributes, the indexing approach with uncontrolled vocabulary dominates the controlled vocabulary in the allocation of a majority of physical attributes, identifiers, functions and other types of perceptual qualities.

Finally, the analysis of the structural relationships took into account four types of relations: generic relations (e.g., desk chair), partitive relations (e.g., toilet seat), instance relations (e.g., Michelin tire) and associative relations (e.g., curling iron and brush). Our analysis indicates that the indexing approach with uncontrolled vocabulary has more relations than the indexing approach with controlled vocabulary.

Furthermore, even if the examined indexing approaches differ from one another, language differences within the same indexing approach (with controlled or uncontrolled vocabulary) were generally not very noticeable. Indeed, the analysis of indexing terms highlighted that the characteristics of the French and English controlled vocabulary approaches and the French and English uncontrolled vocabulary indexing approaches are quite similar. However, the few observed differences between the indexing terms assigned in French and English were generally more pronounced with the indexing approach using controlled vocabulary.

Five factors having an impact on the characteristics of the indexing terms have been examined: the indexing vocabulary, the indexing language, the indexing policy, the type of images and the choice of indexers. Our analysis reveals that the indexing vocabulary is the most influencing factor, while the language seems less important. In addition, the impact of these factors was mostly observed with the perceptual attributes. Finally, it should be noted that the indexing terms attributed to the images are the result of the combined influence of several of

these factors. In short, this analysis of the indexing terms has concluded that the two indexing approaches of this study are fundamentally different. With this, it was allowed to expect different results at the time of retrieval, since it was based on the matches to be made between the indexing terms and the query terms.

Image Retrieval Simulation

The results of the retrieval simulation confirm that the retrieval of ordinary images representing common objects is more effective (proportion of retrieved images) when they are indexed with the approach combining the controlled and uncontrolled vocabularies rather than an approach using only the controlled vocabulary (approximately 22% more effective) or only uncontrolled vocabulary (approximately 20% more effective). This result suggests that, on the one hand, the increased effectiveness of retrieval is due to the combination of the advantages of both indexing vocabularies and, on the other hand, the disadvantages of both vocabularies once combined have a lesser impact on the retrieval effectiveness.

The results of this study indicate that the differences in terms of temporal efficiency (time required to retrieve an image) are not significant, whereas for human efficiency, the results indicate that the indexing approach with controlled vocabulary is more efficient (queries needed to retrieve an image) than the uncontrolled vocabulary indexing approach. In the context of this study, the use of a controlled vocabulary, offering a maximum of consistency for the representation of concepts contained in the image, has a positive effect on the human efficiency. However, the combination of the two vocabularies, while improving effectiveness by providing more opportunities for matches between the query terms and the indexing terms, also increases the number of displayed images, resulting in a loss of temporal efficiency.

With regard to the satisfaction of the image searchers, the results of our analysis confirm that image searchers are more satisfied when images are indexed with the combined approach rather than the indexing approach using controlled vocabulary. Even if it remains a subjective measurement, the evaluation of the image searcher satisfaction reinforces the results obtained with the retrieval effectiveness and efficiency with the image searcher perspective.

The results of the retrieval simulation also reveal that the retrieval is more effective (superior performance of approximately 25%) and more efficient (in terms of queries needed to retrieve an image) in the monolingual context rather than the multilingual context. Furthermore, image searchers are more satisfied when the retrieval is done in a monolingual context rather than in a multilingual context. In general, the semantic ambiguity of translation explains this relatively inferior retrieval performance of the multilingual context, in terms of effectiveness, human efficiency and satisfaction of the image searcher, when compared to the retrieval performance in the monolingual context.

Discussion

One of the main outcomes of this study revealed that retrieval of ordinary images representing common objects should be based on pre-iconographic attributes such as physical attributes and identifiers related specifically to the description of the object illustrated on the image (colour, trademark, etc.). In addition, the indexing of the ordinary images must promote the inclusion of the object's function to increase the probability of their proper retrieval. This research also fills an important gap in the literature on ordinary images. This category of images has been little studied in the past while other types of images, such as artistic or documentary images, have been the subjects of several studies. Nevertheless, the ordinary images of common objects are voluminously represented in information systems.

The results of the retrieval simulation also highlighted that an indexing approach combining the controlled and uncontrolled vocabularies could be the best solution in terms of effectiveness and image searcher satisfaction. However, the combined indexing approach did not perform as well with regards to efficiency. Therefore, if the goal of the searcher is to retrieve images with as little effort as possible, we can speculate if the combined approach for ordinary images indexing is the best solution to adopt. In the same manner, the comparison of the performance of the image retrieval indexed with a controlled or uncontrolled vocabulary did not clearly show that either one of the two indexing approaches could outperform the other, except in the case of human efficiency, where the indexing approach with controlled vocabulary surpassed the indexing approach with uncontrolled vocabulary. Therefore, this study did not formally decide in favour of either one of these two indexing approaches for ordinary images representing common objects.

This study also drew attention to image retrieval in a multilingual context as well as to its inherent difficulties. While this research revealed little variation in the way of indexing from one language to another where the same indexing approach is used, the fact remains that retrieval performance largely depends on the linguistic context, either monolingual (direct retrieval) or multilingual (retrieval following the query translation process). With the exception of the results of the temporal efficiency, which showed no superiority of one language over the other, the results of this study regarding the effectiveness, human efficiency and image searcher satisfaction highlighted the need to consider how to optimize the multilingual retrieval process. In addition, the differences between the two linguistic contexts emphasized that the slower retrieval performance in the multilingual context can largely be attributed to the semantic ambiguity resulting at the time of the translation. However, for the moment, no translation technique seems to be completely able to solve this problem. It is important, therefore, to continue the research to develop a consistent linguistic resource to be used in multilingual information retrieval systems (including machine translation mechanisms) in order to make the retrieval process as successful in the multilingual context as in the monolingual context.

Conclusion

The analysis of the indexing terms has demonstrated that both indexing approaches, with controlled and uncontrolled vocabulary, are inherently different from one another. However, in terms of the indexing language, the analysis of indexing terms confirms that the characteristics of the approaches are quite similar within the same indexing approach, regardless of the indexing language. Differences, if they exist, manifest more in the indexing approach using uncontrolled vocabulary. Depending on the circumstances, the two indexing approaches could coexist. Indeed, uncontrolled vocabulary indexing should not necessarily be considered as an alternative or a solution to replace traditional indexing using controlled vocabulary but rather as an opportunity for improvement. The results of this study suggest that in the near future, information systems could eventually benefit from the combined use of controlled and uncontrolled vocabulary generated by, for example, a collaborative indexing process. Moreover, the dynamic collaboration of image searchers in the indexing process could be a future solution leading to more effective, efficient and satisfactory image retrieval, both in a monolingual and a multilingual context.

Finally, the results of this study are closely related to the image type used. Thus, one can assume that the study of indexing and retrieval of other image categories could adequately complement the results of this research. The analysis of indexing terms of documentary images, such as images related to a specific area (sports news, medical imaging, etc.) or artistic images

(museum objects, famous works, etc.), seems to be a crucial research avenue if a better understanding of the best indexing approach to adopt is to be obtained and the necessary suggestions to improve the organization and, eventually, the retrieval of all image types are to be made.

References

- Arsenault, C. (2006), "L'utilisation des langages documentaires pour la recherche d'information", *Documentation et bibliothèques*, Vol. 52 No. 2, pp. 139-48.
- Boudry, C. and Agostini, C. (2004), "Étude comparative des fonctionnalités des moteurs de recherche d'images sur Internet", *Documentaliste – Science de l'information*, Vol. 41 No. 2, pp. 96-105.
- Chu, H. (2005), *Information representation and retrieval in the digital age*, Information Today, Medford, NJ.
- Corbeil, J.-C. and Archambault, A. (2003), *Le Nouveau dictionnaire visuel multilingue*, Québec Amérique, Montréal.
- Enser, P. et al. (2007), "Facing the reality of semantic image retrieval", *Journal of Documentation*, Vol. 63 No. 4, pp. 465-81.
- Furnas, G.W. et al. (1987), "The vocabulary problem in human-system communication", *Communications of the ACM*, Vol. 30 No. 11, pp. 964-71.
- Jansen, B.J. (2008), "Searching for digital images on the web", *Journal of Documentation*, Vol. 64 No. 1, pp. 81-101.
- Jørgensen, C. (2003). *Image retrieval: Theory and research*. Scarecrow Press, Lanham, MD.
- Hudon, M. (2006). "Le thésaurus : au carrefour des sciences de l'information et de la terminologie", in Mustafa El Hadi, W. (Ed.), *Terminologie et accès à l'information*, Hermès Science, Paris, pp.71-98.
- Kherfi, M.L., Ziou, D. and Bernardi, A. (2004), "Image retrieval from the World Wide Web: issues, techniques, and systems", *ACM Computing Surveys*, Vol. 36 No. 1, pp. 35-67.
- Macgregor, G. and McCulloch, E. (2006), "Collaborative tagging as a knowledge organisation and resource discovery tool", *Library Review*, Vol. 5 No. 5, pp. 291-300.
- Matusiak, K.K. (2006), "Towards user-centered indexing in digital image collections", *OCLC Systems & Services*, Vol. 22 No. 4, pp. 238-298.