

# Image Indexing: How Can I Find a Nice Pair of Italian Shoes?

by Elaine Ménard

Folksonomies

From time immemorial the image has been a communication tool. Images have a multifaceted and very real value. Images play a double role. They serve not only as sources of information but, with the development of more accurate visualization techniques, they also enhance the understanding of that information. In recent years the diffusion of images has increased, mainly because of the development of digital technologies and the unprecedented growth of the World Wide Web. It would be impossible to enumerate all the types of images we encounter on the web; however, we can mention visual collections (paintings, prints, engravings, illustrations), drawings, charts, postcards, photographs (historical, botanical, police, medical, documentary, personal and familial, artistic), as well as images generated by computers, among many examples. Actually, the digital image is an integral part of our daily reality. We now scrutinize image search engines in the same way we consulted encyclopedias or illustrated dictionaries a few years ago. There is nothing easier than to discover a close up of Paris Hilton's latest arrest or a picture of her dog Tinkerbell on the web. Students will search image databases to illustrate a research paper, potential travelers will check travel resorts to have a better idea of the destination of their dreams, sportswriters will enrich their articles with evocative pictures of Maria Sharapova and so on. Google claims in 2007 that their users now have full access to more than 2 billion images.

Elaine Ménard is a Ph.D. candidate in the École de bibliothéconomie et des sciences de l'information at Université de Montréal; C.P. 6128, succ. Centre-ville, Montréal (QC) H3C 3J7, Canada. She can be reached by email at [elaine.menard@umontreal.ca](mailto:elaine.menard@umontreal.ca); or at her website at [www.gin-ebis.umontreal.ca/menarde/](http://www.gin-ebis.umontreal.ca/menarde/)

Confronted with this profusion, individuals now speculate on how to retrieve images with efficacy and efficiency. In general two categories of queries are used to retrieve images on the web: graphic queries or textual queries. In the first category, the individual submits a graphic query (using an image or a drawing), and the system tries to retrieve a similar image by using certain physical characteristics of the image such as color, shape or texture. However, since such content based-image retrieval systems (CBIR systems) have many limitations, the majority of image searches on the web still use textual queries, and the retrieval's success depends on the match between the query terms and the text (ancillary text or indexing terms) associated with the images.

Consequently, to retrieve an image the individual must first translate into words what he is looking for. This conversion is the first challenge. The second obstacle comes from the "language" of the image. By their very nature, images are considered to be language-independent resources. Nevertheless, the text associated with the images gives the image a linguistic status similar to any other textual document, which can significantly affect its retrieval. And given the great linguistic diversity existing on the web, we must expect that the text associated with images exists in many different languages. For example, if a user formulates a query in English, and the images to be retrieved are associated with English text, the cross-lingual problem does not arise. However, if an English query is used, and the associated text is Italian (or any other language different from the query language), the retrieval will not be possible unless the retrieval system includes a cross-language information retrieval (CLIR) mechanism which allows cross-language mapping between the query terms and the associated text.

### Image Indexing: The Case of the Ordinary Image

For many years a great deal of work has been devoted to visual resources such as the image. We usually distinguish three main categories of images: the artistic image, the documentary image and the ordinary image. The artistic image, which is defined as the representation of an artistic expression, is characterized by an elaborate documentary process that includes several levels of significance (pre-iconographic, iconographic and iconological). The documentary image is mainly found in historical files, news services and media files. Family photographs also belong to this second category. Processing this type of image consists primarily of adding legends or information whose objective is to identify the image within a specific collection. Finally, the ordinary image is the image generally being used for commercial purposes or as an illustration. When processing this particular type of image, we generally take little account of such descriptive or analytical metadata as the title or the author, but will rather consider the visual content of the image, that is, its subject. In other words, the ordinary image does not require in-depth processing. A categorization by class such as “animal” or “landscape” could be enough. Nevertheless the advent of the web has highlighted the pressing need to acquire suitable tools for describing ordinary images, since we find them to be the majority of the resources available on the web – personal pages, blogs, virtual libraries, museum collections, services and product catalogues, and governmental information all often fall into this category.

Recently the needs and behaviors of image searchers have changed considerably. We actually perceive an evolution in the manner of formulating the queries for image retrieval. For example, queries containing a single term are less frequent than then they were when web searching began. Individuals tend to use more sophisticated search strategies while including more and more search elements such as proper names, trademarks or colors, in their queries. Image searchers also develop queries containing the relationships between these elements. Consequently, if the manner of searching images is evolving gradually, maybe it is time to consider whether the image indexing methods, and more particularly the controlled

vocabulary traditionally employed for the indexing process, is still well adapted to the real and current needs and behaviors of image searchers.

### The Traditional Way

When users pose textual queries, the success of the retrieval largely depends on the correspondence between the query and the text associated with the images. Since images do not always include a caption or any kind of ancillary text, the indexing process remains crucial. Image indexing has, so far, been divided between two camps: those who concentrate on *controlled* vocabulary and those who concentrate on *uncontrolled* vocabulary. The former focuses on assigning index terms extracted from thesauri, classification schemes or subject heading lists, while the latter focuses on terms drawn from natural language.

Over the years several excellent controlled vocabularies have been developed, varying in scope from the general to the very specific. The main purposes of these vocabularies are to help cataloguers find the right term to describe an object, to categorize concepts into broad topics and to generally improve the retrieval. For example, the *Art and Architecture Thesaurus (AAT)* is a controlled vocabulary intended for indexing physical objects. The AAT provides preferred terms for concepts related to art, architecture, decorative arts, material culture and archival materials. This vocabulary is largely used in museums, libraries, archives, visual resource collections and conservation agencies. Most terms included in the AAT are in English, but terms coming from other languages are progressively being incorporated in that vocabulary.

Two other examples of thesauri used for image description are the *Thesaurus for Graphic Materials I* and the *Thesaurus for Graphic Materials II (TGM I and TGM II)* created by the Library of Congress Prints and Photographs Division to support both cataloguing and retrieval needs. The TGM I is mostly used for subject indexing of graphical materials, including historical photographs, architectural drawings and artwork. Available only in English, the TGM I does not include proper names of people, organizations, events or geographic places. As for the TGM II (also available in English), it was created as a complement to the TGM I. This thesaurus provides headings

for categories of material by genre, vantage point, representation method, production technique, marking, shape or size, purpose, characteristics of the image's creator or publication status.

Another popular vocabulary for image description is *ICONCLASS*, which provides a different perspective. This subject-specific classification scheme is designed for the description and classification of visual resources collections. *ICONCLASS* includes a hierarchically ordered collection of definitions of objects, persons, events and abstract ideas that can be the subject of an image represented in various media such as paintings, drawings or photographs. Primary users of that classification scheme include art historians, researchers and museum curators. For the moment this classification scheme only exists in English, but translations in other languages are currently in progress.

The use of controlled vocabularies for image indexing offers many advantages for retrieval, browsing and interoperability. The control offered by these vocabularies is manifold. It manages the use of synonyms, homonyms, lexical anomalies and so on. However, one of the main disadvantages of controlled vocabularies is that they quickly become outdated. For example, neologisms will often take a long time to appear in controlled vocabularies. As a result, the search will be less accurate because controlled vocabulary will sometimes not allow a specific search. Furthermore, we must consider that the development and the management of controlled vocabularies involve significant costs. But the main difficulties associated with the use of controlled vocabularies for image indexing can be summarized as follows:

1. These vocabularies are not suitable for all image types and certainly not for the majority of ordinary images.
2. The use of the majority of controlled vocabularies is beyond the capacity of the non-expert or less trained professional.
3. Most of the controlled vocabularies exist only in one language (in most cases, in English), which implies that it will not be a great help in all linguistic contexts.

So what can we do to overcome the limitations of controlled

vocabularies? Well, it seems that web users have created their personal solutions by using their own methods to index images as they do with collaborative tagging, the latest trend in image indexing.

### The Latest Trend

Collaborative tagging has recently become very popular on many web services such as [del.icio.us](http://del.icio.us) or [CiteULike](http://CiteULike). Collaborative tagging began with any user assigning his or her own keywords to textual documents, but the same phenomenon was quickly implemented for image resources. Collaborative tagging is now the pillar on which photo sharing sites like [Flickr](http://Flickr) rest. These sites allow massive image storage and web diffusion. In these systems, users upload their own images and index them using their own terms (tags). It is also possible to make these images public; that is, the images can be seen by all users or by a group of people chosen by the system user, thus forming a vast and communal image database. In a Flickr-style system, the user who uploads images can thus determine who will have access to these images by stating certain rules of access control. In parallel other users of the system have the possibility to update the image indexing by adding other keywords or comments to any image they have access to. These annotations assigned by the uploader or by any other user of the Flickr system constitute a form of free indexing. It is this free indexing that is called "collaborative tagging." Obviously, this kind of indexing supposes that the individuals use their own words to describe images. Of course, they could choose to index with a controlled vocabulary, but why bother? Consequently, tags assigned by collaborative indexing generally contain a single term (for example, *house*, *Christmas*, *Lassie*). However, tags sometimes tend to be more descriptive (for example, "covered cat litter box," "SportRack bicycle rack," "black down filled jacket"). Instinctively, the users seem to include what they think is significant and imperative to employ if someone else needs to retrieve an image.

Collaborative tagging may therefore seem very seductive because of its close relationship with the real users and the way they see and describe things. Moreover, neologisms and all forms of newly created terms are quickly

integrated in collaborative indexing. Compared to controlled vocabularies and especially for new topics, collaborative indexing is likely to win hands down since the same words could take months, maybe years, before they are even considered for inclusion in a controlled vocabulary. Besides, in these image tagging systems, the indexing can be done in one language or in a combination several languages, which can ease the user's retrieval problem. However, despite its growing popularity and much like indexing with controlled vocabulary, collaborative tagging also presents several gaps. For example, some ambiguities emerge because the same keyword is often employed by several individuals, but in various contexts. In the same vein, the lack of synonym control results in the use of many different keywords to describe the same concept. Consequently, free indexing is often considered to be of poor quality.

To illustrate the difference between controlled and uncontrolled vocabularies used for image indexing, a database of ordinary images including 3,950 ordinary images drawn from the eight sections of an online

**FIGURE 1. An ordinary image and its indexing terms**

	<b>French Controlled Vocabulary</b>
	pneu à crampons
	<b>French Uncontrolled Vocabulary</b>
	pneu Tiger Paw
	<b>English Controlled Vocabulary</b>
	studded tire
<b>English Uncontrolled Vocabulary</b>	
Tiger Paw tire	

commercial catalogue was created for an in-progress study. Each image was indexed in four different ways (see Figure 1) – with controlled vocabulary (French and English) and with uncontrolled vocabulary (French and English). The indexing process using French uncontrolled vocabulary was carried out by a French-speaking indexer, while the indexing using English uncontrolled vocabulary was carried out by an English-speaking indexer. In order to reproduce the conditions of collaborative tagging, no fixed directive was given to the indexers concerning the number or the form of indexing terms to use for the image's description. For the controlled vocabulary indexing, a bilingual indexer used the *Nouveau dictionnaire visuel multilingue*. This dictionary contains appropriate terms (French and English) for the type of

images contained in the database and offers a form of standardization of the terms that allows a clear and precise identification of the objects while exerting maximum control on word variations.

Following the indexing process, the assigned indexing terms were examined. The objective of this analysis was to identify the specific characteristics of each indexing approach. A grid analysis was developed and applied to the complete set of indexing terms. Three levels of analysis were carried out: terminological, perceptual and interpretative. Preliminary results of this analysis revealed several similarities and differences between the two kinds of indexing.

First, we notice that in most cases, the indexers assigned only one indexing term to the particular type of image we found in this database. Second, among these indexing terms, we observed that uncontrolled vocabularies tended to use indexing terms containing multiple words (84% for French and 94% for English) where controlled vocabularies have a propensity to use uniterms (55% for French and 54% for English). However, perhaps the biggest difference emerging from this partial analysis is that uncontrolled indexing terms often include words referring to size, color, texture, gender or trademarks, contrary to controlled vocabularies which have a tendency to be less graphic and, in many cases, less detailed or descriptive. An extensive examination of the whole database is still in progress and should uncover other significant features differentiating the indexing terms assigned from controlled and uncontrolled vocabularies.

### Conclusion

The preceding analysis could be a good indication of how ordinary images really need to be retrieved. Furthermore, we must take into account that individuals may be interested in finding images indexed in languages other than English. For example, web users may be interested in buying objects from different parts of the world. And before they do so it is probable that they will want to see an image of the object. But they will have to retrieve the image first using some description of the object. As a result, image indexing remains a problem.

MÉNARD, continued

Whatever its virtues, however, there is no reason to consider collaborative tagging as a replacement solution to traditional indexing with controlled vocabulary. The reaction of some veteran information professionals to collaborative tagging is interesting. When this “new” indexing reality is mentioned, some look as if they had just ingested very bad medicine. Of course when you think about it, collaborative tagging is somewhat threatening for information specialists. In fact, there is not really another field that provides an example of how they should approach this usurpation of an important part of their work. We do not encounter this kind of phenomenon in other domains where anybody can walk in and say, “Hey, we think we can do a better job than you do!” It is difficult to imagine a lawyer or a physician confronted with a group of individuals coming up with their own solutions to solve a legal case or a medical diagnosis. But willingly or not, information specialists are now confronted by this kind of rivalry. There is certainly a valuable lesson to learn from collaborative tagging. According to circumstances, these two approaches may co-exist and be very helpful. In the near future we could see more and more information systems allowing the co-existence of controlled vocabularies and uncontrolled vocabularies resulting from collaborative image tagging. But would it be effective, especially if we consider all the images that are now available to the image searcher? We will probably find out soon enough. ■

## Resources

### Websites Mentioned in the Article

CiteULike – [www.citeulike.org/](http://www.citeulike.org/)

del.icio.us – <http://del.icio.us/>

Flickr – [www.flickr.com/](http://www.flickr.com/)

### Indexing Tools Mentioned in the Article

*Art and Architecture Thesaurus* –

[www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)

*ICONCLASS* – [www.iconclass.nl/](http://www.iconclass.nl/)

*Nouveau dictionnaire visuel multilingue* –

[www.ledictionnairevisuel.com/langues/index.htm](http://www.ledictionnairevisuel.com/langues/index.htm)

*Thesaurus for Graphic Materials I* – [www.loc.gov/rr/print/tgm1/](http://www.loc.gov/rr/print/tgm1/)

*Thesaurus for Graphic Materials II* – [www.loc.gov/rr/print/tgm2/](http://www.loc.gov/rr/print/tgm2/)