

# Large Scale Semantic Structures for Image Retrieval

Adrian Popescu  
CEA/LIST – LIC2M  
18 route du Panorama  
92260, Fontenay aux Roses, France  
+33146548013  
adrian.popescu@cea.fr

## ABSTRACT

In this paper, we describe our midterm PhD work that focuses on role of large scale semantic resources in image retrieval. Current Internet picture search engines are based on a raw processing of the text associated to the images and this leads to an important degree of imprecision in the results sets, to an unstructured presentation of the answers and to poor interactivity options. The introduction of semantic layers in the retrieval frameworks may enhance the quality of results obtained with existing systems. One of the main challenges in the field is to develop architectures that include semantics and fit the requirements of real-life applications, like the Internet search engines, and this challenge constitutes a central point in our approach. The expected outcome of our approach is building picture search applications that better fit users' expectancies.

## Categories and Subject Descriptors

**H.3.3 [Information Storage and Retrieval]:** Information Search and Retrieval – *retrieval models, search process.*

## General Terms

Algorithms, Design, Experimentation, Human Factors.

## Keywords

Ontology, CBIR, WordNet, Web.

## 1. INTRODUCTION

Existing Web image search engines have indexed a huge quantity of pictures and provide an important number of answers to a large diversity of queries. The answers are presented using chain matching for the text associated to the pictures and they are often not pertinent for the given query. The current approach to Web image retrieval is fast and robust but raises a number of problems:

- The answers proposed to the user are often not representative for the query because the parsed text does not describe the content
- The results are presented as a plain list and, when the number of returned images is high, the navigation becomes difficult
- With the notable exception of Ask, the major search engines do not propose any interaction possibilities on the answers page
- Although pictures are the object of the retrieval process, the search is text based and image processing techniques are only used to differentiate between black and white and color images

With the introduction of semantic structures in the image retrieval frameworks, it is possible to offer some solutions to the problems listed above. First, the textual indexes associated to images by the search engines can be exploited so as to reduce the noise in the answers sets if semantic structures like WordNet are used. The obtained advantages are: the disambiguation of polysemic queries – separate image sets are presented for the different senses of a term; the automatic reformulation of queries that allows the use of specialized concepts to represent general categories – a semantically structured presentation of results is obtained and the navigation in a plain list is replaced by a navigation between concepts; close categories to the initial query can be presented to the user using knowledge in the ontology.

Second, we consider that keyword and content based search are complementary and propose a dual access to pictures. After a keyword search, which implies an automatic reformulation destined to present results for subtypes of the current query, a content based retrieval inside the specialized classes is proposed.

We propose a lightweight architecture that works directly with Internet images and covers a hefty chunk of the queries expressed by the users [2]. We aim at proving that the introduction of semantic structures in image retrieval can be performed without placing a supplementary burden on the users.

The remainder of this paper is structured as follows: in Section 2, we discuss related work; Section 3 contains a presentation of the accomplished work and, in Section 4, some future directions we intend to follow are described.

## 2. RELATED WORK

An important effort research [3], [10], [11], [12] is directed toward similar goals to ours'. In [3], the authors describe a complex retrieval framework for Web images which employs WordNet to propose related queries but no sense separation is performed. In [10], a small scale and ad-hoc conceptual hierarchy is proposed to the user and enables the navigation in its structure. In addition, content based retrieval is proposed. In [11], a multimedia ontology describing a few hundreds concepts in the animals domains is used to improve Web image search. In [12], WordNet and CBIR techniques are employed but the approach heavily relies on relevance feedback, a technique that is often rejected in large public applications. There are important differences between the approach we propose and the ones mentioned above. Notably:

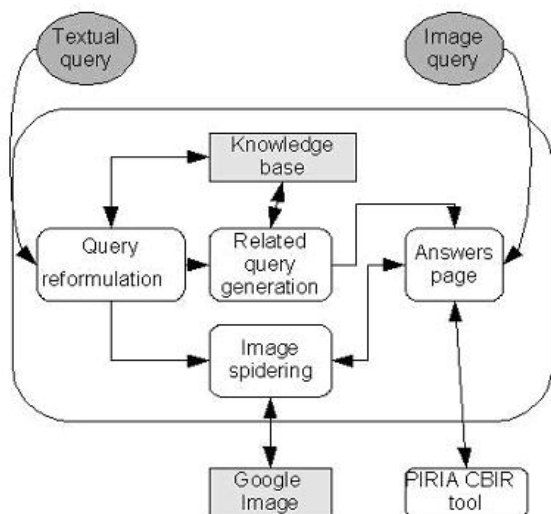
- The size of the semantic structures we employ is notably bigger than that of the resources used in similar approaches and we

work directly on real-life data, that is images that are collected on the fly from Internet search engines.

- The architectural choices we assumed so as to build a system that copes with very large scale databases, like the Web pictures dataset.
- The attention accorded to the way people search for pictures on the Internet and to the design of an retrieval scheme that is as simple as current ones but that proposes more functions.

### 3. REALIZED WORK

The work realized to date was concentrated on the use of WordNet [5] knowledge in keyword picture search [8] and the joint utilization of high and low level semantics for content based image retrieval [9]. In order to test the efficacy of our approach we built a prototype system whose main features are described in the following.



**Figure 1. Functional diagram of the image retrieval system. The user interaction parts are represented in the ellipses; the employed resources are drawn as rectangles and the active components of the application in rounded rectangles.**

Query disambiguation, synonymy treatment and query reformulation are performed using lexical knowledge extracted using WordNet from which we extracted the nouns hierarchy. When one types a query, the system first tests if the term is ambiguous. If so, images for the first WordNet sense (the most frequent) are presented but it is stated that the term has several senses and the user can see a page with different image sets for each sense. A good example of polysemy is *butterfly* which is both *an insect* and *a swimming style*. Results for both meanings of the term can be visualized separately. Images for synonym terms, like *polar bear* and *ice bear*, are presented in a same set as all terms point toward the same entity in the world.

Second, if the concept has subtypes, the most specialized among these last are used to launch queries in Google Image and retrieve images. A query for *bear* returns results for *grizzly*, *ice bear*, *polar bear* or *sloth bear*. The results are displayed as an ordered list of leaf concepts found under the initial query. The order is established using the frequency of each concept on the Web.

Thus, the navigation in a plain list is replaced by one among subtypes of the initial query.

The answers page equally contains suggestions for related image classes of three types: narrower concepts, terms having the same parent and more general categories. For *bear*, the system automatically proposes *brown bear*, *bruin* or *American black bear* as narrower queries; *canine* or *feline* as concepts having *carnivore* as parent; *carnivore*, *animal* or *organism* as extended queries.

A visual similarity search is proposed among the images that represent a leaf node in the hierarchy. The restriction of the search space for the CBIR process ensures that the retrieved images are both conceptually and visually similar to the query. This approach to content based search circumvents two important limitations of CBIR systems: the conceptual incoherency of the results and the access to large scale databases.

Currently, the system is able to answers to queries formed of one concept, which represent around a quarter of the total number of Web picture requests [2]. We recently realized a series of tests and the results are very encouraging as they show that, for keyword queries, the precision in the answers set obtained with our system outperforms that of Google by 16% on a panel of 40 concepts from different conceptual domains (*animals*, *plants*, *natural objects* and *artifacts*). These results confirm those obtained in [10]. An interactivity test was realized with 10 persons and it shows that they prefer the way results are presented in our system. The preference for a structured presentation of the answers is coherent with the results reported in [4]. As for the CBIR function, we compared our system and Cortina [6] which is, to our knowledge, the largest visual search engine. The test was performed on a panel of 80 image example queries belonging to 40 concepts. From a human's point of view, the answers we obtain are by far more relevant than those displayed by Cortina. This result is caused by the fact that we use both conceptual and visual parameters to obtain the answers.

Another line of work we followed was the creation of a multilingual conceptual hierarchy using aligned WordNets in several languages [7]. The ontology we built includes, besides English, Spanish and Italian. This resource is useful in order to perform multilingual disambiguation, to switch languages when there are not enough answers for a concept and to provide means for querying the system in other languages than English.

### 4. FUTURE WORK

Our PhD reached its midterm and we obtained encouraging results. Future work shall focus on aspects like:

- The extension of the semantic structure employed for image retrieval which will be performed in relation to usage trends in picture search and having in mind which are the conceptual relations that are useful in the given task. Search engines are often used to find images of celebrities [2] and there exist freely available sources of information, like Wikipedia, that can be exploited to structure relevant knowledge about famous people. We currently extract Wikipedia information corresponding to *singers* (around 6000 entries), *actors* (14000 entries), *models* (around 1000 pages), *footballers* and *politicians*. This knowledge will be used to complete WordNet and queries for pictures of celebrities will soon be supported. A second extension dimension is to complete the WordNet hierarchy with

conceptual relations that are potentially useful in image retrieval. Information from a large scale semantic map extracted from the Web [1] will be filtered so as to complete our ontology. Given a concept, close categories will be tested to see if they contain valuable information. A noise elimination procedure that uses initial knowledge in WordNet is proposed. For example, a location relation for an *animal* will be retained only if the assessed term is found under location in WordNet. Adapted relations are to be proposed for the different parts of the conceptual hierarchy.

- The treatment of more complex queries is another issue we want to tackle. This will be beneficial for the quantity of queries the system will be able to answer to. Similar reformulation procedures to the ones used for mono-term queries are to be tested for two and three terms queries, which, along with mono-term queries, represent the majority of image demands on the Web [2]. For example, when one types a query formed of two terms, it is possible to search for the most relevant WordNet subtypes of the two concepts and query for images associated to different combinations of these subtypes. Preference can be accorded to combinations that are semantically more plausible and ontological relations (as described above) can constitute a criterion for ranking the combinations of subtypes.
- The redefinition of the answers organization procedure. Subtypes and related classes are now presented using a statistical measure of their presence on the Web. The interaction tests we performed revealed that the users would prefer a presentation of results that accounts both for statistical measures and for the structure of the hierarchy. An example: a query with *animal* should be responded by presenting subtypes from different subdomains like: *mammals*, *fish*, *bird* or *insects* whereas currently *mammals* tend to be overrepresented.
- The introduction of additional image processing techniques in the framework. For example, it is possible to separate pictorial representations by production mode using algorithms that differentiate photographs, cliparts, scanned document and paintings [8]. This feature will allow the system to present separate picture sets for these different types of depictions. For photographs, it is equally possible to detect and filter images containing human faces. These last often represent a source of noise as they have associated text that is not correlated with the image content. Another dimension we want to explore is basic image segmentation that will separate pictures containing too small or too big representations of the desired concept from those containing well good size depictions.
- A final validation cycle of the proposed approach that includes larger scale evaluation of current system features as well as an assessment of the new functions we want to introduce. The tests should address the quality of the knowledge we use to extend the hierarchy, the quality of the answers for automatically reformulated queries, the organization of the answers and the effect of the introduction of more image processing techniques in the framework.

The main challenge we face is to ensure that, with the introduction of new features in the retrieval architecture, the usability in real-world conditions remains untouched. It is also important to ensure that the newly introduced functions improve the quality of the retrieved results when compared to existing approaches. A recurrent difficulty we encountered is the obtaining

of log files that contain real image queries addressed to search engines. The analysis of this type of information is crucial when one wants to characterize user behavior and interests in image retrieval. Finally, it is sometimes difficult to design sound tests that validate system features. This is especially true for interactivity tests, which are intrinsically subjective but necessary so as to test an application in real-world conditions.

## 5. REFERENCES

- [1] Grefenstette, G. Conquering Language: Using NLP on a Massive Scale to Build High Dimensional Language Models from the Web. In Proc of the 8<sup>th</sup> CICLING Conference (Mexico City, Mexico, Feb. 18 – 24, 2007), 35 – 49.
- [2] Jansen, B.J, Spink, A., Pedersen, J. An Analysis of Multimedia Searching on Altavista. In *Proc. of the 5th ACM SIGMM international workshop on Multimedia information retrieval* (Berkeley, CA, USA, 2004).
- [3] Joshi, D., Datta, R., Zhuang, Z., Weiss, W.P., Friedenber, M., Li, J., Wang, J.Z. PARAGrab: a comprehensive architecture for web image management and multimodal querying. In *Proc. of Very large data bases*, 32 (Seoul, Korea, 2006), 1163 – 1166.
- [4] Liu, H., Xie, X., Tang, X., Li, Z.W., Ma, W.Y. Effective browsing of web image search results. In *Proc. of the 6th ACM SIGMM international workshop on Multimedia information retrieval* (New York, NY, USA, 2004), 84-90.
- [5] Miller, G. A., Ed. WordNet: An on-line lexical database. *Int. Journal of Lexicography* 3, 4 (Winter 1990), 235-312.
- [6] Quack, T., Mönich, U., Thiele, L., Manjunath, B.S. Cortina: a system for large scale, content-based web image retrieval. In *Proc of. ACM Multimedia* (New York, NY, USA, 2004). 508-511.
- [7] Popescu A. Image retrieval using a multilingual ontology. In *Proc. of Recherche d'Information Assistée par l'Ordinateur* (Pittsburgh, USA, May 30 – June 1, 2007).
- [8] Popescu A., Grefenstette, G., Moëllic, P.A. Improving Image Retrieval Using Semantic Resources. To appear in *Springer Studies in Computational Intelligence*.
- [9] Popescu, A., Millet, C., Moëllic, P.A. Ontology Driven Content Image Retrieval. In *Proc of the ACM Conference on Image and Video Retrieval* (Amsterdam, The Netherlands, July 9 -11, 2007).
- [10] [cite Smith] Smith, J.R., Chang, S.F. Visually searching the Web for Content, *IEEE Multimedia*, 4, 3 (July-Sept 2000), 12-20.
- [11] Wang, H., Liu, S., Chia L.T., Does ontology help in image retrieval?: a comparison between keyword, text ontology and multimodality ontology approaches. In *Proc. of ACM Multimedia* (Santa Barbara, CA, USA, October 23 – 27, 2006), 109 – 112.
- [12] Yang, J., Wenyin, L., Zhang, H., Zhuang, Y. Thesaurus-aided Approach for Image Browsing and Retrieval, In *Proc. of IEEE International Conference on Multimedia and Expo* (Tokyo, Japan, 2001).