

SemRetriev – an Ontology Driven Image Retrieval System

Adrian Popescu
CEA/LIST - LIC2M
18 route du Panorama
92260 Fontenay aux Roses, France
+330146548013
adrian.popescu@cea.fr

Pierre – Alain Moëllic
CEA/LIST - LIC2M
18 route du Panorama
92260 Fontenay aux Roses, France
+330146549619
pierre-alain.moellic@cea.fr

Christophe Millet
CEA/LIST - LIC2M
18 route du Panorama
92260 Fontenay aux Roses, France
+330146548137
milletc@zoe.cea.fr

ABSTRACT

This paper describes the technical details of SemRetriev, a prototype system for image retrieval which combines the use of an ontology which structures an image repository and of CBIR techniques. The system models a real-world situation by including pictures gathered from the Internet and is designed for exploratory picture search. SemRetriev proposes two methods for retrieving images, using keywords and visual similarity and the ontology is useful in both cases. First, it is employed to reformulate queries and to render propose structured picture sets in response and second, it is used to propose CBIR processes in different subsets of the conceptual hierarchy. The user supplies a query term and the system furnishes images corresponding to the subconcepts of the queried term. It is then possible to narrow or extend the current search, to see detailed image sets or detailed images. It is equally possible to click any displayed image and obtain visually similar answers. The first experiments with SemRetriev give promising results and encourage us to continue the development of the system.

Categories and Subject Descriptors

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

General Terms

Algorithms, Experimentation

Keywords

Concept based image retrieval, content based image retrieval, semantics, ontology, WordNet

1. INTRODUCTION

There exist two main image retrieval (IR) paradigms: the keyword based search and the CBIR. SemRetriev is meant to show that the use of semantic resources (i.e. ontologies) can enhance the retrieval results for the two cited image search methods. We build here on previous work ([5], [6]) and describe a system that combines textual picture search and CBIR techniques.

In [5], we showed that, with the use of the knowledge in an ontology, the precision of image retrieval is fairly improved. Similar results were obtained in [cite Wang – does ontology]. In [6], we described a technique for controlling and ameliorating CBIR processes using a semantic resource.

Copyright is held by the author/owner(s).
CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.
Copyright 2007 ACM 978-1-59593-733-9/07/0007.

Related approaches to ours are described in [1], [9] or [10] but several important differences distinguish our work from those cited above. First, in SemRetriev, no machine learning techniques are employed. This aspect is important when one wants to scale up an IR system because the learning techniques are time consuming and often hard to generalize. Second, to our knowledge, no existing system employs an ontology to control the conceptual neighbourhood where CBIR is performed in the way described here.

As we mentioned, SemRetriev incorporates an ontology which is built by extracting the term hierarchy ranging under *placental* in WordNet [4] and separate the leaf synsets in this ontology. Picture sets are associated to leaf terms and these sets are indirectly related to more general concepts in the ontology using the hypernymy relation. A structured database is obtained and it is possible to respond to textual queries in the domain covered by the ontology. The answers are conceptually structured.

On the user side, the utility of a hierarchy which captures the way people organize entities in the world resides in that it provides intuitive explanation of the results. This feature is particularly important in CBIR, whose basic version calculates similarity considering solely low level image descriptors and the results are often disappointing from a human's point of view. Moreover, in the textual mode, picture sets which are related to the current query can be displayed.

The image processing components of the system address image filtering and indexing. Internet image classes are noisy and it is important to filter unwanted items. We observed that noisy pictures are often cliparts or contain human faces, and employ a filter bank which eliminates those individuals. A similarity matrix is constructed in order to provide visually similar pictures to a query image.

The first experiments with SemRetriev show promising results and this encourages us to refine and extend our ontology based approach to Internet image retrieval. In the following sections, we describe in more detail the elements of the SemRetriev architecture. Before concluding, we present implementation details and a typical walkthrough the system.

2. THE ONTOLOGY

We created a module that automatically extracts regions of the WordNet [4] nouns hierarchy. For demonstration purposes only, we consider the concepts ranging under *placental mammals*. WordNet concepts are structured following commonsense knowledge rather than scientific principles. Our application is directed towards a large public utilization and, with the use of a

commonsense semantic resource, it is probable to better fit users' expectancies. Other reason for using WordNet is that it is one of the largest conceptual hierarchies available and we can easily extend our approach to other domains. The conceptual structure currently includes 1113 nodes, with 841 leaf terms. The latter differ from other nodes in that they have associated picture sets. The depth of the *placental* hierarchy ranges from 1 to 8. For example, *livestock* is a terminal node which inherits directly from the root, while *milking short horn* or *Holstein* are leaf subconcepts of *placental* with the following intermediary nodes: *dairy cattle*, *cattle*, *bovine*, *bovid*, *ruminant*, *even-toed ungulate*, and *ungulate*.

The ontology is exploited in both retrieval methods proposed by SemRetriev. For example, a textual query for dog would result in answers organized in classes like: *pug*, *collie*, or *golden retriever* etc. In the CBIR mode, the conceptual hierarchy controls the area in the database where visually similar images are looked for. If one demands similar images to a particular representation of *Weimaraner*, the system can provide answers from the conceptual ontology subset which has as root: the concept itself, *hound*, *hunting dog*, *dog*, *canine*, *carnivore* and *placental*. These aspects are detailed in the walkthrough section.

3. THE IMAGE DATABASE

Our application runs on a picture repository constituted of Web images. The images were collected using the Ask¹ search engine. This application was used as it proved to render more precise image responses for queries with unique terms. For a set of 20 queries with familiar concepts and 50 images per query, the mean precision is around 80%, while Picsearch, Yahoo! and Google obtain respectively 70%, 63% and 56%. Leaf terms in the concept hierarchy were submitted to Ask so as to populate the image repository. The number of collected images was limited to a maximum of 150 per concept. We initially collected over 33000 images. After the elimination of invalid links and invalid files, there are around 31000 items left. The use of image filtering techniques reduces the number of items in the database to 25470, distributed over the 841 leaf nodes of the hierarchy.

The Web is a rich resource, but linguistic terms are unequally represented. The mean number of items in a class is 30.3, with a standard deviation of 23.8. The number of items associated to a terminal node varies between 0 and 147. The terms which do not have a pictorial representation are either rare ones, like *Pteropus capestratus* or *baronduki*, or secondary senses of known terms, like *doe* or *yearling*, for which the primary meanings do not belong to the *placental* hierarchy. Well represented nodes are related to familiar concepts like *hippopotamus* or *grizzly*. As for the precision of the images in the database, we performed a test on 600 images and found that, after filtering, the mean precision in the tested set is of 86%. This is to be compared with the 80% obtained when using a non-filtered evaluation set.

Image processing focuses on the elimination of noisy image categories. We notably identified as such cliparts and pictures containing human faces.

The algorithm we developed to separate photographs from cliparts and scanned texts is based on the shape of histograms. The

histogram for a clipart is more discrete and made of peaks, whereas the histogram for a photograph is more continuous.

The face detector we used is a multi-stage AdaBoost detector [8]. If the total area of the detected faces in a given picture is small (experimentally, we defined smallness as being less than five percent of the surface), we consider that it is a false positive, and thus ignore the detection and keep that item.

After filtering unwanted pictures, it is necessary to create a similarity matrix which will be used to render visually similar images. Due to speed limitations, the indexing step is currently performed offline. We index the dataset with the border/interior pixel classification algorithm proposed in [7]. It first quantizes each R, G and B component into 4 values. Then, pixels are classified into border pixels or interior pixels: a pixel whose 4 neighbours have the same quantized color is called interior, and border otherwise. Eventually, two 64 bins RGB histograms are built: one for border pixels, and one for interior pixels.

4. IMPLEMENTATION

The application is implemented employing Apache to host the application and PHP to create the Web interface. It is important to propose things that are not completely unknown to the users and SemRetriev includes an interface that is partially inspired by existing search engines. A maximum of 16 images are presented on each response page for the keyword mode. If the query addresses a non terminal node, the images belong to 4 (or less) leaf concepts under the current node. If the category has more than 4 terminal subconcepts, up to 10 response pages are presented.

The presentation order for leaf nodes is a function of the number of Web images that exist for each leaf class. If the query is a terminal node, at most 16 images are presented. The limitations of the number of answers are motivated by works like [2], which show that search engine users rarely go beyond the first answers pages. The same considerations motivate the choice of presenting a maximum of 20 pictures for the visual similarity mode. In this search mode, the user retrieves potentially interesting pictures he would normally ignore because they are not among the first answers. In order to facilitate the interaction, a box describing the possibilities offered to the user is permanently present on the response pages.

The active part of the framework, which includes basic reasoning over the conceptual hierarchy and calculating visual similarities, was implemented in Perl. The conceptual hierarchy was built using a C++ module and implemented in an OWL compliant form but, as we stated above, the reasoning is performed using a dedicated Perl module. The rationale for this choice is that current reasoners are slow when dealing with large ontologies and we are currently scaling-up the conceptual hierarchy in SemRetriev, fact that would render reasoner impractical.

The image indexing was performed using the PIRIA tool [3]. Generally, the represented object is to be found in the center of the image and, consequently, the indexation process was performed on the central zone of the image. We acknowledge that this solution is a compromise but it is necessary as there exist no segmentation methods that perform satisfactorily on images from broad domains and of variable quality.

¹ www.ask.com

5. WALKTHROUGH THE SYSTEM²

We describe in this section a typical interaction with the system and begin by presenting the keyword based retrieval. Suppose that one wants to retrieve images for *dog*. A part of the first answers page is presented in fig. 1.

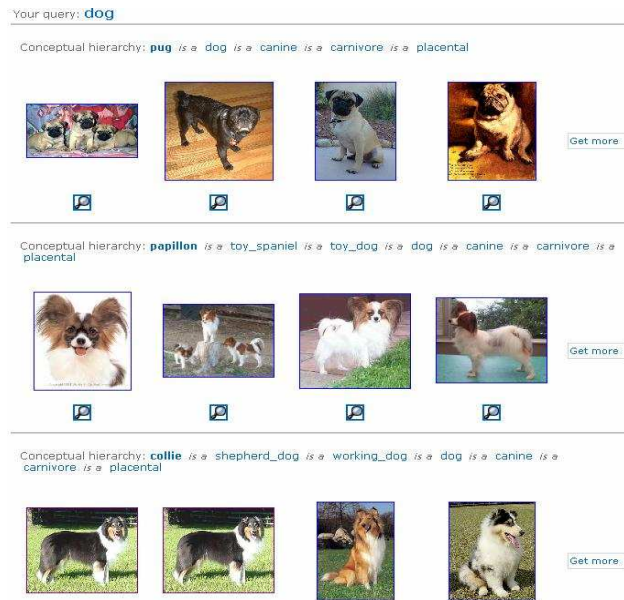


Figure 1. SemRetriev answers for *dog*, part of the first page.

The most common subconcepts of *dog* are *pug*, *papillon*, *collie* and *beagle*. The user can extend or narrow his search using the related categories displayed above each picture set corresponding to terminal nodes. For example, one can refine his query and ask to see images of *hunting dogs*. The first page includes answers for *beagle*, *golden retriever*, *Labrador retriever* and *basset*. As for the second page of responses, it is partially presented in fig. 2.

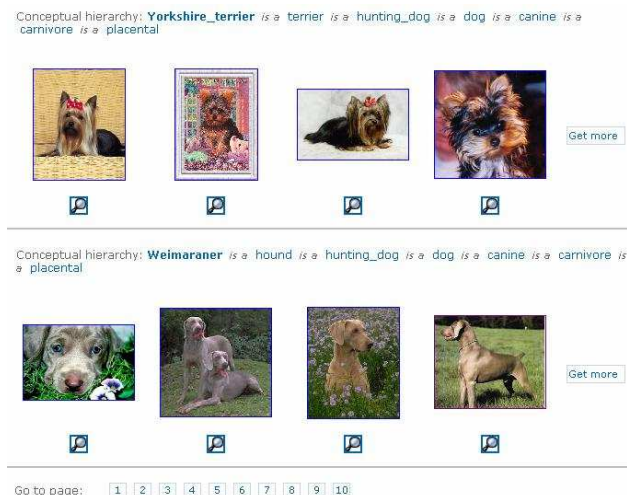


Figure 2. SemRetriev answers for *hunting dog*, part of the second page.

Suppose now that the user finds the images for *Weimaraner* interesting and decides he wants to see an enlarged selection (fig. 3).

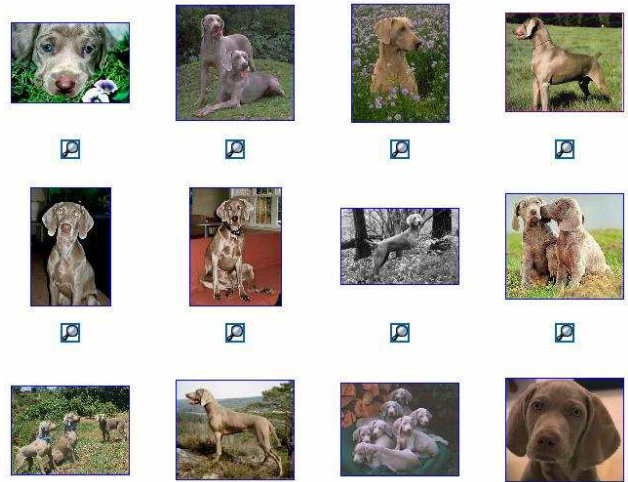


Figure 3. SemRetriev answers for *Weimaraner*, part of the detail page.

It is equally possible to click the images in order to get similar ones (query by example mode). If one clicks the fourth item on the first row in fig. 3, the 19 closest images from the *Weimaraner* category are displayed (fig. 4).

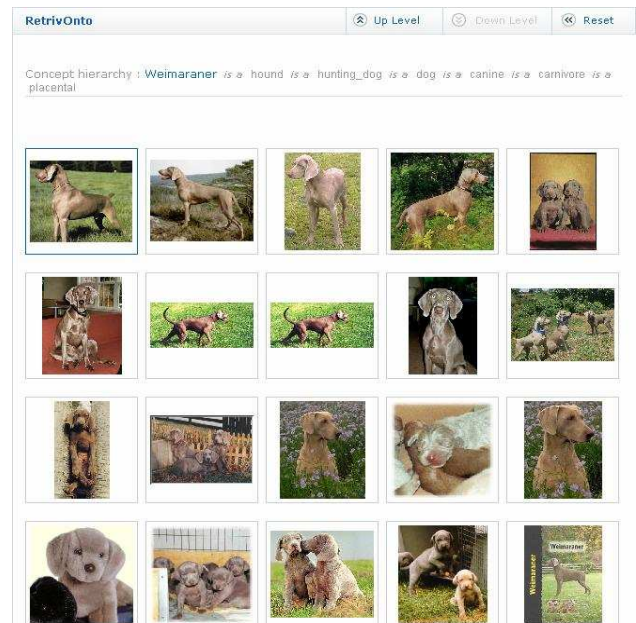


Figure 4. SemRetriev query by example inside the *Weimaraner* class. The example image is placed on the top left corner.

² Video available for download at <http://moromete.net/semretriev.wmv>

The user can navigate in the ontology using either the buttons or the categories displayed above the image set (see figure 4). In fig. 5, we present the results of a CBIR process based solely on low-level parameters and observe that, even if the retrieved pictures are similar in terms of basic descriptors, they do not resemble from a human's point of view. A detailed discussion about the quality of results of the visual similarity search is to be found in [6].

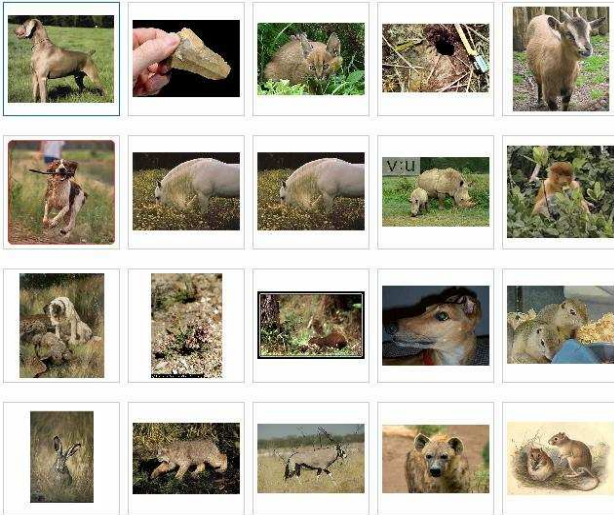


Figure 5. SemRetriev query by example in whole database (classical CBIR). The example image is placed on the top left corner.

6. CONCLUSIONS

In this paper we presented an IR system that supports both keyword-based and query by example retrieval and showed in what the introduction of an ontology in the framework enhances the retrieval process. Namely, the precision of the obtained results is increased and a conceptual navigation is enabled in when using textual search. For the CBIR mode, SemRetriev proposes an ontologically driven IR. In both situations, the results are easily understandable and the user is presented with interaction means which guide him intuitively to continue the retrieval process if the desired images were not found.

The main limitations of the system come from the domain covered by the ontology, the impossibility to provide answers for complex queries and the number of images in the system. Future work will focus on the extension of the image database, correlated with the use of a wider ontology. In a second version, the demo will run with over 1 million images and an ontology of about 10000

concepts. We are equally interested in testing the reaction of users when presented in face of increased interactivity for IR systems which combine semantics and image processing techniques. We are equally interested in testing the reaction of users when presented in face of increased interactivity for IR systems which combine semantics and image processing techniques.

7. ACKNOWLEDGEMENTS

We thank Sofiane Souidi for having developed the PHP interface of the CBIR part application.

8. REFERENCES

- [1] Gao, Y., Luo, H., and Fan, J. Searching and browsing large scale image database using keywords and ontology, In *Proc. of ACM Multimedia 2006* (Santa Barbara, USA, 2006).
- [2] Jansen, J., Spink, A., and Saracevic, T. Real life, real users, and real needs: a study and analysis of user queries on the web. *Int. Journal of Information Management and Processing*, 36,2 (January 2000), 207 - 227.
- [3] Joint, M., Moëllic, P.-A., Hède, P., and Adam, P. PIRIA: A general tool for indexing, search and retrieval of multimedia content. In *Proc. of SPIE Image processing: algorithms and systems* (San Jose, California, January 19 – 21, 2004), 116 - 125.
- [4] Miller, G. A., Ed. WordNet: An on-line lexical database. *Int. Journal of Lexicography* 3, 4 (Winter 1990), 235-312.
- [5] Popescu, A., Grefenstette, G., and Moëllic, P.-A. Using Semantic Commonsense Resources in Image Retrieval, In *Proc. of SMAP 2006* (Athens, Greece, December 4 - 5, 2006).
- [6] Popescu, A., Millet, C., and Moëllic, P.-A. Ontology Driven Content Based Image Retrieval, In *Proc. of CIVR2007*.
- [7] Stehling, R. O., Mario. Nascimento, A., and Falcao, A.X.. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proc. of CIKM '02* (McLean, Virginia, USA, 2002).
- [8] Viola, J. and Jones, M. Robust Real-time Object Detection. In *Proc. of the Second Int. Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling*. (Vancouver, Canada, 2001).
- [9] Wang, H., Liu, S. and Chia L.T. Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches, In *Proc. of ACM Multimedia 2006* (Santa Barbara, USA, 2006).
- [10] Yang, J., Wenyin, L., Zhang, H., Zhuang, Y. Thesaurus-aided Approach for Image Browsing and Retrieval, In *Proc. of IEEE – ICME'01* (Tokyo, Japan, 2001).