Extracting an Ontology of Portrayable Objects from WordNet

S. Zinger [1], C. Millet, B. Mathieu, G. Grefenstette, P. Hède, P.-A. Moëllic

Commissariat à l'Energie Atomique-LIST/Atomic Energy Agency of France-LIST
LIC2M (Multilingual Multimedia Knowledge Engineering Laboratory)
18 Route du Panorama 92265, Fontenay aux Roses, France

tel. +33 1 46 54 96 56, fax +33 1 46 54 75 80
sveta_zinger@yahoo.com, {milletc, mathieub, gregory.grefenstette, hedep, moellicp}@zoe.cea.fr

Abstract

We describe building a large-scale image ontology using the WordNet lexical resources. This ontology is based on English words identifying portrayable objects. We reviewed the upper structure and interconnections of WordNet and selected only the branches leading to portrayable objects. This article explains our pruning approach to WordNet. The words, which represent portrayable objects, are then used as queries in our VIKA (VIsual KAtaloguer) system which acquires images through a web image search engine, performs content-based image indexing and clustering. Coherent images form clusters and others are rejected. So images inside clusters mostly represent the object determined by the query, and in this way image collections representing objects are created. An ontology of portrayable objects with representative images in its nodes will be a useful tool for solving the object recognition task.

1 Introduction

The constantly increasing amount of digital information requires development of more effective image retrieval strategies. While text processing methods have been successfully applied in information search engines, much work remains to be done in the area of web image retrieval [2]. We can state the goal of this work as finding the most relevant images that correspond to a search engine query.

Currently, search engines approach the problem of image retrieval by considering the text surrounding images, or pointing to images, that can be found on web-pages. Both textual and link information is used in this framework. There have been many propositions for improving the analysis of text found connected to images, for example, segmenting web-pages into blocks in order to better localize the information [1]. One can readily understand that object recognition would also be a good strategy for improving content-based image retrieval

[1] This work was carried out during the tenure of a MUSCLE Internal fellowship.

performance. Current research in object recognition consists of assigning sets of words to image regions using labeled training data. In this approach, image annotation is considered as a translation from a visual feature language to a subset of natural language [3]. Another research approach to object recognition is based on a constructing a specific domain ontology and then aligning points in this ontology with several classes of visual concepts, such as object's shape, scene type, texture type and color, using statistical learning methods [7]. In order to improve both of these methods, we have set ourselves the goal of building a large-scale, general image ontology, whose nodes contain not only words, but also sets of images representing the objects which correspond to these words.

In the following sections we will explain our approach to construction of large-scale image ontology, based on a subset of WordNet [8], and how this ontology is populated with images drawn from the web.

2 Image ontology construction

Since our purpose is to create a large ontology of portrayable objects, and WordNet covers both the abstract and concrete vocabulary of English, we must prune WordNet, extracting only those branches containing these objects. In the next subsection, we give several examples of how this was done and explain our choices. In the second subsection we present the VIKA system developed in the LIC2M laboratory for creating sets of images, representing objects, using web-based image search.

2.1 Pruning approach to WordNet

To create an ontology of portrayable objects, we started by examining the top level of the WordNet 2.0 ontology. We chose a large class *entity* at this top level. WordNet defines an *entity* as having "a distinct separate existence (living or nonliving)". There are a number of subclasses within the *entity* class: *location*, *substance*, *object*, *causal agent*. Among these subsets, we chose the class *object,* which is defined in WordNet as "physical object (a tangible and visible entity)". This definition seemed to be suitable for our purposes because portrayable objects are supposed to be visible. We initially expected that most of the portrayable objects would appear within the class "object". But not all the nodes found under this node were necessarily useful. We first removed adjectives from these branches, and we simplified the connections between nodes, in order to produce a tree rather than a lattice. Some classes in WordNet 2.0 are assigned several times to different nodes, sometimes, on different levels of the same branch. In such a case, we left only one connection, that seemed the most logical in our opinion. We found some cases in which we deleted branches of the remaining tree because the node did not seem to correspond to a portrayable object in the applications we are considering. For example, WordNet contains a class *tree*, and inside this class, types of trees. One of the subclasses of tree is *tree of knowledge* which does not correspond to the type of portrayable object that we have in mind for object recognition in images. Other examples of nodes that we pruned are *wildlife*, found in the branch *object - living thing - life*", and *classic* from the branch *object - artifact - creation*. In general, we retained the branches of WordNet which lead to nodes representing physical objects that have either similar shapes or colors or texture. Some WordNet classes which define collections of objects, such as the class “facility, installation”, were not included in our ontology.

After pruning, the resulting upper level ontology of portrayable objects contained 102 nodes, some of which are shown in Figure 1. The lowest nodes below this subset of nodes contain the list of words (more than 24,000 terms) corresponding to objects, which we can use as queries to a web-image search engine to obtain representative images. In order to make the search query more precise, when the term only contains one word, we combined the query term with the words in the node immediately superior to it.

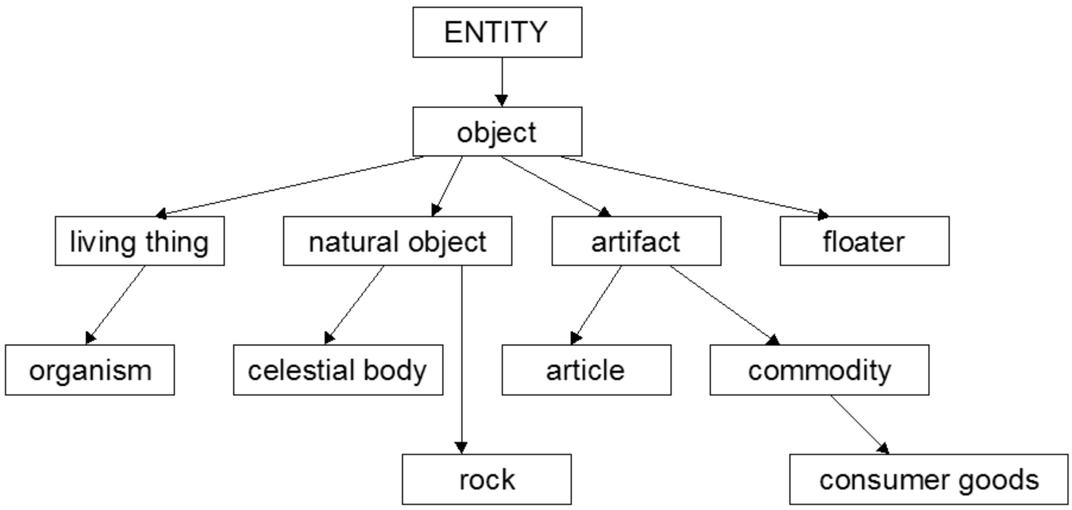

Figure 1. An extraction from the WordNet-based top level ontology of portrayable objects

For example, the queries for the types of trees are composed with the word *tree* from the immediately superior node:

- kino tree
- red sandalwood tree
- carib wood tree
- Japanese pagoda tree
- palm tree
- ...

The query “Japanese pagoda tree” will lead to pictures of trees, while “Japanese pagoda” will lead to images of buildings.

2.2 VIKA system

The approach described above provides us with query terms for web-based image mining for representative images of objects. In this section, we describe the system VIKA developed in our laboratory for creating such sets of images. With this interface, the user can type a query and specify how many images to find (Figure 2).

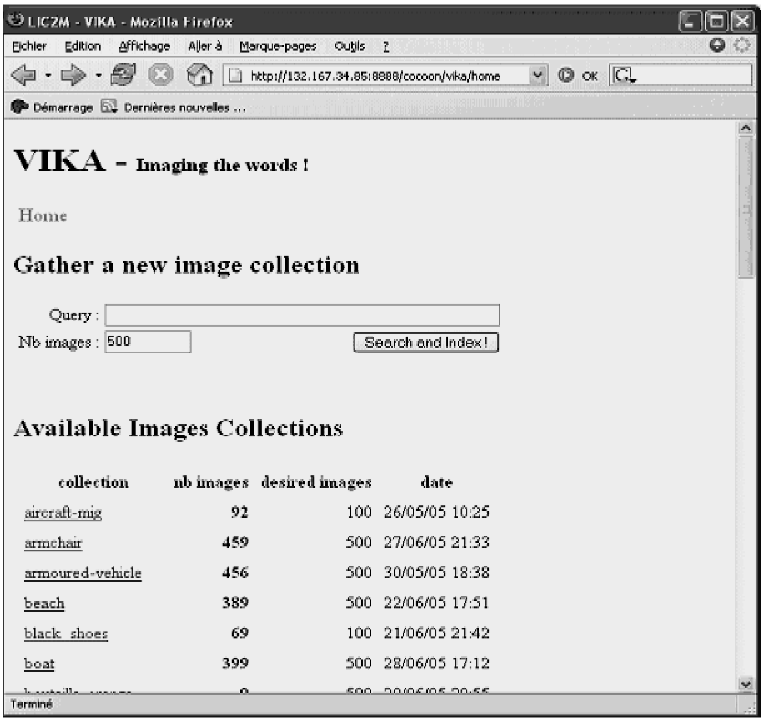

Figure 2. VIKA (VIsual KAtaloguer) system for gathering and clustering images

VIKA contacts the Alltheweb image search engine to acquire images. Retrieved images are then indexed using our in-house image indexing platform PIRIA (Program for the Indexing and Research of Images by Affinity) [6]. Many algorithms are available in PIRIA for extracting image signatures in order to index images. The algorithm that we found most successful for indexing objects, especially man-made objects, is based on border/interior pixel classification [9]. This method builds two histograms for each image. One histogram takes into account only border pixels, the other considers only pixel within this border. Thus, the first step of the algorithm is classifying pixels as being interior or border pixels. This algorithm is fast, simple and provides information not only on colors of an image but also on the sizes of constant color areas within an image. Although this border/interior algorithm gives intuitively pleasing results, we have yet to determine which PIRIA's indexing scheme gives globally optimal results, or to develop a method for choosing the best scheme for a given class of objects.

The next step is to cluster retrieved images in order to find prototype images to illustrate the query words. To perform this task, we used a k-SNN clustering algorithm (Shared Nearest Neighbor), based on ideas from [5] that are developed in [4].
For each image, the algorithm considers only the k most similar neighbor images. The more common neighbors that two images have, the more similar they are. Images that are most similar to their neighbors are considered as topic images. Topic images are used to create clusters, and images that are strongly linked to topic images are aggregated into clusters. Other images remain unclustered. The quantity of unclustered images depends on parameters set previously.

In our application, the initial image collection retrieved from the Internet is often noisy. We hope that the clustering algorithm will ignore off-topic images, supposedly isolated, and can extract some highly coherent clusters that will include prototype images for query words. For this approach, the SNN clustering algorithm has two main advantages. First, it does not predefine the number of clusters. Second, it does not force images to belong to a cluster. A screen-shot of VIKA after clustering for the query "chair" is in Figure 3.

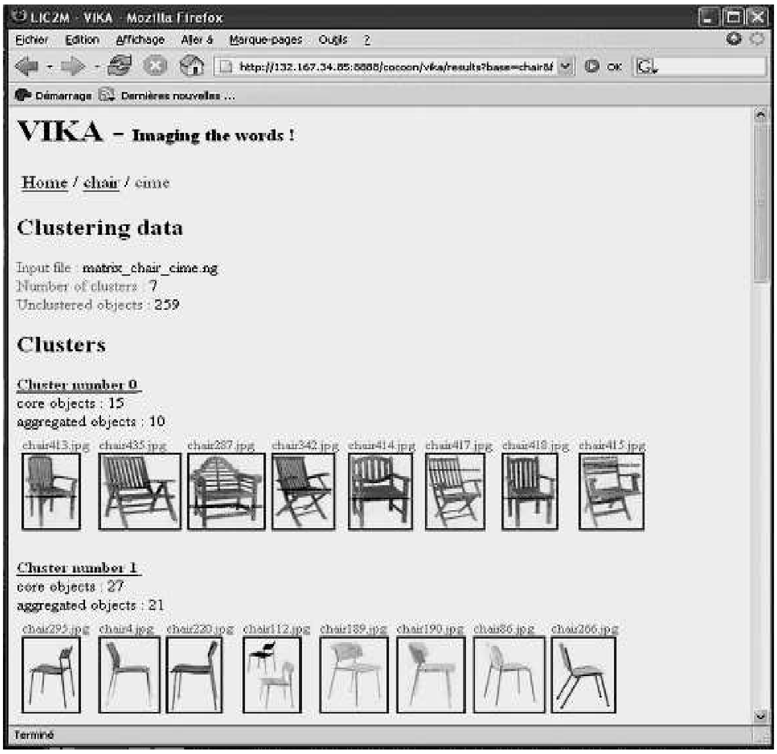

Figure 3. The clustered results of web image search for the query "chair": images of chairs form clusters, representative for the object "chair".

Figure 4 shows the present web image search results currently available in their original retrieval order on the left, and on the right the reordered images found by taking the best clusters produced by VIKA. This example anecdotally illustrates the results we wish to obtain for all objects, though we are still developing techniques for eliminating clusters built form irrelevant images. One observation we can make is that there are many photographs of people on the web and therefore detecting faces will be a good semantic filter for getting cleaner sets of images of objects.

Figure 4. Left image - web image search results currently available, right image – desirable results (clusters from VIKA).

3 Conclusions

In this article we describe our initial approach to building a large-scale image ontology of objects, pruning the WordNet lexical resources to obtain the list of portrayable objects. We also presented our system VIKA for indexing and clustering web image search results in order to obtain sets of relevant images representing these objects. We hope that the large-scale image ontology can serve as a data source for improving object recognition in images. Given this resource, and an appropriate image indexing mechanism, new, unlabelled images can be indexed and matched against the images attached to each node of the ontology, providing the node labels to the new image. Having this ontology should not only allow the identification of objects in images, but also give a more general description of an image using the inheritance of traits present in the entire ontology. Future work on this ontology involves deriving distinctive visual signatures for each node in the ontology, exploiting co-occurrence information, and resolving scaling problems.

References

[1] Cai, D., He, X., Li, Z., Ma, W.-Y., and Wen, J.-R., "Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information", In *Proceedings of the 12th annual ACM international conference on Multimedia,* New York, NY, USA, pp. 952-959, 2004.

[2] Deselaers, T., Keysers D., and Ney, H., "Clustering Visually Similar Images to Improve Image Search Engines", In *Informatiktage 2003 der Gesellschaft für Informatik*, Bad Schussenried, Germany, November 2003.

[3] Duygulu, P., "Translating images to words : A novel approach for object recognition", PhD thesis, Middle East Technical University, Dept. of Computer Engineering, 143 pages, 2003.

[4] Ertöz L., Steibach M. And Kumar V., "Finding topics in Collections of Documents: A Shared Nearest Neighbor Approach". *Proc. Of Text Mine'01, Workshop on Text Mining, First SIAM International Conference on Data mining*, pp. 83-104, 2001.

[5] Jarvis, R. A. and Patrick, E. A., Clustering Using a Similarity Measure Based on Shared Nearest Neighbors. *IEEE Transactions on Computers, vol. C22, No. 11*, pp. 1025-1034, 1973.

[6] Joint, M., Moëllic, P. A., Hède P. and Adam, P., "PIRIA: a general tool for indexing, search, and retrieval of multimedia content", *Journal of Image Processing: Algorithms and Systems III (Proceedings of the SPIE, Volume 5298),* pp. 116-125, 2004.

[7] Maillot, N., M. Thonnat, M., "A Weakly Supervised Approach for Semantic Image Indexing and Retrieval**",** CIVR 2005, Singapore, 10 pages, 2005.

[8] Miller, G. A., "WordNet: a lexical database for English", *Communications of the ACM, Vol. 38, Issue 11*, pp. 39-41, 1995.

[9] Stehling, R. O., Nascimento, M. A., Falcão, A. X., "A Compact and Efficient Image Retrieval Approach Based on Border/Interior Pixel Classification", In *Proceedings of the eleventh international conference on Information and knowledge management.* (McLean, Virginia, USA, 2002)**.** ACM Press, New York, NY, pp. 102-109, 2002.