

A Conceptual Approach to Web Image Retrieval

Adrian Popescu, Gregory Grefenstette

CEA LIST

18 Route du Panorama, 92260 Fontenay aux Roses, France

E-mail: adrian.popescu@cea.fr, gregory.grefenstette@cea.fr

Abstract

People use the Internet to find a wide variety of images. Existing image search engines do not understand the pictures they return. The introduction of semantic layers in information retrieval frameworks may enhance the quality of the results compared to existing systems. One important challenge in the field is to develop architectures that fit the requirements of real-life applications, like the Internet search engines. In this paper, we describe Olive, an image retrieval application that exploits a large scale conceptual hierarchy (extracted from WordNet) to automatically reformulate user queries, search for associated images and present results in an interactive and structured fashion. When searching a concept in the hierarchy, Olive reformulates the query using its deepest subtypes in WordNet. On the answers page, the system displays a selection of related classes and proposes a content based retrieval functionality among the pictures sharing the same linguistic label. In order to validate our approach, we run to series of tests to assess the performances of the application and report the results here. First, two precision evaluations over a panel of concepts from different domains are realized and second, a user test is designed so as to assess the interaction with the system.

1. Introduction

Picture search represents an important part of queries users express when using Internet information retrieval applications and all major search engines propose dedicated facilities. In the current paradigm, the retrieval process is based on the use of text chains related to the image file and the results are only partially relevant for the user query. With the exception of Ask¹, the Internet search engines do not systematically employ semantic resources to propose related queries on the answers page, a fact that limits the interaction to the expression of a query and the browsing of the results. It is interesting to note that major commercial actors do not include image processing techniques like the content based retrieval, in their architectures. This could be an effect of the fact that such techniques are difficultly scalable to the size of the Internet picture repository and that the obtained results do not resemble from a human's point of view. Consistent research efforts (Joshi, 2006), (Wang, 2006), (Liao, 2005) are directed toward the introduction of semantic resources in image search applications and the use of such structures is showed to fit the users' need. One important challenge is to propose semantic frameworks that are appropriated for a use in real-world applications, like Web search engines.

An interesting characteristic of Internet image queries is that they are generally composed of few words (Jansen, 2004). The same study shows that, for the most of the times, the users tend to look only at the first answers pages. These facts should be remembered when designing retrieval architectures.

In this paper, we describe Olive, a picture retrieval framework that employs a large scale conceptual hierarchy to provide a dual access to Web images. Here, we present a working system that builds on (Popescu, 2007a), where we mainly discussed the principles sustaining its construction. The employed semantic structure includes over 117000 English terms, which

cover a hefty chunk of common nouns as well as some representative proper names. Queries can be expressed using keywords or using one of the retrieved items as seed for a conceptually controlled CBIR process. The system employs ontological relations in WordNet (Miller, 1990) to automatically disambiguate and reformulate user queries, to present results in a structured manner and to propose related concepts for browsing. If a term is polysemous, the user is presented with separate image answers sets. For each query, the systems checks if the concept has subtypes in the hierarchy and, if so, images corresponding to leaf nodes are collected on the fly using Google Image and proposed to the user in response to the initial query. As a consequence, Olive proposes a concept-based answers browsing as the user navigates through pages presenting subclasses of the initial concept instead of looking through a plain list of pictorial representations. With the use of the reformulation, the quality of retrieved picture sets is improved (Popescu, 2007a) because the application employs specialized terms and their association to the content of the depiction is less noisy than that of more general concepts. The knowledge in WordNet is equally employed to narrow the search space before performing a content-based retrieval. Visually similar images are searched among images standing for the same leaf node in the hierarchy. This choice is motivated by the fact that, for humans, the conceptual similarity prevails over the visual one (Cox, 2000). Currently, the system is able to answer to around mono-term queries, which represent around a quarter of the total number of Internet image demands (Jansen, 2004).

The remainder of this paper is structured as follows: in Section 2 we discuss related work, in Section 3 we describe the image retrieval architecture we developed and, before concluding, we present an extensive evaluation of the system.

2. Related Work

In (Liu, 2004), the authors evaluate different picture browsing strategies. A similarity based presentation of

¹ <http://www.askx.com>

results is proposed and this approach is compared to a cluster based approach and a plain list display of answers. The structured presentation of results reduces the browsing time and is preferred by the users when compared to the plain list display.

Knowing how the users formulate their queries is an important aspect in image search. Studies like the one we already cited (Jansen, 2004) focus on this multimedia retrieval and it is important for this finding to be considered when one designs retrieval frameworks. In (Jansen, 2004) it is shown that Web image queries are generally short. 22% of them include a single word and 32% are formed of 2 words. Internet image search systems should focus on returning good quality answers to such queries. Another interesting conclusion of (Jansen, 2004) is that people generally want to have the responses with the least effort possible. This translated in the fact that they rarely look beyond the third answers page. Another explanation of this fact could be that generally, the first answers pages contain more representative results than the subsequent ones.

In the following, we stress some of the important common points and differences between our work and the articles cited above. First, similarly to (Smith, 2000), (Wang, 2006), a conceptual hierarchy is used in the framework. One important distinction arises from the difference in size between the knowledge base employed by Olive and the other ones.

In (Joshi, 2006), (Wang, 2006), (Yang, 2001), WordNet is used in different settings to improve image retrieval. None of the above approaches systematically use the type-subtype relation in the semantic structure to perform automatic query reformulation and propose structured answers to the user. Moreover, in (Joshi, 2006), (Wang, 2006), (Yang, 2001) the meaning separation in WordNet is disregarded and the possibility of performing query disambiguation is lost while in Olive this property is preserved. Knowing that the average polysemy of English nouns 1.23² and multiple meanings appear mostly for frequently used terms, polysemy represents an important noise source in image retrieval.

Automatic query reformulation was shown to help image retrieval (Liao, 2005). Here we propose a different approach from that in (Liao, 2005) and use subtypes of a given query to search for corresponding images.

Evidence that a structured presentation of pictures answers outperforms a plain list display is found in (Liao, 2005), (Liu, 2004). The former paper proposes an organization of the results using reformulated queries that include modifiers of the initial one while in the latter work the images are grouped following visual similarity measures. In Olive, we propose a presentation of the results based on query reformulation that is different from that in (Liao, 2005). Instead of using modifiers, we employ leaf nodes in a conceptual hierarchy to display pictures.

The size and the evolution of the employed database is another important factor in Web image retrieval. The use of a locally stored database is rendered necessary by the fact that those frameworks include relevance feedback ((Cox, 2000), (Smith, 2000), (Yang, 2001)) or learning techniques ((Wang, 2004), (Cai, 2004)) to enhance results. We show that it is possible to improve the results without

the use of relevance feedback and learning techniques. This lightweight architecture equally allows the performing of an on the fly image collection and the system evolves at the same rate as the host, Google Image.

3. System Architecture

In this section we describe the retrieval framework we have developed. We start by presenting an architectural overview of Olive (figure 1), to continue with a presentation of its main components and functions.

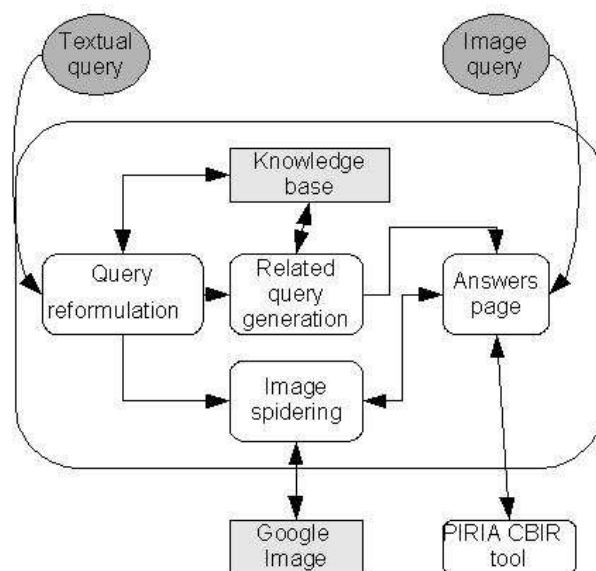


Figure 1. Functional diagram of Olive. The user interaction parts are represented in the ellipses, the resources Olive employs are drawn as rectangles and the active components of the application in rounded rectangles.

A typical interaction in Olive goes as follows: a query is typed and it is reformulated by the system employing WordNet knowledge about the given concept. The output of this step constitutes the entry for the image spidering module, which employs an external application, Google Image to collect Web pictures. In the same time, a list of categories that are close to the query is generated using the knowledge base. Once these two processes are finished, an answers page is generated. For each image on the answers page, it is possible to search for visually related images using the PIRIA visual search engine (Joint, 2004).

3.1 The Knowledge Base

The knowledge base is the central part of the piece in the Olive architecture as it enables other processes like query reformulation of related query generation. We have parsed the WordNet files containing information about the nouns in the hierarchy and extracted information regarding: polysemy, hyponymy, synonymy and classes having the same parent. This pretreatment is necessary in order to speed up the execution time. There are 81426 synsets in the hierarchy and they include 117097 unique English terms standing for 145104 meanings. Around 65000 synsets are leaves in the hierarchy. We already mentioned that the average polysemy in WordNet is 1.23. In figure 2, we present a pseudo-log distribution of the number of senses for each one of the unique terms in the

² <http://wordnet.princeton.edu>

hierarchy. The number of monosemous terms exceeds 100000 and, with the increase of the number of senses, the amount of corresponding terms generally decreases. There are 10257 terms with 2 meanings, 2989 having 3 senses and 1178 with 4 meanings. The English word with the largest number of senses is *point*, which has 26 WordNet entries.

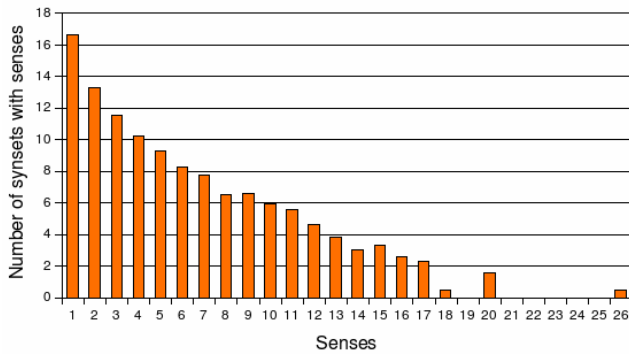


Figure 2. Pseudo-log distribution of the number of senses for nouns in WordNet. For visualization purposes, $\log(0)$ was replaced with 0 and $\log(1)$ with 0.5.

In the knowledge base, each synset has a separate entry whose structure depends on status of the synset in the hierarchy. Entries corresponding to leaves in the hierarchy contain information about: polysemy; the associated Web queries; categories having the same parent and hypernyms of the class. The entries for concepts having subtypes in WordNet include: polysemy information; a list of leaf concepts under the current class; a list of subtypes (with non-leaf concepts presented in priority); categories having the same parent and hypernyms of the class.

The hierarchy includes a large number of categories and it is necessary to propose a way to order them so as to collect images from the Web and present results. WordNet contains some frequency information but it is not detailed enough for our purposes and it was extracted from text corpora. As we work with images, we considered appropriate to use information about this kind of media and we have developed a simple procedure to extract picture frequencies using an Internet search engine. First, it is necessary to disambiguate concepts so as to reduce the influence of polysemy on the obtained results. For each synset, the first member and the immediate parent are used in conjunction. Second, the above mentioned query is launched and the number of associated images is recuperated and stocked into a file which will be used when creating the entries in the knowledge base. In table 1, we present an excerpt from the entry corresponding to the sense of *dog* as *animal* where the related classes are ordered using Web frequency information.

Polysemic	yes
Leave subtypes	pooch, pug, Newfoundland, basset, beagle, cairn, Airedale, Doberman, German shepherd, basenji
Narrower concepts	Poodle, corgi, spitz, cur, hunting dog, working dog, toy dog, Dalmatian, griffon
Related	Wolf, fox, hyena, wild dog, domestic

types	cat, bitch, jackal
Parents	Organism, canine, domestic animal, living thing, physical entity, object

Table 1. Entry for dog in the knowledge base.

The information in table 1 is used query for image with the leave subtypes and to include classes that are related to *dog* in the answers page.

3.2 The PIRIA Visual Search Engine

PIRIA (Joint, 2004) is a tool that performs image indexing and retrieval based on low-level features. It is used as an external module in Olive. The engine includes a wide range of picture indexers and we have chosen to use one which computes texture and color information.

3.3 The Query Reformulation Module

We already mentioned that reformulation is an important feature of our approach as it allows an amelioration of the quality of the results as well a structured and interactive way of presentation. An automation of the reformulation process could implicate a risk if the knowledge in the ontology would not be accurate but, as WordNet was manually constructed by lexicographers, it generally contains good quality knowledge and can be safely used to represent a concept via its subtypes. This module recuperates information about leaf nodes in the hierarchy like the one presented in table 1 and employs it to search for Web images. When a query is launched a disambiguation procedure similar to that described in subsection 3.1 is used to compose the query (the immediate parent is added to the query). If not enough items are obtained this way, a query containing uniquely the concept name is formed.

3.4 The Query Reformulation Module

As shown by the recent introduction of a related queries generation module in the Ask, the proposition of more interactivity options in image search engines is an interesting way to explore. The main challenge that arises is that large scale semantic structures are to be included in the retrieval architectures. The utilization of such resources like WordNet constitutes a solution as they provide access to better structured information than that elicited by Ask. As an anecdotal example, we present the related queries proposed by this search engine when one queries for *dog* (table 2).

Narrower search	Puppy, Alaskan husky
Related names	Rabbit, bird, Scooby Doo
Expanded search	Cat, monkey, elephant, tiger, lion, golden retriever, kittens, Chihuahua, snake

Table 2. Related categories for dog in Ask.

The categories in table 2 are globally not as well connected to the initial query as the ones presented on the last three lines of table 3. While *cat* and *puppy* are close to *dog*, it is unclear why *elephant* and *snake* are proposed. Furthermore, the presentation is semantically unsound as *Alaskan husky* is considered as a narrower class while *golden retriever* and *Chihuahua* are considered as

expanded searches. The same observation is valid for the inclusion of *rabbit* and *bird* in the list of related names and of *cat* and *monkey* in that of enlarged searches.

When querying polysemic concepts, additional meanings are proposed. The items on the first answers page correspond to the first WordNet sense. For example, other senses of *dog* include “informal term for a man” or “a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll”. If one queries for *Angora*, separate image sets are proposed for the term as: a *rabbit*, a *goat*, a *cat* and a synonym of *Ankara*, the Turkish capital.

3.5 The Image Spidering Module

We mentioned, in section 2, that Olive collects the presented pictures are collected on the fly using a parallelized script that employs the output of the query reformulation module and is plugged on the Google Image engine. This phase takes 2-3 seconds for leaf nodes, around 5 seconds for the other concepts and about 10 seconds when recuperating images for performing a CBIR search. These times are obtained for an architecture using only one computer and a 1Mbps Internet connection and they can be linearly reduced if the pass bandwidth of the connection is increased.

3.6 The Answers Page Generation

The outputs of the query reformulation and image spidering modules are aggregated and the results are displayed as in figure 3. The answers page is constructed to include three zones: a results panel, a related query box and a help box which is adapted to the type of query.

In Olive, some interaction means are inherited from current image search engines. We speak notably of the possibility to formulate another query and the navigation among the results pages. In addition, related classes are proposed on the right of the page, as well as the possibility to see a detailed page for each presented leaf concepts. If the concept is polysemic a page containing images for different meanings of the term is proposed. The presentation of close concepts can be of help when the user wants to precise his search (Liao, 2005) and it is favorably assessed by the users. These conclusions are supported by the user test we present in the evaluation section.

If the current concept has subtypes, the images are presented in a structured fashion, using four terminal subconcepts on each page. In figure 3, the first page of responses for *dog* contains images for its most frequent hyponyms: *pooch*, *pug*, *Newfoundland* and *basset*. The second page would include images for *beagle*, *cairn*, *Airedale* and *Doberman*.

The results in figure 3 are to be compared with those obtained if we query Google Image with *dog* directly (figure 4). We observe that the pictures in the first figure are of better quality and are presented in a conceptually structured manner.

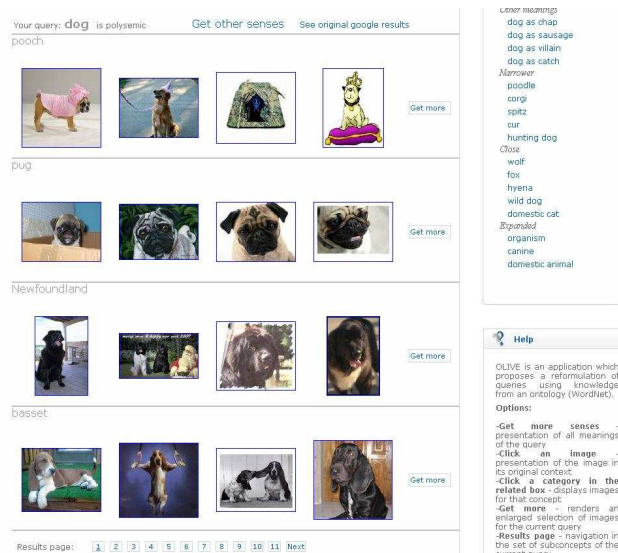


Figure 3. The first answers page for *dog* in Olive.



Figure 4. The first answers page for *dog* in Google Image.

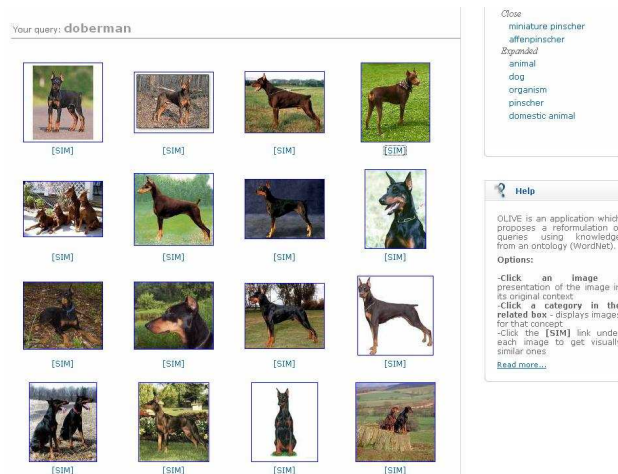


Figure 5. Answers page for Doberman in Olive.

A detailed page with answers for Doberman is presented in figure 5. Note that, as concept is a leaf in the hierarchy, there are no narrower classes proposed in the related box. Moreover, the concept is not ambiguous and there are no other senses to be presented. A notable difference between

the pages in figures 3 and 5 is that in the latter case the user can perform a visual similarity search using any of the displayed images. The results of such a process are presented in figure 6.

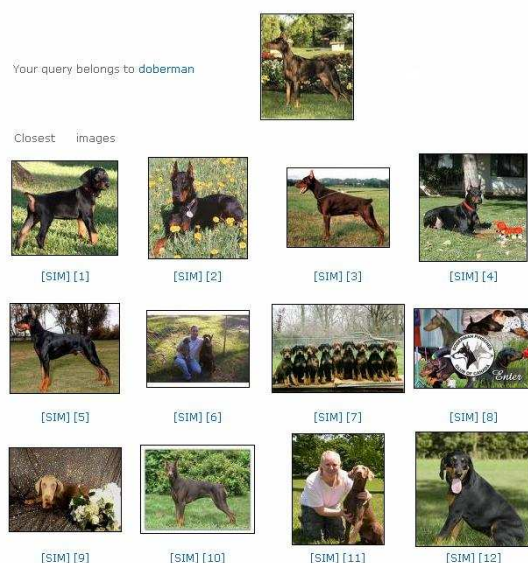


Figure 6. Visual similarity search for an image of Doberman.

We remind the reader that the CBIR process is performed with a set of Web images standing for the same concept as the query (here *Doberman*) and employs color and texture to assess similarity. In (Popescu, 2007b), we showed that a conceptually controlled visual similarity search is by far more efficient than a retrieval process that accounts solely for low-level image parameters. This finding is also conform to the findings in (Yang, 2001), where the authors obtain better results when using both low-level and high-level semantics to find pictures.

4. Evaluation

We assessed Olive on two dimensions: first, a precision measure was employed so as to evaluate the quality of the obtained results and second, a test was designed to test the interaction of the user with the system. This double evaluation is necessary because it is shown in (Turpin, 2006) an increase of results precision in an information retrieval application does not necessarily engender a improved perception of the system from an user's point of view. The precision was evaluated in two settings: a surface test where the results on the first two response pages were tested and an in-depth test where the responses on the first 10 answers pages were assessed. In all tests, we used Google Image as baseline because Olive draws on this search engine.

4.1 Surface Precision Test

The conceptual hierarchy included in Olive covers a wide range of domains. We propose here an evaluation that accounts for a well established separation of categories in the world, where these last are separated into *natural* and *artifacts* (Keil, 1992). Further, *living things* and *natural objects* are subclasses of *natural entities*. In their turn, *living things* separate into *plants* and *animals*. We used the following four general categories to select concepts for testing: *animals*, *plants*, *natural objects* and *artifacts*.

We have chosen to evaluate the performances of Olive and Google over a panel of 40 categories, 10 from each one of the general classes mentioned above. The choice of the concepts is done so as to cover as well as possible the general categories and to be easily recognizable by the evaluator, which was asked to count the number of images that were pertinent for the given query for each response page. Jansen et al. (2004) show that people who are using Web image search engines regard, in a majority of cases, the first two answers pages. An evaluation of the items on these pages will give a fair idea of the performances of the two image retrieval system as they are perceived by the users.

The results of the test are presented in figure 7 and they indicate that, in mean 68.2% of the assessed images were judged pertinent for Google while 84.9% were considered as so for Olive. Over the 40 concepts, Google behaved better than Olive in two cases (*lake* and *computer*) but the differences between the two systems are minimal in these cases (less than 3%). Important differences in favor of Olive are obtained for *apple* (81.2% vs. 19.4%), *table* (90.6% vs. 44.4%) *cloud* (84.4% vs. 58.3%) or *fox* (81.2% vs. 58.3%). Over the four general categories, excellent results are obtained with our approach for *animals* and *plant*. For all these classes, a precision of more 80% is obtained, with peaks at 100% for *butterfly* and *spider*. For Google, there is a noteworthy difference between the quality of the answers on the first page, 80.3% and those on the second page, 56.1%. This distance could be explained by the fact that a relevance feedback procedure is used to rank Google results³. In Olive, the corresponding percentages have values around 85%. This finding is to corroborate with the finding of the in-depth test we describe below.

The results we obtained in this test are coherent to those we reported in (Popescu, 2007a), where a similar test was run over a smaller number of concepts. They are equally to be associated to the outcome of the evaluation in (Wang, 2006), where it is shown that the use of a domain ontology improves image retrieval.

4.2 In-depth Precision Test

We have already mentioned that Olive proposes a concept-driven navigation instead of a plain list one (Google like). We wanted to evaluate the quality of results beyond the second page of results. Eight concepts (2 for each general category) were selected: *dog*, *frog*; *pine*, *fungus*; *star*, *sea*; *ship*, *plaything*. The answers in the two systems were evaluated over the first ten pages and a mean over the eight concepts for each page is presented in figure 8.

The results in figure 8 indicate that the answers one obtains when using concept-based navigation have a constant quality while the Google results are by far better on the first page.

³ <http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnethtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=6,799,176.PN.&OS=PN/6,799,176&RS=PN/6,799,176>

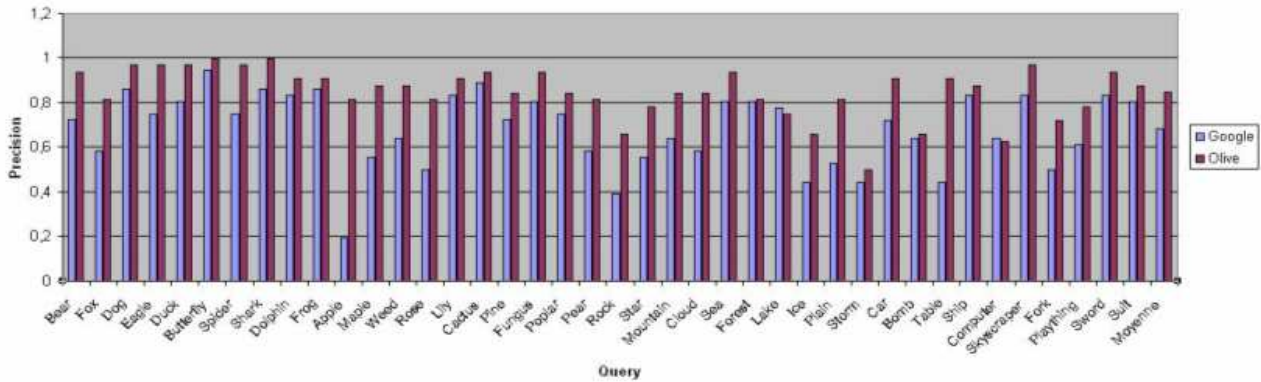


Figure 7. Precision test comparing Google and Olive on a panel of 40 concepts from different domains for the first two answers pages.

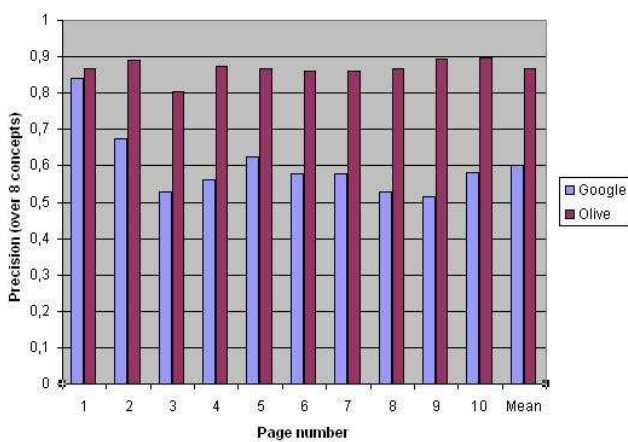


Figure 8. In-depth precision test comparing Google and Olive over a panel of 8 concepts for the first ten pages of results.

In this test, the pertinence of results is of 60.1% for Google and 86.7% for Olive. The difference between the two applications is bigger than the one reported in the surface precision test. This observation is supported by the fact that for Google, the results on the pages from two to ten are significantly worse than those on the first page while for Olive, the quality of the results is relatively constant over the ten pages. For pages three to ten, the precision is around 60% for Google and superior to 85% for Olive. When we look at individual concepts, the precision obtained with Olive is better for all eight concepts. Important differences are obtained for *dog* (96.2% vs. 66.7%), *star* (82.8% vs. 40.6%) or *plaything* (85.6% vs. 51.1%). The smallest distance, 7.2%, is obtained for *frog*.

We stated that, for existing search engines, the users prefer to look at the first answers pages. An obvious explanation for this observation is that reaching the results requires less effort. The results in figure 8 support a second explanation: the users rarely look beyond the second answers page because they are aware that the corresponding results are of poor quality when compared to those displayed first.

4.3 Interactivity Test

We designed a test where ten evaluators were asked to use Olive and rate some of the characteristics of the interaction process. The testers (computer science and

linguistics students) had no experience with the system except for a training query. The interaction scenario consisted in a series of five entry queries (*duck*, *angora*, *apple*, *car*, *rock*) followed by a free exploration of the functions of the system. Each tester passed around 20 minutes to explore Olive and another 10 to answer the questions. In all, the users have looked at 523 results pages. When opening a response page in Olive, the users were asked to do the same for Google. The test included two types of responses: directed questions and free text. The former (see table 3) general characteristics of Olive compared that are either comparable to Google features (GQ1 – GQ6) or not as well as a detailed evaluation of the related class presentation module (RQ1 – RQ4). In the open text part of the test, the evaluators freely expressed their opinions about the strengths and weaknesses of Olive.

G Q	1	You have used Olive and compared the answers to the ones displayed by Google. Please rate the overall quality of the image responses:
	2	You observed that, in Olive, the results are displayed in a structured manner (see the example of duck in the above question). Do you find this presentation:
	3	Do you find that Olive is easy to use (intuitive)?
	4	Olive proposes an automatic reformulation of your queries. Do you find that, with the use of the reformulation the system responds satisfactorily to your question?
	5	Some terms in a language have several senses. Please rate your preference concerning the presentation of results for ambiguous terms in an image search application
	6	In Olive there are more interactivity options than in Google. Do you think that the proposition of these navigation options is useful?
R Q	1	Rate the global pertinence of the propositions in the "Related image data sets" box when reported to the initial query.
	2	Rate the utility of the narrower terms search
	3	Rate the utility of the close terms search
	4	Rate the utility of the more general terms search

Table 2. Interactivity test questions.

Given their different nature, adapted answers were proposed for the questions in table 3:

- For GQ1 and GQ2, answers were on a scale of 1 (strong preference for Google over Olive) to 5 (the inverse).
- For GQ3, GQ4 and GQ6, possible answers were yes and no. For GQ4, the users were asked to choose between a presentation of image results for ambiguous questions together (marked as no in table 4b) or in separate sets (yes ins table 4b) corresponding to each meaning of a word.
- For RQ1, a scale of 1 (completely irrelevant related classes propositions) to 4 (very relevant) was presented.
- For RQ2 – RQ4, a scale from 1 to 4 was proposed. 1 stands for complete inutility of proposing related category and whereas 4 corresponds to a great utility of this presentation.

The results of the test are presented in table 4.

	Mean	St. dev
GQ1	4.5	0.71
GQ2	4	0.94

(a)

	Yes	No
GQ3	10	0
GQ4	9	1
GQ5	9	1
GQ6	10	0

(b)

	Mean	St. dev
RQ1	3	0
RQ2	3.1	0.74
RQ3	2.9	0.57
RQ4	2.9	0.87

(c)

Table 2. Related categories for dog in Ask.

The results in table 4b show that, although inexperienced, the evaluators considered that Olive is easy to use. As for a structured presentation of results and the presentation of results for ambiguous terms in separate picture sets, nine users out of ten found these options preferable to the opposite. Finally, all the users thought that the presentation of more interaction options than it is the case in current search engines helps during image retrieval. These findings are coherent with those in (Liao, 2005) and support the ideas that automatic query reformulation and the presentation of more interactivity options on the answers page in image retrieval applications are favorably seen by the users.

When comparing general quality of the results for the two systems (GQ1), the users find that it is better for Olive than for Google. This finding is in accord with the results of the precision tests we described in subsections 4.1 and 4.2. As for the structure of the results, there testers seem to prefer a semantically organized presentation over a plain list. Only one of the ten evaluators expressed a slight preference for the way Google presents results. The preference for a structured answers display we found here is coherent with the results in (Liao, 2005) and (Liu, 2004).

When asked to rate the related class presentation (RQ1), all users considered that the majority of the presented classes were relevant for the initial concept. For questions RQ2, RQ3, RQ3 the majority of displayed classes was considered helpful. The opinions of the users were most uniform for RQ3 (standard deviation - 0.57) while they differed the most concerning the presentation of more general terms (standard deviation – 0.87). In mean, the users considered that the proposition of narrower classes (RQ2) is slightly more relevant during retrieval than that of close terms (RQ3) and more general concepts (RQ4). This preference for narrower terms is explainable by the fact that they are subtypes of the query and help the user precise what are the images he wants to see. We would expect an even greater difference between the scores for RQ2 and RQ3, RQ4.

The free text part of the interaction test proved to be stimulant for the users and the main ideas for ameliorating Olive they expressed are presented hereafter:

- Extension of the hierarchy so as to include more images for people names.
- Better arrangement of the presented subtypes.
- Separation of the images in photographic, cliparts, maps etc.
- Proposition of a graph view for the related class box.

The ideas presented address some of the current limitations of Olive. The extension of the ontology to include pictures of people is in accord with the findings in (Jansen, 2004), where one of the main usages of an image search engine was discovered to be the search for celebrities. An arrangement of the presented subtypes that accounts, aside term frequency for the structure of the hierarchy that has the current concept as a root brings a conceptual representation problem into light. The separation of different types of depictions following their production mode adds a new dimension to the need for a structured presentation of results. Finally, the reaction time is an important parameter in an interactive application and it should be as short as possible. In section 5, we present some ideas for future work that address the questions raised by the users.

4.4 Discussion of Results

In this section, we presented a series of evaluations which support our claim that the introduction of a semantic structure in image retrieval architectures enhances the search process. While it is arguable that precision tests are not robust, we think that the diversity of the evaluated concepts and the difference between the results obtained with Olive and those elicited by a state of the art system are sufficient to validate our approach. Future precision tests should concentrate on the evaluation of the system by a panel of users.

The results of interactivity tests are intrinsically subjective but these tests are necessary as they represent the only way to directly evaluate the impact the system has on users. We believe that the results presented in section 4.3 sustain our claim that the introduction of semantic structure in image retrieval architectures helps is better fitting the users' needs. Among the enhanced aspects, we cite the results organization, interactivity, overall quality of the rendered picture sets. Further tests should concentrate on a refinement of the methodology and on an extension of the number of evaluators.

The coherence between the parts of the experimental results obtained in our work and those in described in other papers ((Liao, 2005), (Wang, 2006), (Yang, 2001)) provides further support to our approach to image retrieval. It is our belief that, even it is subjective and time costing, the evaluation of picture search applications implicating the participation of human testers is necessary as these systems are finally destined to the users. This opinion is supported by the findings in (Turpin, 2006), where it is shown that there is no automatic correlation between the quantitative increase of the performances and the perception of the application by the users. Olive outperforms the baseline, Google Image, on both dimensions and this allows us to state that a real improvement is obtained.

5. Conclusions

In this paper, we presented an Internet image retrieval system based on the utilization of a large scale conceptual hierarchy. The main contributions of our work can be summarized as follows:

- The presentation of a semantic retrieval architecture that covers a significant part of Web picture queries. The coverage is sustained by the size of the employed knowledge base, which is substantially increased when compared to similar approaches like (Smith, 2000), (Wang, 2004), (Wang, 2006)
- The proposition of a dual access to picture content. In Olive, both textual and visual queries are supported. CBIR processes are performed directly on Web images and in regions that are conceptually coherent and the obtained results are in better accord with the way people assess similarity than classical content based search
- A dynamic structure, where the retrieved pictures are retrieved on the fly. This characteristic is important in a fast-changing environment like the Internet.
- A proof that, with the utilization of semantic structures, existing image indexes can be better used than it is the case in current applications.
- A consistent series of tests where quantitative and qualitative measures are employed and which show that our approach outperforms Google Image, the baseline system.

This paper describes an ongoing work and in the future we shall concentrate on the following aspects:

- The extension of the hierarchy. One idea we shall develop is the use of semi-structured knowledge included in such free access resources like Wikipedia.
- The proposition of a refined algorithm for arranging the displayed concepts. This new procedure should reflect both the frequency associated to each leaf synset and the structure of the conceptual hierarchy.

6. Acknowledgements

The research reported in this paper is a part of the Mediatic project (French regional business clusters Cap Digital et Systematic).

7. References

Cai, D., He, X., Li, Z., Ma, W.-Y., Wen, J.-R. (2004) Hierarchical clustering of WWW image search results using visual, textual and link information. In *Proc. of*

- ACM Multimedia* New York, NY, USA, pp. 952 – 959.
- Cox, I.J., Miller, M.L., Minka, T.P., Papathomas T.V., Yianilos, P.N. (2000) The Bayesian Image Retrieval System, PicHunter. *IEEE Tran. On Image Processing*, 9(1), pp. 20-37.
- Jansen, B.J, Spink, A., Pedersen, J. (2004) An Analysis of Multimedia Searching on Altavista. In *Proc. of the 5th ACM SIGMM MIR Workshop*, Berkeley, CA, USA.
- Joint, M., Moëllic, P.A., Hède, P., and Adam, P. (2004) 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, pp. 116-125.
- Joshi, D., Datta, R., Zhuang, Z., Weiss, W.P., Friedenber, M., Li, J., Wang, J.Z. (2006) PARAGrab: a comprehensive architecture for web image management and multimodal querying. In *Proc. of Very large data bases*, 32, Seoul, Korea, pp. 1163 – 1166.
- Keil, F.C.(1992). Concepts, Kinds, and Cognitive Development. *MIT Press, Cambridge, Massachusetts*.
- Lee, C.C, Prabhakara, R. (2005) Querying Web Images by Topic and Example Specification Methods. In *Proc of Advanced Data Mining and Applications*, Wuhan, China, pp. 515-526.
- Liao, S.P., Cheng, P.J., Chen, R.C., Chien, L.F. (2005) LiveImage: Organizing Web Images by Relevant Concepts. In *Proc. of the Workshop on the Science of the Artificial Hualien*, Taiwan, pp. 210-220.
- Liu, H., Xie, X., Tang, X., Li, Z.W., Ma, W.Y. (2004) Effective browsing of web image search results. In *Proc. of the 6th ACM SIGMM MIR Workshop*, New York, NY, USA, pp. 84-90.
- Miller, G. A., Ed. (1990) WordNet: An on-line lexical database. *Int. Journal of Lexicography* 3, (4), 235-312.
- Popescu A., Grefenstette, G., Moëllic, P.A. (2007a) Improving Image Retrieval Using Semantic Resources. *Springer Studies in Computational Intelligence*, 93.
- Popescu, A., Millet, C., Moëllic, P.A. (2007b) Ontology Driven Content Image Retrieval. In *Proc of the ACM Conference on Image and Video Retrieval* (Amsterdam, The Netherlands, July 9 -11, 2007).
- Smith, J.R., Chang, S.F. (2000) Visually searching the Web for Content, *IEEE Multimedia*, 4(3), pp. 12-20.
- Turpin, A., Scholer, F. (2006) User performance versus precision measures for simple search tasks. In *Proc of ACM SIGIR Conference*, Seattle, WA, USA.
- Wang, H., Liu, S., Chia L.T (2006). 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; pp. 109 – 112.
- Wang, X. J., Ma W. Y. and Li X (2004). Data-driven Approach for Bridging the Cognitive Gap in Image Retrieval. In *Proc. of the IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan.
- Yang, J., Wenyin, L., Zhang, H., Zhuang, Y. (2001). Thesaurus-aided Approach for Image Browsing and Retrieval, In *Proc. of IEEE International Conference on Multimedia and Expo*, Tokyo, Japan.