Semantic Image Retrieval based on Ontology and Relevance Model: A Preliminary Study

Ernest Weke MAINA†, Manabu OHTA‡, Kaoru KATAYAMA‡, Hiroshi ISHIKAWA ‡

Tokyo Metropolitan University 1-1 Minami-Osawa, Hachioji-shi Tokyo, Japan 192-039

E-mail: †ewmaina@ieee.org, ‡{ohta,katayama,ishikawa}@eei.metro-u.ac.jp

Abstract Modern Web search engines index hundreds of millions of images. To search these images is a daunting task for the user who can, realistically, only visually inspect a handful. In general, the way the user responds to an information need depends on the task at hand. Further, some tasks will require browsing, while others are targeted and require a more directed approach. In this paper, we present the preliminary results of research to develop a framework for applying semantics to enhance image retrieval. We consider this problem on two separate levels. First, we consider the application of an Ontology to define the semantic query space for image search and navigation, as well as to approximate the users context for the search. Secondly, in order to further improve upon the search results we apply the Relevance model, using data from the web to train the model. The role of the Relevance Model is to rank images from the search engine. The study also investigates how application of an ontology affects the quantity and quality the retrieved images and also the effects to the experience of the user in image search. We then contrast the results with those obtained by the Relevance Model for exactly similar search terms. The Relevance Model is based on a probabilistic model, which applies user definable language models to the text linking to the image. In our Relevance Model, the relevance of a *HTML* document linking to an image is evaluated and assigned with respect to highly ranked textual documents from the web. The ranking of the *HTML* document, is also assigned to ranking of the respective image. The main advantage here is that the Relevance Model can be learnt from the Web without any preparation of training data and is independent of the underlying algorithm of the image search engines. We show that navigation is indeed a very powerful tool for image browsing and that using the ontology dramatically enhances recall for specialized terms. Relevance feedback mainly improves precision by effective re-ranking.

Keyword: Web Image Search, Language Model, Relevance Model, Wordnet, Ontology

1. Introduction

Recently Multimedia information in the form of images has rapidly proliferated due to digital cameras and mobile telephones equipped with imaging devices. The Google web search engine claims to have 880Million images indexed[1]. These images on the world wide web are accessible mainly through a search engine. The ability of the search engine to ease the access to this information then becomes either an enabler or a restriction and hence defines a practical limit on the accessibility of information that is actually available online. Sensibly, one can only be able to browse a handful of images when performing a search through this gigantic database. For this reason, with the predictable forward charging growth of the web, it is indeed necessary to have powerful tools to assist users in performing the search, which conceivably becomes less trivial as the web grows further in size.

Although capability and coverage vary from system to system, we can categorize the image search engines into three broad groups in terms of how images are indexed. First is the text-based index. The representation of the image includes filename, caption, surrounding text, and text in the HTML document that displays the image. The second one is an image-based index. Here the image is represented using visual features such as color, texture, and shape. The third one is a hybrid of text and image indexing. In practice, the text-based index seems to be the prevailing choice now if anyone plans to build a large-scale web image retrieval system. Possible reasons include text input interface allows users to express their information needs more easily than image interface, (asking users to provide a sample image or drawing a sketch is difficult and error prone), image understanding is still an open research problem, and image-based index are usually of very high dimensionality.

In this paper, we present the preliminary results of a research that focuses on exploring a framework to semantically ground search of unstructured image documents that form the larger part of images on the World Wide Web. This paper presents the first part of this investigation. We investigate the actual effects of application of ontology for semantic query expansion of image search terms. In addition to any new advantages, we measure the traditional figures of merit for this set-up to quantify the result.

We also investigate the application of relevance modeling to image search, using identical search terms to those used in the semantic query expansion research above. Relevance Modeling is remarkable in that while ordinarily one would require prior training data, we apply a technique that does not require preparing any training data. We are able to compare and contrast the two image search enhancements and use the results to suggest future course of action.

2. Previous Research

Many of the research publications on image searches display a focus or bias towards narrow and well-defined application domains, and have applied advances in image processing and machine learning in an attempt to derive semantic information from low-level sensory surface data. While most of these systems perform well in the lab the major problem of general purpose search engine technology that can search arbitrary images at a high semantic level remains.

The application of an ontology to represent semantic concepts in our work is similar to that of Hyvonnen et.al[10]. However, their work focuses on a museum collection and both the annotation for the images and the ontologies are handcrafted. Hollink et al [12] also applies multiple specialized ontologies in addition to WordNet to define and navigation space for the user. In this case too the ontology serves as a source of metadata, which implies that the image collection is a restricted one, at least in its applications. In this case again the application is for annotating digitized art collections.

Suprisingly, have found no literature that investigates the application of the ontology from the users point of view, except when the user is considered an expert in a field for which annotation is desired. Ontology manipulation as an expression of the users context or interest is one of the interesting approaches that we wish to investigate in this work.

Our framework seeks to incorporate maximal semantic information to aid us in image search, but with as little user intervention as possible, and to make up for the information deficit using context information already available around us. We are interested in getting the user to provide their own annotation or relevance information in an unstructured and more natural way. Such as, say, annotation learnt by machine exclusively from the narration of a trip blog with images.

3. Semantic Grouping with WordNET

One approach to improve the user experience is to layer the unstructured information on the web in an intuitive way and hence enable the user to navigate freely but also ensure that they have meaningful data to start with. This can be done by combining the power of the Web with an ontology.

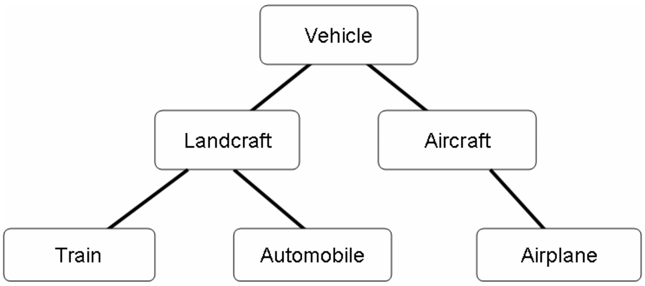


Fig 1:*The Wordnet Ontology is a mature general purpose ontology that can be used for semantic query expansion. Above is a typical noun graph.*

The search term is first input to the Wordnet ontology, which outputs a list of semantically related terms; that is synonyms, hyponyms and hypernyms.

We applied the hyponyms representing IS-A relationships in the ontology to the search engine to expand the search. This has the added advantage of providing a context for the user to formulate his query space interactively as well as for later navigation. At the search engine, the query expansion provided by the ontology serves to expand the search space for the retrieval of the requisite images.

4. Relevance Model

We use the term relevance model to refer to a

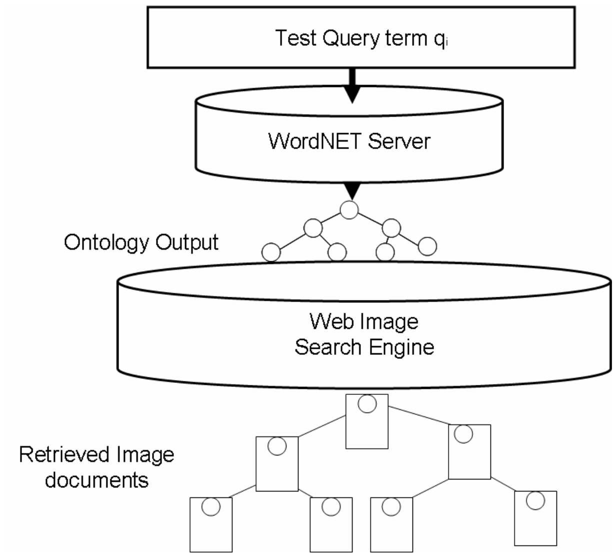


***Fig 2:**. Query expansion using server based WordNet ontology.Actual results of a search are shown in appendix A.*

mechanism that determines the probability P(w|R) of observing a word w in the documents relevant to a particular information need. When applied to information on the web it is an advantage to be able estimate this probability without having to download a lot of training data or to be able to use data without having to adhere to very rigid structure it as is required by several schemes for image annotation such as MPEG7. So we would like to consider a good way to estimate the relevance model.

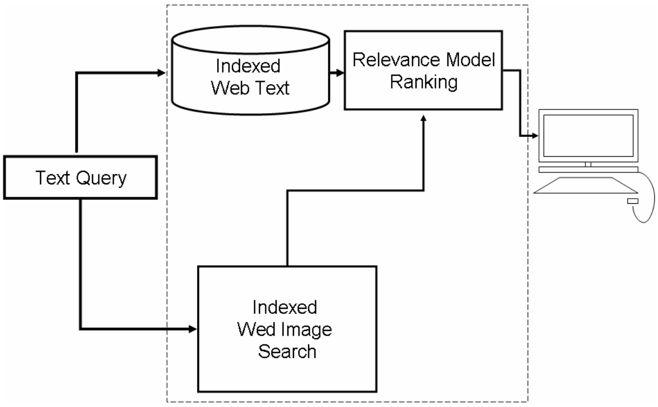


***Fig 3:** Overview of a general scheme for ranking Images*

A method, initially proposed by Lavrenko and Croft [9], offers a solution to approximate the relevance model without preparing any training data. Instead of collecting relevant web pages, we can treat query Q as a short version of relevant document sampling from relevant documents,

$$\Pr(w \mid R) \approx \Pr(w \mid Q) \qquad (1)$$

Suppose the query Q contains k words (q_1,q_2,…,q_k). Expand the conditional probability in Equation 1,

$$\Pr(w \mid Q) = \frac{\Pr(w, q_1, q_2, \ldots, q_k)}{\Pr(q_1, q_2, \ldots, q_k)} \qquad (2)$$

Then the problem is reduced to estimating the probability that word w occurs with query Q, i.e. Pr(w,q_1,q_2,…,q_k). First we expand Pr(w,q_1,q_2,…,q_k) using chain rule,

$$\Pr(w, q_1, q_2, \ldots, q_k) \approx \Pr(w) \prod_{i=1}^{k} \Pr(q_i \mid w, q_{i-1}, \ldots, q_1)$$

(3)

If we further make the assumption that query word q is independent given word w, Equation 3 becomes

$$\Pr(w, q_1, q_2, \ldots, q_k) \approx \Pr(w) \prod_{i=1}^{k} \Pr(q_i \mid w) \qquad (4)$$

We sum over all possible unigram language models M in the unigram universe Ξ to estimate the probability Pr(q|w), as shown in Equation 5. The Unigram language model is designed to assign a probability of every single word. As a result, words that appear often will be assigned higher probabilities. A document will provide a unigram language model to help us estimate the co-occurrence probability of w and q.

$$\Pr(w, q_1, q_2, ..., q_k) = \Pr(w) \prod_{i=1}^{k} \sum_{M \in \Xi} \Pr(q_i, M \mid w)$$

(5)

In practice, we are unable to sum over all possible unigram models in Equation 5, and usually we only consider a subset. In this paper, we fix the unigram models to top ranked p documents returned from a textual web search engine given a query Q.

If we further assume query word q is independent of word w given the model M, Equation 5 can be approximated as follows,

$$\Pr(w, q_1, q_2, ..., q_k) \approx \Pr(w) \prod_{i=1}^{k} \sum_{j=1}^{p} \Pr(M_j \mid w) \Pr(q_i \mid M_j)$$

(6)

The approximation modeled in Equation 6 can be regarded as the following generative process: we pick up a word w according to Pr(w), then select models by conditioning on the word w, i.e. Pr(M|w), and finally select a query word q according to Pr(q|M). There are still some missing pieces before we can actually compute the final goal Pr(D|R). Pr(q_1,q_2,…,q_k) in Equation 4 can be calculated by summing over all words in the vocabulary set V,

$$\Pr(q_1, q_2, ..., q_k) = \sum_{w \in V} \Pr(w, q_1, q_2, ..., q_k) \qquad (7)$$

where Pr(w,q_1,q_2,…,q_k) is obtained from Equation 6, Pr(w) in Equation 8 can estimated by summing over all unigram models,

$$\Pr(w) = \sum_{j=1}^{p} \Pr(M_j, w)$$
$$= \sum_{j=1}^{p} \Pr(M_j) \Pr(w \mid M_j)$$

(8)

It is not a good idea here to estimate the unigram model Pr(w|Mj) directly using maximum likelihood estimation, i.e. the number of times that word w occurs in the document j divided by the total number of words in the document, and some degree of smoothing is usually required. One simple smoothing method is to interpolate the probability with a background unigram model, expressed as follows in equation (9):

$$\Pr(w \mid \mathrm{M_j}) = l \frac{\mathrm{c}(w, \mathrm{j})}{\sum_{v \in V(j)} c(v, j)} + (1 - l) \frac{c(w, G)}{\sum_{v \in V(G)} c(v, G)}$$

(9)

where G is the collection of all documents, c(w, j) is the number of times that word w occurs in the document j, $V(j)$ is the vocabulary in the document j, and ? is the smoothing parameter whose value lies between between zero and one.

Now it is possible to estimate Pr(w|R) as shown above and re-rank the image list in decreasing order of Pr(D|R). However this results in longer documents having more product terms hence a smaller Pr(D|R). This will result in short documents being favorably ranked. We use the Kullback -Leibler (KL) divergence to avoid the short number document bias. The KL divergence measures the similarity between probability distributions p and q and is defined in the following equation:

$$D(\Pr(w \mid D_i) \,\|\, \Pr(w \mid R) = \sum_{v \in V} \Pr(v \mid D_i) \log \frac{\Pr(v \mid D_i)}{\Pr(v \mid R)}$$

(10)

where Pr(w|D_i) is the unigram model from the document associated with rank i image on the list and Pr(w|R) is the relevance model, and V is the vocabulary. Hence we take the following steps to determine the relevance/rank of the documents that we have search and wish to rerank:

a) The unigram model is estimated for each document associated with an image.

b) and then we calculate the KL divergence between Pr(w| D_i) and Pr(w|R).

c) Finally, since the KL divergence is effectively the "distance" between the unigram and relevance models, smaller divergence means the document is likely to be more relevant.

These three steps are summarized in Figure 4.

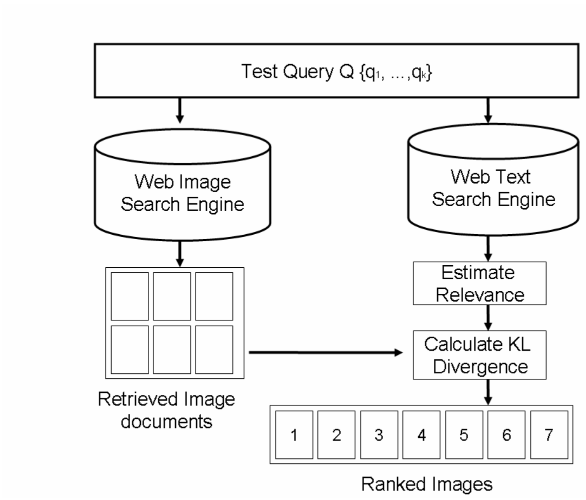


Fig 4. *Summary of procedure for relevance model estimation*

Query No.	Text Query	Relevant Images for Relevance Model	Relevant Images for Wordnet
1	Birds	52	18-113*
2	Food	115	20-160**
3	Fish	72	52-164***
4	Fruits and Vegetables	115	0-170❖
5	Sky	77	36-164♦
6	Flowers	94	120-192♦♦

Table 1*: Results showing the number of relevant documents for the first 200 results of an image search.*

*lower limit for Night Raven, higher limit for cock
**lower limit for pabulum, higher limit for chocolate
***Lower limit for rough fish, upper limit for trough
❖lower limit for hip fruit and leek vegetable. Upper limit for pumpkin and papaya
♦lower limit for Blue air and Upper limit for blue sky
♦♦Lower limit for blazing star and upper limit for Phaius

5. Experiments

The data for the investigation was based on concept groups in the Corel Stock Image Series. We selected six terms to use in the experiments. Then we performed semantic query expansion as shown in fig. 2 and fed the results to the image search engine. We noted the precision of the first 200 terms for all the terms. The term count per query was not constant and ranged from about 5 to more than 50.

We then performed a second set of experiments, this time using relevance modeling, as shown in fig 3. The images obtained from the search engine were re-ranked with respect to how statistically similar they were to the relevance model. The relevance model is created using the same respective search terms, applied to the textual search engine alone as shown in fig 4. Afterward, the html pages linked to the image results were retrieved and used and assessed for relevance. The calculated relevance was assigned to the respective image. No image processing was performed. The results of these experiments are shown in table 1. For the relevant images from semantic query expansion, we show a range of values. This is because for each of the six original search terms, WordNet outputted as many as 50 expansion phrases. Each of these terms was entered into an image search engine and the precision was noted for the first 200 results (10pages). The range shown here represents the worst result and the best result, respectively.

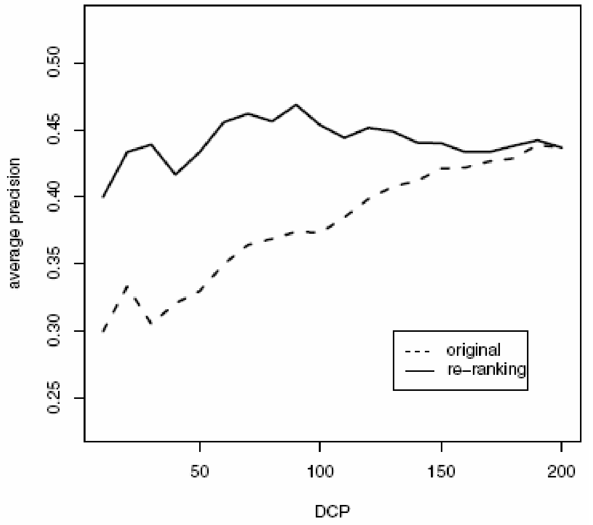


Fig 5. *Average document precision at various Document Cut-off Points [7]*

6. Conclusion

We have shown that preliminary components of our image search framework can augment each other in the enhancement of image search on the web. We applied an ontology for query expansion as well as to provide some context for the user for the navigation of the search results. The application of the ontology for query expansion was very instructive in that the computer came up with search terms that would never have have ordinarily occurred to us to consider for searching but that were very relevant to the search. On the other hand, considered as in totality, the results of the query expanded search also included many improbable terms such that their relevancy was poor. However, for navigational purposes, the ontology enhanced search proved to very effective as evidenced by the "best case" scenarios in Table 1. A typical scenario where this situation is likely to occur is when one is browsing and navigating search results. In such cases, query expansion, which effectively casts of a wider net on the search space, outperforms relevance models in terms maximum of precision potential. On the other hand, since for our application, ontology is fixed at this stage, there are certain terms that will give bad precision. This tendency can be controlled by the provision of manual feedback by the user, to allow or disallow expanded ontology terms from the image search. This can be done before the search engine is invoked in which case the user is expressing his "interests" or context. Another scenario is when the user selectively pursues certain expanded terms but not others. This would be the equivalent of pruning the results of the search engine. The freedom of the user to interact and make these is most important for such an ontology enhanced search. The user interaction becomes an important feature since the selection can enable the system to limit or prevent the "worst case" scenario results being over represented in experience of actual users.

However we think that we have a natural solution for pruning the unpopular terms in an automated way, without user intervention which we will consider in the following discussion.

The second experiment using the relevance Model provided very high precision for low document cut-off points. This illustrates the effectiveness of the relevance model obtained from the web. Effectively, re-ranking pushes the noise to the back of the queue, allowing the user to experience increased effective precision at minimum mouse clicks.

Since the relevance model works by measuring the "distance" or KL-divergence between the relevance model and search result document, and rewards similarity, we think that we can use this property reduce the influence of low precision terms arising from query expansion using WordNet. This investigation will form our next quest. Further, since the relevance models are really independent of the query themselves, we hope to investigate other applications such as multi-lingual or cross-lingual image search.

In summary, have shown that image retrieval from the image search engines can be enhanced in navigational tasks by providing an intuitive semantic surface on which the user can traverse the search results space. For browsing this provides a more intuitive experience than the raw search results. For browsing tasks, it can provide the best precision but one has to avoid a few, very low precision terms. This is easy to do interactively when one is browsing.

In applications where moderate recall can be tolerated for a little more precision, the relevance model is effective as the re-ranking pushes the low unpopular terms to the rear of the queue. The one possible disadvantage for the use or semantic query expansion is that number of search sessions per query per user will increase dramatically (more than a ten fold). On the user side the details of the query process and the larger number of files being processes can be automated, and so is not an issue to the user.

In order to more fully represent the users context for a specific image search we would like to implement a system that allows choice of ontology depending on ones interest. This would enable use to study the effects of such of multiple ontologies on image retrieval and especially the for users of a general purpose image search engine.

In this present research we only investigate manipulation in a fixed space; a single, general purpose ontology.

This research was partially supported by Grant-in-Aid for Science Research on Priority Areas "Informatics" No. 16016273 from the Ministry of Education, Culture, Sports, Science and Technology of Japan

References

1. Google Image Search, 2005. http://images.google.com

2. The Google Web API http://www.google.com/apis/

3. Google Web Search. http://www.google.com/
4. V. Lavrenko and W. B. Croft. Relevance Bas ed Language Models. In *Proceedings of the International ACM SIGIR Conference*, 2001.

5. WordNet http://www.cogsci.princeton.edu/~wn/

6. G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller. ”Introduction to WordNet: an on-line lexical database.” *International Journal of lexicography*. Vol.3. PP.235-244,1990

7. W. Lin, R. Jin, A. Hauptmann. Web Image Retrieval Re-Ranking with Relevance Model. In *Proceedings of the IEEE/WIC Internation Conference on Web Intelligence* (WI’03)

8. R. Lempel, A. Soffer. PicASHOW: Pictorial Authority Search by Hyperlinks on the Web. In Proceedings of the Tenth International World Wide Web Conference, 2001

9. V. Lavrenko and W. B. Croft. Relevance-based language models. In Proceedings of the International ACM SIGIR Conference, 2001.

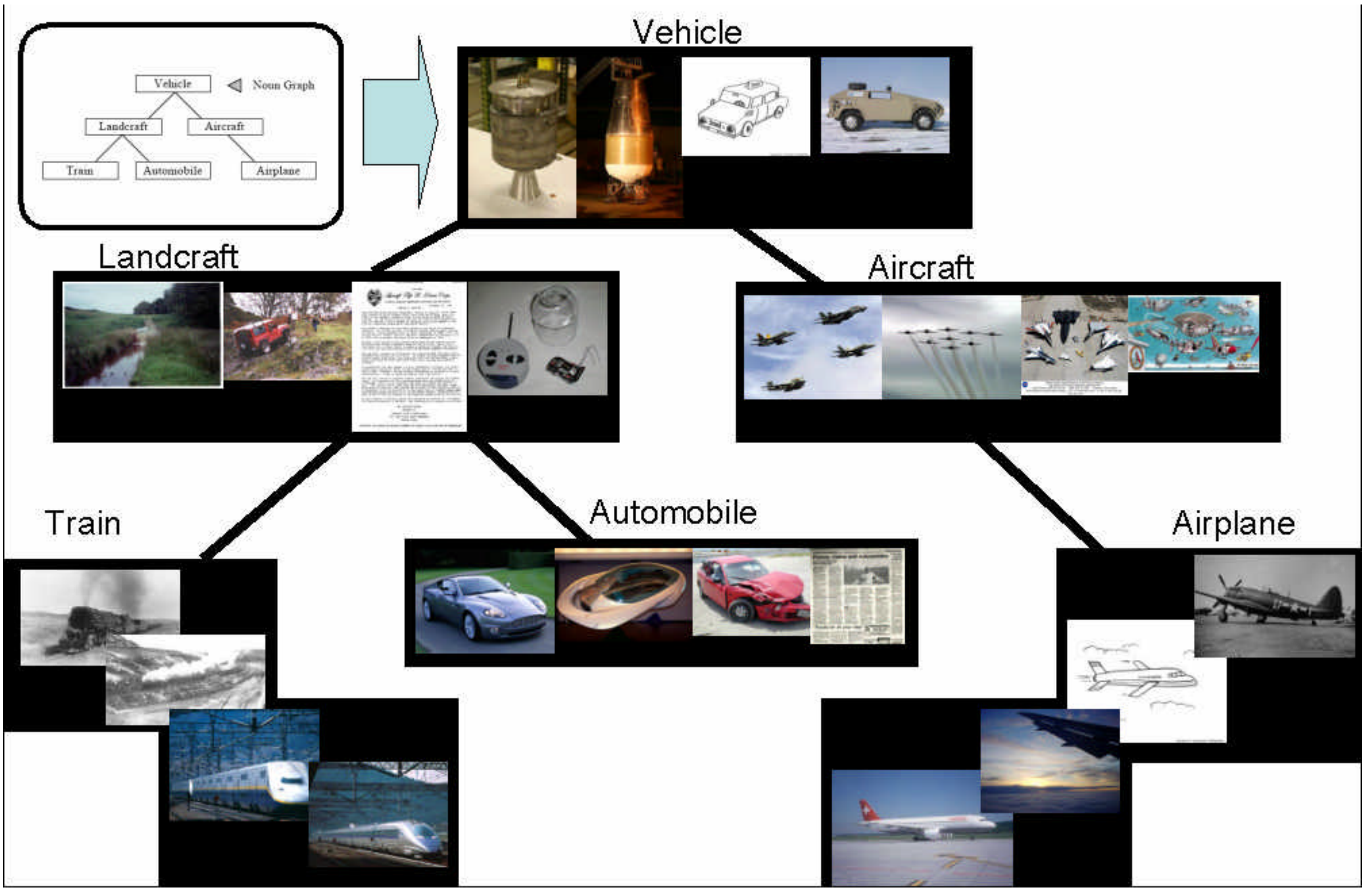
10. E. HyvÄonen, A. Styrman, and S. Saarela. Ontology-based image retrieval. Number 2002-03 in HIIT Publications, pages 43-45. Helsinki Institute for Information Technology (HIIT), Helsinki, Finland, 2002. http://www.hiit.fi

11. L.Hollink, A.Th.Schreiber, J.Wielemaker, and B.Wielinga. Semantic annotation of image collections.

In *Proceedings of the KCAP'03 Workshop on Knowledge Markup and Semantic Annotation*, Florida, USA, October 2003.

Appendix A

An example of image information layered on top of an ontology for intuitive navigation of the image space as well as visual feedback of the semantic context of the images.


Vehicle
Noun Graph
Landcraft
Aircraft
Train
Automobile
Airplane
Vehicle
Landcraft
Aircraft
Train
Automobile
Airplane