10-2008

Ontology enhanced web image retrieval: Aided by Wikipedia and spreading activation theory

Huan WANG

Xing JIANG

Liang-Tien CHIA

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Ontology Enhanced Web Image Retrieval: Aided by Wikipedia & Spreading Activation Theory

Huan Wang, Xing Jiang, Liang-Tien Chia, and Ah-Hwee Tan
School of Computer Engineering Nanyang Technological University, Singapore
{wa0004an, jian0008, asltchia, asahtan}@ntu.edu.sg

ABSTRACT

Ontology, as an effective approach to bridge the *semantic gap* in various domains, has attracted a lot of interests from multimedia researchers. Among the numerous possibilities enabled by ontology, we are particularly interested in exploiting ontology for a better understanding of media task (particularly, images) on the World Wide Web.

To achieve our goal, two open issues are inevitably involved: 1) How to avoid the tedious manual work for ontology construction? 2) What are the effective inference models when using an ontology? Recent works[11, 16] about ontology learned from Wikipedia has been reported in conferences targeting the areas of knowledge management and artificial intelligent. There are also reports of different inference models being investigated[5, 13, 15]. However, so far there has not been any comprehensive solution.

In this paper, we look at these challenges and attempt to provide a general solution to both questions. Through a careful analysis of the online encyclopedia Wikipedia's categorization and page content, we choose it as our knowledge source and propose an automatic ontology construction approach. We prove that it is a viable way to build ontology under various domains. To address the inference model issue, we provide a novel understanding of the ontology and consider it as a type of semantic network, which is similar to brain models in the cognitive research field. Spreading Activation Techniques, which have been proved to be a correct information processing model in the semantic network, are consequently introduced for inference. We have implemented a prototype system with the developed solutions for web image retrieval. By comprehensive experiments on the *canine* category of the animal kingdom, we show that this is a scalable architecture for our proposed methods.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

General Terms

Design, Experimentation

Keywords

Ontology, Wikipedia, Spreading Activation

MIR'08, October 30–31, 2008, Vancouver, British Columbia, Canada.

1. INTRODUCTION

Concept ontology has been one of the latest trends in today's image retrieval researches[2, 5, 8, 15], especially in the field of web image retrieval[3, 13, 14]. The large volume of web information, from both text and image cues, makes it easier for computers to *understand* things in the way that human-beings do. In this paper, we focus on the solution of two open issues in ontology-based web image retrieval field, which are the ontology construction and ontology inference problems.

The first inevitable question is how to efficiently build an ontology that fully exploits the information contained in the web images and their corresponding web pages. Currently, a common approach would be using WordNet [2] as a lexical resource to build concept ontology. However, as the most important feature of WordNet is to group words into synset and connect them through hypernymy/hyponymy (ISA) and meronymy (PARTOF) relationships, the ontology generated from WordNet can only come with *concepts and their hierarchical relationships*. However, previous works, e.g, [13], have shown that the retrieval performance can benefit more from ontologies with other types of relations besides the hierarchical relations. Therefore, an ontology with rich relations is more meaningful. But scalability has become a new problem if we want to build an ontology with rich relations, as manual ontology is tedious and time consuming. In this paper we consider both the ontology completeness and scalability problems, and provide our original solution which is to build ontology automatically from Wikipedia.

Secondly, given an ontology, how can we use it for image retrieval? In our case, after a query is submitted, we want to find the most relevant images to the query. As each image is associated with query by particular concepts of the ontology, we would perform inference on the ontology to calculate their relevance values. For this purpose, we abandon the traditional reasoners (semantic matchmakers), whose inference results are somehow roughly categorized according to human cognition (e.g.:exact match, subsume match), and will fail in tasks which involve further measurement (e.g.: ranking). We consider our multi-modality ontology as a type of semantic network, since they are of similar structures.

Then, we propose to use the spreading activation procedure for ontology inference, as the spreading activation inference model has been proved to be effective for inferencing in the semantic network of the cognitive field.

We have implemented a prototype system called OntoEnhanced that integrates our developed solutions for web image retrieval. Web images (and their corresponding web pages) are crawled from the top 200 hits returned by Google Image Search and indexed in the system. For evaluation, we concentrate on 20 different classes of animals under the *canine* family. Our experimental results demonstrate the efficacy of our system for web image retrieval.

The rest of the paper is organized as follows: Section 2 generally describes the components and the work flow of the prototype system. Section 3 discusses how we build the multi-modality ontology in two stages. Section 4 gives a detailed description of our proposed SAT-based inference model. Experiment results are shown in Section 5. Finally, we conclude our work and discuss about potential future works in Section 6.

2. THE WEB IMAGE RETRIEVAL PROTOTYPE SYSTEM

In this section, we generally present the components and the work flow of the proposed OntoEnhanced system. This system includes two components, namely an offline ontology construction part and an online image retrieval part, as shown in Figure 1.

The offline part generates the multi-modality ontology as follows: Firstly, a set of relevant concepts to the target domain with their associated semantic relations, including taxonomy and non-taxonomy relations, are extracted from Wikipedia. Then, other concepts related to the low-level features of the images, together with their relations are obtained from a set of training images. Finally, the concepts and relations extracted in the two steps are combined to form the final multi-modality ontology. A detailed description of the ontology construction part will be given in Section 3.

For the online image retrieval part, the users will first submit queries to the system. For the moment, the queries would be particular concepts in the multi-modality ontology (However, it is easy to enable users to submit free texts and many methods have been proposed to convert the free texts into particular concepts in the ontology [4, 6]). Then, the search engine fetches all the crawled images in the database, analyses these images, and produces the required inputs for the spreading activation procedure. After inferencing on the multi-modality ontology, the relevance of each image to the submitted query is calculated. Finally, these images are ranked according to their relevance values and returned to the users.

3. ONTOLOGY BUILDING

The ontology building process is divided into two separate phases: Wikipedia-based high-level text description ontology construction and visual feature ontology construction. We define the *canine* classes as our target concepts. Each target concept is to be associated with a set of common concepts, both from high-level and low-level information. In the following subsection, we will first discuss the concept ontology extracted from Wikipedia.

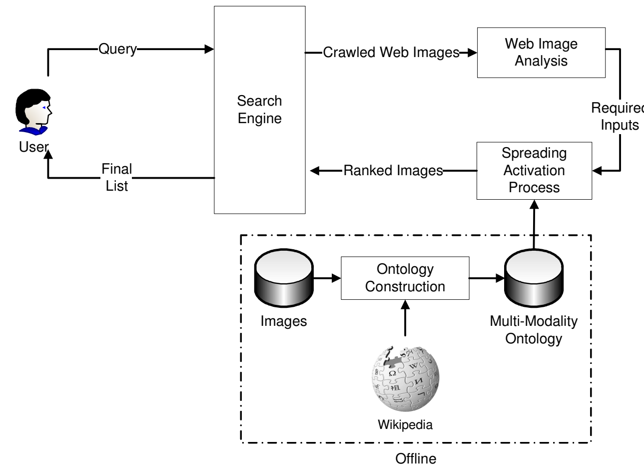


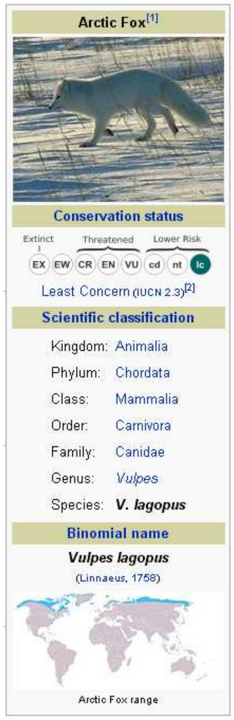
Figure 1: Illustration of the prototype system for web image retrieval.

3.1 Wikipedia-based Text Ontology Construction

Knowledge sharing has become a popular activity on today's Internet, where groups of people get together to contribute and exchange information. Among such communities, Wikipedia is one of the best for both its well maintained knowledge system and popularity. The emergence of such online encyclopedia has gradually taken the place of traditional web directories, which are usually domain specific and loosely structured, and further helped the machine learning of human knowledge. [9, 16] has successfully applied Wikipedia in tasks like concept disambiguation and ontology refinement.

We find Wikipedia an ideal source for knowledge acquisition and ontology construction for various domains. First of all, it is an online collaborative encyclopedia project which offers definitions and elaborations of more than 2 millions words and phrases. This number is considerable large compared to WordNet's word coverage. Besides its board knowledge coverage and up-to-date vocabulary, Wikipedia gives one page of definition for each concept. And whenever the concept is referred to it is automatically linked to this definition page. Wikipedia also has its own categorization system which provides a fundamental taxonomy of all the concepts. This taxonomy information is much like the relationships provided by WordNet. Figure 2(a) shows one such example, a taxonomy box for *Arctic Fox*. And Figure 3 is the underlying data to generate the taxonomy information. An automatic hierarchical ontology structure can be directly extracted from such taxonomy box, where the super node is *Animalia*, with a series of descendant concepts following the definition of the scientific classification. And the leaf node of this hierarchical tree is the concept *Arctic Fox*, which is directly under the instance *V.lagopus* of *Species* class. Wikipedia also provides necessary disambiguation for concepts, which, indirectly, solves the problem of *concept disambiguation*. Redirection function which automatically directs the concept query to more agreeable synonym concept makes it easier to build the synset.

Other than the taxonomy information, Wikipedia's elabo-



(a)

hasName	Arctic_fox
	white_fox
	snow_fox
hasDistribution	Russia
	Canada
	Alaska
	Svalbard
	sub-Arctic
	alpine
	Iceland
	Scandinavia
	Norway
	Sweden
	Finland
hasDiet	lemming
	Arctic_hare
	reptile
	amphibians
	eggs
	carrion
	ringed_seal
	polar_bear

(b)

Figure 2: An example of taxonomy information and other extractable concepts and relationships in *Arctic fox* Wikipedia page.

```
{Taxobox
 name = Arctic Fox<ref name=msw3>{{MSW3 Wozencraft | pages = |id=14000873 }}</ref>
 status = LC
 status_system = iucn2.3
 status_ref = <ref name=iucn>{{IUCN2007 | assessors = Angerbjörn, A., Hersteinsson , P.
 year = 2004 | title = Alopex lagopus | id = 899 | downloaded = [[2008-04-23]]}}</ref>
 trend = stable
 image = Polarfuchs 1 2004-11-17.jpg
 image_width = 200px
 regnum = [[Animal]]ia
 phylum = [[Chordate|Chordata]]
 classis = [[Mammal]]ia
 ordo = [[Carnivora]]
 familia = [[Canidae]]
 genus = ''[[Vulpes]]''
 species = '''''V. lagopus'''''
 binomial = ''Vulpes lagopus''
 binomial_authority = ([[Carolus Linnaeus|Linnaeus]], [[Systema Naturae|1758]])
 range_map = Distribution arctic fox.jpg
 range_map_width = 200px
 range_map_caption = Arctic Fox range
```

Figure 3: The underlying data of the taxonomy box in Figure 2.

rate definition makes it easier to drawn various related concepts and relationships. We show the same concept *Arctic Fox* as an example. The Wikipedia page contains one introduction paragraph and seven sections. In the introduction paragraph, two homonymy concepts *white fox* and *snow fox* is given in bold font. The remaining sections give information on various aspects: adaptations, reproduction system, diet, size, subspecies, population and distribution, and reference. There are a few paragraphs under each section title, where we can extract related concepts. The basic idea is that we look into certain sections and extract all the words which are hyperlinked to their own Wikipedia definition pages. For those plain text on the Wikipedia web page, we believe it is of trivial importance to our final ontology. We connect the concepts with our target concept by the relationships which are explicitly defined by the section title. For a simplified example, concepts *Canada*, *Alaska*, and *Svalbard* are extracted from the "population and distribution section" and connected with *Arctic fox* by *hasDistribution* relationship; Concept *lemmings, Arctic Hare, reptiles* and *amphibians* are extracted from "Diet" section and connected with *Arctic fox* by *hasDiet* relationship. A full list of extract concepts from Arctic Fox Wikipedia page is shown in Figure 2(b).

We do further hierarchical construction among all the extracted concepts. This step is done based on the Wikipedia category structure, which offers a systematic categorization of all the concepts. The category information is given as a separate section at the bottom of each Wikipedia web page. In most cases one Wikipedia concept belongs to several categories, some of which serves for Wikipedia administration purposes, such as *Wikipedia administration*. We remove these categories and keep the rest, which follows different categorical classification. And for each related category, we move one step further to find its parent category. In our current implementation we do five iterations, and construct a hierarchical structure of five levels for each concept. This step helps to formulate the information and introduce more structured concepts on top of the current ontology.

3.2 Visual Description Ontology Construction

Concepts from image features are also an essential part in our multi-modality ontology. For our experiment, we use recognition techniques to build a visual vocabulary and train classifiers using support vector machine (SVM). We do not generate our own object detection techniques as these techniques have been extensively discussed in computer vision literatures. Our objective is to show that an extended ontology will help to improve on the current context-understanding of existing object detection techniques.

We have also been studying object detection techniques and we are aware of the latest development ([17] provides substantial survey of existing techniques). For our work, we first use Harris-Laplace detector[10] which is scale invariant and detects corner-like regions in the images as interest point, and then use SIFT[7] descriptor to represent the shape information around the interest point. Color descriptor is also combined with SIFT descriptor. Different from [8], we use *Opponent Angle*[12] instead of *hue* color description. *Opponent color* consists of 3 opponent channels: an achromatic channel which responses to gray scale changes along with two chromatic channels which range from red to green and blue to yellow, respectively. *Opponent angle* is proved to be invariant to specular variations. Research [12] has shown that for web images which have less saturated colors and presence of diffuse lighting, it is advisable to use *opponent angle* descriptor to describe color information around the interest point. A 20 by 20 image patch around the center of the interest point is generated to extract *opponent angle* features. In addition, a shift along the horizontal or vertical axis is made when boundary is within the patch range. After combining *opponent angle* with SIFT descriptor, we do normalization and the final descriptor is a vector of dimension 164, where 128 dimensions are from SIFT descriptor and 36 dimensions are from *opponent angle* descriptor, with 3 scales and 4 orientations for each opponent channel .

We build a vocabulary of 1,000 visual words based on

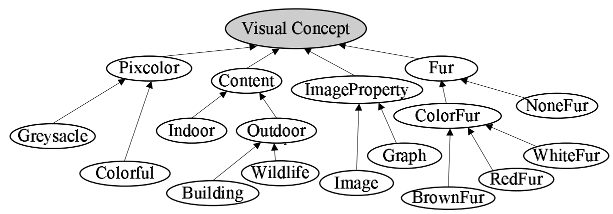


Figure 4: A simplified Example of Visual Description Ontology.

k-means clustering result of feature vectors from all images. The size of visual words is experimentally found to give good result. For each image in the data set, a histogram of visual words is calculated and then each image is represented by a vector whose dimension is 1,000. After feature space construction, the next step is to build image classifiers. We do a cross-training on our data set, In each round of experiment we use two thirds of the data as training data, and the rest is used as testing data. Sixteen concepts are defined to describe the color and texture information from image features. These concepts are linked to the target concept through low-level relationships. Figure 4 shows a simplified example of the final visual description ontology structure. And this structure is connected to the target concept as part of the whole multi-modality ontology.

4. ONTOLOGY INFERENCING

After the users submit the queries to the system, we need to compute the relevance of each image to the particular concept c_q of the query. Images are then ranked according to their corresponding relevance values of c_q and returned to the users. In our system, we use the spreading activation theory (SAT) based inference procedure [1] to compute the relevance values.

In the field of cognitive science, one popular form of storing knowledge in long term memory is semantic network. Concepts are represented as nodes in the network and linked through relations. Information processing in the semantic network typically follows the spreading activation theory, in which the activation value of each and every node spreads to its neighboring nodes. Given an initial input activating specific nodes of the network, after the spreading activation process finishes, each and every concept in the network will be activated with certain values depending on its relations to neighboring nodes. As an ontology is structurally similar to a semantic network and spreading activation theory has been proved to be efficient for inferencing in an ontology in the previous work, e.g., [6], it is adopted as a natural choice of inferencing in the multi-modality ontology here.

An illustration of the spreading activation procedure in the ontology is given below. Referring to Figure 5, the node *Jackal* has been activated with an activation value of 1.0. Its activation then propagates across the entire semantic network following the spreading activation procedure. When the network stabilizes, the nodes in the network will be activated with certain activation values such as those shown in Figure 7. Note that the activation value of each node does not depend solely on its distance from the initial node. For instance, the concept *Gazelle* obtains a higher activation value than that of *Goat* following the network configuration, which means *Gazelle* is considered more related to *Jackal* in this semantic network.

The mechanism of the spreading activation theory is hereby defined formally as below: Given a source node x and a destination node y, the activation propagation process follows the formula:

$$I_y(t_{i+1}) = O_x(t_i) \times w_{xy} \times (1-\alpha), \quad \alpha \in (0,1) \tag{1}$$

where $I_y(t_{i+1})$ is the input of node y at time t_{i+1}, $O_x(t_i)$ is the output of node x at time t_i, w_{xy} is the link between nodes y and x, and α is a decay factor to represent the energy loss in the spreading activation process. A simplified spreading activation theory is that the output of the node y at time t_i is the input of the node y at time t_i, $O_y(t_i) = I_y(t_i)$. Thus, the entire spreading activation process can be summarized into the following formula:

$$O = \left[\mathcal{E} - (1-\alpha)w^T\right]^{-1} I, \tag{2}$$

where $I = [I_1, \ldots, I_n]^T$ is the initial input to the network, w is the matrix representation of the user ontology whose element w_{ij} represents the link between concepts c_i and c_j, α is the decay factor, $\mathcal{E}$ is an $n \times n$ identity matrix of order n, and $O = [O_1, \ldots, O_n]^T$ is the final output vector of the spreading activation process in which O_i is the value of concept c_i obtained from the spreading activation process.

In our case, the surrounding texts and low level features of each image have been converted to particular concepts of the multi-modality ontology. Given these associated concepts with their frequencies obtained, we can form a vector $I = [I_1, I_2, \ldots, I_n]^T$ as the input of the spreading activation process to inference an image's relevance to c_q, where I_n, the input to the concept c_n, is calculated by

$$I_n = \frac{freq(c_n)}{\sum_{all\ c_n} freq(c_n)}, \tag{3}$$

where $freq(c_n)$ represents the frequency of the concept c_n obtained in the image's surrounding texts or low level features.

Upon receiving the input vector I, the spreading activation procedure is performed on the semantic network w to infer the relevance of the image to c_q. In our system, the configuration of the matrix representation w is described as follows. We first extract all the semantic relations from a set of sample data set. The element $w'_{ij}s$ value of the matrix w is the frequency of the semantic relation r_{ij} in the data set. Then, we normalize the matrix using the following formula:

$$w_{ij} = \frac{freq(r_{ij})}{\sum_{all\ j} freq(r_{ij})}, \tag{4}$$

where $freq(r_{ij})$ represents the frequency of the relation r_{ij} in the data set. Using the spreading activation formula (2), we calculate the activation value obtained for concept c_q. A high value obtained represent that this image is more relevant to this concept c_q. We can thus rank these images according to their relevance values of c_q from the spreading activation procedure and return the result to the users.

5. EXPERIMENT

To illustrate the improvement introduced by our proposed methods, we take the initial ranking from Google Image Retrieval result initial as the baseline and Figure 6 shows details of the returned ranking results for four *canine* classes.

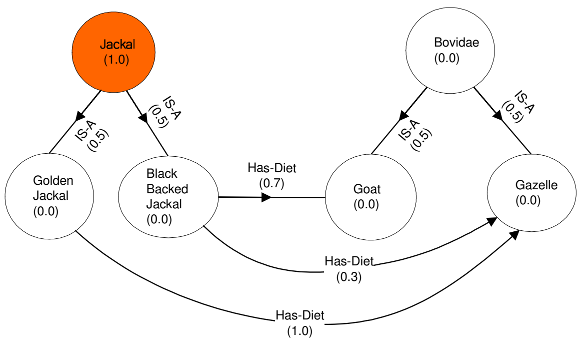


Figure 5: Initial stage of the spreading activation process.

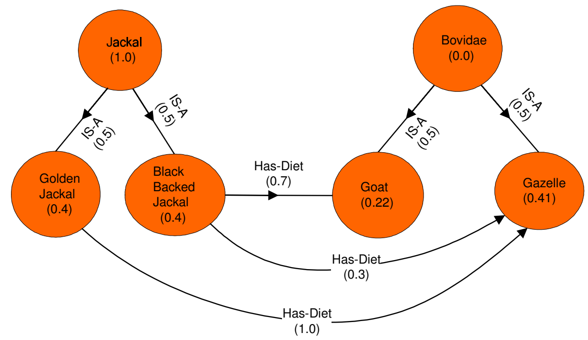


Figure 7: Final stage of the spreading activation process.

This figure plots the number of correct images in top N retrievals versus the total number of images returned (in ranking order). Therefore an optimal result would be a line with 45 degrees to the x-axis. From Figure 6 We can see that the performance improves by our proposed ontology model.

To give a comparison of the overall performance, we also evaluate the performance of our method using Average Precision (AP), which is defined as the averages (interpolated) precisions at certain recalls. As a most frequently used summary measure of a ranked retrieval, AP is defined as: $AP = \frac{1}{min(R,k)} \sum_{j=1}^{k} P(r_i) I_j$, where R is the total number of correct images in the ground truth, k is the number of current retrievals, and $I_j = 1$ if image ranked at jth position is correct and $I_j = 0$ otherwise. $P(r_i) = \frac{R_j}{j}$ is the interpolated precision. And {r,P(r)} are the available recall-precision pairs from the retrieval results. By using AP, the PR curve can be characterized in a scalar. A better retrieval performance, with a PR curve staying at the upper-right corner of the PR plane, will have a higher AP, and vice versa. In the current experiment, we set $j = 200$. In Table 1, we show the AP from both Google retrievals and ontology enhanced web image retrieval (OntoEnhanced). And it is clear that the proposed method performs consistently well.

6. CONCLUSIONS AND DISCUSSIONS

In this paper, we study the possibility of using ontology for effective web image retrieval. In view of the existing problems, we first introduce an approach of automatically constructing multi-modality ontology from Wikipedia and image features. And the built ontology is the basis of our solution and its quality greatly determines the final image retrieval performance. Then, we describe the spreading activation theory based procedure for inferencing on the ontology. We have implemented a prototype system that combines the two proposed methods for web image retrieval. The experimental results, on the *canine* category of the animal kingdom, demonstrate the efficacy of the ontology based web image retrieval. Finally, it is shown that our OntoEnhanced approach will help to improve the retrieval performance for images of various domains. The proposed approach largely dispenses with the conflict between cost and precision in ontology-based applications. We would like to conclude by drawing the attention of the readers to Figure 8. The results from our OntoEnhanced search for input query *bush dog* further show the power of ontology in better understanding of multimedia.

There are potential works for future development. Firstly, the construction of the semantic network with the multi-modality ontology is only a preliminary experiment. The weights of the relations are only determined by their frequencies. More advanced methods that consider the users' preference could be used for this task. Meanwhile, now we only compare with Google for image retrieval on the animal data sets. In our future work, we would extend to a larger scale image retrieval system and test on different data sets.

7. REFERENCES

[1] R. J. Anderson. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22:261–295, 1983.

[2] Y. A. Aslandogan, C. Thier, C. T. Yu, J. Zou, and N. Rishe. Using semantic contents and wordnet in image retrieval. In *Proc. of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 286–295, 1997.

[3] T. L. Berg and D. A. Forsyth. Animals on the web. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1463–1470, 2006.

[4] J. Contreras, V. R. Benjamins, M. Blázquez, S. Losada, R. Salla, J. Sevilla, D. Navarro, J. Casillas, A. Mompó, D. Patón, Ó. Corcho, P. Tena, and I. Martos. A semantic portal for the international affairs sector. In *Proc. of the 14th International Conference on Engineering Knowledge in the Age of the Semantic Web*, pages 203–215, 2004.

[5] Y. Gao and J. Fan. Incorporating concept ontology to enable probabilistic concept reasoning for multi-level image annotation. In *Proc. of the 8th ACM international workshop on Multimedia information retrieval*, pages 79–88, 2006.

[6] X. Jiang and A.-H. Tan. Ontosearch: A full-text search engine for the semantic web. In *Proc. of the 21st National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, pages 1325–1330, 2006.

[7] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

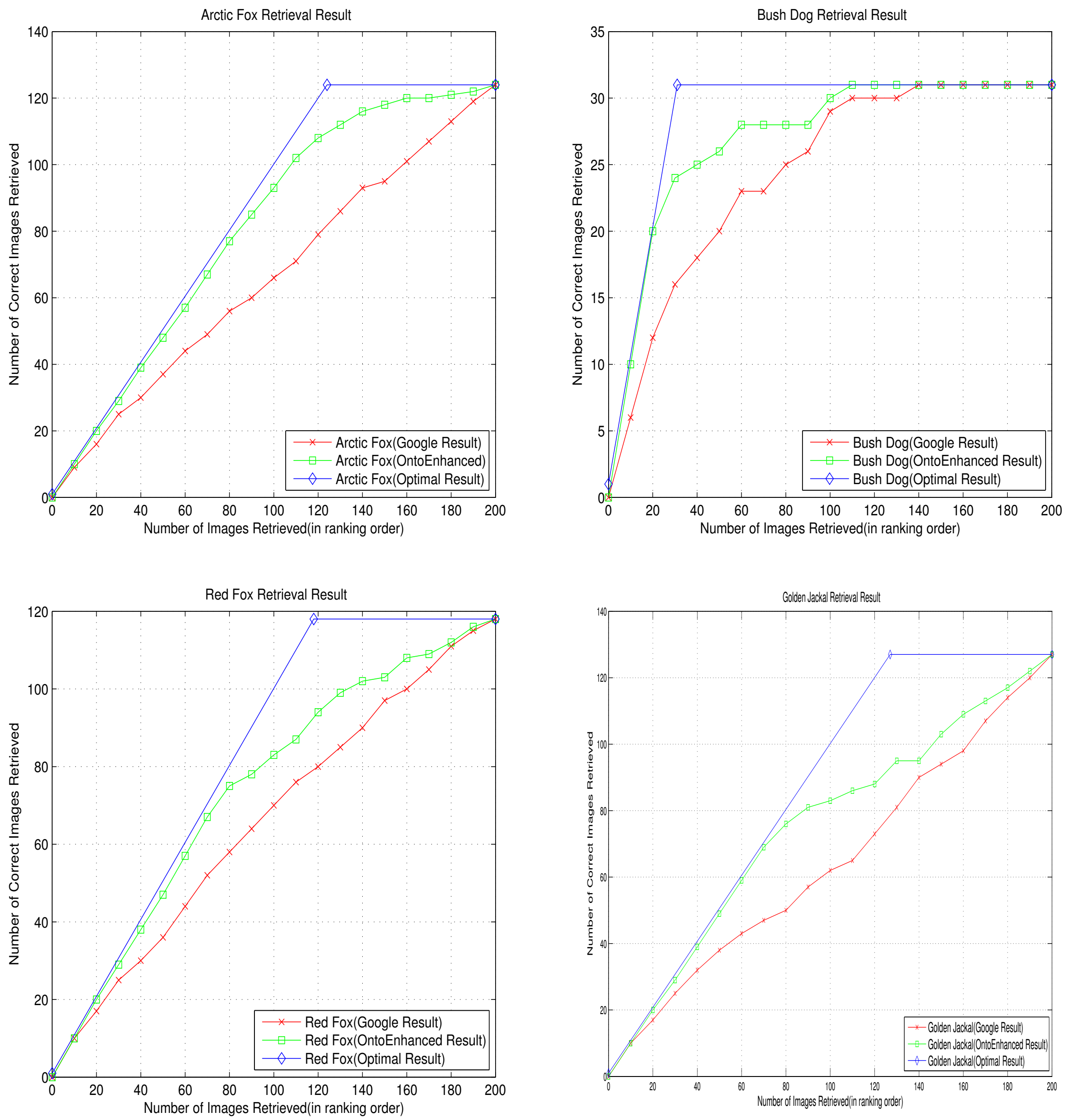


Figure 6: A comparison of image retrieval results between different approaches.

Table 1: Performance of the OntoEnhanced system for web image retrieval.

Class	Aardwolf	CapeFox	BushDog	ArcticFox	EthiopianWolf	Coyote	GrayWolf	GrayFox	FennecFox	SpottedHyena
Google	0.5801	0.4958	0.4695	0.715	0.7516	0.5042	0.7513	0.7183	0.8181	0.8365
OntoEnhanced	0.7167	0.5914	0.8611	0.9404	0.862	0.7694	0.8003	0.8767	0.8237	0.8591

Class	Dhole	RedFox	ManedWolf	BlackJackal	Bat-EaredFox	Dingo	KitFox	RedWolf	GoldenJackal	AfricanWildDog
Google	0.6342	0.744	0.7949	0.8872	0.7967	0.67	0.6698	0.7669	0.7092	0.7844
OntoEnhanced	0.7546	0.889	0.8662	0.9271	0.8856	0.7109	0.7981	0.8768	0.8793	0.8753

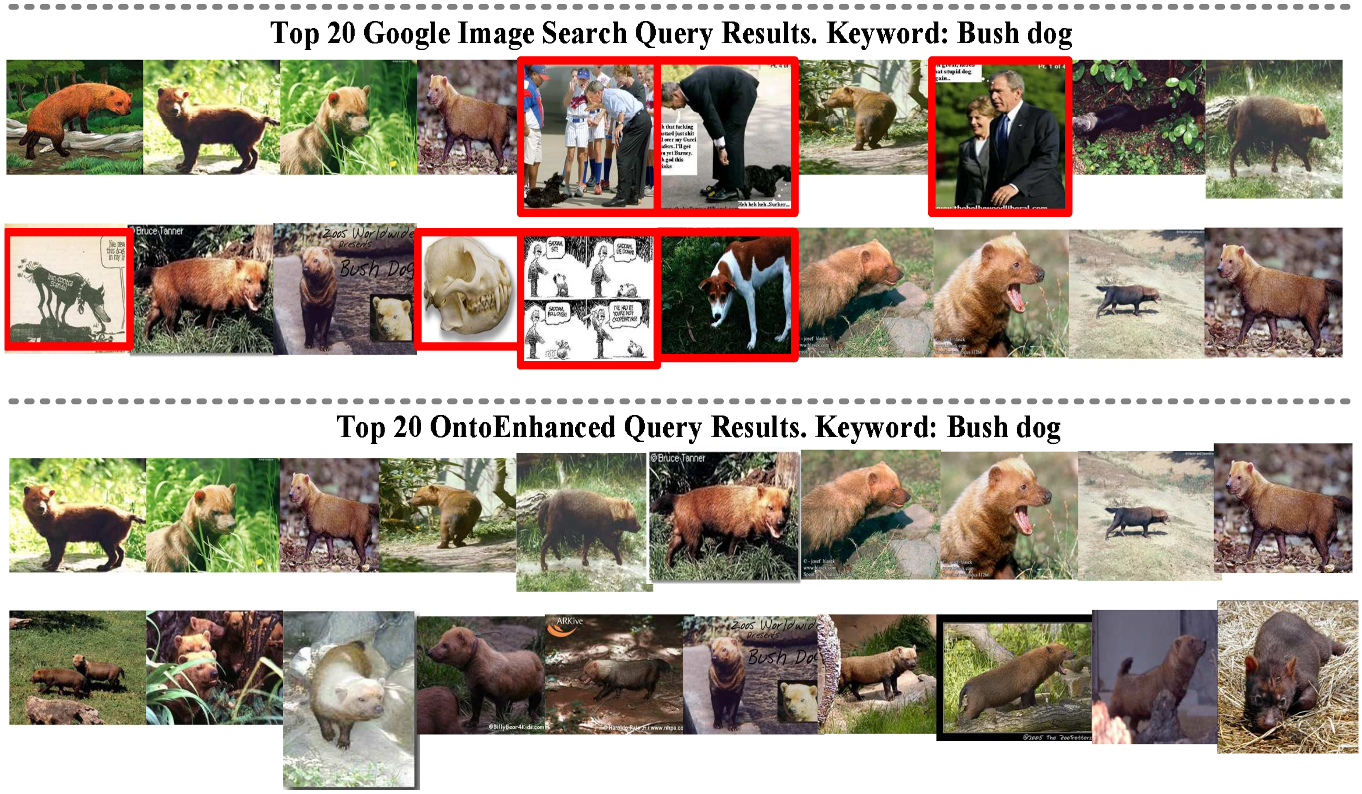


Figure 8: An example of the top 20 image returned by Google and OntoEnhanced given the Query *bush dog*.

[8] M. Marszalek and C. Schmid. Semantic hierarchies for visual object recognition. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.

[9] R. Mihalcea and A. Csomai. Wikify!: linking documents to encyclopedic knowledge. In *Proc. of the 16th ACM conference on Conference on Information and Knowledge Management*, pages 233–242, 2007.

[10] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[11] D. Milne, I. Witten, and D. Nichols. A knowledge-based search engine powered by wikipedia. In *Proc. of the 16th ACM conference on Conference on Information and Knowledge Management*, pages 445–454. ACM New York, NY, USA, 2007.

[12] J. van de Weijer and C. Schmid. Coloring local feature extraction. In *Proc. of the 9th European Conference on Computer Vision*, pages 334–348, 2006.

[13] H. Wang, S. Liu, and L.-T. Chia. Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches. In *Proc. of the 14th annual ACM international conference on Multimedia*, pages 109–112, 2006.

[14] H. Wang, S. Liu, and L.-T. Chia. Image retrieval with a multi-modality ontology. *Multimedia Syst.*, 13(5-6):379–390, 2008.

[15] X.-Y. Wei and C.-W. Ngo. Ontology-enriched semantic space for video search. In *Proc. of the 15th annual ACM international conference on Multimedia*, pages 981–990, 2007.

[16] F. Wu and D. S. Weld. Autonomously semantifying wikipedia. In *Proc. of the 16th ACM Conference on Information and Knowledge Management*, pages 41–50, 2007.

[17] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, 73(2):213–238, 2007.