Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

1-2010

Wikipedia2Onto: Building concept ontology automatically, experimenting with web image retrieval

Huan WANG
Nanyang Technological University

Xing JIANG
Nanyang Technological University

Liang-Tien CHIA
Nanyang Technological University

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

Citation

Wikipedia2Onto – Building Concept Ontology Automatically, Experimenting with Web Image Retrieval

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

Given its effectiveness to better understand data, ontology has been used in various domains including artificial intelligence, biomedical informatics and library science. What we have tried to promote is the use of ontology to better understand media (in particular, images) on the World Wide Web. This paper describes our preliminary attempt to construct a large-scale multi-modality ontology, called AutoMMOnto, for web image classification. Particularly, to enable the automation of text ontology construction, we take advantage of both structural and content features of Wikipedia and formalize real world objects in terms of concepts and relationships. For visual part, we train classifiers according to both global and local features, and generate middle-level concepts from the training images. A variant of the association rule mining algorithm is further developed to refine the built ontology. Our experimental results show that our method allows automatic construction of large-scale multi-modality ontology with high accuracy from challenging web image data set.

Povzetek: Prispevek opisuje izgradnjo velike multimodalne spletne ontologije AutoMMOnto.

1 Introduction

Real-world images always involve pictures with various backgrounds, object aspects, poses and appearances. Taking the animal classes in Figure 1 as an example, human-beings can easily differentiate the four classes. However, computers are not able to identify the difference in the same way. The varied background environment of the same *Arctic Fox* class can introduce great variance in global image features, while the subtle fur color difference between *Arctic Fox* and *Fennec Fox* makes it difficult to classify from local image features. It is also hard to identify the different distribution of colors over *Maned Wolf* or *Dhole* from spatial features. On the other hand, cues from the text on the corresponding web page could make a substantial contribution to the performance of image classification. For example, even a single keyword *Kashmir* could indicate the *Dhole* class, as *Kashmir* is the habitat of *Dhole*. Similar useful relationships which help to narrow down the final concepts include name, diet, and distribution relationships. Therefore, an effective way is to combine the images features with the text information for image retrieval, where ontology is utilized for this purpose.

Ontology, which clearly defines concepts and their relationships in a domain, has been widely used in many information retrieval fields, including document indexing, i.e., extracting semantic contents from a set of text document, image retrieval and classification, i.e., using concepts either from image features or surrounding text for content representation, and video retrieval, i.e.,

using text in video captions for semantic concept detection. Note that most of the approaches involve external lexical dictionary or online category as ontologies. They certainly improve the performance. However, they also introduce the following major questions:
1. Is ontology just as same as a hierarchical collection of concepts?
2. Ontology has to be manually built, which is extremely time consuming. Can it be done automatically?
3. Can Ontology be scalable when it is extended to large domains?
Through research on the use of ontology to better understand media information, we have provided our answers to the aforementioned questions:
1. Ontology is not just a hierarchical collection of concepts with parent-child relation. Details of the differences will be apparent as you read the details in this paper.
2. We do agree that one main difficulty that hedges against the development of ontology approaches is the extra work required in ontology construction and annotation. But there is a hope, and this paper describes our original attempt to use both structural and content features of Wikipedia to build a proposed hierarchy with not only hyponymy(*is-a*) or meronymy(*part-of*) relationships but also more real-life relationships. Therefore, the resulting semantic concept hierarchy of the built ontology, called AutoOnto, is consistent with

real world knowledge and can be used to map text information on the web page to detect semantic concepts. 3. Scalability is indeed a problem when a single party or a consortium tries to create a whole ontology structure. However, the problem could be solved when we can import existing ontologies or newly created ontologies to merge with other ontologies. The important issue here is to understand and handle the similarities/dissimilarities of concepts existing in the respective ontologies, which is current interest to relevant groups in AI and ontology-related areas.

Figure 1: An example of web image classes in our data set. Even though these images are portraying 4 difficult animal classes, it is easy for human-beings to identify the classes: *Arctic fox* has light-colored fur; *Fennec fox* has a pair of grotesque ears and ET-Style face; *Maned wolf* is featured with its black long legs; and *Dhole* has white fur spreading from its jaw to abdomen. However, it is not easy for image processing approaches to tell the classes apart due to the lack of discriminative low-level features.

Note that we have manually built a text ontology, called ManuOnto, and shown that it can effectively help machine understand multimedia in a better way in our previous work [9]. In this paper, we first show that AutoOnto captures the relationships between concepts as well as, if not better than the manually-built ontology with bigger knowledge coverage and higher efficiency. Then, we train classifiers according to our 164 dimensional features (SIFT with opponent color angle) and generate middle-level concepts from the training result and integrate the AutoOnto to form AutoMMOnto (Auto Multi-Modality Ontology). The MAP results of our experiment on Google (top 200 retrievals) Image search, AutoOnto and AutoMMOnto (AutoOnto+ visual descriptions) are 0.7049, 0.8942 and 0.9125 respectively. We have therefore shown that our method allows automatic construction of large-scale multi-modality ontology with high accuracy from a challenging web image data set.

Our contribution in this paper is concluded as follows: We propose a method to build large scale concept ontology from Wikipedia in a cost effective way. The generated ontology is able to extract additional information from the web pages and increase the concept detection accuracy. We also propose an association rule mining algorithm to refine relationships in the ontology. The resulting relationship set are more concise with higher precision.

The rest of this paper is organized as follows. Section 2 introduces the related works. Section 3 discusses the Wikipedia category and structure information. Section 4 discusses how we use Wikipedia to automatically build the concept ontology. In Section 5 we propose an association rule mining algorithm to discover the key semantic relationships. Experiment results on our collected web image database are given in Section 6. We conclude this paper in Section 7.

2   Related works

Due to the dependency between the external knowledge source and semantic concepts, the chosen knowledge source will affect the derived concepts and relationships for ontology construction, which is used to classify the existing ontology construction approaches.

WordNet [1], developed at Princeton University, has been commonly used as such a lexical dictionary as it is able to group words into sets of synsets. The structured network makes it easy to derive semantic hierarchies for various domains. Some typical examples which practise WordNet directly as an ontology for object recognition and video search include [2, 3, 4]. However, the reason why WordNet works fine in these experiments is that only common concepts (e.g., car, dog, grass, tree) and relationships (hypernymy, meronymy) are employed. If concepts and relationships outside the scope of WordNet, they will not be included and cannot be utilized. For example, WordNet has limited coverage of less popular or more specific concepts like "mountain bike" or "bush dog". This limitation decides that WordNet only works on sparse or general concept domains. Also, WordNet is also disconnected from the update of natural language vocabulary which changes with almost everyday. Therefore, it is not able to work on domains with novel topics and concepts.

Besides the above approaches, WordNet has also been used for assisting ontology building. For example, in [16], WordNet is dynamically included to extend a knowledge acquisition tool CGKAT. Particularly, top-level concept of WordNet ontology is subordinated into Sowa Ontology for finalizing an ontology. Similar approach is also proposed in [17], which incorporates general purpose resources like WordNet and open web directory to build large scale ontology for document indexing and categorization. Particularly, it takes two steps to build an ontology. Firstly, an initial ontology consists of two sub-trees from both the web directory and WordNet synset graph respectively. Then, it iteratively fills the gaps and enriches the existing ontology with new concepts and relationships until the ontology is verified by domain expert to be usable. As a result, the ontology building process is only semi-automatic. Also, the proposed method lacks support from solid experiment results and the performance in real application is yet to be evaluated.

With the rapid development of Internet, online categories seem to be a better choice for ontology construction, especially when some popular online categories also provide easy access [5, 6]. By indexing a

huge number of web pages/topics, online categories cover most real world objects, activities, news and documents in a timely manner. Besides the hierarchical structure offered by these categories, web page submitters and category indexers also provide more related concepts with varied relationships, which further extend the coverage. Some approaches which use online resources to construct knowledge base include [7], wherein specific domain knowledge of animal is extracted from online animal category and image features to construct ontology for web image classification. The application indicates that with the evolution of ontology-based applications, finding a proper knowledge source has become an important issue.

Among the existing online categories, there is an increasing interest in using Wikipedia as the resource for knowledge mining. In [18], a refinement on the Wikipedia category network is implemented step by step to generate taxonomy of *is-a* semantic links. All syntax-based, connectivity-based, lexicon-syntactic based and inference-based methods are used to remove noisy links and set up correct *is-a* links. To compare and analyse the performance of the Wikipedia-based ontology, the manually built ontology ResearchCyc [19] and WordNet are used as the performance baseline. The evaluation shows that the automatically-built taxonomy is comparable with the two existing ontologies. However, the construction is discontinued at the level of domains. While the authors put more emphasis on drawing out a taxonomy of 105,418 *is-a* links, the broad coverage also makes the taxonomy insensitive to specific applications, as different applications need different emphasis on the domain knowledge.

Other than using external resources to construct ontology, existing large-scale ontology constructions usually involve mass manual work. For example, LSCOM [8] aims to design a taxonomy with a coverage of around 1,000 concepts for broadcast news video retrieval. This approach is hampered by the tens of millions of human judgments required, which has been proved to be very ineffective and costly. In a word, most ontology constructions are either constructed on dependent domain or still involve mass manual work. And even those semi-automatic construction processes rely heavily on external knowledge resources, like the aforementioned lexical dictionary and online categories. Another disadvantage is apparent as the important merit of either dictionary or category is a hierarchical graph which connects concepts together. As a result, only shallow relationships like hypernymy/hyponymy(*is-a*) or meronymy(*part-of*) could be mined. These relations are not sufficient enough to support information mining from web images, which are usually attached to web pages with text information. Mining through such kind of text corpus involves more than the aforementioned semantic relationships. An ontology with enriched knowledge provides more discriminative information in web image retrieval, classification and annotation.

Referring to the existing work, we can see that an advanced ontology for multimedia research and applications should meet the following requirements: 1) The ontology should be constructed automatically, so that when it is applied to extended domains, the scalability will not become the bottleneck. 2) The ontology should involve more than domain-specific concepts. Also, besides *is-a* or *part-of* relationships, deeper semantic relationships should also be included so that the ontology is a better imitation of human general knowledge.

3 Wikipedia concepts and structure

Wikipedia is by far the biggest online free encyclopedia. It provides definitions for more than 2 million words and phrase concepts. This number is still growing as Wikipedia is based on online collaborative work and anyone can freely access, create and edit the page content of each concept. This open feature makes Wikipedia an up-to-date knowledge source, where even the latest concepts can be found. It also covers many concepts which are not commonly used and included in other electronic lexical dictionaries. In the following subsections, we will introduce some of Wikipedia's features which make it suitable for ontology construction.

3.1 Wikipedia category

The underlying structure of Wikipedia can be described in two network graphs: category graph and article graph. In both graphs, nodes represent articles and edges represent links between articles. Basically, all the Wikipedia web pages are put into a subject category according to general knowledge. This structure is depicted as the category graph which has been proved to be a scale-free, small world graph by graph-theoretic analysis[20]. The category graph is formed following the taxonomy of concepts. Therefore, the links in category graph indicate either *is-a* or *part-of* relationships between the two connected concepts (a sample of the category graph is given in Figure 2). In this sense, the semantic relationships provided by the category graph is quite similar to the relationships provided by WordNet. When referring to specific article, the Wikipedia classification is listed in a separate Categories section. Besides the category graph, there is also an article graph which indicates the cross-references between Wikipedia web pages. In particular, the articles are nodes of the graph, which are hyperlinked to corresponding Wikipedia articles. These links indicate a direct semantic relationship between the two connected concepts. Compared with WordNet which mainly organizes word concepts according to synset, Wikipedia category provides a more formal classification of concepts. As a result, the extracted concepts and relationships are closer to a formal ontology with various semantic relationships.

Figure 3: An example of Wikipedia web page with corresponding extracted concept. The extracted concept definition is: (define-concept concept_gray_wolf(or Some animal(all hasName(or gray_wolf timber_wolf wolf))(all has Distribution(or Canada Ireland Kazakhstan the_Middle_East North_America Russia Europe the_United_States India Asia Finland))(all hasDiet (or Herbivore Coyote American_Bison Deer Caribou Moose Yak Ungulate Rodent)))).

3.2 Wikipedia web page

In Wikipedia, each web page defines one concept according to general knowledge. Disambiguation is removed by separating different senses in different web pages. The searching in Wikipedia is straightforward as each web page has already been associated with the keywords. In most cases, the page title is the indexed keywords. The text information on the web page is divided into sections. Each section describes one aspect of the concept in details. Taking the concept Aardwolf as an example (see Figure 3), the main web page content includes physical characteristics, distribution and habitat, behaviour, and interaction with humans. From the viewpoint of concepts, each section is connected to the main concept with semantic relationships depicted as section titles. A concept graph is easily drawn from this web page content structure. On the right, the web page also provides a section of Scientific classification, which lists the zoology taxonomy of the animal. By integrating different concepts under the same domain *Animalia*, a big hierarchy picture can be easily constructed with the concepts positioned under corresponding branches. Compared to our manually built Animal Domain Ontology [7], the hierarchy generated from Wikipedia Scientific classification is more formally defined, and is considered to contain rigid domain information.

3.3 Concept coverage

In comparison to WordNet, whose total number of words is limited to around 147,278, Wikipedia certainly contains more information. For our case, only 12 out of the 20 class names are covered by WordNet. Class names such as *African wild dog, bat-eared fox, black jackal, bush dog, cape fox, Ethiopian wolf, fennec fox, golden jackal* are all missing from WordNet. Such limitations make WordNet an incomplete appropriate resource for ontology learning. On the contrary, Wikipedia is more suitable for this task. The total number of words has

Figure 2: An example of the *Canines* Wikipedia category.

reached 2 million and it keeps increasing significantly daily. It can cover almost all the relevant concepts in our experiment.

4 Automatic ontology construction – Wikipedia2Onto

In this section we discuss the construction of our multi-modality ontology. Similarly to our previous manually construction process, the automatic process includes 3 steps. Particularly, the key concepts in the animal domain and the taxonomic relations are firstly extracted from Wikipedia. Then, the narrative descriptions of particular animals, including relevant concepts and non-taxonomic relations, are extracted. Finally, the *visual descriptions* of each concept are added. Note that we do not use the XML corpus provided by Wikipedia directly for construction. Instead, we use a web page crawler to download relevant concept web pages before ontology building in advance. Such an approach makes it more flexible to build ontology for specific domain. Meanwhile, a dynamic connection to Wikipedia can ensure "freshness" of our concepts as Wikipedia web pages are edited from time to time.

4.1 Key concepts and taxonomic relations extraction

Wikipedia has provided an entire category of many meaningful concepts, which is formed according to hypernymy relationships between concepts. In other words, Wikipedia category provides taxonomy of general concepts in natural language, which is much more precise than our in-door manually built one. Therefore, our *Animal Domain Ontology*, which is used to describe the taxonomy information of animal concepts, can be directly obtained from Wikipedia category. However, as the Wikipedia concepts under *animal* domain have some special content features, we use the *Scientific Classification* entry on each concept page as a shortcut.
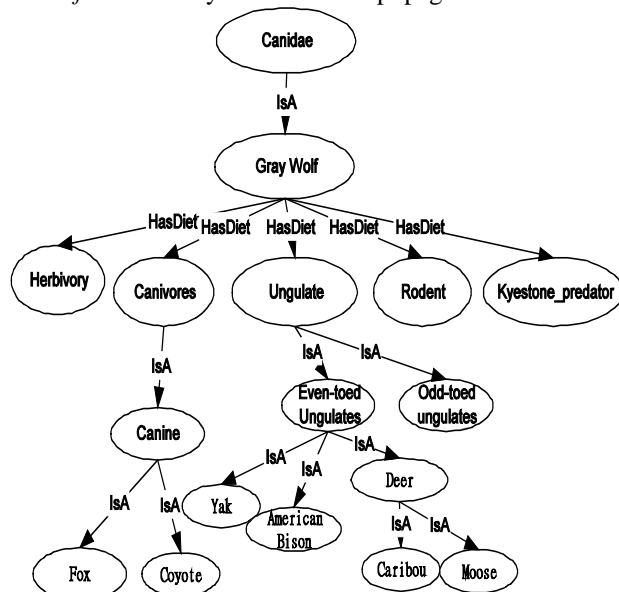

Figure 4: Knowledge resource structure in our system.

This entry provides animal taxonomy in a top-down manner, from *Kingdom*, *Phylum*, *Class*, *Order, Family, Subfamily, Genus* to Species. We then extract the hierarchy structure from this entry and form our *Animal Domain Ontology*. For example, *Phylum* is defined as a sub–class of *Kingdom*, while *Class* is defined as a subclass of *Phylum*. Since our ontology is only defined for general web image classification, we stop at *Family* level and do not go beyond *Subfamily*. Taking *Aardwolf* as an example, this concept belongs to the family of *Hyaenidae*. So when an input query suggests a concept of *Hyaenidae*, Aardwolf will also be considered as a matched concept.

4.2 Narrative descriptions extraction

In the definition of *ontology*, what is alluded to but not formally stated is the modelling of concept relationships. In order to show that ontology is more than a set of related keywords, we have to prove that every concept in the ontology is different from a plain word. It should be understood as concepts supported by structures. When building the *Textual Description Ontology*, our main concerns are twofold: an ontology, which depicts the real world, should contain more descriptive concepts and relationships. These relationships convey general knowledge according to domain knowledge. On the other hand, the related concepts should contain a hierarchical structure, so that when we do concept inference, additional facts could be generated. Here is an example to illustrate the above concerns. *South Africa* is where the species *cape fox* lives. Therefore, *South Africa* is linked to *cape fox* with a named relationship *hasDistribution*. Given two other relations *Zimbabwe* is a part of *South Africa* and *South Africa* is part of *Africa*, one could reasonably infer that *cape fox* can also be found in *Zimbabwe*. And this possibility increases when additional information matches. Therefore, the first step is to find all the important terms. Some pre-process includes crawling Wikipedia web page of relevant concept and using HTML parser to filter irrelevant HTML codes. After that, we analyse the web page content to extract useful concepts and relationships. It is worth noticing that at the beginning of each web page, where a short paragraph is given as a brief introduction of the particular concept, some words are emboldened as alternative name or synonymy to the main concept. By extracting these words, a synonymous set is first constructed for the original concept. We use a *hasName* relationship to link it to the original concept. This relationship extends the naming information. In the next step, by analysing HTML tags of document title, section title and links to other pages, we locate the title of each section. Before we look into the details of the section content, we exam the section title to see if it contains relevant semantic relationships, like information about *Distribution*, *Habitat*, *Diet*, etc. Once the relevant keywords are discovered in the section title, we look into the details of the section and find candidate concepts for that particular relationship. Candidate concepts are defined as those that have their own Wikipedia web pages. For the normal

plain text on the Wikipedia web page, we believe it is of trivial importance, thus has less contribution to the concept detection. Based on this assumption, we extract a set of concepts from the section for each relationship. While not all the candidate concepts are correct, an association rule mining is discussed later to improve the accuracy of the generated ontology.

After the relationships and related concepts are collected, we do further hierarchical construction among all the concepts. This step is done based on the Wikipedia category structure, which offers a systematic categorization of all the concepts. The category information is listed 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 serve for Wikipedia administration purposes, such as *Wikipedia administration*. We remove these categories and keep the rest, which follow 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. To evaluate the performance of the proposed ontology system with other textual aware methods, we also follow the text processing part of [10] and use Latent Dirichlet Allocation(LDA) to find 10 latent topics from the web page text. And we take the top 20 words from each latent topic as the topic representation. However, the resulting clusters of words do not show explicit semantic meanings. We presume that it is due to the relative smaller size of text corpus. Therefore, the ontology approach is more appropriate on our median-sized data set.

4.3   Visual descriptions extractions

In this section we discuss the visual description features for our concept ontology. We collect a median size collection of 4,000 animal web images together with the corresponding web pages as our experiment data set. More specifically, the data set contains 20 animal categories under the domain of canine. 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 researches. Our aim is also to show that instead we follow the object detection techniques whose superiority has been proved in the latest researches[11]. We first use Harris-Laplace detector[12] which is scale invariant and detects corner-like regions in the images as interest point and then use SIFT[13] descriptor to represent the shape information around the interest point. Color descriptor is also combined with SIFT descriptor. A 20 by 20 image patch around the centre 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. The final descriptor is a vector of

dimension 164, where 128 dimensions are from SIFT descriptor and 36 dimensions are from *opponent angle* descriptor. We build a vocabulary of 1,000 visual words based on k-means clustering result of feature vectors from all images. 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, half of the data set is used as training sample, which is of size 2,000. The training set is further divided into 5 parts for cross validation. After training, the relations between image feature concepts and the *animal* concepts are obtained.

After construction, we use association rule mining to refine the initial ontology.

5   Association rule mining for ontology

Wikipedia is an online collaborative work and the content is maintained by users, therefore a certain level of inherent noise must be expected. When we extract real-world relations besides hypernymy and meronymy relations into the ontology, we are extracting those relations from the Wikipedia web pages with text analysis techniques. A small set of wrong relations could be extracted either due to the complexity or correctness of the texts and the strategy we used for relation extraction. For association rule mining[14], the research has evolved from a flat structure with a fixed support value to variances that consider complex tree or graph structure with different support values. In order to enhance the correctness of semantic relations extracted, we develop a variant of association rule mining method which considers the hierarchical structure of the ontology and propose a new quality measure called Q measure for relation pruning.

Here, we use Figure 5 to illustrate the idea of the *Q* measure. We can see concept *Even-toed Ungulates* has three children in the ontology, namely *Deer*, *Yak*, and *American Bison*. If the relation *Gray_Wolf hasDiet Even-toed Ungulates* is correct, the three relations *Gray_Wolf*
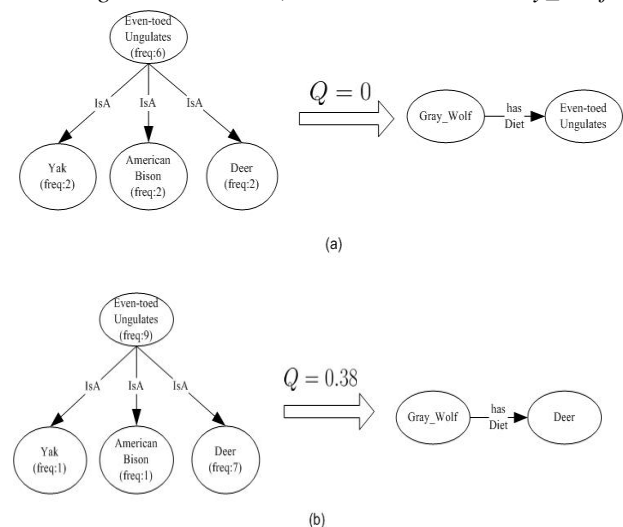

Figure 5: An example for association rule mining.

hasDiet Deer, *Gray_Wolf hasDiet Yak*, and *Gray_Wolf hasDiet American Bison* should also be correct if a minimum support level is present. Given a sufficient large number of documents collected, the three relations should have the same frequencies (i.e., the expected value of 1/3). But in realty and with a smaller number of documents, the three relations have different frequencies. We therefore could compute a variance-like value *Q* by

$$Q = \sum_{i=1}^{N}[\frac{freq(C_i)}{freq(R)} - \frac{1}{N}]^2 \qquad (1)$$

where C_i represents a child rule of a generalized rule R, and N is the number of children rules of R.

For those relations with parent concepts, they would have a lower Q value although they have high frequencies. We therefore can efficiently remove those relations by looking at the Q value and the predefined support threshold.

6   Experiment result

In this section, we evaluate the performance of our approach by using the built AutoOnto and AutoMMOnto for image retrieval.

The matchmaking of concept ontology is defined as a process that requires the user specified domain concept repository to take an image's detected concept as the input, and return all the matched domain concepts which are compatible with the concept generated from the input concept. From the matchmaking result we can conclude which predefined concept the generated image concept corresponds to and what relationship can be find between two given concepts. In this step, reasoners(semantic matchmakers) are used to derive additional facts which are entailed in any optional ontologies and predefined rules, through process and reason over the knowledge encoded in the ontology language. We use both the description logic reasoner RACER[15] and an enhanced ranking algorithm [9] in the experiment. The matched concepts are attached with the web images as semantic labels.

There are 20 classes of web images in our database and each class has 200 web images downloaded from Google Image Retrieval. The performance is computed using Average Precision (AP), which is defined as the average (interpolated) precisions at certain recalls

$$AP = \frac{1}{\min(R,k)}\sum_{j=1}^{k}P(r_i)I_j ,$$

where R is the total number of correct images in the ground truth, k is the is number of current retrievals, $I_j = 1$ if image ranked at j^{th} 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 by 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. As MAP is sensitive to the entire ranking with both recall and precision reflected in this measurement, we will also give Mean Average Precision (MAP)

We compare our result to both the Google Image Retrieval results and the manually built ontology results(namely ManuOnto and ManuMMOnto). The corresponding comparisons are shown in Table 1 and Table 2 respectively, where the Average Precision (AP) values for each class using different approaches are presented. From Table 1, we can conclude that text ontology improves the retrieval performance by formulating the text information into structured concepts. From Table 2, we can observe that the AutoMMOnto approach gives comparable performance to the ManuMMOnto approach. And in most classes, AutoMMOnto generates even better results by extracting more concepts from the web page text. And the MAP of the Google, ManuMMOnto and AutoMMOnto are 0.7049, 0.8942 and 0.9125, respectively. The result of MAP also shows an overall improvement. It is worth adding, AutoMMOnto requires minimal level of human involvement: Only the main domain concepts, which are the image classes in our case, are given by users according to experimental domain to build up the whole concept hierarchy in the domain. The result is encouraging, as it proves that it is viable to build large-scale concept ontology from Wikipedia automatically for effective web image retrieval. Ranking results from several sample classes are also shown in Figure 6.

7   Conclusion and future works

In this paper we have proposed Wikipedia2Onto, an approach that uses the content and structure features of the online encyclopaedia Wikipedia to build large-scale concept ontology automatically. The constructed ontology has automatically extracted more descriptive semantic relationships than most existing ontologies. More importantly, this ontology is a ready structure that can be used in semantic inference. Through association rule mining, our approach has detected 743 concepts with high accurate corresponding relations.

Finally, it is shown that our approach will help to improve precise retrieval for images (with free text information) for various domains. The proposed approach largely dispenses with the conflict between cost and precision in ontology-based applications. We would also like to conclude by drawing the attention of the readers to Figure 7. The results from our AutoMMOnto search for "wild dog in Kashmir region" further show the potential of ontology in the better understanding of multimedia.

References

[1]   Y. A. Aslandogan, C. Their, C. T. Yu, and N. Rishe (1997) Using semantic contents and wordnet in image retrieval. In SIGIR'97: Proceedings of the 20th annual international ACM SIGIR conference

on research and development in information retrieval, pp. 286 - 295, New York, USA.

[2] M. Marszalek and C. Schmid (2007) Semantic hierarchies for visual object recognition.) In *CVPR'07: Proceedings of IEEE conference on computer vision and pattern recognition*, pp. 1-7, Minnesota, USA,.

[3] X.-Y. Wei and C.-W. Ngo (2007) Ontology-enriched semantic space for video search. In MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia, pp. 981-990, New York, USA.

[4] A. Popescu, P.-A. MoÄellic, and C. Millet (2007) Semretriev: an ontology driven image retrieval system. In CIVR'07: Proceedings of the 6th ACM international conference on image and video retrieval, pp. 113 -116, New York, USA.

[5] L. Khan and F. Luo (2002) Ontology construction for information selection. In ICTAI'02: Proceedings of 14th IEEE International Conference on Tools with Artificial Intelligence, pp. 122-127, Washington, USA.

[6] Y. Labrou and T. Finin. Yahoo! as an ontology: using Yahoo! categories to describe documents. In CIKM'99: Proceedings of the 8th international conference on Information and knowledge management, pp. 180-187, New York, USA, 1999.

[7] H. Wang, S. Liu, and L.-T. Chia (2006) Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches. In MULTIMEDIA'06: Proceedings of the 14th annual ACM international conference on Multimedia, pp. 109-112, New York, USA.

[8] M. Naphade, J. Smith, J. Tesic, S. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis (2006) Large-scale concept ontology for multimedia. IEEE MultiMedia Magazine, 13(3), pp. 86-91.

[9] H. Wang, L.-T Chia and S. Liu (2007) Semantic Retrieval with Enhanced Matchmaking and Multi-Modality Ontology. In ICME'07: Proceedings of IEEE International Conference on Multimedia and Expo, pp. 516-518, Beijing, China.

[10] T. L. Berg and D. A. Forsyth (2006) . Animals on the web In CVPR'06: Proceedings of IEEE conference on computer vision and pattern recognition, pages 1463-1470, New York, USA, 2006.

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

[12] K. Mikolajczyk and C. Schmid (2004) Scale & invariant interest point detectors. In International Journal of Comput Vision, 60(1), pp. 63-86.

[13] D. Lowe. (2003) Distinctive image features from scale-invariant keypoints. In International Journal of Computer Vision, volume 20, pp. 91-110.

[14] R. Srikant and R. Agrawal (1995) Mining Generalized Association Rules. In VLDB'95: Proceedings of the International conference on very large data bases , pp. 407-718, Zurich, Switzerland.

[15] V. Haarslev and R. Moller (2001) Racer system description. In IJCAR'01: Proceedings of International Joint Conference on Automated Reasoning, pp. 701-705, Siena, Italy.

[16] P. Martin (1995) Using the WordNet Concept Catalog and a Relation Hierarchy for Knowledge Acquisition. In Proceedings of the 4th Peirce Workshop, Santa Cruz, USA.

[17] V. Varma. (2002) Building large scale ontology networks. In Proceedings of Language Engineering Conference, pages 121-127, Hyderabad, India.

[18] S. Ponzetto and M. Strube (2007) Deriving a Large Scale Taxonomy from Wikipedia. In AAAI'07: Proceedings of the 22nd National Conference on Artificial Intelligence, pp. 22-26, Vancouver, CA, 2007.

[19] D. Lenat and R. Guha (1989) Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

[20] T. Zesch and I. Gurevych (2007) Analysis of the Wikipedia Category Graph for NLP Applications. Proceedings of the TextGraphs-2 Workshop, NAACL-HLT, pp. 1-8, New York, USA.

Table 1: Performance of image classification on single-modality text ontology

Class	Aardwolf	CapeFox	BushDog	ArcticFox	Ethiopian Wolf	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
ManuOnto	0.6209	0.5446	0.7881	0.7905	0.0.844	0.5421	0.7196	0.6336	0.8145	0.8683
AutoOnto	0.6472	0.5231	0.8422	0.7983	0.7973	0.496	0.7316	0.6465	0.8214	0.9024
Class	Dhole	RedFox	ManedWolf	BlackJackal	Bat-Eared Fox	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
ManuOnto	0.6598	0.781	0.8565	0.8805	0.7914	0.6799	0.6844	0.715	0.7252	0.7723
AutoOnto	0.6835	0.7522	0.8193	0.8959	0.8396	0.7196	0.6791	0.8175	0.7528	0.7869

Table 2: Performance of image classification on multi-modality ontology

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
ManuMMOnto	0.8332	0.8911	0.8087	0.9955	0.9218	0.9058	0.8267	0.935	0.857	0.9391
AutoMMOnto	0.8552	0.8835	0.9302	0.9938	0.9447	0.884	0.8561	0.9766	0.8981	0.942
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
ManuMMOnto	0.8184	0.966	0.9508	0.9498	0.9134	0.8108	0.9483	0.8537	0.8962	0.8627
AutoMMOnto	0.8535	0.9526	0.938	0.9555	0.9333	0.8334	0.8821	0.9034	0.9166	0.9185

Top Retrievals from Google Image Search using keyword: "dhole"

Top Retrievals from Google Image Search using keyword: "wild dog in Kashmir region"

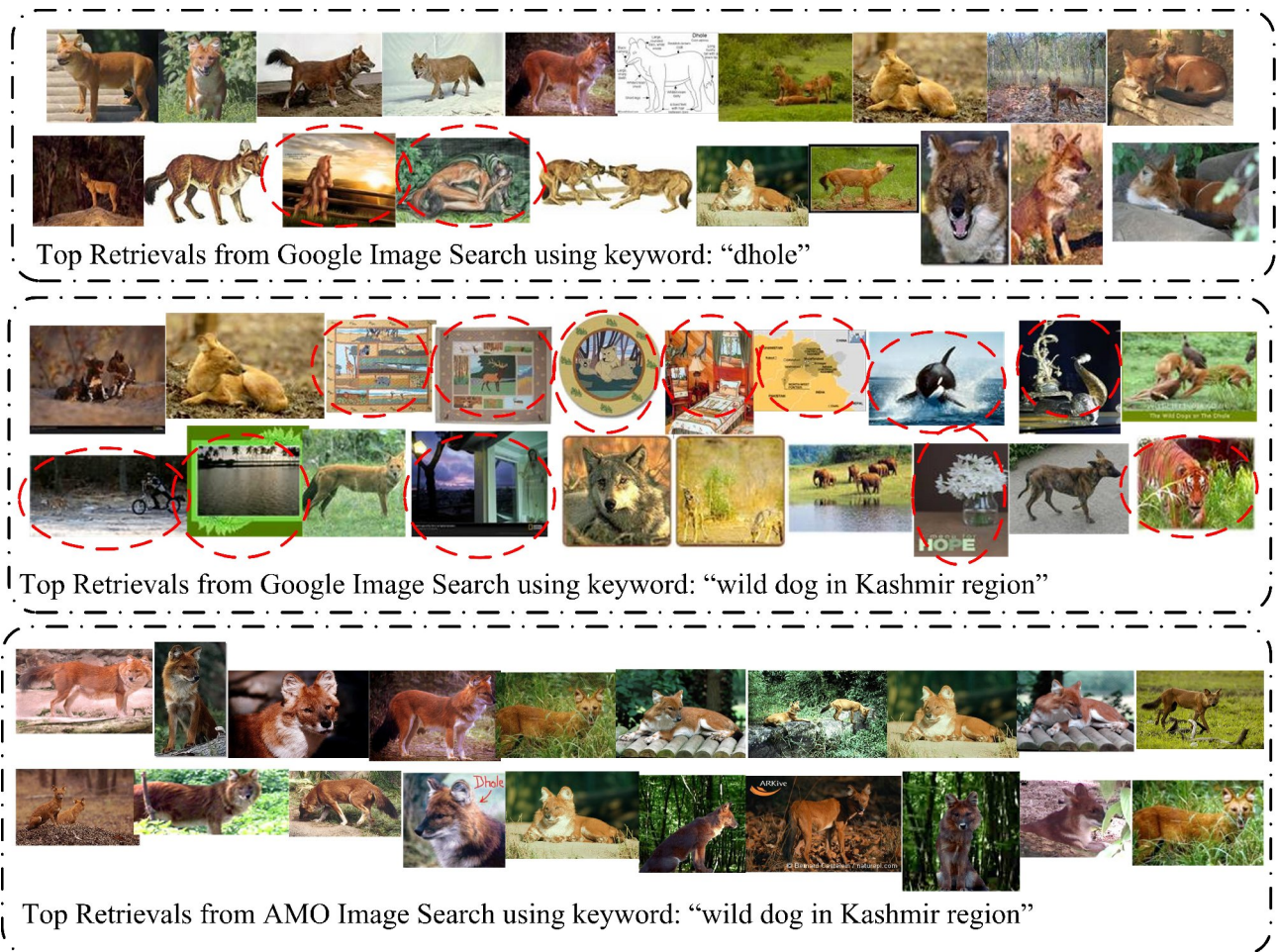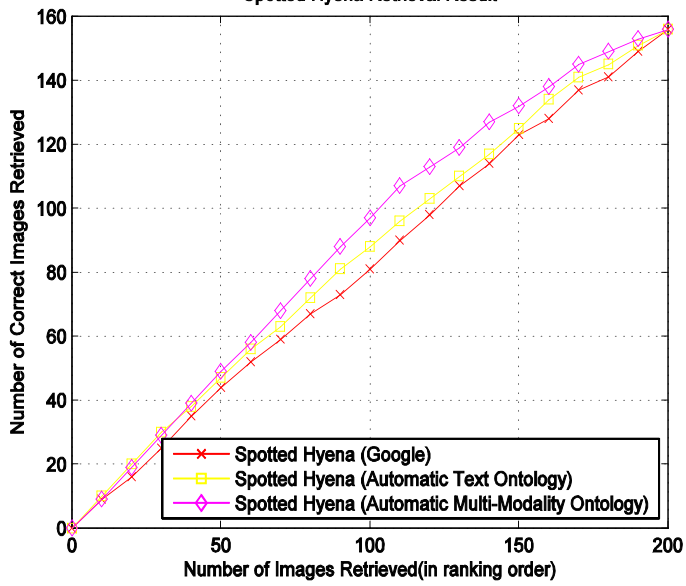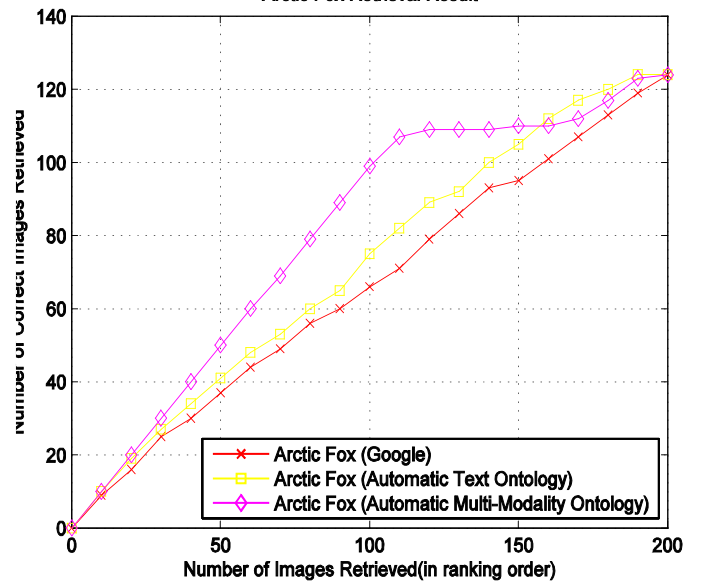Top Retrievals from AMO Image Search using keyword: "wild dog in Kashmir region"

Figure 6: An example of web image classes in our data set. Different results returned for different keywords but it is the same animal from the canine family: Dhole - a wild dog in the Kashimar region.
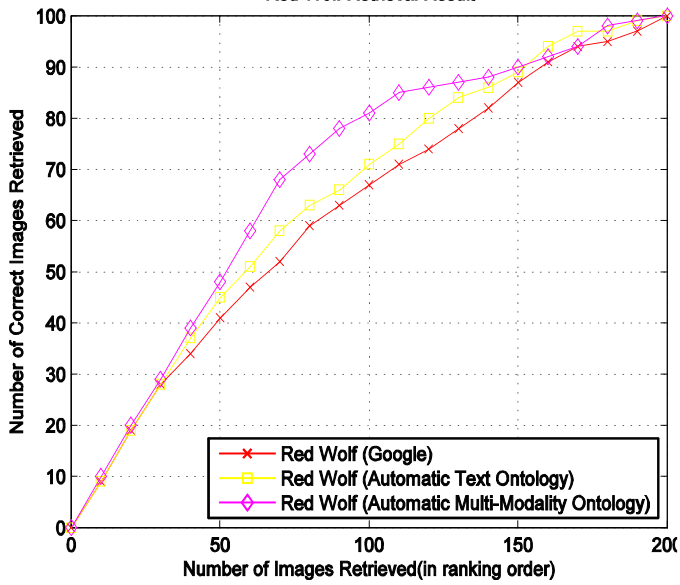
Spotted Hyena Retrieval Result

Arctic Fox Retrieval Result

Red Wolf Retrieval Result

Bush Dog Retrieval Result