Multipedia: Enriching DBpedia with Multimedia Information

Andrés García-Silva
Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
hgarcia@fi.upm.es

Max Jakob, Pablo N. Mendes and Christian Bizer
Web-based Systems Group, Freie Universität Berlin, Germany
first.last@fu-berlin.de

ABSTRACT

Enriching knowledge bases with multimedia information makes it possible to complement textual descriptions with visual and audio information. Such complementary information can help users to understand the meaning of assertions, and in general improve the user experience with the knowledge base. In this paper we address the problem of how to enrich ontology instances with candidate images retrieved from existing Web search engines. DBpedia has evolved into a major hub in the Linked Data cloud, interconnecting millions of entities organized under a consistent ontology. Our approach taps into the Wikipedia corpus to gather context information for DBpedia instances and takes advantage of image tagging information when this is available to calculate semantic relatedness between instances and candidate images. We performed experiments with focus on the particularly challenging problem of highly ambiguous names. Both methods presented in this work outperformed the baseline. Our best method leveraged context words from Wikipedia, tags from Flickr and type information from DBpedia to achieve an average precision of 80%.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*Knowledge acquisition*

General Terms

Algorithms,Design,Experimentation

Keywords

Ontology, Multimedia, DBpedia, Linked Data

1. INTRODUCTION

Enriching knowledge bases with multimedia information makes it possible to complement and improve results of knowledge consuming tasks including question and answering systems and recommendation processes among others. Multimodal knowledge bases have been successfully used in the past for several knowledge consuming tasks including semantic browsing of video collections [3] and query interpretation for multimodal information retrieval [20], among others. However, retrieving relevant images from the Web for instances in a knowledge base is not a trivial task.

The prevalent information retrieval paradigm on the Web is keyword-based search. Naturally, multimedia content has been particularly challenging in this context, since images, video, etc. are generally opaque to keyword searches. The most common approaches for multimedia retrieval have relied on matching search keywords to metadata associated to multimedia content such as the filename, title, amongst others [6].

Words appearing near a multimedia item on Web pages have also been used as targets for matching the search terms [1]. In addition, websites such as Flickr and Youtube have incorporated content tagging as a way to let users describe and interconnect related media. Tags are words associated to media that can be used in a later stage for categorizing, retrieving and interconnecting content [14].

However, the ambiguity in the words (metadata, text, tags) used as descriptions of multimedia items makes the retrieval task particularly difficult. For instance, take the resource `dbpedia:Hornet`[1], which refers to a wasp in the DBpedia knowledge base [5]. If we query Flickr or Google Images for pictures related to the entity name '*hornet*', we can see in Figure 1 that both Flickr and Google return images related to other meanings of the word. Flickr shows images of a plane (*F/A-18 Hornet*) and a fictional character (*The Green Hornet*), while Google displays images of a motorcycle (*Honda CB600F*). Consequently, currently available multimedia search engines are not readily apt to collect relevant images for ontology entities.

Our work presents a contribution to the task of populating an ontology with images from the Web. We focus on retrieving relevant images for entities extracted from Wikipedia, the world's largest source of encyclopedic knowledge. The DBpedia project collects facts from Wikipedia containing 3.5 million entities, their attributes and relationships with other entities [5]. DBpedia is classified in a consistent cross-domain ontology with classes such as persons, organisations or populated places; as well as more fine-grained classifications like basketball player or flowering plant. The DBpedia project has evolved to one of the center pieces of the

K-CAP'11, June 26–29, 2011, Banff, Alberta, Canada.

[1]The prefix `dbpedia:` refers to `http://dbpedia.org/resource/`

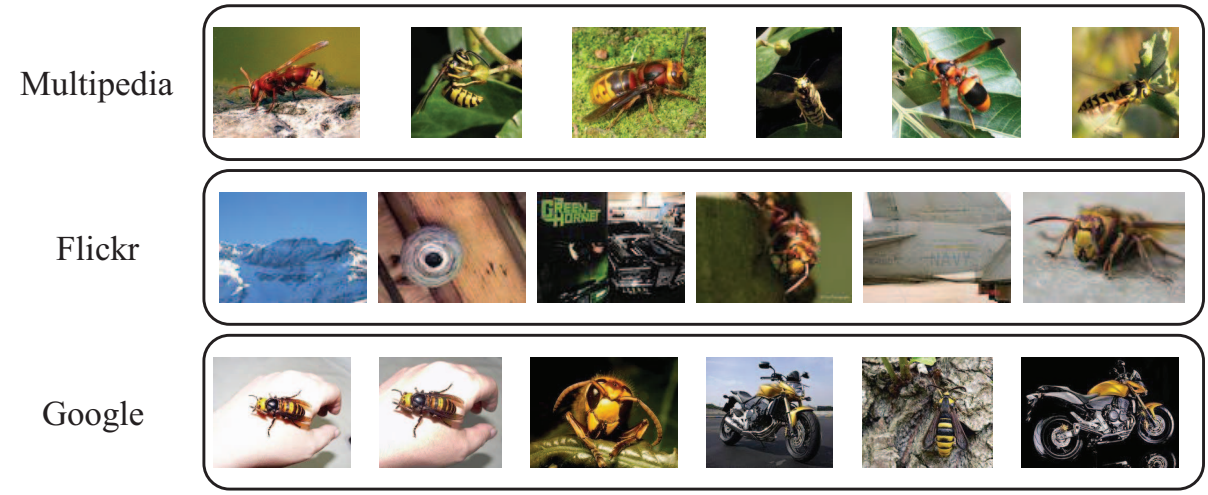

Figure 1: Querying the Web for images related to the resource `dbpedia:Hornet`

Linking Open Data (LOD) project[2], which seeks to enable a Web of Data where information can be effectively exchanged as structured facts in addition to natural language text [4]. As such, our work extends the encyclopedic knowledge in the Linked Data cloud with relevant images of DBpedia resources.

We introduce **Multipedia**, a system for collecting multimedia information for DBpedia. Our approach leverages existing image search engines and improves their ability to retrieve images for DBpedia resources with ambiguous names. Multipedia achieves this by: (i) expanding the semantic neighborhood of DBpedia resources with 'context words' – words that occurred around DBpedia resources mentioned in Wikipedia text; (ii) performing query expansion with context words and searching existing engines; (iii) computing semantic relatedness between tagging information and DBpedia resources; (iv) aggregating the results into a final rank using a ranking aggregation method.

We evaluate the effectiveness of our approach with a user study involving 15 people and resulting in 2250 image relevance judgments. We use commercial Web search engines as a baseline and present how the algorithms introduced in this work offer improvements of 8.9% and 9.4% over the baseline.

This paper is organized as follows. Section 2 describes related work in the context of disambiguated image retrieval and their hierarchical organisation. Section 3 presents our approach to this task. It includes the description of how we acquire various sets of ranked images as well as the method of how these rankings are combined into the result set of images. In section 4, we evaluate our approach for ambiguous entity names. We discuss our conclusions in section 5, presenting our plans for future work.

2. RELATED WORK

We address the problem of acquiring images for resources in the DBpedia knowledge base from the Web beyond the images that are attached to Wikipedia articles since this multimedia data is already part of DBpedia[3].

[2]`http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData`

[3]with the relation `foaf:depiction`

The main obstacle is ambiguity of resource names. The task of retrieving images for resources in the presence of ambiguity has been approached in various ways. In general, there are mainly three types of features that can be utilized in this endeavor. A number of approaches use contextual data in which the image is found [19, 27, 28, 29]. Other works rely on image meta data such as date, GPS information or tags [7, 15]. Lastly, visual similarity features are employed by some authors [9, 17, 22, 27]. Datta et al. [8] offer a survey on image retrieval and image classification, focusing on visual similarity features.

The work that is most closely related to ours was done by Taneva et al. [27]. They take the YAGO knowledge base [26] as their source of resources and develop a supervised learning method based on ranking aggregation to gather images. They use the properties of the YAGO resources to iteratively query a number of search engines. The result rankings from these queries are merged into one, while recognizing duplicates based on the URL and visual similarity. In contrast, our approach does not rely on any training data which, in general, is expensive to gather. Furthermore, we use context words of Wikipedia page links to expand the image queries instead of ontology properties. Context words around page links offer better coverage of resources because ontology properties are extracted from infoboxes that are not part of every Wikipedia page. We additionally employ a tag similarity measure in order to increase the precision.

ImageNet [9] also addresses the problem of populating a knowledge base with images. They chose the semantic classes in WordNet [12] for linking the images, while our work includes finding images for ontology instances. They use hierarchical relations in WordNet and visual features to find images related to the classes and therefore employ a significantly different strategy from ours. RetrievOnto [19] is a system that also uses WordNet, but only a small part of the typed term hierarchy, inducing images at new leaf nodes. The ontological relation then allow for controlling conceptual neighborhoods in order to increase precision in a use case of semantic, content-based image retrieval. However, their approach is evaluated on a small subset of instances of a specific type of concept. The large image collection

LabelMe [22] offer ground truth labels to be used in object recognition research, mainly for recognizing objects embedded in a scene. They also link on class level to WordNet concepts.

There are other approaches that attempt to organize an image collection in some sort of semantic category system, not in a typed ontology. The OPTIMOL system [17] collects pictures from the web and incrementally learns a category model. It uses object recognition techniques and aims at providing data for computer vision research. Crandall et al. [7] organize a large image collection collected from Flickr into a hierarchical structure of places while exploiting GPS data of the images. They also use the tags given by the uploader if GPS data is not available. Wang et al. [29] construct an ontology from the Wikipedia category hierarchy and populate it with related images by viewing the structure as a semantic network. They show how spreading activation techniques help to improve performance in image retrieval. Medialife [15] is a system that uses ontological information to facilitate the generation of user specified image collection subsets that represent a chronicle of life, for instance a collection of pictures of family members at a specific social event. These kind of queries are only possible in the context of a personalized world model. This differs from our approach that attempts to populate a general world knowledge ontology.

3. MULTIPEDIA

In order to retrieve relevant images for DBpedia resources we propose an approach that takes advantage of existing image search engines and of tagging information when it is available. We propose to query the Web using the resource label plus some other context phrases extracted from Wikipedia. This is done iteratively, resulting in one query per context word. Then we carry out two activities simultaneously. First we aggregate the rankings produced by each context word query in a new context-based ranking. Second we create a new tag-based ranking taking into account the semantic similarity between each one of the retrieved pictures and the current DBpedia resource. This semantic similarity is calculated by comparing the picture tags and the DBpedia resource context terms. Finally, we merge both the context-based and the tag-based rankings in a final ranking from which we take the top n results as images relevant to the resource. In the following we present the details of this process.

3.1 Resource Context

Although DBpedia resource URIs are unambiguous, i.e. each URI refers to one and only one resource[4], DBpedia resource names may be ambiguous when searching for information about them on the Web. In this work, we use 'name' (as in resource name) to refer to the value of the property `rdfs:label` for each DBpedia resource. Examples of ambiguous resource names are '*Hornet*' as presented on Figure 1, as well as '*Apple*' and names of many other resource.

Humans are capable of easily identifying the meaning of ambiguous names based on the context – by using their background knowledge and the understanding of the surrounding text. However disambiguation is a hard problem for computers. Natural Language Processing (NLP) research has attempted to model context of ambiguous terms by collecting surrounding words, part of speech information, etc. [18].

As DBpedia resources correspond to Wikipedia articles, we can tap into the Wikipedia corpus to find mentions of Wikipedia articles and collect context information. We consider that a DBpedia resource has been mentioned whenever we find its corresponding Wikipedia article as the target of a wikilink (i.e., link between Wikipedia articles). In this work, **context words** are any terms (excluding stopwords) appearing before and after the wikilink representing a mention of a DBpedia resource. Thus, we have created an index in which for each article we have the set of words appearing along with an article mention and their frequency. For instance, the context for `dbpedia:Apple` consists of words such as '*fruit*' or '*juice*'. In contrast, `dbpedia:Apple_Inc.` context contains words such as '*software*' or '*mac*'.

In order to complement the context information we are using information from the DBpedia Ontology. Currently the DBpedia ontology classifies 1.6 million resources. We use the class name as an additional feature to add to the resource context. In the case of our example, we add the class name '*flowering plant*' to the `dbpedia:Apple` context and to the `dbpedia:Apple_Inc.`context the class name '*public company*'.

Thus, for a given DBpedia resource d we create a set C of context terms c_i collected following the procedure mentioned above.

3.2 Gathering images

In order to collect an initial set of images, we query the Web for candidate images for a DBpedia resource. To do so we rely on existing image search engines and image sharing sites. First, we pose a query to an image sharing site using the name for a resource, if we do not get results then we use the search engine. In order to cope with ambiguity, we pose new queries using the resource name plus one term extracted from the context in the hope that these query results produce more accurate results. For instance, querying images for '*apple*' and '*fruit*' produces mostly `dbpedia:Apple` images. We repeat this procedure for the top N frequent context terms. In Section 4 we experiment with $N = 3, 4, 5, 6, 7$. Henceforth C refers to the context subset of size N.

Thus, given a DBpedia resource d, the output of this task is a set R of image rankings r_j with $1 <= j <= |C|+1$, that is a ranked list for each query using the resource name and a context term plus the initial query using just the resource name. In addition, we produce a set P of unique images with the union of all images in each ranking r_j.

3.3 Aggregating query results

We rank and aggregate the rankings produced in the previous step using Borda's count [23]. Borda's count was developed initially to elect members to an organization. In an election with X candidates, each voter awards X points to his first choice, X-1 to his second choice, and so on. The results are added up and the candidate with the most points wins. Borda's count is a positional method [10]. That is, it assigns a weight corresponding to the position in which a candidate appears within each voter list. The main advantage of Borda's method is that it is very easy computation-

[4]Resources may be duplicates [11], i.e. two URIs identify distinct resources representing the same real world object. Nonetheless, each URI refers to one and only one resource.

ally since this method can be implemented to run in linear time [21].

This method has been adapted to rank and aggregate the results gathered by metasearches on the web [21]. Voters are search engines used by the metasearch and candidates are the documents retrieved by each search engine. Following with this idea, we use Borda's count to merge in a unique list the rankings r_j. In this case, each query is a voter and images are the candidates.

Borda's count considers that all candidate images p_k in P are ranked in all lists $r_1, \ldots, r_j, \ldots, r_{|C|+1}$. For each candidate p_k in r_j, the method assigns a score $S_j(p_k)$ equal to the number of candidates ranked below p_k in r_j. The total Borda score for this candidate is calculated according to equation 1.

$$S(p_k) = \sum_{j=1}^{|C|+1} S_j(p_k) \quad (1)$$

Finally, the fused ranked list is created by sorting the candidates p_k in decreasing order of total Borda score. Note that Borda's count can be extended to deal with partial lists. That is, when not all the candidate images appear in all ranked lists [21]. Let us suppose we have a ranked list r_j so that the number of candidate images ranked in this list is less than the number of candidate images ($|r_j| < |P|$). Thus the Borda score for all candidates not belonging to r_j is $|P| - |r_j| - 1$.

We apply Borda's count to the query results obtained from the previous step and call the new list context-based ranking.

3.4 Tag-based ranking

With the advent of the Web 2.0, users started to provide a wealth of metadata about the information they post on the Web. These metadata take the form of geo-localization information and tags among others. In this respect, image sharing social networks encourage users to tag images to improve resource visibility within the community, as well as a mean of self organization. A possible use of tags is to describe the content of the annotated resource. Thus, we have the advantage of using tagging information in order to measure the relatedness of a specific image and a DBpedia resource.

Our relatedness measure between a DBpedia resource and an image is calculated based on the overlapping of terms between the context of the former and the tags of the latter. To do so, we follow a Vector Space Model [25] to represent the DBpedia resource and the images, and then compare them using a standard metric.

First we create the $Vocabulary$ set as the union of the context terms related to the DBpedia resource. For each candidate image we create a vector in $\Re^{|Vocabulary|}$ where each position corresponds to an element in an ordered version of the $Vocabulary$ set. The value w_i associated with the i-th position in the vector is calculated using TF-IDF[5] [24] for the corresponding i-th term in the ordered set.

Similarly, we create a vector for the DBpedia resource and its context. In this case, w_i takes as value the term frequency calculated as how often the term appears along a mention of the DBpedia resource in Wikipedia. We compare the keyword vector and each one of the image vectors using as similarity measure the *cosine* function. Finally, we sort all the candidate images in decreasing order of similarity, and produce a new list called tag-based ranking.

[5]Term Frequency and Inverse Document Frequency

3.5 Fusing final ranks

Finally, we fuse both the tag-based and the context-based rankings in a final ranking using Borda's count. We expect that this last fusion raises relevant images, according to the tagging information, in the final list.

4. EXPERIMENTS

The experiments presented in this section were carried out using Flickr and Bing Image Search due to the convenience of their Web APIs, but they could be easily adapted to use other search engines or image sharing sites.

In Section 3.2 we described our approach to gather images for the top N context terms for a resource. Our first experiment investigated how many context words to use in order to guide the image retrieval towards a specific sense of an ambiguous word. We designed an initial experiment where the dataset was manually selected, taking care of including unambiguous and ambiguous resources names and varying the number of context words N = 3, 4, 5, 6 and 7. Results are shown in Figure 2.

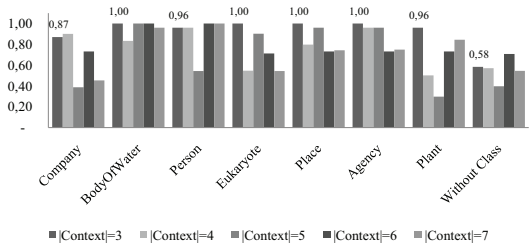

Figure 2: Average precision for different numbers of context terms. Precision values are shown for context of size 3.

A context containing 3 terms produces the best results in terms of average precision achieving a 0.92 value. Using more than 3 context words seems to decrease the average precision. This number is similar to the findings of an earlier experiment about word sense disambiguation presented by Kaplan [16] that found 4 as the number of words above which the context does not add more resolving power to the disambiguation. For instance, in our running example the context of size 7 for '*apple*' consists of the following words '*juice, fruit, apples, capital, michigan, orange*'. One can see that longer contexts start to include words – such as '*capital*' – which may be less helpful to identify the meaning of the resource name '*apple*'.

In the following, we present details of an experiment carried out to evaluate our proposal using 3 context words.

4.1 Dataset

We have constructed an evaluation dataset to assess the ability of Multipedia to retrieve images for ambiguous DBpedia resource names. The highest ambiguity happens when a name can be used to refer to many resources with no dominant sense. A dominant sense is a resource that is by large the most common use of an ambiguous name. Dominance reduces ambiguity in practice since randomly choosing images

is more likely to find the dominant resource, even without any other information.

Therefore, the first criterion employed was to select resource names that are linked from a disambiguation page. This information can be queried in DBpedia using the relation `dbpo:wikiPageDisambiguates`[6]. This relation allows us to detect that this resource may be confused with other resources with the same or similar names. However, from this relation alone it is not possible to measure to what degree this confusion between the resources actually happens in practice. For instance, a name such as '*stonehenge*' is ambiguous, although most of the time it refers to the prehistoric monument `dbpedia:Stonehenge`. Consequently, querying the web for images using the name '*stonehenge*' will retrieve mostly images about the monument.

We have defined a measure of **dominance** (Equation 2) to calculate how common is the most frequent sense of an ambiguous word with respect to all other senses. In this equation w_i is the ambiguous name, S is the set of possible senses, $freq()$ is a function returning the number of times that w_i has been used in Wikipedia to refer to a specific sense. Hence, a value close to 1 means that there is a dominant sense (one resource is much more common than other confusable resources), while a value close to 0 means that there is not a dominant sense.

$$dom(w_i) = \frac{Max(freq(s_j))}{\sum_{j=0}^{|S|} freq(s_j)} \quad (2)$$

We created a program to automatically gather the dataset. We first selected 10 classes from the DBpedia Ontology in order to ensure diversity. For each class, we randomly picked up 15 popular resources with an ambiguous name and a *dom* value below 0.7. Popularity was required so that DBpedia resources can be easily assessed by human evaluators. A resource was considered popular if there were more than 100 wikilinks to its corresponding Wikipedia article. We found resources fulfilling these requirements classified under the classes `dbpo:Mammal`, `dbpo:Bird` and `dbpo:Insect`. For the rest of the classes we had to increase the $dom(w_i)$ limit to 0.9.

4.2 Evaluation

We asked a group of 15 people, students and researchers from the Freie Universität Berlin and the Universidad Politécnica de Madrid, to evaluate the top 5 results of three methods. The experiment was conducted as a blind evaluation, i.e., the results were conflated into a ranking with 15 images per DBpedia resource, without telling the raters which result came from which method. Each evaluator rated the image as *Highly Related*, *Related* or *Not Related* with the DBpedia resource. If they could not take a decision regarding the current image, e.g. due to low picture quality, evaluators could select the *Don't Know* option.

We presented to each evaluator additional information of the image as available tags, textual description and title. We made sure every image was rated by three evaluators so that we can take into account the decisions taken by majority.

We have measured the reliability of agreement between our evaluators using Fleiss' Kappa [13]. It measures how much of the observed agreement exceeds what would be expected if all raters made their ratings completely randomly.

[6]The prefix `dbpo:` refers to `http://dbpedia.org/ontology/`

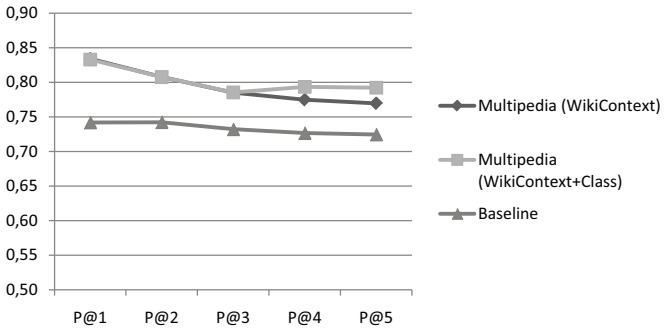

Figure 3: Precision at n of the evaluated approaches

If a fixed number of people assign ratings to a number of items, then the kappa can be seen as a measure for the consistency of ratings. The scoring range is between 0 and 1. Using all ratings in our evaluation we obtained $\kappa = 0.445$ with $z = 53.6$. There was a total of 2250 ratings. In 49.93% of the cases, all three users agreed exactly on the rating (unanimous decision). When collapsing 'Highly Related' and 'Related' into one category, 76.82% of the ratings were unanimous. In 93.46% of the cases, at least two raters agreed.

The images presented to the raters were obtained from two versions of our approach and a baseline (5 images from each). The first version, which we call **Multipedia WikiContext**, used the top 3 most frequent words appearing along a mention of the resource in Wikipedia as the context words for computing relatedness. The second version, which we call **Multipedia WikiContext+Class**, extended the context with the class name to which the resource belongs. The baseline was defined as querying an image sharing site using just the resource name. In case the image sharing site search does not produce any result we pose a query to an image search engine.

From all the evaluated images, the 81.96% corresponds to images extracted from the image sharing site, and the 18.03% was extracted from the image search engine. The evaluated dataset is publicly available[7].

Precision (P) is the fraction of relevant images to the images retrieved by each approach given a DBpedia resource. We have measured $P@N$ with $N = 1, 2, 3, 4, 5$ (precision at N rank position [2]). The Average Precision (AP) is defined as the average of $P@N$ values. Precision values were calculated from those evaluations where users were able to take a decision.

Figure 3 depicts $P@N$ values achieved by each approach. Multipedia approaches produce more precise results than the baseline along all the values of N. We can observe that WikiContext+Class is better than WikiContext starting from N=3. This means that the class names are an important factor in the context to help in the selection of relevant images. Table 1 shows that WikiContext+Class was the best approach with a $AP = 0.80$. Both Multipedia approaches were able to increase AP value (%inc) regarding the baseline. WikiContext increased AP in 8.9% and WikiContext+Class in 9.4%.

Figure 4 shows AP values per each Ontology Class. Note

[7]`http://delicias.dia.fi.upm.es/wiki/images/b/b2/MultipediaEvaluation.zip`

Table 1: Average precision (AP) per class and percentage increase (%inc) with respect to the baseline.

Class	Baseline	Wiki Context	%inc	WikiCont. +Class	%inc
Athlete	0.65	0.74	14.6%	0.68	5.0%
Bird	0.66	0.86	**31.8**%	0.90	**37.2**%
Building	0.84	0.83	*-0,6%*	0.84	0.6%
Insect	0.89	0.89	0.4%	0.93	5.2%
Mammals	0.68	0.73	7.6%	0.86	**27.2**%
MeanOfTrans	0.86	0.86	*-0.2%*	0.93	7.4%
Mountain	0.66	0.85	**29.8**%	0.85	**28.8**%
Politician	0.57	0.48	*-16.4%*	0.39	*-31.2%*
Sport	0.72	0.93	**29.3**%	0.80	10.5%
WorldHeritage	0.81	0.75	*-7.2%*	0.84	3.0%
Average	0.73	0.79	8.9%	**0.80**	**9.4**%

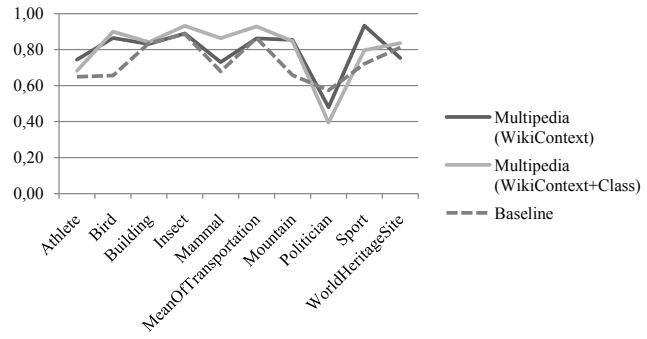

Figure 4: Average precision per class

that WikiContext+Class increases AP values in all classes except `dbpo:Politician`. WikiContext+Class achieved the best results with `dbpo:Bird`, `dbpo:Mammal` and `dbpo:Mountain` with improvements of 37.2%, 27.2% and 28.8% respectively. Recall that $dom(w_i)$ for names of birds and mammals used in this dataset was 0.7, indicating that these names do not have a strong dominant sense. Thus, for these two classes we have validated that 1) the baseline fails when dealing with ambiguous names lacking of a dominant sense and 2) that our approach produces better results for this sort of names.

Nevertheless, the class `dbpo:Mountain` did have a dominant sense in Wikipedia. What we found on the Web for some mountains was that their names were actually used to refer to things related to the mountain such as hotels, resorts or restaurants. Therefore the baseline erroneously retrieved images with regard to those other resources, while the use of Wikipedia-based context helps Multipedia to find the correct images. In addition, this means that despite those mountains having a dominant sense in Wikipedia, they do not have it on the Web. Thus, the Wikipedia corpus is a starting point to measure ambiguity degrees as the dominant sense ratio, though more evidence information should be taken from other sources or the Web itself.

For the class `dbpo:Politician`, on the other hand, Multipedia approaches present worse results than the baseline. The use of context words did not seem to help reduce ambiguity. We found that many images have been included along text related to political issues, although the images do not depict a specific politician. In our dataset 24% of images retrieved for the three approaches contain a description with more than 150 characters including the politician name (14% of images have description longer than 500 characters). For instance, an image depicting the Brandenburg Gate[8] presented in our dataset is described (and annotated) with a long text showing different events and mentioning different politicians taking part in those events. So, when we were retrieving pictures for `dbpedia:Helmut_Kohl` former chancellor of Germany, we found pictures of the Brandenburg Gate where he was mentioned. The use of context words, such as '*Minister*' does not help to get rid of these pictures because usually those descriptions are well contextualized including positions of the politicians and locations. Further research is needed in order to develop methods to deal with this kind of misleading metadata.

Since it is impossible to know the set of all relevant images for a DBpedia resource that are available on the Web in advance, it is not possible to compute recall. Nevertheless, we can report coverage per each approach defined as the number of retrieved images divided by the number of expected images. All three approaches have an almost perfect coverage since just for one DBpedia resource we could not find images on the Web.

5. CONCLUSIONS

In this paper we addressed the problem of how to enrich ontology instances with links to images. We focused on the particularly challenging problem of ambiguity in instance names. We collected resources belonging to diverse types from DBpedia, one of the most prominent knowledge bases in the Linked Data cloud. We relied on mentions of DBpedia resources in Wikipedia text in order to gather contextual information for those resources. Our approach takes advantage of existing image search engines on the Web, and retrieves images using the collected context information for a resource. We measured the relatedness of each image to a DBpedia resource by calculating a semantic similarity between the image metadata information and the resource context. As a final step we produce a ranking using the Borda's count, a well known method for ranking aggregations.

We have carried out a human-driven evaluation of the approach involving 15 users and a total of 2250 image ratings containing DBpedia resources from several classes. The dataset was selected so that all of the instance names where ambiguous. A variation of Multipedia using Wikipedia textual information plus the ontology class as context achieved the best results, improving average precision by 9.4% over a baseline of keyword queries to commercial image search engines. We have validated that in contrast to the baseline our approach achieves the highest precision values with ambiguous names lacking a dominant sense.

As future work we plan to improve the precision for images with misleading textual descriptions as the ones found in our experiment for Politicians. In addition, some images have metadata that can be considered as spam (e.g., sometimes users in social networks add popular metadata to their images so that they can appear first in the search results). Therefore new techniques have to be developed to cope with these challenges.

6. ACKNOWLEDGMENTS

Our work has been partially funded by the Project CENIT España Virtual (ALT0317), an FPI grant (BES-2008-007622) of the Spanish Ministry of Science and Innovation, Neofonie

[8] `http://dbpedia.org/resource/Brandenburg_Gate`

GmbH, a Berlin-based company offering leading technologies in the area of Web search, social media and mobile applications (http://www.neofonie.de/) and the European Commission through the project LOD2 – Creating Knowledge out of Linked Data (http://lod2.eu/).

7. REFERENCES

[1] Google's Peter Linsley Interviewed by Eric Enge, 2009. http://www.stonetemple.com/articles/interview-peter-linsley.shtml.

[2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1st edition, May 1999.

[3] M. Bertini, G. D'Amico, A. Ferracani, M. Meoni, and G. Serra. Web-based semantic browsing of video collections using multimedia ontologies. In *Proceedings of the international conference on Multimedia*, MM '10, pages 1629–1632, New York, NY, USA, 2010. ACM.

[4] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009.

[5] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantic*, 7(3):154–165, 2009.

[6] S. Chang and A. Hsu. Image information systems: where do we go from here? *Knowledge and Data Engineering, IEEE Transactions on*, 4(5):431–442, 1992.

[7] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 761–770, New York, NY, USA, 2009. ACM.

[8] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, April 2008.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:248–255, 2009.

[10] C. Dwork, R. S. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the Web. In *World Wide Web*, pages 613–622, 2001.

[11] A. Elmagarmid, P. Ipeirotis, and V. Verykios. Duplicate record detection: A survey. *Knowledge and Data Engineering, IEEE Transactions on*, 19(1):1–16, Jan. 2007.

[12] C. Fellbaum, editor. *WordNet: an electronic lexical database*. MIT Press, 1998.

[13] J. L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378 – 382, 1971.

[14] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198Ű208, April 2006. cite arxiv:cs/0508082.

[15] A. Gupta, S. Rafatirad, M. Gao, and R. Jain. MEDIALIFE: from images to a life chronicle. In *SIGMOD Conference*, pages 1119–1122, 2009.

[16] A. Kaplan. An experimental study of ambiguity and context. *Mechanical Translation*, 2:39–46, 1955.

[17] L.-J. Li, G. Wang, and L. Fei-Fei. OPTIMOL: automatic Online Picture collecTion via Incremental MOdel Learning. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007.*, pages 1–8, June 2007.

[18] R. Mihalcea and A. Csomai. Wikify!: linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 233–242, New York, NY, USA, 2007. ACM.

[19] A. Popescu, C. Millet, and P.-A. Moëllic. Ontology driven content based image retrieval. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, CIVR '07, pages 387–394, New York, NY, USA, 2007. ACM.

[20] S. Radhouani, J.-H. Lim, J.-P. Chevallet, and G. Falquet. Combining Textual and Visual Ontologies to Solve Medical Multimodal Queries. In *ICME*, pages 1853–1856, 2006.

[21] E. M. Renda and U. Straccia. Web metasearch: rank vs. score based rank aggregation methods. In *SAC '03: Proceedings of the 2003 ACM symposium on Applied computing*, pages 841–846, New York, NY, USA, 2003. ACM Press.

[22] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vision*, 77:157–173, May 2008.

[23] D. G. Saari. The mathematics of voting: Democratic symmetry. *The Economist*, page 83, March 2000.

[24] G. Salton and M. J. Mcgill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.

[25] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18:613–620, November 1975.

[26] F. M. Suchanek, G. Kasneci, and G. Weikum. YAGO: A Large Ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 6(3):203–217, 2008.

[27] B. Taneva, M. Kacimi, and G. Weikum. Gathering and ranking photos of named entities with high precision, high recall, and diversity. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 431–440, New York, NY, USA, 2010. ACM.

[28] H. Wang, L.-T. Chia, and S. Gao. Wikipedia-assisted concept thesaurus for better web media understanding. In *Proceedings of the international conference on Multimedia information retrieval*, MIR '10, pages 349–358, New York, NY, USA, 2010. ACM.

[29] H. Wang, X. Jiang, L.-T. Chia, and A.-H. Tan. Ontology enhanced web image retrieval: aided by Wikipedia & spreading activation theory. In *Proceeding of the 1st ACM international conference on Multimedia information retrieval*, MIR '08, pages 195–201, New York, NY, USA, 2008. ACM.