

# DAEDALUS at ImageCLEF Wikipedia Retrieval 2010: Expanding with Semantic Information from Context

Sara Lana-Serrano<sup>1,3</sup>, Julio Villena-Román<sup>2,3</sup>, José Carlos González-Cristóbal<sup>1,3</sup>

<sup>1</sup> Universidad Politécnica de Madrid

<sup>2</sup> Universidad Carlos III de Madrid

<sup>3</sup> DAEDALUS - Data, Decisions and Language, S.A.

[slana@diatel.upm.es](mailto:slana@diatel.upm.es), [jvillena@it.uc3m.es](mailto:jvillena@it.uc3m.es),  
[josecarlos.gonzalez@upm.es](mailto:josecarlos.gonzalez@upm.es)

**Abstract.** This paper describes the participation of DAEDALUS at the ImageCLEF 2010 Wikipedia Retrieval task. The main focus of our experiments is to evaluate the impact in the image retrieval process of the incorporation of semantic information extracted only from the textual information provided as metadata of the image itself, as compared to expanding with contextual information gathered from the document where the image is referred. For the semantic annotation, DBpedia ontology and YAGO classification schema are used. As expected, the obtained results show that, in general, the textual information attached to a given image is not able to fully represent certain features of the image. Furthermore, the use of semantic information in the process of multimedia information extraction poses two hard challenges still to solve: how to automatically extract the high level features associated to a multimedia resource, and, once the resource has been semantically tagged, which features must be used in the retrieval process to best model the actual and complete meaning of the user query.

**Keywords:** Image retrieval, domain-specific vocabulary, ontology, semantic expansion, information retrieval, indexing, topic expansion, context.

## 1 Introduction

The basic goal of the ImageCLEF 2010 Wikipedia Retrieval task [1] was, similar to previous campaigns, given a textual query and/or sample images describing a user's multimedia information need, find as many relevant images as possible from the Wikipedia images collection. Each image in the collection is tagged with both its user-provided annotation consisting of unstructured and noisy textual annotations in English, French, and German, and also links to the article(s) that contain the image.

This paper describes the participation of DAEDALUS team at the ImageCLEF 2010 Wikipedia Retrieval task. We are a research group led by and named after DAEDALUS, a small private company in the field of Information and Telecommunication Technologies and a leading provider of language-based solutions

in Spain, and research groups of two universities, Universidad Politécnica de Madrid and Universidad Carlos III de Madrid. We have taken part in CLEF since 2003 in many different tracks and tasks, as part of the MIRACLE team till last year.

This year, the main objective of our experiments is to evaluate and compare the results achieved by the application of techniques that are based on the computational similarity between the metadata associated to the images and the query itself, as opposed to other techniques based on the semantic description of the image based on the contextual information provided by the Wikipedia article in which the image is referred. For this purpose, the DBpedia ontology [2] and the YAGO [3] classification schema have been used as the knowledge base to annotate the semantic content,

## 2 System Description

Based on our experience in previous campaigns in CLEF and other forums, we designed a flexible system in order to be able to execute a large number of runs that exhaustively cover many combinations of different techniques. Our system is composed of a set of small components that are easily combined in different configurations and executed sequentially to build the final result set. Specifically, our system is composed of four modules:

- **Linguistic processing module**, which extract, parses and prepares the input text for subsequent modules.
- **Semantic module**, which expands documents and/or topics with semantic information retrieved from knowledge base.
- **Textual (text-based) retrieval module**, which indexes image annotations in order to search and find the list of images that are most relevant to the text of the topic.
- **Result combination module**, which uses the OR operator to combine, if necessary, two different result lists.

A common baseline algorithm was used in all experiments to process the collection, following these steps:

1. **Text Extraction:** Ad-hoc scripts are run on the files that contain image annotations, on the Wikipedia articles and on the topics. The purpose of this process is to generate the different collections or topics that set up the different specific features of each experiment.
2. **Tokenization:** This process extracts the basic textual components in the annotations. Some basic entities are also detected, such as numbers, initials, abbreviations, and years. So far, compounds, proper nouns, acronyms or other types of entity are not specifically considered. The outcomes of this process are single words, multi-words, years in numbers and tagged entities resulting from the application of the semantic module.
3. **Conversion to lowercase:** All document terms are normalized by changing all letters to lowercase.

4. **Filtering:** All words recognized as stopwords are filtered out. Stopwords in the target languages were initially obtained from the University of Neuchatel's resources page [4] and afterwards extended using our own developed resources.
5. **Stemming:** This process is applied to each one of the words to be indexed or used for retrieval. Standard Porter stemmers [5] for each considered language have been used.
6. **Indexing and retrieval:** Lucene [6] was used as the information retrieval engine for the whole textual indexing and retrieval task.

### 3 Experiments and Results

The main idea behind our experiments is to evaluate and compare the results achieved by the application of techniques that are based on the computational similarity between the metadata associated to the images and the query itself, opposed to other techniques based on the semantic description of the image using the contextual information provided by the Wikipedia article in which the image is referred.

The following fields have been considered as contextual information:

- the metadata associated to the image itself (C),
- the title of the article (T), and
- the first paragraph in the article (S).

The core knowledge base for the semantic expansion is a subset of the DBpedia ontology [2], conveniently adapted and formatted to our purposes, and using the YAGO classification schema. YAGO [3] is a huge semantic knowledge base, part of the YAGO-NAGA project at the Max-Planck Institute for Informatics in Saarbrücken (Germany). It currently, holds more than 2 million entities (persons, organizations, cities, etc.) with over 20 million facts about these entities. YAGO has a manually confirmed accuracy of 95%, unlike many other automatically assembled knowledge bases.

Our resulting knowledge base contains 1,651,225 entities, 226,087 hierarchically related classes by means of 225,781 *subClassOf* relations and 4,121,043 *typeOf* relations among entities and classes.

Afterwards, an entity identification process is run using the information contained in the knowledge base using a parser specifically developed for this task. Last, the semantic information generated as the output is built by adding up the information about the entity itself, the information about its class(es) and all the ancestors of its class(es).

Finally we submitted 6 experiments to be evaluated, described in **Table 1**.

**Table 1.** Description of the experiment set.

Run Identifier	Contextual Expansion	Semantic Expansion
DAEDALUS_Bas	NO	NO

<b>DAEDALUS_NER_Bas</b>	NO	YES
<b>DAEDALUS_W_CT</b>	C + T	NO
<b>DAEDALUS_NER_W_CT</b>	C + T	YES
<b>DAEDALUS_W_CTS</b>	C + T + S	NO
<b>DAEDALUS_NER_W_CTS</b>	C + T + S	YES

The results achieved after the evaluation of these experiments are shown in next **Table 2**. The highest figures are highlighted in bold.

**Table 2.** Evaluation of experiments.

Run	MAP	P@10	P@20	R-prec	NDCG	Relevant Retrieved
<b>DAEDALUS_Bas</b>	0.1492	0.3971	0.3529	0.2377	0.3556	6088
<b>DAEDALUS_NER_Bas</b>	0.1249	0.3643	0.3236	0.2115	0.3255	5628
<b>DAEDALUS_W_CT</b>	<b>0.1820</b>	<b>0.4471</b>	<b>0.4029</b>	<b>0.2662</b>	0.4055	6453
<b>DAEDALUS_NER_W_CT</b>	0.1610	0.4043	0.3736	0.2514	0.3801	6038
<b>DAEDALUS_W_CTS</b>	0.1737	0.3943	0.3521	0.2478	<b>0.4315</b>	<b>6871</b>
<b>DAEDALUS_NER_W_CTS</b>	0.1593	0.4029	0.3593	0.2342	0.4036	6105

A first preliminary evaluation of these figures shows that, globally, the contextual expansion greatly helps to improve the retrieval results and the semantic expansion tends to make them worse. However, this conclusion is not completely true because of the fact that in the retrieval process, the semantic terms have been boosted with respect to the contextual terms by assigning the first ones a higher relevance factor. This was initially done to be able to better analyze the impact of the semantic expansion, but finally it turned out not to be a good idea. Actually, this issue causes that the final results for the semantic experiments don't exactly reflect the behavior of an actual retrieval system.

If a deeper analysis is done, it is interesting to notice that, independently of the experiment, the precision levels are quite low when the queries include any reference to primitive features of the image (“*white house with garden*”, “*red fruits*”, “*yellow buses*”, “*close up of antenna*”) or to high-level semantic features such as actions (“*people playing guitar*”, “*people laughing*”) or perceptions.

Moreover, we can also notice that, regardless of the precision level of the results, in general, the incorporation of semantic information for a given topic always produces similar effects (improvements or reductions) independently of the type of contextual information that has been applied. Considering the impact that the use of semantic information has produced in the retrieval process, the following groups of queries have been identified:

- Queries that couldn't be annotated with semantic information (“*lightning in the sky*”).

- Queries in which the use of semantic information produces slight improvements in the results (“horseman”, “civil airplane”).
- Queries in which the use of semantic information significantly improves the results (see **Table 3**). In those queries, the weighting of the semantic terms with respect to the contextual ones, has turned out to be successful to model the specific semantic features that better represent the full meaning of the original query.
- Queries in which the use of semantic information significantly reduces the precision of the results (see **Table 4**). The semantic representation of the query has extracted one or more features that are too general (mainly associated to the first levels of the ontology in the knowledge base), thus contributing with search terms that are not very precise, which in turn produce a very high volume of relevant documents. This fact, combined to the semantic terms boosting described before, have caused a significant decrease in the precision of the results.

**Table 3.** Some examples of the improvements in precision (measured in percentage over the average value) when using semantic information in the experiment setting.

	<b>Experiments</b>		
	<b>Bas</b>	<b>W_CT</b>	<b>W_CTS</b>
<b>Topic 8: tennis player on court</b>			
MAP	385.43%	<b>590.53%</b>	37.88%
R-prec.	297.59%	<b>390.76%</b>	12.81%
<b>Topic 15: cyclist</b>			
MAP	<b>370.50%</b>	307.05%	50.42%
R-prec.	<b>453.59%</b>	329.05%	12.51%
<b>Topic 16: spider with cobweb</b>			
MAP	<b>708.47%</b>	511.69%	140.14%
R-prec.	<b>500.54%</b>	400.54%	100.00%
<b>Topic 55: building site</b>			
MAP	348.65%	<b>352.29%</b>	402.13%
R-prec.	<b>150.00%</b>	110.53%	100.00%
<b>Topic 70: close up of trees</b>			
MAP	409.69%	<b>429.28%</b>	213.17%
R-prec.	<b>164.25%</b>	137.56%	63.20%

**Table 4.** Some examples of the decrease in precision (measured in percentage over the average value) when using semantic information in the experiment setting.

	<b>Experiments</b>		
	<b>Bas</b>	<b>W_CT</b>	<b>W_CTS</b>
<b>Topic 30: <i>harbour</i></b>			
MAP	-82.69%	<b>-88.34%</b>	-64.14%
R-prec.	-69.74%	<b>-75.20%</b>	-48.72%
<b>Topic 50: <i>portraits of people</i></b>			
MAP	-56.40%	<b>-82.74%</b>	-36.28%
R-prec.	-36.39%	<b>-56.91%</b>	-28.96%
<b>Topic 52: <i>satellite image</i></b>			
MAP	-57.80%	-76.62%	<b>-91.53%</b>
R-prec.	-35.08%	-46.31%	<b>-65.16%</b>
<b>Topic 59: <i>cities at night</i></b>			
MAP	<b>-99.96%</b>	-99.88%	-99.81%
R-prec.	<b>-98.14%</b>	-97.36%	-97.72%
<b>Topic 67: <i>white house with garden</i></b>			
MAP	<b>-96.49%</b>	-93.93%	-94.31%
R-prec.	<b>-79.98%</b>	<b>-79.98%</b>	-79.97%

## 4 Conclusions and Future Work

After the detailed analysis of the achieved results for each of the topics, we can point out that the text-based information retrieval techniques applied to image retrieval only provide good results when the formulated queries exactly make reference to the semantic or contextual content of the image (images including *something* o located *somewhere*), but tend to be of no application for the extraction of primitive features (such as color, brightness, texture, shapes, corner points or its spatial distribution) or high-level semantic features about the meaning and purpose of the objects or scenes depicted (sentiments, emotions, actions, perceptions).

For the first case, the incorporation of semantic information, based on the contextual information of the article in which the image is referred, usually improves the results for those queries in which the semantic information contributes with specific terms that narrow the search. For instance, the semantic information corresponding to the “tennis player on court” topic may help to select images associated to the “tennis player” class; however, the semantic information corresponding to the “cities at night” topic broadens the search to all images that show any of the subclasses extending from “city”, which turns out to be very noisy.

Consequently, it seems that our future efforts should be focused first to study how to better apply any content-based image retrieval technique that helps us to extract the semantics of the image itself, and, on the other hand, to try and find the answers to the following open issues: 1) Should the semantic information be taken into account for all queries during the retrieval process? 2) In any case, should it have a specific processing depending on the query type? 3) Would it be a good idea to assign the same weight during the retrieval process to the semantic information associated to a given entity, or is it better to make this value dependent on the information class and/or the query type?

## Acknowledgements

This work has been partially supported by the Spanish Center for Industry Technological Development (CDTI, Ministry of Industry, Tourism and Trade), through the CONTENIDOS A LA CARTA Project, INGENIO 2010 Programme, AVANZA I+D 2008. Other partners in the Project are Agencia EFE, Germinus XXI, 11870.com and Universidad Politécnica de Madrid.

## References

1. Popescu. A.; Tsirikia. T. and Kludas. J. Overview of the Wikipedia Retrieval task at ImageCLEF 2010. Working Notes of CLEF 2010. Padova. Italy. 2010.
2. The DBpedia Knowledge Base. <http://wiki.dbpedia.org/>
3. YAGO: A Core of Semantic Knowledge. <http://www.mpi-inf.mpg.de/yago-naga/yago/>
4. University of Neuchatel. IR Multilingual Resources at UniNE. <http://members.unine.ch/jacques.savoy/clef/index.html>
5. Porter. M. Snowball stemmers and resources page. <http://www.snowball.tartarus.org>
6. Apache Lucene project. <http://lucene.apache.org>