DISA at ImageCLEF 2014: The search-based solution for scalable image annotation

Petra Budikova, Jan Botorek, Michal Batko, and Pavel Zezula

Masaryk University, Brno, Czech Republic
{budikova,botorek,batko,zezula}@fi.muni.cz

Abstract. This paper presents an annotation tool developed by the DISA Laboratory for the ImageCLEF 2014 Scalable Concept Image Annotation challenge. Our solution exploits the search-based annotation paradigm and utilizes several sources of semantic information to determine the relevance of candidate concepts. Rather than relying on the quality of training data, our approach profits from the large quantities of information available in large image collections and semantic knowledge bases. The results achieved by our system confirm that this approach is very promising for scalable image annotation.

1 Introduction

While modern technologies allow people to create and store data in many forms (e.g. images, video, etc.), the most natural way of expressing one's need for a specific piece of data is still a text query. Natural language remains the primary means of information transfer both for person-to-person communication and person-to-computer interactions. However, a lot of existing digital data is not associated with any text information that would help users access or categorize the data. The purpose of automatic annotation is to increase the findability of information by bridging the gap between data and query representations.

Since 2006, the ImageCLEF initiative has been encouraging the development of image annotation tools by organizing competitions on automatic image classification and annotation. In the ImageCLEF 2014 Scalable Concept Image Annotation challenge, participants were required to develop solutions that can annotate common personal images using only automatically obtained training data. Utilization of manually prepared training samples was forbidden to ensure that the solutions would easily scale to larger concept sets.

This paper presents the annotation tool developed for this task by the DISA Laboratory[1] at Masaryk University. Our solution exploits the search-based annotation paradigm and utilizes several sources of semantic information to determine the probability of candidate concepts. Rather than relying on the quality of training data, our approach profits from the large quantities of information available in large image collections and semantic knowledge bases. The results achieved by our system confirm the strengths of this approach.

---

[1] http://disa.fi.muni.cz

The rest of the paper is structured as follows. First, we briefly review the task definition and discuss which resources can be used. Next, we describe our approach and individual components of our solution. Analysis of results is provided in Section 4. Section 5 concludes the paper and outlines our future work.

2 Scalable Concept Image Annotation Task

The problem offered by this year's Scalable Concept Image Annotation (SCIA) challenge [6, 13] is basically a standard annotation task, where an input image needs to be connected to relevant concepts from a fixed set of candidate concepts. The *input images* are not accompanied by any descriptive metadata such as EXIF or GPS, so that only the visual image content can serve as annotation input. For each test image, there is a *list of SCIA concepts* from which the relevant ones need to be selected. Each concept is defined by one keyword, a link to relevant WordNet nodes, and, in most cases, a link to a relevant Wikipedia page.

Annotation tasks of this type have been studied for more than a decade and some impressive results have already been achieved [9]. However, most existing solutions rely on large amounts of manually labeled training data, which limits the concept-wise scalability of such methods and their applicability to many real-world scenarios. As its name suggests, the Scalable Concept Image Annotation task takes into consideration not only the annotation precision and recall, but also the scalability of annotation techniques. The proposed solutions should be able to adapt easily when the list of concepts is changed, and the performance should generalize well to concepts not observed during development. Therefore, participants were not provided with hand-labeled training data and were not allowed to use resources that require significant manual preprocessing. Instead, they were encouraged to exploit data that can be crawled from the web or otherwise easily obtained.

Accordingly, the training dataset provided by organizers consists of 500K images downloaded from the web, and the accompanying web pages. The images were obtained by querying popular image search engines (namely Google, Bing and Yahoo) using words in the English dictionary. For each image, the web page that contained the image was downloaded and processed to extract selected textual features. An effort was made to avoid including near duplicates and message images (such as "deleted image") in the dataset, however the dataset can be considered and is supposed to be very noisy. The raw images and web pages were further preprocessed by competition organizers to ease the participation in the task, resulting in several visual and text descriptors as detailed in [13].

The actual competition task consists of annotating 7291 images with different concept lists. Altogether, there are 207 concepts, with the size of individual concept lists ranging from 40 to 207 concepts. Prior to releasing the test image set, which became available a month before the competition deadline, participants were provided with a development set of query images and concept lists, for which a ground truth of relevant concepts was also published. The development set contains 1940 images and only 107 concepts out of the final 207.

2.1 Utilization of Additional Resources

Apart from the 500K set of web images, participants were encouraged to exploit additional knowledge sources such as ontologies, language models, etc., as long as these were not manually prepared and were easily available. Since we found it difficult to decide what level of manual effort is acceptable (e.g. most ontologies are created with significant human participation), we discussed several resources that we were considering with the SCIA task organizers:

WordNet The WordNet lexical database [8] is a comprehensive semantic tool interlinking dictionary, thesaurus and language grammar book. The basic building block of WordNet hierarchy is a *synset*, an object which unifies synonymous words into a single item. On top of synsets, different semantic relations are encoded in the WordNet structure, e.g. hypernymy/hyponymy (super-type and sub-type relation) or meronymy (part-whole relation). Currently, 117 000 synsets are available in the English WordNet 3.0.

The WordNet is developed manually by language experts, which is not in accordance with the SCIA task rules. However, it can be used to solve the SCIA challenge as it is an existing resource with wide coverage that does not limit the concept-wise scalability of annotation tools. Indeed, many last year's solutions of the SCIA task utilized WordNet to learn about semantic relationships between concepts [14].

ImageNet The ImageNet [7] is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds or even a few thousands of images. Currently, there are about 14M images illustrating 22 000 synsets from selected branches of the WordNet. The images are collected from the web by text queries formulated from the words in the particular synset. In the next step, a crowdsourcing platform is utilized for manual cleaning of the downloaded images.

According to the organizers, the ImageNet should not be used for solving the SCIA challenge. Although it is easily available, its scope is limited and extending it to other concepts is very expensive in terms of human labor.

Profiset The Profiset [4] is a large collection of annotated images available for research purposes. The collection contains 20M high-quality images with rich keyword annotations, which were obtained from a web-site that sells stock images produced by photographers from all over the world. The data contained in the Profiset collection was created manually, however this labor was not focused on providing training data for annotation learning. The Profiset is thus a by-product of another activity and can be seen as ordinary web data downloaded from a well-chosen site. It is also important to mention that the image annotations in Profiset have no fixed vocabulary and their quality is not centrally supervised. At the same time, however, the photographers are interested in selling their photos and are thus motivated to provide rich sets of relevant keywords.

The organizers agreed that the Profiset can be used as a resource for the SCIA task.

3 Our Approach

Similar to our previous participation in an ImageCLEF annotation competition in 2011 [5], the DISA solution is based on the MUFIN Image Annotation software, a tool for general-purpose image annotation which we have been developing for several years now [1]. The MUFIN Image Annotation tool follows the search-based approach to image annotation, exploiting content-based retrieval in a very large image collection and a subsequent analysis of descriptions of similar images. In 2011, we were experimenting with applying this approach in a task more suited for traditional machine learning (the 2011 Annotation Task offered manually labeled training data). However, we believe that the search-based approach is extremely suitable for the 2014 Scalable Concept Image Annotation.

The general overview of the solution developed for the SCIA task is provided in Figure 1. In the first phase, the annotation tool retrieves visually similar images from a suitable image collection. Next, textual descriptions of similar images are analyzed with the help of various semantic resources. The text is split into separate words and transformed into synsets, which are expanded and enhanced by semantic relations. The probability of relevance of each synset is computed with respect to the initial probability value assigned to that synset and the types and amount of relations formed with other synsets. Finally, synsets linked to the candidate concept words (i.e. the words in the list of concepts provided with the particular test image) are ordered by probability and a fixed number of top-ranking ones is selected as the final image description.

3.1 Retrieval of Similar Images

The search-based approach to image annotation is based on the assumption that in a sufficiently large collection, images with similar content to any given query image are likely to appear. If these can be identified by a suitable content-based retrieval technique, their metadata such as accompanying texts, labels, etc. can be exploited to obtain text information about the query image.

In our solution, we utilize the MUFIN similarity search system [2] to index and search images. The MUFIN system exploits state-of-the-art metric indexing structures [12] and enables fast retrieval of similar images from very large collections. The visual similarity of images is measured by a weighted combination of five MPEG7 global visual descriptors as detailed in [11]. For each test image, the k most similar images are selected; if more datasets are used, the most similar images from all searches are merged, sorted by visual distance, and the k best are selected. The values of k are discussed in Section 3.3.

Image Collections The choice of image collection(s) over which the content-based retrieval is evaluated is a crucial factor of the whole annotation process. There should be as many images as possible in the chosen collection, the images should be relevant for the domain of the queries, and their descriptions should be rich and precise. Naturally, these requirements are in a conflict – while it is
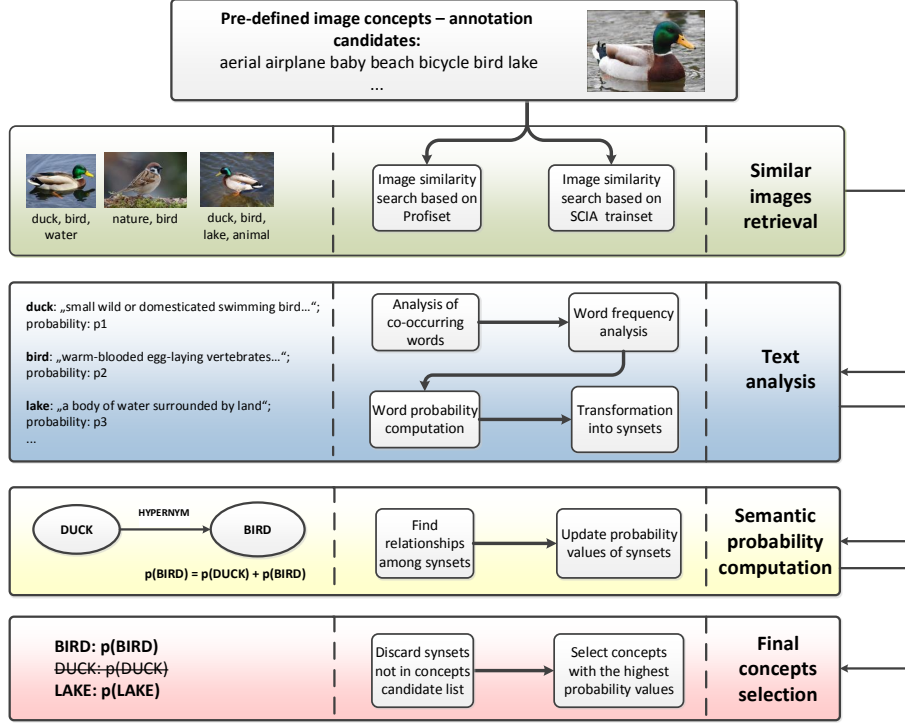
Fig. 1. Architecture of the DISA solution

relatively easy to obtain large collections of image data (at least in the domain of general-purpose images appearing in personal photo-galleries), it is very difficult to automatically collect images with high-quality descriptions.

Our solution utilizes two annotated image collections – the 20M Profiset database (introduced in Section 2) and the 500K set of training images provided by organizers (we denote this collection as the *SCIA trainset*). The Profiset represents a large collection of general-purpose images with as precise annotations as can be achieved in a non-controlled environment. The SCIA trainset is smaller and the quality of text data is much lower; on the other hand, it has been designed to contain images for all keywords from the SCIA task concept lists, which makes it a very good fallback for topics not sufficiently covered in Profiset.

We further considered the 14M ImageNet collection which provides reliable linking between visual content and semantics of images, but we found out that this resource is not acceptable for the SCIA task due to its low scalability (as discussed in Section 2). We also experimented with the 100M CoPhIR image dataset [3] that was built automatically by downloading Flickr photos, but we found the text metadata to be too noisy for the purpose of automatic image annotation.

3.2 From Similar Images to SCIA concepts

In the second phase of the annotation process, the descriptions of images returned by content-based retrieval need to be analyzed and linked to SCIA concepts of a given query to decide about their (ir)relevance. During this phase, our solution relies mainly on the WordNet semantic structure, but we also employ several other resources. The following sections explain how we link keywords from similar images' annotations to WordNet synsets and how the probability of individual synsets is computed. Various parameters of the whole process are summarized in Table 1.

Selection of Initial Keywords Having retrieved the set of similar images, we first divide their text metadata into separate words and compute the frequency of each word. In case of Profiset data, we use directly the keyword annotations of individual images, whereas for SCIA trainset we utilize the *scofeat* descriptors extracted from the respective web pages [13]. This way, we obtain a set of *initial keywords*.

This set can be further enriched by adding a fixed number of most frequently co-occurring words for each initial word. The lists of co-occurring words were obtained using the method described in [10] applied to the ukWac corpus[2], which contains about 2 billion words crawled from the .uk Web domain. Only words which occur at least 5000 times in the corpus and do not begin with a capital letter (indicating a name) were eligible for the co-occurrence lists.

For each keyword in the extended set, we then compute its *initial probability*, which depends on the frequency of the keyword in descriptions of similar images and eventually the probability of its co-occurrence with other initial keywords. Finally, only the n most probable keywords are kept for further processing.

Matching Keywords to WordNet The set of keywords with their associated probabilities contains rich information about query image content, but it is difficult to work with this representation since we have no information about semantic connections between individual words. Therefore, we need to transform the keywords into semantically connected objects. Since we have chosen the WordNet hierarchy as a corner stone for our analysis, each initial keyword is mapped to a relevant WordNet synset. However, there are often more possible meanings of a given word and thus more candidate synsets. Therefore, we use a probability measure based on the cntlist[3] frequency values to select the most probable synset for each keyword. This measure is based on the frequency of words in a particular sense in semantically tagged corpora and expresses a relative frequency of a given synset in general text. To avoid false dismissals, several highly probable synsets may be selected for each keyword (see Table 1). Each selected synset is assigned a probability value computed as a product of the WordNet normalized frequency and the respective keyword's initial probability.

---

[2] http://wacky.sslmit.unibo.it
[3] https://wordnet.princeton.edu/wordnet/man/cntlist.5WN.html

Exploitation of WordNet Relationships By transforming keywords into synsets, we are able to group words with the same meaning and thus increase the probability of recognizing a significant topic. Naturally, this can be further improved by analyzing semantic relationships between the candidate synsets. In the DISA solution of the SCIA task, we exploit the following four WordNet relationships to create a *candidate synset graph*:

- *Hypernymy* (generalization, IS-A relationship): the fundamental relationship utilized in WordNet to build a hierarchy of nouns and some verb groups. It represents upward direction in the generalization/specialization object tree organization. E.g. *dog* is a hypernym of words *poodle* and *Dalmatian*.
- *Hyponymy* (specialization relationship, the opposite of hypernymy): downward direction in the generalization/specialization tree. E.g. *car* is a hyponym of *motor vehicle*.
- *Holonymy* (has-parts relationship): upward direction in the part/whole hierarchy. E.g. *wheeled vehicle* is a holonym of *wheel*.
- *Meronymy* (is-a-part-of relationship, the opposite of holonymy): downward direction in the part/whole tree. E.g. *steering wheel* is a meronym of *car*.

To build the candidate synset graph, we first apply the upward-direction relationships (i.e. hypernymy and holonymy) in a so-called *expansion mode*, when all synsets that are linked to any candidate synset by these relationships are added to the graph; this way, the candidate graph is enriched by upper level synsets in the potentially relevant WordNet subtrees. However, we are not interested in some of the upper-most levels that contain very general concepts such as *entity*, *physical entity*, etc. Therefore, we also utilize the Visual Concept Ontology (VCO)[4] in this step, which is designed as a complementary tool to WordNet and provides a more compact hierarchy of concepts related to image content. Synsets not covered by the VCO are considered to be too general and therefore are not included in the candidate graph. The VCO was created semi-automatically on top of WordNet and its structure is independent of the SCIA task, therefore its utilization is not in conflict with the SCIA scalability requirement.

After the expansion step, the other two relationships are utilized in an *enhancement mode* that only adds new links to the graph based on the relationships between synsets that already are present in the graph. Finally, the candidate graph is submitted to an iterative algorithm that updates the probabilities of individual synsets so that synsets with high number of links receive higher probabilities and vice versa.

Final Concept Selection At the end of the candidate graph processing, the system produces a set of candidate synsets with updated probabilities. The final annotation result is then formed by the k most probable concepts from the intersection of this set with the list of SCIA concepts provided for the particular query image. The matching between candidate synsets and the SCIA concepts

---

[4] http://disa.fi.muni.cz/vco/

Table 1. Annotation tool parameters

Annotation phase	Parameter	Tested values	Development best
Similar images retrieval	datasets	Profiset, SCIA trainset, both	*both*
	# of similar images	10, 15, 20, 25	*25*
Text analysis	# of co-occurring words	0-5	*0*
	max # of synsets per word	1-10	7
	# of initial synsets	100-500	200
Semantic probability computation	relationships	hypernymy, hyponymy, holonymy, meronymy	*all*
Final concepts selection	extended concept definition	true/false	*true*
	# of best results	5-30	7

is based on the definition of SCIA concepts provided by the organizers, which contains links to WordNet. However, as we detected some missing links (e.g. concept *water* was linked with meaning H_2O but not with *body of water*), we manually added several links to this definition, thus creating the *extended concept definition*.

3.3 Tuning of the System

As described in the previous sections, there are many parameters in the DISA annotation system for which suitable values need to be selected. To determine these values, we performed many experiments using the development data and annotation quality measures provided by the SCIA organizers. To increase the reliability of experimental results, we utilized three different query sets: 1) the whole development set of 1940 images as provided by the organizers, 2) a subset of 100 queries randomly selected from the development set, and 3) a manually selected subset of 100 images for which the visual search provided semantically relevant results. These three test sets are significantly different, which was reflected in the absolute values of quality measures, but the overall trends observed in all experiments were consistent. Table 1 summarizes the values of parameters that were tested and the optimal values determined by the experiments.

3.4 DISA Submissions at ImageCLEF

For the actual SCIA competition, we submitted five results produced by different variants of our system. Apart from the optimal set of parameters determined by experiments on development data, we chose several other settings to verify the influence of selected parameters on the overall performance. The values that

were modified in some competition runs are highlighted by italics in Table 1. The individual run settings were as follows:

- DISA-MU_01 – the baseline DISA solution: content-based retrieval only on Profiset collection, 25 similar images, hypernymy and hyponymy relationships only, original definition of SCIA concepts.
- DISA-MU_02: the same configuration as DISA-MU_01, but with extended definition of SCIA concepts, which should improve the final selection of concepts for annotation.
- DISA-MU_03: content-based retrieval on both Profiset and SCIA trainset, otherwise same as DISA-MU_02.
- DISA-MU_04 – the primary run: content-based retrieval on both datasets, 25 similar images, hypernymy, hyponymy, holonymy and meronymy relationships, extended definition of SCIA concepts.
- DISA-MU_05: the same configuration as DISA-MU_04, but only 15 similar images were utilized.

Originally, we planned to submit two more runs but we didn't manage to prepare them in time due to some technical difficulties. The SCIA organizers kindly allowed us to evaluate these runs as well, even though they are not included in the official result list:

- DISA-MU_06: the same configuration as DISA-MU_04, but 35 similar images were utilized.
- DISA-MU_07: the same configuration as DISA-MU_04, with 3 co-occurring words added to each initial word during the selection of keywords.

4   Discussion of Results

The global evaluation of our submissions is presented in Table 2. As expected, the best results were achieved by the primary run DISA-MU_04. Using both our observations from the development phase and the competition results, we can conclude the following facts about search-based annotation:

- The search-based approach is a suitable solution for the SCIA task. To achieve good annotation quality, the search-based approach requires a large dataset with rich annotations, which was in our case represented by Profiset. In a comparison between solutions based only on Profiset and only on SCIA trainset, Profiset clearly dominates due to its size. However, the best results were achieved when both datasets were utilized. The optimal number of similar images is 20-25.
- The utilization of statistic data for expansion of image descriptions did not improve the quality of annotations. Evidently, the addition of frequently co-occurring words rather introduced noise into the descriptions of image content. A straightforward utilization of keyword co-occurrence statistics such as suggested here is thus not a viable way to annotation system improvement.

Table 2. DISA results in SCIA 2014: mean F-measure for the samples (MF-samples); mean F-measure for the concepts (MF-concepts); and the mean average precision for the samples (MAP-samples). The values between the square brackets correspond to the 95 % confidence intervals.

Run	MF-samples	MF-concepts	MAP-samples
DISA-MU_01	27.9 [27.4–28.5]	15.4 [14.0–18.1]	31.6 [31.0–32.2]
DISA-MU_02	27.5 [27.0–28.1]	15.3 [14.0–18.0]	31.9 [31.3–32.5]
DISA-MU_03	28.5 [28.0–29.1]	18.9 [17.4–21.6]	32.9 [32.3–33.5]
DISA-MU_04	**29.7 [29.2–30.3]**	19.1 [17.5–21.8]	**34.3 [33.8–35.0]**
DISA-MU_05	28.4 [27.9–29.0]	**20.3 [18.8–23.0]**	32.3 [31.7–32.9]
DISA-MU_06	*28.7 [28.2–29.3]*	*18.2 [16.7–20.9]*	*33.4 [32.8–34.0]*
DISA-MU_07	*27.9 [27.4–28.4]*	*17.8 [16.3–20.4]*	*32.9 [32.4–33.5]*
best result (kdevir_09)	37.7 [37.0–38.5]	54.7 [50.9–58.3]	36.8 [36.1–37.5]

– The utilization of semantic relationships between candidate concepts helps to improve the quality of annotations. All relationships that we examined have proved their usefulness. Also, a careful mapping between the WordNet-based output of our annotation system and SCIA concepts is important for the precision of final annotation. Unfortunately, we did not manage to tune the mapping very well, as our extended concept definition slightly decreased the quality of competition results (DISA-MU_02 vs. DISA-MU_01).

In comparison with other competing groups, our best solution ranked rather high in both sample-based mean F-measure and sample-based MAP. Especially the sample-based MAP achieved by the run DISA-MU_04 was very close to the overall best result (DISA-MU_04 – MAP 34.3, best result kdevir_09 – MAP 36.8). The results for concept-based mean F-measure are less competitive, which does not come as a surprise. In general, the search-based approach works well for frequent terms, whereas concepts for which there are few examples are difficult to recognize. Furthermore, the MPEG7 similarity is more suitable for scenes and dominant objects than for details which were sometimes required by SCIA (e.g. a park photo with a very small bench was labeled as *furniture* in the development data). Overall, the best results were obtained for scenes (*sunrise/sunset*, *sky*, *forest*, *outdoor*) and more general concepts (*mammal*, *fruit*, *flower*).

The set of query images utilized in the SCIA competition is composed of four distinct subsets of images that also deserve to be examined in more detail. Subset1 consists of 1000 images that were present in the development set, Subset2 contains 2000 new images. Subset3 and Subset4 contain a mix of development and new images, and consist of 2226 and 2065 images, respectively. Each subset is accompanied by a different list of concepts, as detailed in Table 3. The differences between individual subsets allow us to asses the concept-wise scalability of solutions by comparing the annotation results over these subsets. In

Table 3. Performance by query type for DISA-MU_04: mean sample-based precision, recall, F-measure, MAP.

Query set	Concepts	MP-s	MR-s	MF-s	MAP-s
Subset1	107 old + 9 new	31.7	38.4	32.1	37.5
Subset2	107 old + 9 new	31.4	38.2	31.8	36.7
Subset3	40-51 new	37.3	49.2	38.6	45.1
Subset4	107 old + 100 new	16.2	21.0	16.9	19.0

case of DISA, the trends for all runs are similar to those of the primary run DISA-MU_04, shown in Table 3. We can observe that the DISA annotation system can adapt very well to previously unseen concepts, which is demonstrated by Subset3 results. The lower annotation quality observed for Subset4 is caused by increased difficulty of the annotation task, which grows with the number of candidate concepts.

5 Conclusions and Future Work

In this study, we have described the DISA solution of the 2014 Scalable Concept Image Annotation challenge. The presented annotation tool applies similarity-based retrieval on annotated image collections to retrieve images similar to a given query, and then utilizes semantic resources to detect dominant topics in the descriptions of similar images. The DISA annotation tool utilizes the Profiset collection of annotated images, word occurrence statistics automatically extracted from large text corpora, the WordNet lexical database, and the VCO ontology. All of these resources are freely available and were created independently of the SCIA task, so the scalability objective is achieved.

The competition results show that the search-based approach to annotation applied by DISA can be successfully used to identify dominant concepts in images. While the quality of results achieved by the DISA annotation tool is not as high as we would wish, especially in the view of concept-based precision and recall, the strong advantages of our solution lie in the fact that it requires minimum training and easily scales to new concepts. The mean average precision of annotation per sample achieved by our system was only slightly worse than the overall best result.

The semantic search-based annotation can be further developed in several directions. First, we would like to find better measures of visual similarity that could be used in the similarity-search phase, since the relevance of retrieved images is crucial for the whole annotation process. Second, we plan to extend the set of semantic relationships exploited in the annotation process, using e.g. specialized ontologies or Wikipedia. Finally, we also intend to develop a more sophisticated method of final results selection.

Acknowledgments

References

1. Batko, M., Botorek, J., Budikova, P., Zezula, P.: Content-based annotation and classification framework: a general multi-purpose approach. In: 17th International Database Engineering & Applications Symposium (IDEAS 2013). pp. 58–67 (2013)
2. Batko, M., Falchi, F., Lucchese, C., Novak, D., Perego, R., Rabitti, F., Sedmidubský, J., Zezula, P.: Building a web-scale image similarity search system. Multimedia Tools and Applications 47(3), 599–629 (2010)
3. Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., Rabitti, F.: CoPhIR: a test collection for content-based image retrieval. CoRR abs/0905.4627v2 (2009), http://cophir.isti.cnr.it
4. Budikova, P., Batko, M., Zezula, P.: Evaluation platform for content-based image retrieval systems. In: International Conference on Theory and Practice of Digital Libraries (TPDL 2011). pp. 130–142 (2011)
5. Budikova, P., Batko, M., Zezula, P.: MUFIN at ImageCLEF 2011: Success or Failure? In: CLEF 2011 Labs and Workshop (Notebook Papers) (2011)
6. Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., Garcia-Varea, I., Morell, V.: ImageCLEF 2014: Overview and analysis of the results. In: CLEF proceedings. Lecture Notes in Computer Science, Springer Berlin Heidelberg (2014)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F.: ImageNet: A large-scale hierarchical image database. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009). pp. 248–255 (2009)
8. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. The MIT Press (1998)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NIPS 2012). pp. 1106–1114 (2012)
10. Krčmář, L., Ježek, K., Pecina, P.: Determining compositionality of expresssions using various word space models and methods. In: Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality. pp. 64–73 (2013)
11. Lokoc, J., Novák, D., Batko, M., Skopal, T.: Visual image search: Feature signatures or/and global descriptors. In: 5th International Conference on Similarity Search and Applications (SISAP 2012). pp. 177–191 (2012)
12. Novak, D., Batko, M., Zezula, P.: Large-scale similarity data management with distributed metric index. Information Processing & Management 48(5), 855–872 (2012)
13. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2014 Scalable Concept Image Annotation Task. In: CLEF 2014 Evaluation Labs and Workshop, Online Working Notes (2014)
14. Villegas, M., Paredes, R., Thomee, B.: Overview of the ImageCLEF 2013 Scalable Concept Image Annotation Subtask. CLEF 2013 Evaluation Labs and Workshop, Online Working Notes (2013)