Overview of the WikipediaMM Task at ImageCLEF 2008

Theodora Tsikrika[1] and Jana Kludas[2]

[1] CWI, Amsterdam, The Netherlands
Theodora.Tsikrika@cwi.nl
[2] CUI, University of Geneva, Switzerland
jana.kludas@unige.ch

Abstract. The wikipediaMM task provides a testbed for the system-oriented evaluation of ad-hoc retrieval from a large collection of Wikipedia images. It became a part of the ImageCLEF evaluation campaign in 2008 with the aim of investigating the use of visual and textual sources in combination for improving the retrieval performance. This paper presents an overview of the task's resources, topics, assessments, participants' approaches, and main results.

1 Introduction

The wikipediaMM task provides a testbed for the system-oriented evaluation of multimedia information retrieval from a collection of Wikipedia (`http://www.wikipedia.org/`) images. This collection has been previously used in the INEX 2006-2007 Multimedia track [5, 4]. The aim is to investigate mono-media and cross-media retrieval approaches in the context of a large and heterogeneous collection of images (similar to those encountered on the Web) that are accompanied by unstructured and noisy textual annotations in English, and are searched for by users with diverse information needs.

It is an ad-hoc image retrieval task with an evaluation scenario similar to the classic TREC ad-hoc retrieval task and the ImageCLEFphoto task: simulation of the situation in which a system knows the set of documents to be searched, but cannot anticipate the particular topic that will be investigated (i.e., the topics are not known to the system in advance). Given a textual query (and/or sample images and/or concepts) describing a user's multimedia information need, the aim is to find as many relevant images as possible from the Wikipedia image collection. A multimedia retrieval approach in that case should aim at combining the evidence of relevance from different media types into a single ranking that is presented to the user. In this first year of the task, the focus in on monolingual retrieval.

The paper is organised as follows. First, we introduce the task resources (the image collection and other available resources), the topics, and assessments (Sections 2–4). Section 5 presents the approaches employed by the different participants, while Section 6 summarises their main results. Section 7 concludes the paper and provides an outlook on next year's task.

2 Task Resources

The resources used for the wikipediaMM task are based on Wikipedia data. The following resources were made available to the participants:

(INEX MM) wikipedia image collection: The collection consists of approximately 150,000 JPEG and PNG images provided by Wikipedia's users. Each image is associated with user generated, alphanumeric, unstructured metadata in English. These metadata typically contain a brief caption or description of the image, the Wikipedia user who uploaded the image, and the copyright information (see Figure 1 for an example). These descriptions are highly heterogeneous and of varying length. Further information about the image collection can be found in [5].

Image classification scores: For each image, the classification scores for the 101 different MediaMill concepts were provided by UvA [3]. The UvA classifier had been trained on manually annotated TRECVID video data and the concepts were selected for the broadcast news domain.

Image features: For each image, the set of the 120D feature vectors used to derive the above image classification scores [1] was also made available. Participants could use these feature vectors to custom-build a content-based image retrieval (CBIR) system, without having to pre-process the image collection.

These resources had also been provided in the INEX 2006-2007 Multimedia track. The latter two resources are beneficial to researchers who wish to exploit visual evidence without performing image analysis.

3 Topics

The topics for the ImageCLEF 2008 wikipediaMM task include (i) topics previously used in INEX 2006-2007 Multimedia track, and (ii) topics created by this year's task participants. They are descriptions of multimedia information needs that may contain not only textual, but also visual evidence, in the form of sample images and concepts.

3.1 Topic Format

The wikipediaMM topics are multimedia queries that can consist of a textual, visual, and conceptual part, with the latter two parts being optional.

<**title**> query by keywords
<**concept**> query by concept (optional)
<**image**> query by image content (optional)
<**narrative**> description in which the definition of relevance and irrelevance are given

Fig. 1. Example Wikipedia image+metadata from the (INEX MM) wikipedia image collection.

<title> The topic <title> simulates a user who does not have (or does not want to use) example images or other visual information. The query expressed in the topic <title> is therefore a text-only query. This profile is likely to fit most users searching digital libraries.

Upon discovering that a <title>-only query returns many irrelevant hits, users might decide to reformulate it by adding visual information.

<concept> This field is directly related to the concepts for which classification results are provided as an additional source of information (see Section 2), i.e., they are restricted to the 101 MediaMill concepts.

<image> The second type of visual evidence are example images, which can be taken from outside or inside Wikipedia and can be of any common format.

<narrative> A clear and precise description of the information need is required in order to unambiguously determine whether or not a given image fulfils the given need. In a test collection setting, this description is known as the narrative. It is the only true and accurate interpretation of a user's need. Precise recording of the narrative is important for scientific repeatability - there must exist a definitive description of what is and is not relevant to the user. To aid this, the <narrative> should explain not only what information is being sought, but also

the context and motivation of the information need, i.e., why the information is being sought and what work-task it might help to solve.

The three different types of information sources (textual terms, visual examples, and concepts) can be used in any combination. For each field more than one entry can be specified. It is up to the systems how to use, combine or ignore this information; the relevance of a result item does not directly depend on them, but it is decided by manual assessments based on the <narrative>.

3.2 Topic Development

The topics in the wikipediaMM task have been mainly developed by the participants. Altogether, 12 of the participating groups submitted 70 candidate topics. The 35 topics used in INEX 2006-2007 Multimedia were also added to the candidate topic pool. The task organisers judged all topics in the pool in terms of their "visuality" as proposed in [2] (replacing though the so-called "neutral" option with a "textual" one). This led to the following classification of candidate topics:

visual: topics that have visual properties that are highly discriminating for the problem (e.g., "blue flower"). Therefore, it is highly likely that CBIR systems would be able to deal with them.

textual: topics that often consist of proper nouns of persons, buildings, locations, etc. (e.g., "Da Vinci paintings"). As long as the images are correctly annotated, text-only approaches are likely to suffice.

semantic: topics that have a complex set of constraints, need world knowledge, or contain ambiguous terms (e.g., "labor demonstrations"). It is highly likely that no modality alone is effective.

The candidate topics were classified by the organisers; for the old INEX topics, the results of the INEX participants' runs were also used to aid this classification. The final topic set is listed in Table 1 and consists of 75 topics: 5 visual (topic IDs: 1-5), 35 textual (topic IDs: 6-40), and 35 semantic (topic IDs: 41-75). Table 2 shows some statistics on the topics. Not all topics contain visual/multimedia information (i.e., image examples or concepts); this corresponds well with realistic scenarios, since users who express multimedia information needs do not necessarily want to employ visual evidence.

4 Assessments

The wikipediaMM task is an image retrieval task, where an image with its metadata is either relevant or not (binary relevance). We adopted TREC-style pooling of the retrieved images with a pool depth of 100, resulting in pools of between 753 and 1850 images with a mean and median both around 1290. The evaluation was performed by the participants of the task within a period of 4 weeks after

Table 1. Topics for the ImageCLEF 2008 wikipediaMM task: their IDs, titles, and whether they include visual information (Yes/No) in the form of image examples (IMG) and concepts (CON).

ID	Topic title	IMG	CON	ID	Topic title	IMG	CON
1	blue flower	Y	Y	2	sea sunset	N	N
3	ferrari red	Y	Y	4	white cat	Y	Y
5	silver race car	N	Y	6	potato chips	N	N
7	spider web	Y	N	8	beach volleyball	Y	Y
9	surfing	Y	Y	10	portrait of Jintao Hu	Y	Y
11	map of the United States	N	Y	12	rabbit in cartoons	Y	Y
13	DNA helix	Y	Y	14	people playing guitar	Y	Y
15	sars china	Y	N	16	Roads in California	Y	N
17	race car	N	Y	18	can or bottle of beer	N	N
19	war with guns	N	N	20	hunting dog	N	Y
21	oak tree	Y	Y	22	car game covers	Y	N
23	british trains	Y	N	24	peace anti-war protest	Y	Y
25	daily show	N	Y	26	house architecture	N	Y
27	baseball game	Y	N	28	cactus in desert	Y	Y
29	pyramid	Y	Y	30	video games	N	N
31	bridges	N	N	32	mickey mouse	Y	N
33	Big Ben	N	N	34	polar bear	N	N
35	George W Bush	Y	N	36	Eiffel tower	N	N
37	Golden gate bridge	Y	N	38	Da Vinci paintings	Y	N
39	skyscraper	Y	Y	40	saturn	Y	N
41	ice hockey players	N	Y	42	labor demonstrations	N	Y
43	mountains under sky	N	Y	44	graph of a convex function	Y	Y
45	paintings related to cubism	Y	Y	46	London parks in daylight	Y	Y
47	maple leaf	Y	Y	48	a white house with a garden	Y	Y
49	plant	Y	Y	50	stars and nebulae in the dark sky	Y	N
51	views of Scottish lochs	Y	N	52	Cambridge university buildings	Y	N
53	military aircraft	N	Y	54	winter landscape	N	N
55	animated cartoon	N	Y	56	London city palaces	N	Y
57	people riding bicycles	Y	N	58	sail boat	Y	Y
59	dancing couple	N	Y	60	atomic bomb	Y	Y
61	Singapore	N	N	62	cities by night	Y	Y
63	star galaxy	N	N	64	football stadium	N	N
65	famous buildings of Paris	N	Y	66	historic castle	N	N
67	bees with flowers	Y	Y	68	pyramids in Egypt	Y	Y
69	mountains with snow under sky	N	Y	70	female players beach volleyball	Y	Y
71	children riding bicycles	N	N	72	civil aircraft	N	Y
73	bridges at night	Y	Y	74	gothic cathedral	Y	Y
75	manga female character	N	Y				

Table 2. ImageCLEF 2008 wikipediaMM topics

	all	visual	textual	semantic
Number of topics	75	5	35	35
Average number of terms in title	2.64	2.2	2.3	2.9
Number of topics with image(s)	43	3	22	18
Number of topics with concept(s)	45	4	16	25
Number of topics with both image and concept	28	3	11	14
Number of topics with text only	15	1	8	6

Table 3. Resources used by the 77 submitted runs

Resource modality		# runs using it
textual	Txt	35
visual	Img	5
concept	Con	0
textual/visual	TxtImg	22
textual/concept	TxtCon	13
textual/visual/concept	TxtImgCon	2

the submission of runs. The 13 groups that participated in the evaluation process used a web-based interface previously employed in the INEX Multimedia and TREC Enterprise tracks. Each participating group was assigned 4-5 topics and an effort was made to ensure that most of the topics were assigned to their creators. This was achieved in 76% of the assignments for the topics created by this year's participants.

5 Participants

A total of 12 groups submitted 77 runs. Table 3 gives an overview of the resources used by the submitted runs[3]. Most of the runs are textual only approaches, but compared to the INEX Multimedia track, there is a rise in fusion approaches that combine text and images, text and concepts, and all three modalities.

Below we briefly describe the approaches investigated by the participating groups:

Digital Media Institute, Peking University, China (7 runs). They investigated the following three approaches: (i) a text-based approach with query expansion where the expansion terms are automatically selected from a knowledge base that is (semi-)automatically constructed from Wikipedia, (ii) a content-based visual approach, where they first train 1-vs-all classifiers for all queries by using the training images obtained by Yahoo! search, and then treat the retrieval task as a visual concept detection in the given Wikipedia image set, and (iii) a cross-media approach that combines and reranks the text- and content-based retrieval results.

[3] Our analysis is based on the runs' descriptions given by the participants themselves.

CEA LIST, France (2 runs). Their approach was based on query reformulation using concepts considered to be semantically related to those in the initial query. For each interesting entity in the query, they used Wikipedia and WordNet to extract related concepts, which were further ranked based on their perceived usefulness. A small list of visual concepts was used to rerank the answers to queries that included them, implicitly or explicitly. They submitted two automatic runs, one based on query reformulation only, and one combining query reformulation and visual concepts.

NLP and Information Systems, University of Alicante, Spain (24 runs). They employed their textual passage-based retrieval system as their baseline approach, which was enhanced by a module that decomposes the (compound) file names in camel case notation into single terms, and by a module that performs geographical query expansion. They also investigated Probabilistic Relevance Feedback and Local Context Analysis techniques.

Data Mining and Web Search, Hungarian Academy of Sciences (8 runs). They used their own retrieval system to experiment with a text-based approach, that uses BM25 and query expansion based on blind relevance feedback, and its combination with a segment-based visual approach.

Database Architectures and Information Access, CWI, Netherlands (2 runs). They used a language modelling approach based on purely textual evidence and also incorporated a length prior to bias retrieval towards images with longer descriptions than the ones retrieved by the language model.

Laboratoire Hubert Curien, Université Jean Monnet, Saint-Etienne, France (6 runs). They used a vector space model to compute similarities between vectors of both textual and visual terms; the textual part is a term vector of the terms' BM25 weights and the visual part a 6-dimensional vector of clusters of colour features. The applied both manual and blind relevance feedback to a text-based run in order to expand the query with visual terms.

Dept. of Computer Science and Media, Chemnitz University of Technology (4 runs). They used their Xtrieval framework based on both textual and visual features, and also made use of the provided resources (concepts and features). They experimented with text-based retrieval, its combination with a visual approach, the combination of all three modalities, and thesaurus-based query expansion. They also investigated the efficiency of the employed approaches.

Multimedia and Information Systems, Imperial College, UK (6 runs). They examined textual features, visual features, and their combination. Their text-based approach was combined with evidence derived from a geographic co-occurrence model mined from Wikipedia which aimed at disambiguating geographic references either in a context-dependent or a context-independent manner. Their visual-based approach, employing Gabor texture features and the Cityblock distance as a similarity measure, was combined with the text-based approach and with blind relevance feedback using a convex combination of ranks.

SIG-IRIT, Toulouse, France (4 runs). They explored the use of images' names as evidence in text-based image retrieval. They used them in isolation,

by computing a similarity score between the query and the name of images using the vector space model, and in combination with textual evidence, either by fusing the ranking of a text-based approach (based on the XFIRM retrieval system) with the ranking produced by the name-based technique, or by using the text-based approach with an increase in the weight of terms in the image name.

Computer Vision and Multimedia, Université de Genève, Switzerland (2 runs). Their approach was based on the preference ranking option of the SVM light library developed by Cornell University. Their first run employed a text-based retrieval approach. Their second run applied a feature selection to the high dimensional textual feature vector based on the features relevant to each query.

UPMC/LIP6 - Computer Science Lab, Paris, France (7 runs). They investigated (i) text-based retrieval, using a *tf.idf* approach, a language modelling framework, and their combination based on the ranks of retrieved images, and (ii) the combination of textual and visual evidence, by reranking the text-based results using visual similarity scores (Euclidean distance and a manifold-based technique, both based on HSV features).

LSIS, UMR CNRS & Université Sud Toulon-Var, France (5 runs). They applied the same techniques they used for the Visual Concept Detection task at ImageCLEF 2008, by relating each of the wikipediaMM topics to one or more visual concepts from that task. They also fused these visual-based rankings with the results of a text-based approach.

6 Results

The analysis presented in this section takes into account only the top 75% of all submitted runs that perform best in terms of Mean Average Precision (MAP), so as to exclude noise by removing erroneous and buggy runs. Table 4 shows the top 30 runs ranked by their MAP; the complete result list of results can be found at: `http://www.imageclef.org/2008/wikimm-results`.

We first analyse the performance per modality by comparing the average performance over runs that employ the same type of resources. Table 5 shows the average performance and standard deviation with respect to modality. There were no automatic image-only runs in the top 75% of the runs. According to the MAP measure, the best performing runs fuse text and concepts, followed by the runs that fuse text, concepts and images, and the text-only baseline. The latter two perform almost identically. A similar result is obtained with the precision after R (= number of relevant) documents are retrieved (R-prec.). Again, the TxtCon runs outperform all others. The runs that fuse textual and visual information (TxtImg) perform worst for all measures, so this remains an open research issue. But, in general, it can be said that this year the fusion approaches perform surprisingly well, mostly due to the incorporation of concepts.

Next, we analyse the performance per topic. Figure 2 shows for each topic the average MAP over all (top 75% of) runs, over only the text-based ones among

Table 4. Results for the top 30 submitted runs ranked by their MAP.

Group	Run	Modality	FB/QE	MAP	P@10	P@20	R-prec.
1 upeking	zzhou3	Txt	QE	0.3444	0.4760	0.3993	0.3794
2 cea	ceaTxtCon	TxtCon	QE	0.2735	0.4653	0.3840	0.3225
3 ualicante	IRnNoCamel	Txt	NOFB	0.2700	0.3893	0.3040	0.3075
4 cea	ceaTxt	Txt	QE	0.2632	0.4427	0.3673	0.3080
5 ualicante	IRnNoCamelLca	Txt	FB	0.2614	0.3587	0.3167	0.2950
6 ualicante	IRnNoCamelGeog	Txt	QE	0.2605	0.3640	0.2913	0.3000
7 ualicante	IRnConcSinCamLca	TxtCon	FB	0.2593	0.3493	0.2900	0.3016
8 ualicante	IRnConcSinCam	TxtCon	NOFB	0.2587	0.3627	0.2900	0.2975
9 ualicante	IRnNoCamelLcaGeog	Txt	FBQE	0.2583	0.3613	0.3140	0.2922
10 sztaki	bp_acad_avg5	TxtImg	NOFB	0.2551	0.3653	0.2773	0.3020
11 sztaki	bp_acad_textonly_qe	Txt	QE	0.2546	0.3720	0.2907	0.2993
12 ualicante	IRnConcSinCamLcaGeog	TxtCon	FBQE	0.2537	0.3440	0.2853	0.2940
13 cwi	cwi_lm_txt	Txt	NOFB	0.2528	0.3427	0.2833	0.3080
14 sztaki	bp_acad_avgw_glob10	TxtImg	NOFB	0.2526	0.3640	0.2793	0.2955
15 sztaki	bp_acad_avgw_glob10_qe	TxtImg	QE	0.2514	0.3693	0.2833	0.2939
16 ualicante	IRnConcSinCamGeog	TxtCon	QE	0.2509	0.3427	0.2787	0.2924
17 sztaki	bp_acad_glob10_qe	TxtImg	QE	0.2502	0.3653	0.2833	0.2905
18 sztaki	bp_acad_glob10	TxtImg	NOFB	0.2497	0.3627	0.2780	0.2955
19 cwi	cwi_lm_lprior_txt	Txt	NOFB	0.2493	0.3467	0.2787	0.2965
20 sztaki	bp_acad_avg5	TxtImg	NOFB	0.2491	0.3640	0.2773	0.2970
21 sztaki	bp_acad_avgw_qe	TxtImg	QE	0.2465	0.3640	0.2780	0.2887
22 curien	LaHC_run01	Txt	NOFB	0.2453	0.3680	0.2860	0.2905
23 ualicante	IRnConcSinCamPrf	TxtCon	FB	0.2326	0.2840	0.2700	0.2673
24 ualicante	IRnNoCamelPrf	Txt	FB	0.2321	0.3107	0.2800	0.2665
25 ualicante	IRnNoCamelPrfGeog	Txt	FBQE	0.2287	0.3120	0.2787	0.2611
26 ualicante	IRnConcSinCamPrfGeog	TxtCon	FBQE	0.2238	0.2853	0.2673	0.2561
27 chemnitz	cut-mix-qe	TxtImgCon	QE	0.2195	0.3627	0.2747	0.2734
28 ualicante	IRnConcepto	TxtCon	NOFB	0.2183	0.3213	0.2520	0.2574
29 ualicante	IRn	Txt	NOFB	0.2178	0.3760	0.2507	0.2569
30 chemnitz	cut-txt-a	Txt	NOFB	0.2166	0.3440	0.2833	0.2695

Table 5. Results per modality over all topics.

Modality	MAP		P@20		R-prec.	
	Mean	SD	Mean	SD	Mean	SD
All top 75% runs	0.2149	0.049	0.2676	0.047	0.2566	0.050
Txt in top 75% runs	0.2104	0.053	0.2643	0.052	0.2510	0.055
Img in top 75% runs	–	–	–	–	–	–
TxtCon in top 75% runs	**0.2316**	0.025	**0.2874**	0.033	**0.2742**	0.026
TxtImg in top 75% runs	0.2078	0.061	0.2522	0.047	0.2516	0.060
TxtImgCon in top 75% runs	0.2122	0.010	0.2683	0.014	0.2559	0.012

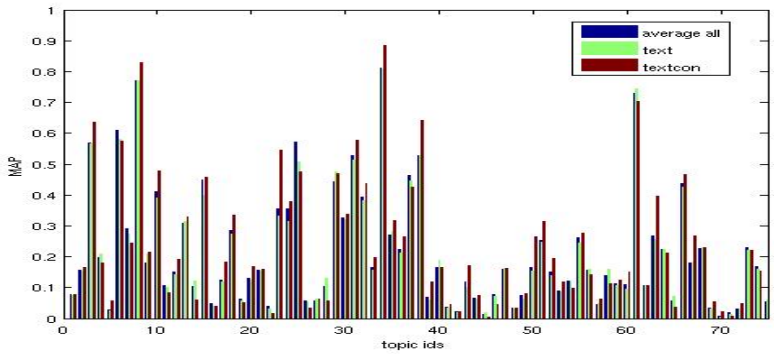

Fig. 2. Performance of different modalities per topic.

Table 6. Top 10 topics for which text/concept runs outperform text-only runs.

text/concept
(23) british trains
(63) star galaxy
(38) Da Vinci Paintings
(50) stars/nebulae dark sky
(10) portrait of Jintao Hu
(67) bees with flowers
(34) polar bear
(43) mountains under sky
(51) Views of Scottish lochs
(3) ferrari red

them, and over only the text/concept-based ones among them. The text/concept-based runs outperform the text-based ones for 64% of the topics with a maximum improvement of 21% in a single topic. Table 6 presents the top 10 topics, where the text/concept runs outperform the textual. The list includes a mixture of topics with proper nouns and complex concepts. Oddly, this top 10 also includes topics that do not have an example concept (see Table 1), which suggests that some conceptual query expansion or a concept schema different from the provided one has been used by some of the participants. These could also be topics that are not well annotated in the collection.

Figure 3 compares the text baseline with the other 2 fusion approaches: (1) text/visual and (2) text/concept/visual. The fusion of text and images outperforms the text-based runs only for less than half of the topics (44%) and the fusion of all 3 modalities for 56%. The maximum improvement of a topic over the baseline is 29% for TextImg and 38% for TextConImg respectively. We again create a list of the top 10 topics for which each fusion approach outperforms the text-only baseline. The lists for the fusion of text/visual and text/concept/visual

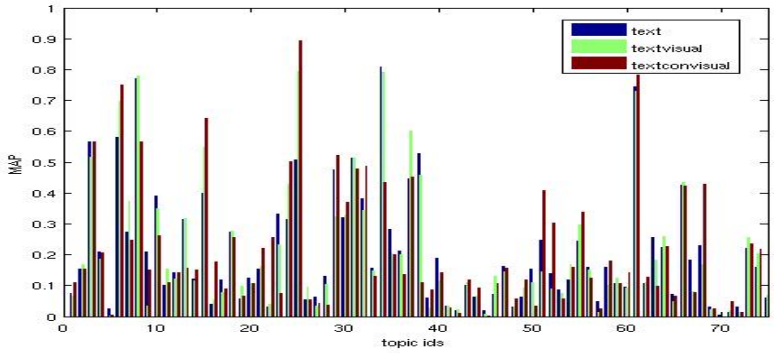

Fig. 3. Performance of different fusion approaches per topic.

Table 7. Top 10 topics for which fusion runs outperform textl-only runs.

text/visual	text/concept/visual
(25) daily show	(25) daily show
(37) Golden gate bridge	(15) sars china
(15) sars china	(22) car game covers
(6) potato chips	(68) pyramids in egypt
(24) peace anti-war protest	(24) peace anti-war protest
(7) spider web	(6) potato chips
(46) London parks in daylight	(52) Cambridge University Buildings
(11) map of the united states	(51) Views of Scottish lochs
(55) animated cartoon	(16) Roads in California
(54) winter landscape	(32) Mickey Mouse

have many entries in common, and contain mostly topics with complex concepts that have an image and/or a concept defined.

In summary, the results indicate that fusion approaches are catching up with the text-based approaches. Especially the text/concept fusion approaches perform particularly well. The visual hints help mainly for topics that incorporate an image example, but can also improve the overall performance.

7 Conclusion and Outlook

Our debut in ImageCLEF 2008 attracted much interest from groups researching multimedia retrieval and significantly more participants than the INEX 2006-2007 Multimedia track. With the help of our participants, we both developed and assessed a diverse set of 75 multimedia topics. The results indicate that the dominance of text-based image retrieval is coming to an end; multi-modal fusion approaches help to improve the retrieval performance in this domain.

Our main focus for next year remains the same: researching the combination of evidence from different modalities in a “standard” ad-hoc image retrieval

task. Possible new directions for 2009 include the addition of multilinguality in form of multi-lingual topics (and if possible annotations), and access to the context of the images, i.e., the Wikipedia web pages that contain them. We also aim at providing new sets of classification scores and low-level features, so that participants can concentrate their research effort on information fusion.

8 Acknowledgements

Theodora Tsikrika was supported by the European Union via the European Commission project VITALAS (contract no. 045389). Jana Kludas was funded by the European project MultiMATCH (EU-IST-STREP#033104). The authors would also like to thank Thomas Deselaers for invaluable technical support and all the groups participating in the relevance assessment process.

References

1. Jan C. van Gemert, Jan-Mark Geusebroek, Cor J. Veenman, Cees G. M. Snoek, and Arnold W. M. Smeulders. Robust scene categorization by learning image statistics in context. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 105, Washington, DC, USA, 2006. IEEE Computer Society.
2. Michael Grubinger and Paul D. Clough. On the creation of query topics for imageclefphoto. In *Proceedings of the Third MUSCLE / ImageCLEF Workshop on Image and Video Retrieval Evaluation (2007)*, 2007.
3. Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek, and Arnold W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 421–430, New York, NY, USA, 2006. ACM Press.
4. Theodora Tsikrika and Thijs Westerveld. The INEX 2007 multimedia track. In Norbert Fuhr, Mounia Lalmas, Andrew Trotman, and Jaap Kamps, editors, *Focused access to XML documents, 6th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2007, Revised and Selected Papers*. Springer, 2008.
5. Thijs Westerveld and Roelof van Zwol. The INEX 2006 multimedia track. In Norbert Fuhr, Mounia Lalmas, and Andrew Trotman, editors, *Advances in XML Information Retrieval: 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Revised Selected Papers*, volume 4518, pages 331–344. Springer, 2007.