The INEX 2007 Multimedia Track

Theodora Tsikrika[1] and Thijs Westerveld[2⋆]

[1] CWI, Amsterdam, The Netherlands
[2] Teezir Search Solutions, Ede, The Netherlands

Abstract. The INEX Multimedia track focuses on using the structure of XML documents to extract, relate, and combine the relevance of different multimedia fragments. This paper presents a brief overview of the track for INEX 2007, including the track's test collection, tasks, and goals. We also report the approaches of the participating groups and their main results.

1 Introduction

Structured document retrieval from XML documents allows for the retrieval of XML document fragments, i.e., XML elements or passages, that contain relevant information. The main INEX Ad Hoc task focuses on text-based XML retrieval. Although text is dominantly present in most XML document collections, other types of media can also be found in those collections. Existing research on multimedia information retrieval has already shown that it is far from trivial to determine the combined relevance of a document that contains several multimedia objects [5].

The objective of the INEX Multimedia track is to exploit the XML structure that provides a logical level at which multimedia objects are connected, in order to improve the retrieval performance of an XML-driven multimedia information retrieval system. To this end, it provides an evaluation platform for the retrieval of multimedia documents and document fragments. In addition, it creates a discussion forum where the participating groups can exchange their ideas on different aspects of the multimedia XML retrieval task.

This paper reports on the INEX 2007 Multimedia track and is organised as follows. First, we introduce the main parts of the test collection: documents, tasks, topics, and assessments (Sections 2–5). Section 6 presents the approaches employed by the different participants and Section 7 summarises their main results. Section 8 concludes the paper and provides an outlook on next year's track.

2 Wikipedia collections and additional resources

In INEX 2007, the Multimedia track employed the following two Wikipedia-based collections (the same as in 2006):

⋆ Part of this work was carried out when the author was at CWI, Amsterdam, The Netherlands.

Wikipedia XML collection: This is a structured collection of 659,388 Wikitext pages from the English part of Wikipedia, the free content encyclopedia (`http://en.wikipedia.org`), that have been converted to XML [1]. This collection has been created for the Ad Hoc track. Given, though, its multimedia nature (as indicated by its statistics listed in Table 1), it is also being used as the target collection for a multimedia task that aims at finding relevant XML fragments given a multimedia information need (see Section 3).

Table 1. Wikipedia XML collection statistics

Total number of XML documents	659,388
Total number of images	344,642
Number of unique images	246,730
Average number of images per document	0.52
Average depth of XML structure	6.72
Average number of XML nodes per document	161.35

Wikipedia image XML collection: This is a collection consisting of the images in the Wikipedia XML collection, together with their metadata which have been formatted in XML. These metadata usually contain a brief caption or description of the image, the Wikipedia user who uploaded the image, and the copyright information. Figure 1 shows an example of such a document consisting of an image and its associated metadata. Some images from the Wikipedia XML collection have been removed due to copyright issues or parsing problems with their metadata, leaving us with a collection of 170,370 images with metadata. This collection is used as the target collection for a multimedia/image retrieval task that aims at finding images (with metadata) given a multimedia information need (see Section 3).

Although the above two Wikipedia-based collections are the main search collections, additional sources of information are also provided to help participants in the retrieval tasks. These resources are:

Image classification scores: For each image, the classification scores for the 101 different MediaMill concepts are provided by UvA [6]. The UvA classifier is trained on manually annotated TRECVID video data and the concepts are selected for the broadcast news domain.

Image features: For each image, the set of the 120D feature vectors that has been used to derive the above image classification scores is available [3]. Participants can use these feature vectors to custom-build a content-based image retrieval (CBIR) system, without having to pre-process the image collection.

These resources were also provided in 2006, together with an online CBIR system that is no longer available. The above resources are beneficial to researchers who wish to exploit visual evidence without performing image analysis.

Fig. 1. Example Wikipedia image+metadata document from the Wikipedia image XML collection.

3 Retrieval Tasks

The aim of the retrieval tasks in the Multimedia track is to retrieve relevant (multimedia) information, based on an information need with a (structured) multimedia character. To this end, a structured document retrieval approach should be able to combine the relevance of different media types into a single ranking that is presented to the user.

For INEX 2007, we define the same two tasks as last year:

MMfragments task: Find relevant XML fragments in the **Wikipedia XML collection** given a multimedia information need. These XML fragments can correspond not only to XML elements (as it was in INEX 2006), but also to passages. This is similar to the direction taken by the INEX Ad Hoc track. In addition, since MMfragments is in essence comparable to the ad hoc retrieval of XML fragments, this year it ran along the Ad Hoc tasks. As a result, the three subtasks of the Ad Hoc track (see [2] for detailed descriptions) are also defined as subtasks of the MMfragments task:

1. FOCUSED TASK asks systems to return a ranked list of elements or passages to the user.
2. RELEVANT IN CONTEXT TASK asks systems to return relevant elements or passages clustered per article to the user.
3. BEST IN CONTEXT TASK asks systems to return articles with one best entry point to the user.

The difference is that MMfragments topics ask for multimedia fragments (i.e., fragments containing at least one image) and may also contain visual hints (see Section 4).

MMimages task: Find relevant images in the **Wikipedia image XML collection** given a multimedia information need. Given an information need, a retrieval system should return a ranked list of documents(=image+metadata) from this collection. Here, the type of the target element is defined, so basically this is closer to an image retrieval (or a document retrieval) task, rather than XML element or passage retrieval. Still, the structure of (supporting) documents, together with the visual content and context of the images, could be exploited to get to the relevant images (+their metadata).

All track resources (see Section 2) can be used for both tasks, but the track encourages participating groups to also submit a baseline run that uses no sources of information except for the target collection. This way, we hope to learn how the various sources of information contribute to the retrieval results. Furthermore, we also encourage each group to submit a run that is based on only the `<mmtitle>` field of the topic description (see Section 4). All other submissions may use any combination of the `<title>`, `<castitle>`, `<mmtitle>` and `<description>` fields (see Section 4). The fields used need to be reported.

4 Topics

The topics used in the INEX Multimedia track are descriptions of (structured) multimedia information needs that may contain not only textual, but also structural and multimedia hints. The structural hints specify the desirable elements to return to the user and where to look for relevant information, whereas the multimedia hints allow the user to indicate that results should have images similar to a given example image or be of a given concept. These hints are expressed in the NEXI query language [8].

The original NEXI specification determines how structural hints can be expressed, but does not make any provision for the expression of multimedia hints. These have been introduced as NEXI extensions during the INEX 2005 and 2006 Multimedia tracks [9, 10]:

- To indicate that results should have images similar to a given example image, an *about* clause with the keyword *src:* is used. For example, to find images of cityscapes similar to the image at `http://www.bushland.de/hksky2.jpg`, one could type:

  ```
  //image[about(.,cityscape) and
                  about(.,src:http://www.bushland.de/hksky2.jpg)]
  ```

  In 2006, only example images from within the Wikipedia image XML collection were allowed, but this year it was required that the example images came from outside the Wikipedia collections.

– To indicate that the results should be of a given concept, an *about* clause with the keyword *concept:* is used. For example, to search for cityscapes, one could decide to use the concept "building":

```
//image[about(.,cityscape) and about(.,concept:building)]
```

This feature is directly related to the concept classifications that are provided as an additional source of information (see Section 2). Therefore, terms following the keyword *concept:* are obviously restricted to the 101 concepts for which classification results are provided.

It is important to realise that all structural, textual and visual filters in the query should be interpreted loosely. It is up to the retrieval systems to decide how to use, combine or even ignore this information. The relevance of a document, element or passage does not directly depend on these hints, but is determined by manual assessments.

4.1 Topic format

The INEX Multimedia track topics are similar to the Content Only + Structure (CO+S) topics of the INEX Ad Hoc track. In INEX, "Content" refers to the textual or semantic content of a document part, and "Content-Only" to topics or queries that use no structural hints. The Ad Hoc CO+S topics include structural hints, whereas the Multimedia CO+S topics may also include visual hints.

The 2007 Multimedia CO+S topics consist of the following parts:

`<title>` The topic `<title>` simulates a user who does not know (or does not want to use) the actual structure of the XML documents in a query and who does not have (or want to use) example images or other visual hints. The query expressed in the topic `<title>` is, therefore, a Content Only (CO) query. This profile is likely to fit most users searching XML digital libraries and also corresponds to the standard web search type of keyword search.

`<castitle>` A NEXI expression with structural hints.

`<mmtitle>` A NEXI expression with structural and visual hints.

`<description>` A brief, matter of fact, description of the information need. Like a natural language description one might give to a librarian.

`<narrative>` A clear and precise description of the information need. The narrative unambiguously determines whether or not a given document or document part fulfils the given need. It is the only true and accurate interpretation of a user's needs. Precise recording of the narrative is important for scientific repeatability - there must exist, somewhere, 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.

In previous years, both structural and visual/multimedia hints were expressed in the `<castitle>` field. This year, the `<castitle>` contains only structural hints, while the `<mmtitle>` is an extension of the `<castitle>` that also

incorporates the additional visual hints (if any). The introduction of a separate `<mmtitle>` is particularly useful, since it makes it easier for systems to compare runs using structural hints to those using structural+visual hints, without having to modify the query expression. In addition, Multimedia CO+S topics can now also be used in Ad Hoc tasks, since they contain fields (all, except `<mmtitle>`) that can be directly processed by an Ad Hoc system.

The fact that the MMfragments task is similar to ad hoc retrieval, not only led to the decision to run the MMfragments tasks along the Ad Hoc ones, but also to include the MMfragments topics as a subset of the Ad Hoc ones. This means that submissions for the INEX 2007 Ad Hoc track also considered the subset of topics used for the MMfragments task. This allows us to compare ad hoc XML retrieval systems submissions on the MMfragments topic subset (i.e., submissions that retrieve XML document parts by using any of the available fields except `<mmtitle>`) to multimedia XML retrieval submissions on the same topic subset (i.e., to submissions that can use any of the topic fields, together with the knowledge that a multimedia XML fragment is required as a retrieval result).

MMimages, on the other hand, runs as a separate task with a separate set of topics. Given that MMimages requires retrieval at the document level, rather than elements or passages, the queries in the `<castitle>` and `<mmtitle>` fields are restricted to: `//article[X]`, where `X` is a predicate using one or more *about* functions with textual and/or multimedia hints.

4.2 Topic development

The topics in the Multimedia track are developed by the participants. Each participating group has to create 2 multimedia topics for the MMfragments task and 4 topics for MMimages. Topic creators first create a 1-2 sentence description of the information they are seeking. Then, in an exploration phase, they obtain an estimate of the amount of relevant information in the collection. For this, they can use any retrieval system, including their own system or the TopX system [7] provided through the INEX organisation. The topic creator then assesses the top 25 results and abandons the search if fewer than two or more than 20 relevant fragments are found. If between 2 and 20 fragments are found to be relevant, the topic creator should have a good idea of what query terms should be used, and the `<title>` is formulated. Using this title a new search is performed and the top 100 elements are assessed. Having judged these 100 documents, topic creators should have a clear idea of what makes a fragment relevant or not. Based on that, they could then first write the narrative and then the other parts of the topic. After each created topic, participants are asked to fill a questionnaire that gathers information about the users familiarity with the topic, the expected number of relevant fragments in the collection, the expected size of relevant fragments and the realism of the topic. The submitted topics are analysed by the INEX Multimedia organisers who check for duplicates and inconsistencies before distributing the full set of topics among the participants.

Table 2 shows the distribution over tasks as well as some statistics on the topics. The MMfragments topics correspond to Ad Hoc topics 525-543. Their average number of terms in `<title>` (3.21) is slightly lower than the average number of terms in the remaining 80 Ad Hoc topics (3.92). This is to be expected, since users who submit multimedia topics express their requirements not only by textual, but also by visual hints. Table 2 indicates that not all topics contain visual/multimedia hints; this corresponds well with realistic scenarios, since users who express multimedia information needs do not necessarily want to employ visual hints.

Table 2. Statistics for the INEX 2007 MM topics

	MMfragments	**MMimages**	**All**
Number of topics	19	20	39
Average number of terms in `<title>`	3.21	2.35	2.77
Number of topics with `<mmtitle>`	6	10	16
Number of topics with src:	2	7	9
Number of topics with concept:	4	6	10
Number of topics with both src: and concept:	0	3	3

5 Assessments

Since XML retrieval requires assessments at a sub-document level, a simple binary judgement at the document level is not sufficient. Still, for ease of assessment, retrieved fragments are grouped by document. Since the INEX 2007 MMfragments task was run in parallel with the Ad Hoc track, the assessments for this task were arranged by the Ad Hoc track organization as follows. Once all participants have submitted their runs, the top N fragments for each topic are pooled and grouped by document. The documents are alphabetised so that the assessors do not know how many runs retrieved fragments from a certain document or at what rank(s) the fragments were found. Assessors then look at the documents in the pool and highlight the relevant parts of each document. The assessment system stores the relevance or non-relevance of the underlying XML elements and passages.

We did not give any additional instructions to the assessors of multimedia topics, but assumed that topic creators who indicated that their topics have a clear multimedia character would only judge elements relevant if they contain at least one image. We analysed the assessed fragments to verify this. We looked at the number of <image> elements in highlighted passages and contrasted the findings for the MMfragments topics with the findings for other Ad Hoc topics, and found that indeed the fragments assessed relevant for MMfragments topics contain many more images than the relevant fragments for Ad Hoc topics. On average, a relevant passage for an Ad Hoc topic contains 0.14 images. An

average relevant passage for a MMfragments topic contains 0.62 images. The box plot in Figure 2 shows the minimum, median and maximum of the average number of images per highlighted passage over the topics in the category, the box shows the data falling between the 25th and 75th percentile. Even though not all highlighted MMfragment passages contain images, the difference with Ad Hoc topics is clear. For around 25% of the MMfragments topics the average number of images per passage is above 1, for half the topics this number is greater than 0.5.

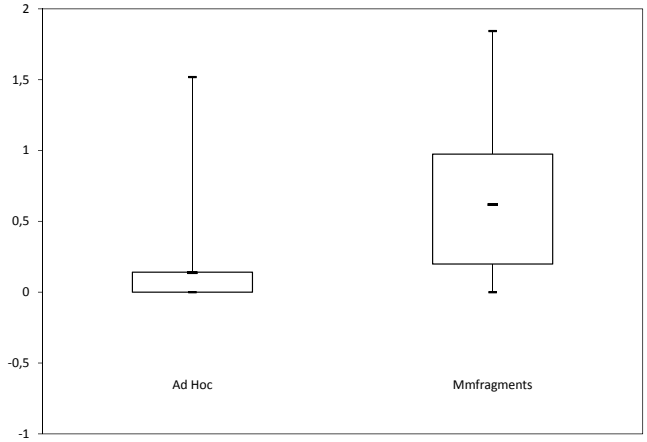

Fig. 2. Number of images per highlighted passage (minimum, median, maximum and 25th to 75th percentile).

The MMimages task is a document retrieval task. A document, i.e., an image with its metadata, is either relevant or not. For this task, we adopted TREC style document pooling of the documents and binary assessments at the document (i.e., image with metadata) level. In 2006, the pool depth was set to 500 for the MMimages task, with post-hoc analysis showing that pooling up to 200 or 300 would have given the same system ordering [10]. This led to the decision to pool this year's submissions up to rank 300, resulting in pools of between 348 and 1865 images per topic, with both mean and median around 1000 (roughly the same size as 2006).

6 Participants

Only four participants submitted runs for the INEX 2007 Multimedia track: CWI together with the University of Twente (CWI/UTwente), IRIT (IRIT), Queensland University of Technology in Australia (QUTAU), and University of Geneva (UGeneva). For the MMfragments task, three of the participants (CWI/UTwente, IRIT and QUTAU) submitted a total of 12 runs, whereas for the MMimages task, all four participants submitted a total of 13 runs.

Table 3 gives an overview of the topic fields used by the submitted runs. For MMfragments, six submissions used the topics' `<title>` field, and six submissions used the `<castitle>` field; the mmitle field was not used by any paricipant. For MMimages, seven submissions used the topics' `<title>` field, and six submissions used the `<mmtitle>` field; no submissions used the `<castitle>` field which is to be expected since this is a document retrieval task.

Table 3. Topic fields used by the submitted runs

topic field	*#MMfragments* runs using it	*#MMimages* runs using it
title	6	7
castitle	6	0
mmtitle	0	6
description	0	0
narrative	0	0

Table 4 gives an overview of the resources used by the submitted runs. Not all groups detailed the resources they used, but judging from the descriptions it appears most submissions only used the target Wikipedia collection of the task at hand. It seems the Wikipedia images collection and the UvA features and classification scores have not been used in the MMfragments task this year. In the MMimages task, the visual resources provided are used by IRIT and UGeneva, whereas some runs also used the main Wikipedia XML collection.

Table 4. Resources used by the submitted runs

resource	*#MMfragments* runs using it	*#MMimages* runs using it
wikipedia	12	4
wikipedia_IMG	0	8
UvAfeatures	0	1
UvAconcepts	0	2

Below we briefly discuss the appproaches taken by the groups that participated in the Multimedia track at INEX 2007.

CWI/UTwente CWI/UTwente participated in both MMfragments and MMimages tasks of the INEX 2007 Multimedia track. For MMfragments, they limited their system to return only fragments that contain at least one image that was part of the Wikipedia images XML collection. They did not use any further multimedia processing and experimented with traditional text based approaches based on the language modelling approach and different length priors. For MMimages, they represented each image either by its textual metadata in the Wikipedia image XML collection, or by its textual context when that image appears as part of a document in the (Ad Hoc) Wikipedia XML collection. Retrieval was then based on purely text-based approaches.

IRIT IRIT participated in both the MMfragments and MMimages tasks of the INEX 2007 Multimedia track, with methods based on the context (text and structure) of images to retrieve multimedia elements. For MMimages topics, the "MMI" method proposed last year that uses 3 sources of evidence (descendant, sibling, and ascendant nodes) is compared to a new method "MMIConc" that uses in addition images classification scores. For the MMfragments task, the "MMF" method based on the "XFIRM Content and Structured" method and the "MMI" method were evaluated. In future work, IRIT plan to extend images context by using links.

QUTAU No description of their approaches has been provided.

UGeneva For their first participation at INEX MM, they submitted three runs to the MMimages task: (1) a baseline run based only on text-based retrieval, (2) an improvement of (1) with additional proper noun detection, and (3) a multi modal fusion approach using a hierarchical SVM approach.

For the simple text-based baseline run (1), the ready-to-use Matlab library TMG [11] is applied to the MMimages collection. It creates a term-document matrix filled with term frequencies of the textual input. The retrieval is done based on the Vector Space Model (VSM). In (2) the simple baseline run is improved by adding to the approach a proper noun detection based on Google result counts. This proved to be an easy and inexpensive way to reliably detect proper nouns. The multi modal fusion run (3) used all available features: textual and visual (color and texture histogram) low level features, plus the visual concepts provided by the University of Amsterdam. The approach was set up hierarchically. First a VSM-based retrieval on the extended term-document matrix was executed. Then the result list was classified into N classes with the k-NN algorithm of the TMG library. The documents of the cluster containing the most relevant documents were taken as input for a hierarchical Support Vector Machine (SVM) classification, which processes first each modality alone, before fusing all result lists in a final step.

Université de Saint-Etienne/JustSystems These two groups did not submit any official runs for the track, but they did help with assessments for the MMimages task, and plan to use the track's data for future studies.

7 Results

This section presents the results for the submitted runs in each of the tasks.

7.1 MMfragments

Three participating groups (CWI/UTwente, IRIT and QUTAU) submitted a total of 12 MMfragments runs (5 Focused, 2 Relevant in Context and 5 Best in Context runs). Of these submissions, 6 used the topics' title field and 6 used the castitle field; the mmitle field was not used by any paricipant in the MMfragments task. Not all groups detailed the resources they used, but judging by the descriptions it appears that all submissions only used the main wikipedia collection for this task. It seems that the wikipedia images collection and the UvA features and classification scores have not been used in the MMfragments task this year.

These runs have been evaluated using the standard measures as used in the Ad Hoc track [4]: interpolated Precision (iP) and Mean Average interpolated Precision (MAiP) for the Focused task and non-interpolated generalized precision at early ranks gP[r] and non-interpolated mean average generalized precision MAgP). Tables 5-7 show the results.

Table 5. MMfragments Results for Focused task.

MAiP	iP[0.00]	iP[0.01]	iP[0.05]	iP[0.10]	Group	Run
0.1169	0.4158	0.3389	0.2921	0.2546	utwente	article_MM
0.0910	0.3744	0.3039	0.2160	0.1713	qutau	COS_Focused
0.1218	0.2989	0.2947	0.2790	0.2382	qutau	CO_Focused
0.0991	0.2471	0.2467	0.2422	0.2294	utwente	star_loglength_MM
0.0042	0.3448	0.0595	0.0000	0.0000	utwente	star_lognormal_MM

Table 6. MMfragments Results for Relevant in Context task.

MAgP	gP[5]	gP[10]	gP[25]	gP[50]	Group	Run
0.1043	0.1729	0.1763	0.1528	0.1193	qutau	CO_RelevantInContext
0.0900	0.2072	0.1787	0.1441	0.1085	qutau	COS_RelevantInContext

Since the MMfragments topics were mixed with the Ad Hoc topics we received many more submissions that were not tailored to answering information

Table 7. MMfragments Results for Best in Context task.

MAgP	gP[5]	gP[10]	gP[25]	gP[50]	Group	Run
0.1783	0.3210	0.3039	0.2558	0.2099	qutau	CO_BestInContext
0.1533	0.3671	0.3084	0.2334	0.1761	qutau	COS_BestInContext
0.0541	0.1423	0.1394	0.0784	0.0437	irit	iritmmf06V2_BIC
0.0506	0.1133	0.1319	0.1267	0.0943	irit	iritmmf06V1
0.0458	0.1164	0.1316	0.1114	0.0876	irit	iritmmf06V3_BIC

needs with a multimedia character. We evaluated these runs on the subset of 19 multimedia topics. Tables 8–10 show the results of these runs for the top 5 performing groups. Compared to the tables above, for none of the tasks the best performing run was an official multimedia submission. That shows that for this task standard text retrieval techniques are competitive. This does not necessarily lead to the conclusion that specific treatment of multimedia topics is ineffective. It may still be the case that a combination of techniques from the top performing Ad Hoc and Multimedia submissions would give better results on these topics than either alone.

Table 8. Ad Hoc runs for the MMfragments topics for Focused task.

MAiP	iP[0.00]	iP[0.01]	iP[0.05]	iP[0.10]	Group	Run
0.0649	0.5367	0.4435	0.1960	0.1393	mines	EMSE,boolean,Prox200NF,0010
0.1059	0.4494	0.4219	0.2952	0.2272	qutau	FOC_02
0.1175	0.3961	0.3856	0.3176	0.2888	justsystem	VSM_CO_02
0.1338	0.3962	0.3853	0.3199	0.2558	unigordon	Focused-LM
0.1050	0.5793	0.3715	0.2990	0.2796	maxplanck	TOPX-CAS-Focused-exp-all

Table 9. Ad Hoc runs for the MMfragments topics for Relevant in Context task.

MAgP	gP[5]	gP[10]	gP[25]	gP[50]	Group	Run
0.1323	0.1838	0.2035	0.1740	0.1438	udalian	DUT_03_Relevant
0.1120	0.2129	0.2151	0.1467	0.1152	rmit	zet-okapi-RiC
0.1044	0.1729	0.1763	0.1528	0.1193	qutau	CO_RelevantInContext
0.0951	0.1282	0.1500	0.1281	0.0980	utwente	star_logLP_RinC
0.0949	0.1987	0.1919	0.1346	0.1049	unigordon	RelevantInContext-LM

7.2 MMimages

The four participating groups (CWI/UTwente, IRIT, QUTAU, and UGeneva) submitted a total of 13 MMimages runs. Figure 3 shows the interpolated recall precision graphs of these runs and Table 11 shows their mean average precision

Table 10. Ad Hoc runs for the MMfragments topics for Best in Context task.

MAgP	gP[5]	gP[10]	gP[25]	gP[50]	Group	Run
0.2275	0.4306	0.3610	0.2725	0.2090	rmit	zet-okapi-BiC
0.1889	0.4306	0.3610	0.2857	0.2210	inria	ent-ZM-BiC
0.1879	0.2505	0.2377	0.1949	0.1660	udalian	DUT_02_Best
0.1852	0.3588	0.3324	0.2243	0.1647	justsystem	VSM_CO_14
0.1839	0.3381	0.3052	0.2244	0.1838	unigordon	BestInContext-LM

scores. Similarly to last year, the top performing runs do not use any image analysis or visual processing; they are purely text-based.

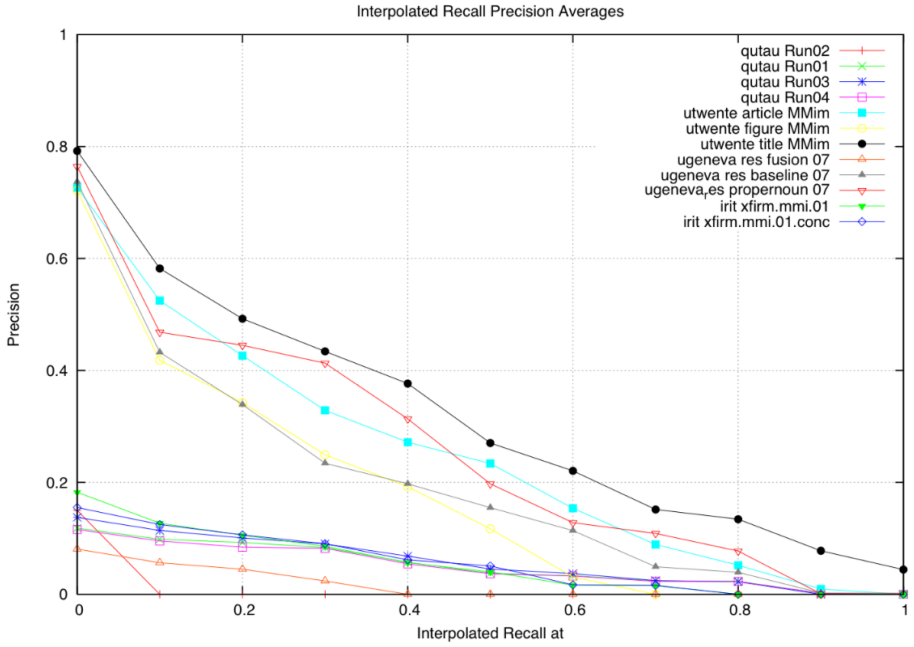

Fig. 3. MMimages: Interpolated Recall Precision Averages

8 Conclusions and Outlook

The INEX 2007 Multimedia track provides a nice collection of related resources (Wikipedia-based collections, together with a set of resources that are either starting points for or results of visual processing) to be used in the track's two retrieval tasks: MMfragments and MMimages. The main research questions these tasks aimed at addressing are the following: Do textual and structural hints

Table 11. Mean average precision (MAP) for submitted MMimages runs

group	run	MAP
utwente	title_MMim	0.2998
ugeneva	res_propernoun_07	0.2375
utwente	article_MMim	0.2240
ugeneva	res_baseline_07	0.1792
utwente	figure_MMim	0.1551
qutau	Run03	0.0482
irit	xfirm.mmi.01	0.0448
qutau	Run01	0.0447
irit	xfirm.mmi.01.conc	0.0445
qutau	Run04	0.0411
ugeneva	res_fusion_07	0.0165
qutau	Run02	0.0011

need to be interpreted differently for the MMfragments compared to the Ad Hoc tasks? How do visual hints in the query help image and XML document fragment retrieval? Since the number of participants in the multimedia track was disappointing with only four groups submitting runs, it is hard to draw general conclusions from the results. What we could see so far is that the top runs in both tasks did not make use of any of the provided visual resources.

The Multimedia track will not run in INEX 2008. Instead the MMimages task will run under the auspices of ImageCLEF 2008, where it is renamed as wikipediaMM task. This decision has been made in an attempt to attract more participants, since ImageCLEF provides a more natural habitat for such an image retrieval task. The set of related collections and resources, makes this task an interesting playground, both for groups with a background in information retrieval, and for groups with a deeper understanding of computer vision or image analysis.

9 Acknowledgements

Theodora Tsikrika and Thijs Westerveld (while he was at CWI, Amsterdam, The Netherlands) were supported by the European Union via the European Commission project VITALAS (contract no. 045389). The authors would also like to thank Jaap Kamps for producing the evaluation results for the Ad Hoc runs on the MMfragments topics, Saadia Malik for valuable technical support, and the reviewer for useful comments that helped us improve this paper.

References

1. L. Denoyer and P. Gallinari. The Wikipedia XML Corpus. *SIGIR Forum*, 40(1):64–69, 2006.

2. N. Fuhr, J. Kamps, M. Lalmas, S. Malik, and A. Trotman. Overview of the INEX 2007 ad hoc track. In N. Fuhr, M. Lalmas, A. Trotman, and J. 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.
3. J. C. v. Gemert, J.-M. Geusebroek, C. J. Veenman, C. G. M. Snoek, and A. 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.
4. J. Kamps, J. Pehcevski, G. Kazai, M. Lalmas, and S. Robertson. INEX 2007 evaluation measures. In N. Fuhr, M. Lalmas, A. Trotman, and J. 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. P. Over, W. Awad, G. Kraaij, and A. F. Smeaton. TRECVID 2007 Overview. In *TREC Video Retrieval Evaluation Online Proceedings*, 2007.
6. C. G. M. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. 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.
7. M. Theobald, R. Schenkel, and G. Weikum. An efficient and versatile query engine for topx search. In *VLDB '05: Proceedings of the 31st international conference on Very large data bases*, pages 625–636. VLDB Endowment, 2005.
8. A. Trotman and B. Sigurbjörnsson. Narrowed Extended XPath I (NEXI). In N. Fuhr, M. Lalmas, S. Malik, and Z. Szlavik, editors, *Advances in XML Information Retrieval: 3rd International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004, Revised Selected Papers*, volume 3493, pages 16–40. Springer, 2005.
9. R. van Zwol, G. Kazai, and M. Lalmas. INEX 2005 multimedia track. In N. Fuhr, M. Lalmas, S. Malik, and G. Kazai, editors, *Advances in XML Information Retrieval and Evaluation, 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Revised Selected Papers*, volume 3977, pages 497–510. Springer, 2006.
10. T. Westerveld and R. van Zwol. The INEX 2006 multimedia track. In N. Fuhr, M. Lalmas, and A. 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.
11. D. Zeimpekis and E. Gallopoulos. TMG : A MATLAB toolbox for generating term-document matrices from text collections. In J. Kogan, C. Nicholas, and M. Teboulle, editors, *Grouping Multidimensional Data: Recent Advances in Clustering*, pages 187–210. Springer, 2006.