Multimedia Retrieval at INEX 2007

Theodora Tsikrika
CWI, Amsterdam, The Netherlands
Theodora.Tsikrika@cwi.nl

Thijs Westerveld*
Teezir Search Solutions, Ede, The Netherlands
thijs.westerveld@teezir.com

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. The INEX Multimedia track considers types of media other than text and its objective 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.

2 Wikipedia collections and additional resources

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

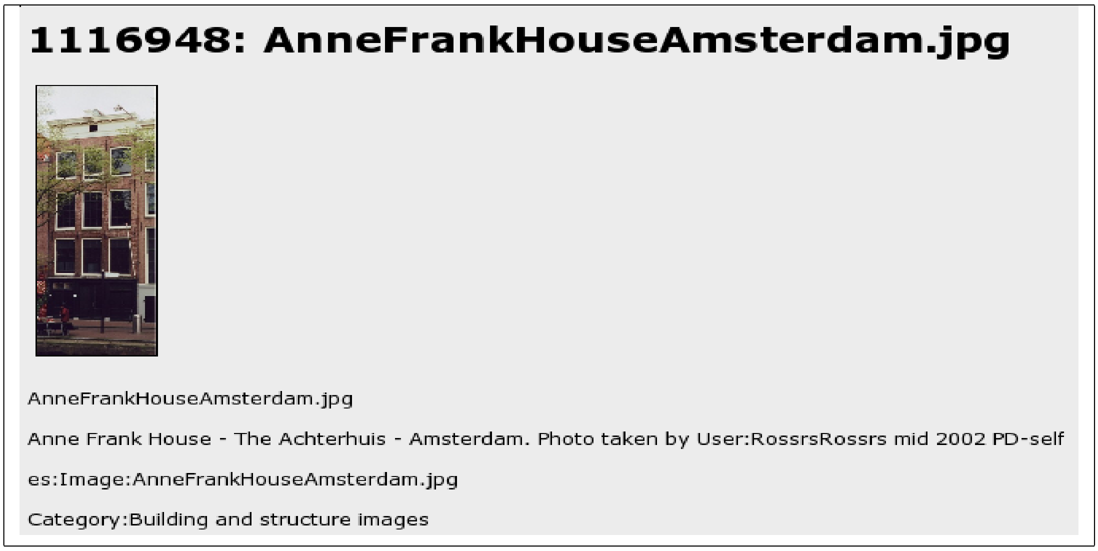
Wikipedia XML collection: This is a structured collection of 659,388 Wikitext pages from the English part of Wikipedia, the free content encyclopedia (`http://en.wikipedia.org`), that have been converted to XML [1]. This collection has been created for the Ad Hoc track. Given, though, its multimedia 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. These metadata, usually containing a brief caption

*Part of this work was carried out when the author was at CWI, Amsterdam, The Netherlands or description of the image, the Wikipedia user who uploaded the image, and the copyright information, have been formatted in XML. 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).



1116948: AnneFrankHouseAmsterdam.jpg

AnneFrankHouseAmsterdam.jpg

Anne Frank House - The Achterhuis - Amsterdam. Photo taken by User:RossrsRossrs mid 2002 PD-self

es:Image:AnneFrankHouseAmsterdam.jpg

Category:Building and structure images

Figure 1: Example of image+metadata document from the Wikipedia image XML collection.

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 are:

Image classification scores: For each image, the classification scores for the 101 different MediaMill concepts are provided by UvA [5]. 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 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.

3 Retrieval Tasks

For INEX 2007, the same two tasks as in 2006 were evaluated:

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 (similarly to 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, with MMfragments topics being a subset of the Ad Hoc ones. 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).

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 [6].

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 [8, 9]:

- 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.

Topics for both tasks are developed by the participants and consist of the usual title, description, and narrative fields. In addition, there are two fields, castitle and mmtitle, that contain NEXI expressions of the title with additional structural and/or visual hints. 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). Table 2 shows the distribution over tasks as well as some statistics on the topics, where the MMfragments topics correspond to Ad Hoc topics 525-543.

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 the INEX 2007 MMfragments task was run in parallel with the Ad Hoc track, the assessments for this task were performed in the context of the Ad Hoc track [2]. No additional instructions were given 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 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, whereas for a MMfragments topic it contains 0.62 images.

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 [9]. 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 Approaches and Results

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). In total we received 12 MMfragment submissions and 13 MMimages submissions.

For MMfragments, six submissions used the topics' <title> field, and six submissions used the <castitle> field; the <mmtitle> field was not used by any participant. 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. 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.

CWI/UTwente participated in both tasks and experimented with traditional text-based approaches based on the language modelling approach and different length priors, without any visual processing. IRIT participated in both tasks with methods based on the context of images to retrieve multimedia elements. They studied the use of collateral text and structure (descendant, sibling and ascendant nodes) as well as the incorporation of image classification scores. UGeneva participated in the MMimages task with a text-only baseline, an extension with proper noun detection and a multi modal fusion approach combining textual terms, color and texture histograms, and the concepts classification scores using a hierarchical SVM approach. QUTAU participated in both tasks, but did not provide descriptions of their submissions.

The MMfragments runs have been evaluated using the standard measures as used in the Ad Hoc track [4]. Since the MMfragments topics were mixed with the Ad Hoc topics we received many more submissions that were not tailored to answering information needs with a multimedia character. We evaluated these runs on the subset of 19 multimedia topics. For none of the MMfragments 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. For the MMimages task we have reported standard interpolated recall-precision graphs and mean average precision measures for all submissions. Also in this task, the top performing runs do not use any image analysis or visual processing; they are purely text-based. The evaluation results and more detailed analysis can be found in the INEX 2007 Multimedia Track overview paper [7].

7 Conclusions

The INEX 2007 Multimedia track provides a nice collection of related resources to be used in the track's two retrieval tasks: MMfragments and MMimages. 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.

8 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).

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] 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.

[6] 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.

[7] T. Tsikrika and T. Westerveld. The INEX 2007 multimedia 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.

[8] 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.

[9] 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.