Combining Image and Structured Text Retrieval

D.N.F. Awang Iskandar, Jovan Pehcevski, James A. Thom, and
S. M. M. Tahaghoghi

School of Computer Science and Information Technology, RMIT University
Melbourne, Australia
{dayang, jovanp, jat, saied}@cs.rmit.edu.au

Abstract. Two common approaches in retrieving images from a collection are retrieval by text keywords and retrieval by visual content. However, it is widely recognised that it is impossible for keywords alone to fully describe visual content. This paper reports on the participation of the RMIT University group in the INEX 2005 multimedia track, where we investigated our approach of combining evidence from a content-oriented XML retrieval system and a content-based image retrieval system using a linear combination of evidence. Our approach yielded the best overall result for the INEX 2005 Multimedia track using the standard evaluation measures. We have extended our work by varying the parameter for the linear combination of evidence, and we have also examined the performance of runs submitted by participants by using the newly proposed HiXEval evaluation metric. We show that using CBIR in conjunction with text search leads to better retrieval performance.

1 Introduction

In a large document collection, it is common to find multimedia elements such as images, audio, and video. Describing these multimedia elements in a standard way is beneficial as it can assist the retrieval process. The eXtensible Markup Language (XML) is a standard developed by the World Wide Web Consortium to describe data in a structured manner, allowing the description of multimedia elements to be represented. The INitiative for the Evaluation of XML Retrieval (INEX) provides a platform for participants to evaluate the effectiveness of their XML retrieval techniques using uniform scoring procedures, and a forum to compare results. INEX 2005 comprised seven tracks. The multimedia track was established with the aim of retrieving relevant XML document fragments containing various types of multimedia[1], of which only text and images were used. Besides RMIT University, four other groups participated in the multimedia track — Queensland University of Technology (QUTAU), Utrecht University (UTRECHT), University of Twente (UTWENTE) and Queen Mary University of London (QMUL).

The aim of the RMIT University group in participating in the INEX 2005 MM track was to explore and analyse methods for combining evidence from content-based image retrieval (CBIR) with content-oriented XML retrieval. In this paper, we describe a

[1] Multimedia Track @ INEX,
http://inex.is.informatik.uni-duisburg.de:2004/presentations/INEX-MM-track.pdf

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE inex_topic SYSTEM "topic.dtd">
<inex_topic topic_id="mm6" inex_track="MM" query_type="CAS" ct_no="14">
<castitle>
//destination[about(., Europe) and about(.//culture//history, king queen)]
//images//image[about(., royal palace residence src:/images/BN7386_10.jpg)]
</castitle>
<description> From all European destinations that were ruled by either
a king or a queen in their cultural history, find images depicting a royal
palace residence. </description>
<narrative>We are a group of historians interested in royal palaces. We
want to visit destinations that contain at least one royal palace. We are
focused on European destinations that were ruled by either a king or a
queen in their cultural history. From these destinations, we want to find
images depicting a royal palace residence.</narrative>
</inex_topic>
```

Fig. 1. Example of a multimedia CAS query with image BN7386_10.jpg, the Royal Palace in Norway (original in colour), in the target element of the query

fusion system that combines evidence and ranks the query results based on text and image similarity. The fusion system consists of two subsystems: the GNU Image Finding Tool (GIFT), and the hybrid XML retrieval system. A technique for linear combination of evidence is used to merge the relevance scores from the two subsystems. Six runs submitted by our group are considered to evaluate the relative importance of image and content-based text components, and these are also compared against the approaches of other participants. The TREC evaluation metric (TRECeval) is used as the official assessment method. We evaluate the performance of our approaches in the INEX 2005 MM track using the standard TRECeval measures: `P@1`, `P@5`, `P@10`, `MAP` and `R-Prec`.

We extend our work on the initially submitted runs to further examine the parameter that influences the weighting scheme between the two subsystems. We evaluate the performance of runs submitted by all the INEX 2005 MM track participants using a newly proposed evaluation metric, namely HiXEval [5]. We also discuss results obtained from the extended work and the HiXEval evaluation in this paper.

The remainder of this paper is organised as follows. In Section 2, we present the multimedia topics and their corresponding relevance judgements. We describe our approach to retrieve the XML document fragments and the associated images based on these multimedia topics in Section 3. In Section 4, we present results obtained from our experiments. Related work on combination of evidence for retrieving image and text are briefly explained in Section 5. We conclude in Section 6 with a discussion of our findings and suggestions for future work.

2 Multimedia Topics and Relevance Judgements

The INEX 2005 multimedia retrieval task focuses on combination of text and images. The *WorldGuide* collection — referred as the Lonely Planet collection in the MM track — was utilised, which was provided by the Lonely Planet organisation[2]. As an initial task, multimedia track participants were asked to propose several topics that might represent typical information needs expressed by users of the collection. As an example, one of the topics proposed by our group is "European destinations ruled by a king or a queen that have a palace". The full specification of this topic and the query image that depicts the royal palace in Norway is shown in Fig. 1.

Two types of queries are explored in INEX using the Narrowed Extended XPath I (NEXI) query: content-only (CO) and content-and-structure (CAS). CO queries are free text queries, while CAS queries contain explicit structural constraints of the desired target and support elements. The multimedia track uses the latter query type to represent a topic. The multimedia query is contained in the `castitle` element which represents the information to be retrieved from the Lonely Planet collection.

The CAS query consists of two elements: target and support. The target element of the query is the last node in the query path, and specifies the element that should be returned as the result. Support elements specify additional structural constraints that should be met. For the topic in Fig. 1, the target element of the query

```
//destination//images//image
```

indicates that the element to be retrieved is an `image` element which contains an image reference in the source (`src`). The support elements of the query are:

```
//destination
//destination//culture//history
//destination//images//image
```

In total, twenty-three multimedia topics that have corresponding relevance judgements were formulated for this collection. These belong to three categories:

1. Topics that contain only text. This topic category does not include any image references in either the target or support elements;
2. Topics that contain a mixture of images and text, where the image reference is explicitly given in the `about` clause of the support elements; and

[2] `http://www.lonelyplanet.com/`

Table 1. Topic category, number of topics and retrieval systems used, and collection involved

Topic category	**1**	**2**	**3**
Number of *official* topics	8	4	7
Number of *extended* topics	12	4	7
Retrieval system used	Hybrid XML	Hybrid XML and GIFT	Hybrid XML and GIFT
Collection involved	Text only	Text and image	Text and image

3. Topics that contain a mixture of images and text, except that here the image reference is explicitly stated in the `about` clause of the target element.

The number of multimedia topics in each category is shown in Table 1. The example given in Fig. 1 belongs to the third topic category.

Relevance judgements for the multimedia topics are divided into two sets: *official* and *extended*. The *official* assessment set includes 19 topics that contain results that match the relevance judgements. The *extended* assessment set has 23 topics, of which the additional four topics contain results that do not match the relevance judgements. These four topics were misinterpreted during the relevance judgements procedure. The *official* assessment set is used for comparing the submitted runs from the INEX 2005 MM track participants.

3 Our Approach

In this section, we describe our fusion system that consists of two subsystems to obtain the results for the multimedia queries. Since the XML document structure serves as a semantic backbone for retrieval of the multimedia fragments, we use a content-oriented hybrid XML retrieval system [4] to retrieve the relevant document fragments. The GNU Image Finding Tool (GIFT)[3], a content-based image retrieval system, is used to retrieve the results based on the visual features of the images.

We aim to achieve the *chorus effect*. According to Vogt and Cottrell [9], "The chorus effect occurs when several retrieval approaches suggest that an item is relevant to a query ... this tends to be stronger evidence for relevance than a single approach doing so". To achieve this, we use data fusion techniques to combine the evidence from GIFT and the content-oriented hybrid XML retrieval system in three phases [8]:

1. The *collection selection* phase identifies the document collection that is most likely to contain relevant document fragments for the user queries.
2. The *document fragment selection* phase determines the number of relevant document fragments to be retrieved from the document collection.
3. The *merging* (or *fusion*) phase combines the evidence from multiple retrieval systems.

[3] `http://www.gnu.org/software/gift/`

3.1 Phase One: Collection Selection

We view the Lonely Planet collection as having three different groups of information items that are related to one another. The first group contains the XML text documents, the second contains images, and the third contains maps. As illustrated in Table 1, the XML text documents are used to process all the queries, while the image data is used for only the queries in topic categories 2 and 3. The map data was not used, since the topic which specified the map as the target element was not assessed.

3.2 Phase Two: Document Fragment Selection

In this phase, each subsystem retrieves document fragments (text or images) and returns a list of retrieval status values (RSVs) presented in descending order. First 250 top-ranked document fragments are returned from our content-oriented hybrid XML retrieval system. For GIFT, the RSVs of all the images in the collection are returned. The following sections explain how each subsystem is used to generate RSVs for each multimedia query; the lists are later merged in phase three to produce the final results.

Content-Based Image Retrieval Content-based image retrieval aims to retrieve images on the basis of features automatically extracted from the images themselves. The GIFT system indexes an image collection by extracting image features and indexing them using an inverted file data structure [7].

GIFT uses the HSV (Hue-Saturation-Value) colour space for local and global colour features [7]. For extracting the image texture, a bank of circularly symmetric Gabor filters is used. GIFT evaluates and calculates the query image and the target image feature similarity based on the data from the inverted file. The results of a query are presented to the user in the form of a ranked list. GIFT also provides the mechanism to perform relevance feedback. We did not perform any relevance feedback in this work.

For the multimedia topics, we presented the images listed in the source (`src`) element of the multimedia CAS query as the query image to GIFT. We used the default Classical IDF algorithm and set the search pruning option to 100%. This allows us to perform a complete feature evaluation for the query image, even though the query processing time is longer. We retrieved and ranked all the images in the Lonely Planet collection.

Referring to the multimedia topic presented earlier, the query image of Fig. 1 is provided to GIFT and Fig. 2 is a screenshot of the query. The query results are presented in Fig. 3, where the RSVs are ranked in descending order from left to right, and top to bottom.

Content-Oriented Hybrid XML Retrieval The second subsystem we used for text retrieval in the INEX 2005 MM track follows a *hybrid* XML retrieval approach [4], combining information retrieval features from Zettair[4] (a full-text search engine) with XML-specific retrieval features from eXist[5] (a native XML database).

[4] `http://www.seg.rmit.edu.au/zettair/`
[5] `http://exist-db.org/`

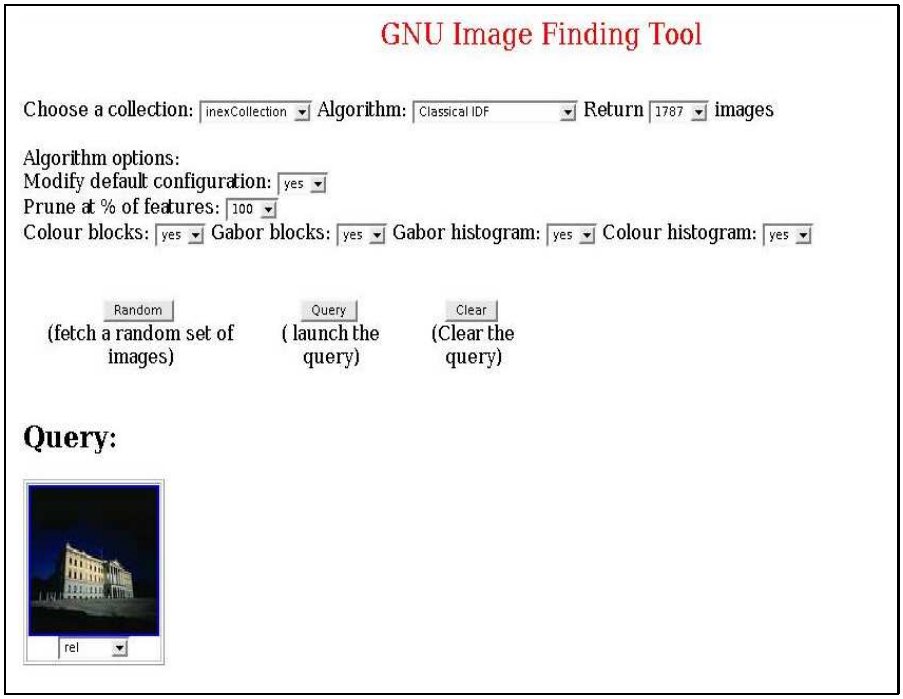

Fig. 2. Querying image BN7386_10.jpg into GIFT (original in colour)

Each multimedia topic was first automatically translated into a Zettair query. Terms that appear in the `castitle` part of the topic (with all structural query constraints and image references completely removed) were used to formulate the Zettair query. A list of (up to) 250 `destination` elements were presented in a descending order according to their estimated likelihood of relevance. To retrieve *elements* rather than full articles, a second topic translation module was used to formulate a query to eXist. As the support and target parts of each multimedia query were strictly matched, both the terms and the structural query constraints from the topic (without the actual image references) were used to formulate the eXist query. We used the eXist OR query operator to generate the element answer list for a given topic. The answer list contains (up to) 250 matching elements, taken from articles that were highly ranked in the list of articles previously returned by Zettair.

Lastly, a post-processing retrieval module with an XML-specific ranking heuristic (TPF) [6] was used to rank and produce the final list of RSVs.

3.3 Phase Three: Merging Evidence of CBIR and Hybrid XML Retrieval

To fuse the two RSV lists into a single ranked result list R for the multimedia queries, we use a simple linear combination of evidence [1]:

$$R = \alpha \cdot S_I + (1 - \alpha) \cdot S_T$$

Fig. 3. First twenty results of a GIFT image query (best viewed in colour)

Here, α is a weighting parameter (determines the weight of GIFT versus hybrid XML retrieval), S_I represents the image RSV obtained from GIFT, and S_T is the RSV of the same image obtained from the hybrid XML retrieval system.

To investigate the effect of giving certain biases to a system, we vary the α value between 0 to 1. When the value of α is set to 1, only the RSVs from GIFT are used. On the other hand, only the hybrid XML retrieval RSVs are used when the value of α is set to 0. For the INEX 2005 MM track, we submitted six runs with the α value set to 0.0, 0.1, 0.3, 0.5, 0.9 and 1.0, respectively. Results obtained from these runs, which we denote as `rmit-0` to `rmit-5`, are shown in Table 2.

4 Experiments and Results

In this section we provide a description of the evaluation metrics used, and we analyse results obtained from the official INEX 2005 multimedia runs submitted by each participating group, including those obtained from the additional RMIT runs.

Table 2. The values for `P@n`, `MAP` and `R-Prec` using TRECeval and HiXEval for each run submitted by INEX 2005 MM track participants.
Italic values – best performance among runs for each participating group and each measure
Bold values – best overall performance among all runs for each measure

Run	TRECeval					HiXEval				
	P@1	P@5	P@10	MAP	R-Prec	P@1	P@5	P@10	MAP	R-Prec
RMIT										
rmit-0	0.4737	***0.3684***	0.3053	0.2759	***0.3267***	0.3498	0.2668	0.2179	0.1952	***0.2485***
rmit-1	0.4737	***0.3684***	0.3053	0.2771	***0.3267***	0.3491	***0.2669***	0.2177	0.1958	0.2485
rmit-2	0.4737	***0.3684***	***0.3105***	***0.2779***	0.3259	0.3465	0.2664	*0.2216*	*0.1960*	0.2479
rmit-3	0.4737	***0.3684***	0.3053	0.2764	0.3259	0.3488	0.2563	0.2176	0.1953	0.2479
rmit-4	*0.5263*	0.3368	0.2579	0.2664	0.3168	*0.4014*	0.2358	0.1938	0.1930	0.2429
rmit-5	0.4737	0.2737	0.2105	0.2244	0.2525	0.3626	0.2150	0.1671	0.1700	0.1935
QUTAU										
qutau-0	0.4211	0.2737	0.1947	0.1995	0.2094	0.3098	0.1970	0.1457	0.1557	0.1445
qutau-1	*0.4737*	0.2737	0.2053	0.2064	0.2116	0.3135	0.2046	0.1538	0.1582	0.1473
qutau-2	*0.4737*	*0.3579*	*0.2842*	*0.2711*	*0.2641*	0.3161	*0.2602*	*0.2132*	*0.1937*	*0.1871*
qutau-3	0.3684	0.2842	0.1895	0.1844	0.1892	0.2600	0.2037	0.1519	0.1429	0.1360
qutau-4	0.4211	0.3053	0.2105	0.2037	0.1986	0.2575	0.2475	0.1715	0.1532	0.1535
qutau-5	*0.4737*	0.2842	0.2053	0.2066	0.2181	***0.4181***	0.2210	0.1715	0.1744	0.1751
UTRECHT										
utrecht-0	0.4615	0.3385	0.2615	0.2329	*0.2776*	0.3278	0.2007	0.1537	0.1229	0.1627
utrecht-1	***0.5294***	*0.3529*	0.2706	*0.2392*	0.2747	*0.3481*	*0.2497*	*0.1965*	*0.1581*	*0.1974*
utrecht-2	0.3529	0.2941	0.2235	0.1769	0.2073	0.2678	0.2094	0.1487	0.1165	0.1519
utrecht-3	***0.5294***	0.3294	*0.2824*	0.2324	0.2648	*0.3481*	0.2462	0.1914	0.1477	0.1864
utrecht-4	***0.5294***	0.3294	*0.2824*	0.2324	0.2648	*0.3481*	0.2462	0.1914	0.1477	0.1864
utrecht-5	0.1579	0.0632	0.0737	0.0554	0.0697	0.1313	0.0524	0.0593	0.0440	0.0567
UTWENTE										
utwente-0	*0.4211*	0.3053	0.2789	*0.2751*	*0.2799*	*0.3559*	*0.2555*	***0.2255***	***0.2208***	*0.2266*
utwente-1	*0.4211*	0.2947	0.2579	0.26	0.2692	*0.3559*	0.2545	0.2246	0.2129	0.2216
utwente-2	0.3889	0.3444	0.2667	0.2567	0.2434	0.2894	0.2346	0.1738	0.1689	0.1681
utwente-3	0.2105	0.2211	0.2263	0.211	0.2227	0.1773	0.1877	0.1739	0.1400	0.1549
utwente-4	0.3889	*0.3556*	*0.2833*	0.2627	0.2458	0.2894	0.2475	0.1896	0.1739	0.1680
utwente-5	0.2105	0.2211	0.2263	0.2133	0.2196	0.1773	0.1877	0.1740	0.1423	0.1518
QMUL										
qmul-0	*0.0526*	*0.0211*	*0.0368*	*0.0412*	*0.0423*	*0.0526*	*0.0211*	*0.0354*	*0.0376*	*0.0409*

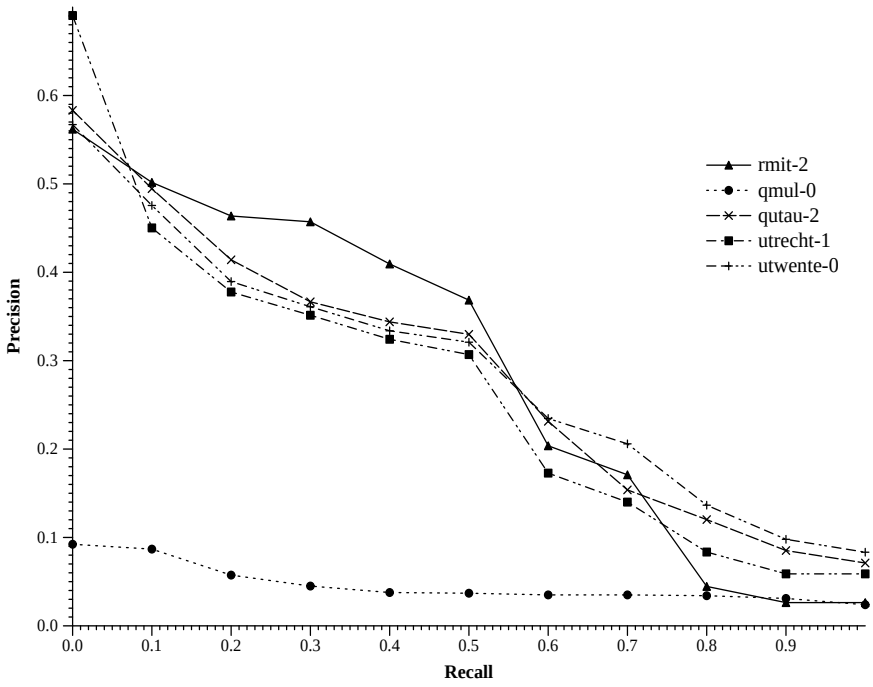

Fig. 4. Interpolated average precision at 11 standard recall levels using TRECeval for the best performing runs submitted by the INEX 2005 MM track participants

4.1 Evaluation Metrics

The TREC evaluation metric was adopted for the multimedia track assessment in INEX 2005. Binary relevance judgements were used to evaluate the runs. We evaluated our results based on the standard recall and precision retrieval performance measures. The following measures were used:

- Precision at cut-off (`P@n`): Precision after *n* document fragments have been retrieved.
- Mean Average Precision (`MAP`): The mean of the average precisions calculated for each topic. Average precision is the average of the precisions calculated at each natural recall level.
- Recall-precision (`R-prec`): Precision after the total number of relevant document fragments have been retrieved.
- Average interpolated precision at 11 standard recall levels (0%-100%).

In addition to the above evaluation measures, we also report values obtained with HiXEval, an alternative evaluation metric for XML retrieval that is solely based on the amount of highlighted relevant information [5]. The reported values are: `P@n`, which measures the proportion of relevant information to all the information retrieved at a rank `n`; `MAP`, the mean average precision calculated at natural recall levels; and `R-prec`,

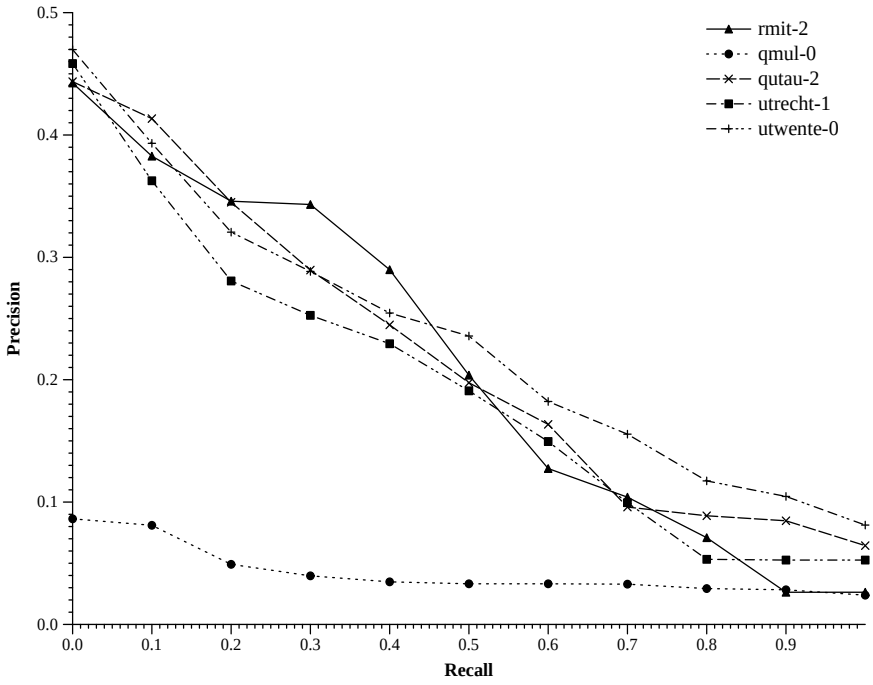

Fig. 5. Interpolated average precision at 11 standard recall levels using HiXEval for the best performing runs submitted by the INEX 2005 MM track participants

which reflects the measured precision after the total number of relevant document fragments have been retrieved.

4.2 Result Analysis

We present the analysis and evaluation for INEX 2005 multimedia runs and additional RMIT runs based on the 19 topics that belong to the *official* multimedia assessment set.

Official INEX 2005 Multimedia Runs We analyse the runs submitted by the INEX MM track participants using both TRECeval and HiXEval evaluation metrics. As presented in Table 2, we report results obtained with precision at cut-offs 1, 5, 10, and with `MAP` and `R-Prec` for both metrics. For each participating group, the best performance under each measure is shown in italics. For each evaluation measure, the best run performance observed among all participants is shown in bold.

Using the TRECeval evaluation metric, UTRECHT performed best for `P@1`. However, our best run outperformed the others for `P@5`, `P@10`, `MAP` and `R-prec`. Using HiXEval as an evaluation metric, QUTAU performed best for `P@1`, our best run again outperformed the others for `P@5` and `R-Prec`, while UTWENTE performed best for `P@10` and `MAP`. The difference in the observed behaviour between the two metrics can be

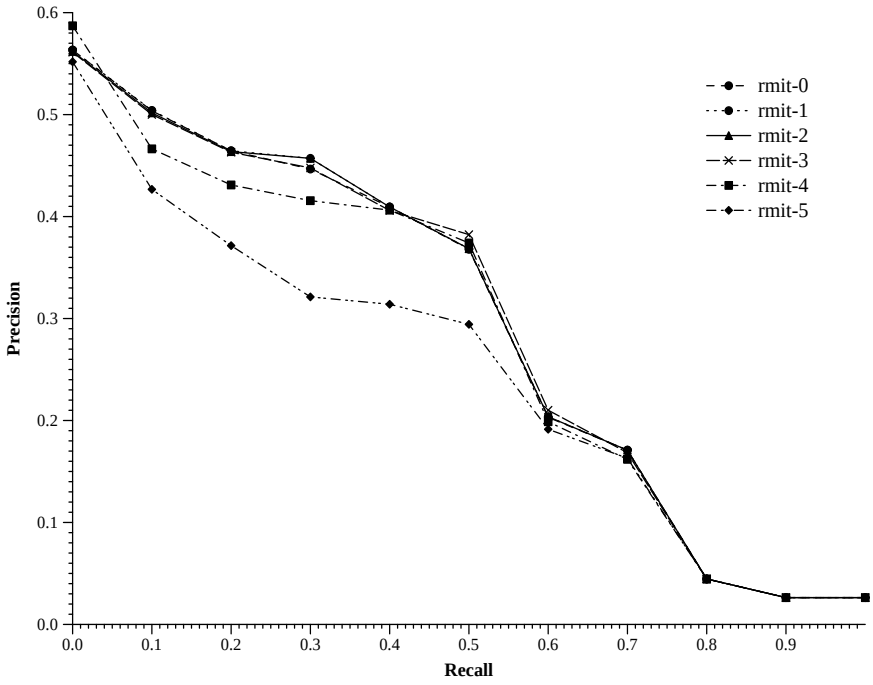

Fig. 6. Interpolated average precision at 11 standard recall levels using TRECeval for the six official RMIT runs submitted to the INEX 2005 MM track

explained by the fact that the two metrics are based on different evaluation methodologies. Indeed, recall under TRECeval is measured as the fraction of relevant *elements* retrieved, whereas HiXEval uses the fraction of relevant *information contained by the elements* retrieved [5]. Arguably, a finer level of evaluation detail is captured by HiXEval which is not captured by TRECeval. This, in turn, suggests that, for the `MAP` measure of HiXEval, on average the best performing UTWENTE run is indeed capable of retrieving larger quantities of relevant information than our best performing run.

Figure 4 illustrates the performance for the multimedia track participants based on the highest `MAP` values of the runs using TRECeval. Figure 5 shows the same graph pattern when using HiXEval as the evaluation metric. Both graphs show that, with the exception of QMUL, the observed average performance among the best runs submitted by participants was similar.

Additional RMIT Runs As shown in Table 2, at one document fragment retrieved the highest precision among the RMIT runs is observed for run `rmit-4` (with the value for $\alpha = 0.9$). There is no visible difference in precision for all the other runs with `P@1`. With `P@5`, combining evidence from text and image at the same weight (α = 0.5) leads to similar performance as when α values of 0.0, 0.1 and 0.3 are used (reflected by the observed performance of runs `rmit-0` to `rmit-3`). The precision values drop as the α value is increased. With `MAP`, the run `rmit-2` ($\alpha = 0.3$) produces the best

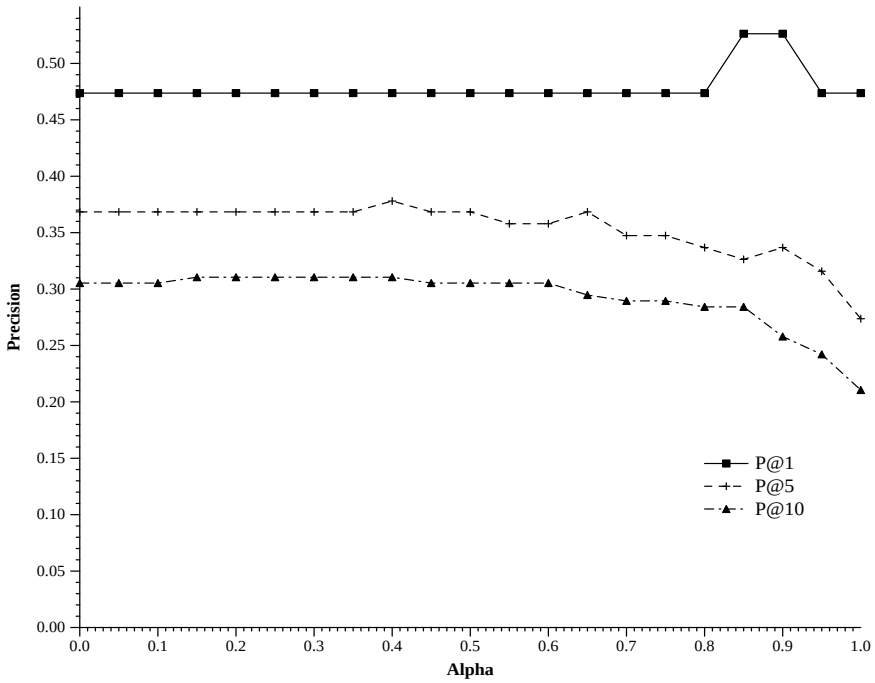

Fig. 7. Precision at cut-off 1, 5 and 10 for α values between 0.0 to 1.0

performance. With `R-prec`, runs `rmit-0` and `rmit-1` perform best and exhibit almost the same performance.

Based on Fig. 6, three RMIT runs (`rmit-0`, `rmit-1` and `rmit-2`) produce the best overall interpolated precision averages. Run `rmit-4` performed best at low recall levels. A constant performance can be seen for all the runs for recall level of 0.8 and above.

To analyse the changes in performance when the parameter α varies between 0 and 1, we performed additional runs at α intervals of 0.05. Figure 7 shows the performance of our runs for twenty different values of the parameter α, as measured by `P@1`, `P@5` and `P@10`. We observe that, to achieve the best performance under `P@1`, values for 0.85 and 0.9 should be used for the parameter α. On the other hand, the best performance under `P@5` and `P@10` is achieved when $\alpha = 0.4$.

The highest `MAP` performance is observed when $\alpha = 0.25$, which can be seen in Fig. 8. On the other hand, the highest `R-prec` performance is obtained when α is less than 0.15. Figure 9 illustrates the `R-Prec` performance when all the 20 α values are used.

We conclude that the content-oriented XML retrieval system benefits by using some evidence from a CBIR system; indeed, as measured by `MAP` and `R-prec`, increasing the weight of the hybrid XML retrieval system component in the fusion system yields better performance than when any of the two subsystems are used in isolation. When only a CBIR system is used to retrieve multimedia document fragments, precision is poor.

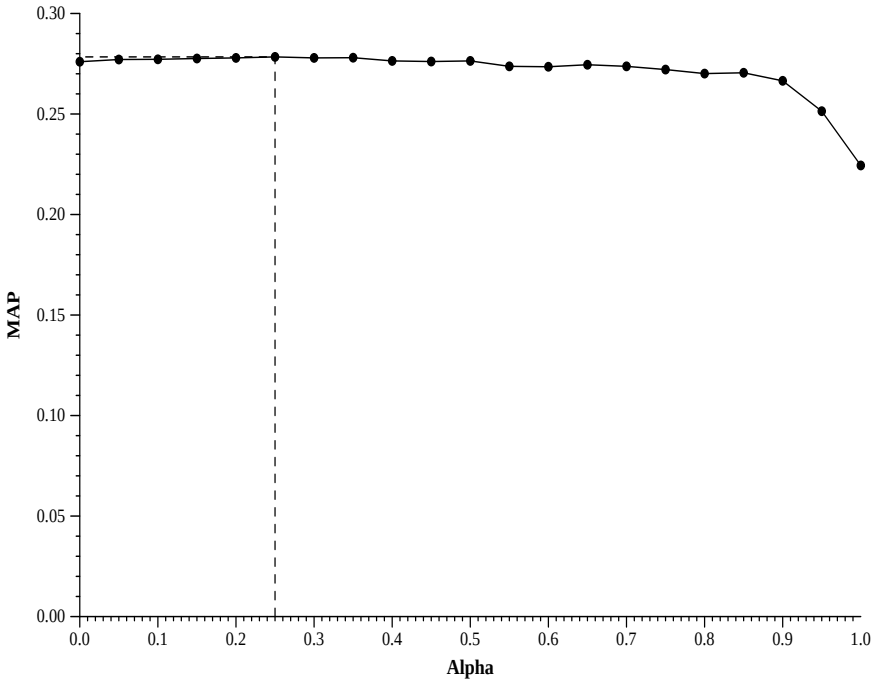

Fig. 8. Mean Average Precision for α values between 0.0 to 1.0

5 Related Work

Data fusion, also known as combination of evidence, is a method of merging multiple sources of evidence. In information retrieval, data fusion has been shown to improve the retrieval effectiveness when compared to using a single retrieval strategy [3, 8, 9].

Multimedia retrieval using combination of evidence has been studied by Haque [2], who compared the retrieval performance of using only image and multimedia (combination of text and image). He conducted experiments using three types of combining algorithms: *feature merging*, *weighted sum of ranking score*, and *weighted sum of inverse rank position*. Haque concluded that using a combination of evidence, multimedia retrieval performs better than image retrieval. The *weighted sum of inverse rank position* algorithm is shown to have the highest eleven point average precision in the multimedia retrieval, while the *weighted sum of ranking score* algorithm performed slightly lower than the *weighted sum of inverse rank position* algorithm.

Aslandogan and Yu [1] have compared the retrieval performance of indexing images of people on the Web using four approaches: text evidence followed by face detection, face detection and recognition, linear combination of evidence, and Dempster-Shafer theory of evidence. They reported that linear combination of evidence and the Dempster-Shafer theory of evidence yielded the same retrieval performance.

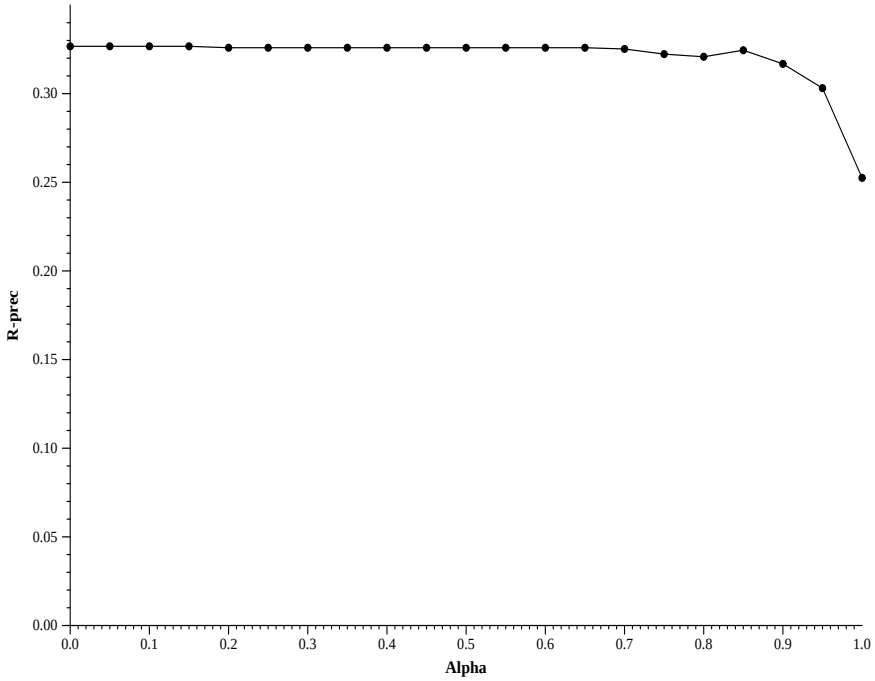

Fig. 9. R-Prec for α values between 0.0 to 1.0

6 Conclusions and Future Work

In this paper we have reported on our participation in the multimedia track of INEX 2005. As part of the XML-multimedia retrieval task, we submitted six runs for the official evaluation by the multimedia track organisers. These runs reflect the various relative weights of 0 to 1. Our approach demonstrated the overall best performance for `P@5`, `P@10`, `MAP` and `R-Prec` in the INEX 2005 MM track using the standard evaluation measures.

We have used the linear combination of evidence to merge the RSVs from two retrieval subsystems for retrieving multimedia information from structured documents. We also carried out additional runs to examine the effect of varying the parameter used for the linear combination of evidence (α). Having $\alpha = 0.25$ leads to the highest `MAP`, and the best `R-prec` values are when α is less than 0.15. We have also evaluated the submitted runs from the participants using HiXEval, where we observed a slightly different performance behaviour.

We conclude that a CBIR system needs a substantial support from a text-based system to effectively retrieve the desired images in a collection. Conversely, retrieving images based only on the surrounding text can be achieved without using a CBIR system, but better retrieval performance will be observed if some evidence from a CBIR system is incorporated.

We plan to extend this work by investigating different evidence combination methods for retrieving structured text and multimedia elements.

Acknowledgements

This research was undertaken using facilities supported by the Australian Research Council and an RMIT VRII grant. We acknowledge Lonely Planet for the permission to publish the images, and thank Jonathan Yu for his assistance in proposing and assessing one of the INEX 2005 multimedia topics.

References

1. Y. A. Aslandogan and C. T. Yu. Evaluating Strategies and Systems for Content-Based Indexing of Person Images on the Web. In *MULTIMEDIA 2000: Proceedings of the Eighth ACM International Conference on Multimedia*, pages 313–321, New York, NY, USA, 2000. ACM Press.
2. N. Haque. *Image Ranking for Multimedia Retrieval*. Ph.D. thesis, School of Computer Science and Information Technology, Royal Melbourne Institute of Technology, 2003.
3. J. H. Lee. Analyses of Multiple Evidence Combination. In *SIGIR 1997: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 267–276, New York, NY, USA, 1997. ACM Press.
4. J. Pehcevski, J. A. Thom and A-M. Vercoustre. Hybrid XML Retrieval: Combining Information Retrieval and a Native XML Database. *Information Retrieval*, Volume 8, Number 4, pages 571–600, 2005.
5. Jovan Pehcevski and James A. Thom. HiXEval: Highlighting XML Retrieval Evaluation. In *INEX 2005 Workshop Pre-Proceedings, Dagstuhl, Germany, November 28–30, 2005*, pages 11–24, 2005.
6. Jovan Pehcevski, James A. Thom and S. M. M. Tahaghoghi. RMIT University at INEX 2005. In *INEX 2005 Workshop Pre-Proceedings, Dagstuhl, Germany, November 28–30, 2005*, pages 217–233, 2005.
7. D. M. Squire, W. Müller, H. Müller and T. Pun. Content-based Query of Image Databases: Inspirations from Text Retrieval. *Pattern Recognition Letters*, Volume 21, Number 13–14, pages 1193–1198, 2000. (special edition for SCIA'99).
8. T. Tsikrika and M. Lalmas. Merging Techniques for Performing Data Fusion on the Web. In *CIKM 2004: Proceedings of the Tenth International Conference on Information and Knowledge Management*, pages 127–134, New York, NY, USA, 2001. ACM Press.
9. C. C. Vogt and G. W. Cottrell. Predicting the Performance of Linearly Combined IR Systems. In *SIGIR 1998: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 190–196, New York, NY, USA, 1998. ACM Press.