Towards Comment-based Cross-Media Retrieval

Martin Potthast   Benno Stein   Steffen Becker

Bauhaus-Universität Weimar
99421 Weimar, Germany
<first name>.<last name>@uni-weimar.de

ABSTRACT

This paper investigates whether Web comments can be exploited for cross-media retrieval. Comparing Web items such as texts, images, videos, music, products, or personal profiles can be done at various levels of detail; our focus is on topic similarity. We propose to compare user-supplied comments on Web items in lieu of the commented items themselves. If this approach is feasible, the task of extracting and mapping features between arbitrary pairs of item types can be circumvented, and well-known text retrieval models can be applied instead—given that comments are available. We report on results of a preliminary, but nonetheless large-scale experiment which shows that, if comments on textual items are compared with comments on video items, topically similar pairs achieve a sufficiently high cross-media similarity.

Categories and Subject Descriptors: H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing—*Abstracting Methods*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Retrieval Models*;

General Terms: Experimentation

Keywords: Web Comments, Cross-Media Retrieval

1. INTRODUCTION

Cross-media retrieval is a subproblem of multimedia information retrieval, which, in essence, builds on a solution to the following problem: given two items of different media types, quantify their topical similarity. Therefore, one of the primary goals of cross-media retrieval is the construction of retrieval models that bridge the gap between different media types by means of identifying correlations between low-level features and semantic annotations. We approach this problem from a different angle by the use of comments in lieu of the commented item. This way, model construction is not an issue since well-known text retrieval models can be applied directly. Although the text surrounding a non-textual item has always been used to extract annotations in multimedia IR [1, 2, 3], comments have not been considered in this respect. In terms of keyword retrieval within a certain media type, however, the value of comments for blog retrieval [4] and video retrieval [7] has been observed, recently.

An important premise of our approach is that comments actually describe the commented item to some extent, which is what we have investigated earlier [5]. In short, we found that comments on *text* are indeed descriptive: 10 comments are sufficient to reach a considerable similarity between a text and its comments, which

is not rooted in duplication, while 100-500 comments contain a measurable contribution of the commenters beyond the commented text. We proceed in this direction, based on the observation that frequent neutral terms in comments on a YouTube video describe it adequately [6].

2. CROSS-MEDIA SIMILARITY

To test our hypothesis, we have set up two large-scale evaluation corpora for items of different types, including the comments each item received. Then, pairs of items, one from each corpus, are compared using a well-known retrieval model, and the pairs that achieve a high cross-media similarity are evaluated manually.

2.1 Evaluation Corpora

Slashdot Corpus. Slashdot is a news Web site for publishing and commenting technology-related news articles. The publishing process is based on a moderation system where users can submit an article d. The editors of Slashdot decide whether or not d will be published, and for each published article a comment board D is available. We have downloaded all Slashdot articles from January 2006 to June 2008, including all comments. In total 17 948 articles were published during this period, and about 3.8 million comments were posted. Comments are organized as discussion threads, so that not all comments are direct responses to an article but responses to other comments. Therefore, we restrict our experiments to the 311 167 direct responses. Together the second and third quartile of the articles get between 16 to 41 direct comments, and the second and third quartile of the comments range from 1 to 45 words.

YouTube Corpus. YouTube is a video sharing Web site for home-made videos. Quite often, a video gets up to thousands of comments, and long explanations or discussions are less frequently observed than on Slashdot. We downloaded 9.8 million comments which were posted on 64 830 videos. Due to limitations in the YouTube API at most 1 000 comments per video were retrieved.

2.2 Retrieval Model

As retrieval model we employ a standard vector space model with $tf \cdot idf$ term weighting. Given a Web item d and its associated set of comments D, d is represented as a term vector $\mathbf{d}$ based on the index terms found in D, while applying stop word reduction and stemming. Two representations $\mathbf{d}$ and $\mathbf{d}'$ are compared using the cosine similarity. Note further that, as a matter of course, all the different kinds of retrieval models which have been developed for text-based information retrieval can be employed in this task. The reason we chose the simplest of all models is to determine how robust a cross-media similarity assessment can be accomplished.
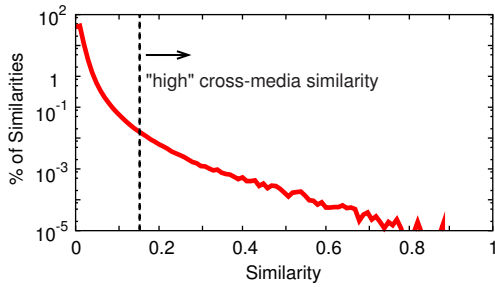
Figure 1: Distribution of comment-based cross-media similarities between YouTube videos and Slashdot articles.

2.3 Experiments

Given the evaluation corpora described above, we sampled 6 000 videos from the YouTube corpus and compared each video with each of the 17 948 Slashdot articles, i.e., about 107.7 million similarities were computed. Slashdot and YouTube are similar in that they are community-driven Web sites, so that at least some topical overlaps can be expected. However, since both corpora have been compiled independent of each other we were not aware of any existing overlaps. Figure 1 shows the similarity distribution as percentage of similarities over similarity intervals; the interval resolution is 0.01. The expected similarity between a YouTube video and a Slashdot article is about 0.009, so that a similarity of about 0.15 may already be considered a "high" cross-media similarity for its considerable positive deviation from the expectation. We found that at this similarity topic overlaps begin to appear more often. 0.1% of the similarities are above this threshold, which is a total of 115 589 pairs of items from YouTube and Slashdot.

The item pairs were sorted in descending order of similarity, and then the top 100 pairs were manually inspected in order to verify whether or not they have similar topics. We classified the degree of topical match into the fuzzy categories "equal," "related," and "unrelated." For instance, if a given Slashdot article and a YouTube video are about the same incident they are considered equal, if they are about similar incidents that happened independent of one another they are considered related, and otherwise they are considered unrelated. Table 1 gives an overview of the categorized item pairs, and Table 2 shows a small selection of matching item pairs. A remarkable portion of 91% of the top item pairs match in their topic. The similarity values in the table give an idea of the similarities measured and their standard deviation (stdev). The few false positives, however, also achieve high similarities, but based on a lot more comments on the side of Slashdot. It seems that the number of comments correlates with the similarity and that more comments result in topic drift. Since the title of a YouTube video is often, yet not always, very descriptive we have determined the percentage of pairs where the video title overlaps with the Slashdot article. On average, this is the case in 60% of the examined item pairs which means that with comment-based cross-medial retrieval

Table 1: Overview of the inspected cross-media similarities.

Fuzzy Match	Share	Similarity				Avg. # Comments		Title Match
		min	avg.	max	stdev	Slashdot	YouTube	
equal	36%	0.71	0.78	0.91	0.06	53	927	72%
related	55%	0.71	0.76	0.91	0.04	81	683	62%
unrelated	9%	0.72	0.78	0.87	0.05	104	872	–
Σ	100%	0.71	0.77	0.91	0.05	74	790	60%

the recall increases by 40% compared to title-based retrieval. Note that the publisher-supplied tags on the videos have not yet been investigated. In sum, the precision at rank 100 is 0.91, when equal and related matches are considered relevant.

3. CONCLUSION

The experiments demonstrate the feasibility of comment-based cross-media retrieval when the focus is on the topical similarity of Web items. This may be the starting point to cluster items across media. It is important to mention that comments will not necessarily replace other types of annotations, such as titles, captions, or tags. Instead, they are yet another information type that complements the others. What sets comments apart is that the act of commenting is not perceived as labor, much unlike tagging: commenters follow their desire to express their positive or negative position to an item. Our approach to utilize comments for the quantification of cross-media similarity is straightforward. I.e., on the one hand it can be assumed that the application of more sophisticated retrieval models will increase the recall, but on the other hand, our choice of a retrieval model demonstrates the robustness by which comment-based cross-media similarity can be measured.

4. REFERENCES

[1] K. Deschacht and M.-F. Moens. Finding the Best Picture: Cross-Media Retrieval of Content. In *Proc. of ECIR'08*.

[2] M. Inoue. On the Need for Annotation-based Image Retrieval. In *Proc. of IRiX at SIGIR'04*.

[3] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based Multimedia Information Retrieval: State of the Art and Challenges. *ACM TOMCCAP*, 2(1):1–19, 2006.

[4] G. Mishne and N. Glance. Leave a Reply: An Analysis of Weblog Comments. In *Proc. of WWE'06*, May 2006.

[5] M. Potthast. Measuring the Descriptiveness of Web Comments. In *Proc. of SIGIR'09*.

[6] M. Potthast and S. Becker. Opinion Summarization of Web Comments. In *Proc. ECIR'10*.

[7] W. G. Yee, A. Yates, S. Liu, and O. Frieder. Are Web User Comments Useful for Search? In *Proc. of LSDS-IR at SIGIR'09*.

Table 2: Selection of matching Web items found with comment-based cross-media retrieval.

Cross-Media Similarity	Slashdot Comments	YouTube Comments	Slashdot URL (In Acrobat the URLs are clickable)	YouTube URL
0.91	83	950	http://slashdot.org/story/07/03/15/2056210	http://www.youtube.com/watch?v=RuWVMB7OxbM
0.82	69	950	http://slashdot.org/story/08/02/05/1511225	http://www.youtube.com/watch?v=Z_gKOCb4QBA
0.81	102	950	http://slashdot.org/story/08/01/02/1611240	http://www.youtube.com/watch?v=tLlHibrFATg
0.76	41	950	http://slashdot.org/story/07/10/16/1526257	http://www.youtube.com/watch?v=TluRVBhmf8w
0.74	40	950	http://slashdot.org/story/07/07/11/1246250	http://www.youtube.com/watch?v=DLxq90xmYUs
0.74	79	766	http://slashdot.org/story/07/08/13/1347253	http://www.youtube.com/watch?v=BWQ5ZMnz25I
0.74	66	78	http://slashdot.org/story/06/02/02/0024235	http://www.youtube.com/watch?v=F0uq21xjMCw
0.73	75	950	http://slashdot.org/story/08/06/04/1159207	http://www.youtube.com/watch?v=adc3MSS5Ydc