

File ID 191360
Filename Chapter 12: Conclusions to the thesis

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type Dissertation
Title Search in audiovisual broadcast archives
Author B. Huurnink
Faculty Faculty of Science
Year 2010
Pages viii, 198
ISBN 978-90-786-7599-0

FULL BIBLIOGRAPHIC DETAILS:

<http://dare.uva.nl/record/358972>

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use.

Chapter 12

Conclusions to the Thesis

12.1 Answers to Research Questions

Part I of this thesis was devoted to deepening our understanding of the searcher in the audiovisual broadcast archive. We started our exploration by asking

RQ 1 *What kinds of content do media professionals search for, and in what manner do they search for it?*

Our transaction log analysis in Chapter 3 showed that, when considering the program titles and thesaurus entries of clicked results, professionals were searching for program names, person names, general subject words, locations, and other names, in that order. An additional analysis in Table 4.1 of Chapter 4 included extra data about matches between search terms and text in the technical metadata fields from purchased results, and showed that the professionals also search for document identifier codes and other technical metadata. In addition, date filters were also frequently used to select specific broadcast dates. From this we conclude that a significant portion of the searches that media professionals perform are for known items, where the searcher knows exactly which program they are looking for. Further, we examined the orders for audiovisual material that media professionals placed to the archive, and found that two thirds of them were not for entire programs, but rather for video stories and fragments.

We used the knowledge that we gained about the types of content that users were searching for in addressing our next research question,

RQ 2 *Can we recreate those searches by media professionals that result in*

*purchases, and use them to create an artificial testbed for retrieval evaluation?*⁶

Using the simulation framework that we developed in Chapter 4, we were able to create simulated pairs of queries and purchases that, when used for retrieval evaluation, approached the system rankings of real queries and purchases. One benefit of the simulator was that we were able to use it to create additional queries and relevance judgments for our archive-based test collection, where due to the limited number of programs only a small amount of real-world transaction log information was available.

In Part II of the thesis we studied how automatically generated content metadata could be used to improve the search for content in audiovisual broadcast archives. Upon a review of the literature, we asked questions about specific problems in content-based video retrieval, firstly,

RQ 3 *Given that search by visual concept detector is a valuable method for content-based video retrieval, how can we identify the correct visual concepts to use for a query?*

In answering this question we investigated the use of a concept selection benchmarks, which defined “correct” query-to-concept mappings on a per query basis, to assess whether a candidate set of selected concepts would do well for retrieval. We found that identifying the set of concepts to use for a query is no easy task; we compared two orthogonal, knowledge-intensive methods for concept selection benchmark creation, and found that neither of these approaches produced mappings that gave consistently the best performance when applied to the retrieval task. However, when applied to the task of assessing concept selection, we found that a benchmark created by mining collection knowledge produced assessments of concept selection that agreed with final retrieval performance. So to answer our research question, a set of benchmark query-to-concept mappings, created by mining an annotated collection, can be used to assess selection of the correct visual concepts for a query. This comes with the caveat that the concept of a “correct concept” is an elusive one when applied to this assessment task; we leave further investigation of this to future work.

Moving on to the problem of the temporal mismatch when searching with transcripts, we posed the question,

RQ 4 *Within a video broadcast, the same object may appear multiple times within and across the video and audio signal, for example being mentioned in speech and then appearing in the visual signal. How can this phenomenon be characterized, and can we model and use this characteristic so as to improve cross-stream retrieval of visual items using transcripts?*

We investigated the temporal mismatch, and the redundant occurrence of visual items in the video signal, in two domains: broadcast news, the audiovisual archive setting (with documentaries, current affairs magazine shows and educational programs). We characterized how occurrences of items were distributed across neighbouring shots, and used these distributions to inform a retrieval model that expands the transcripts associated with a shot with transcripts from neighboring shots. We found marked differences in the responsiveness of queries in the two domains to the document expansion process. The narrative structure of broadcast news, where news stories are presented in tight units one by one, meant that in broadcast news, queries benefited from expanding shots with transcripts from about ten adjacent shots, while in the archive domain retrieval performance kept improving as shots were expanded with transcripts from 60 adjacent shots. We feel this illustrates the difference in narrative structure across the two domains; broadcast news consists of clearly delineated short story units, while programs in the archive domain often do not contain clear story boundaries, and changes in visual content can be unpredictable.

Finally, we moved from our examination of problems in specific content-based video retrieval methods to study the effect of combining state-of-the-art content-based video retrieval methods with the retrieval capabilities of the archive, asking,

RQ 5 *What is the potential impact of content-based video retrieval in the audiovisual broadcast archive, taking into account both the needs of professional users, and the manually created data already present in the archive?*

Through our analysis of transaction logs from the Netherlands Institute of Sound and Vision we had already created two sets of queries that reflected the needs of media professionals. As the number of queries that was created through this process was relatively small, at a total of 65, we also included 2,190 simulated queries created by our best simulator. Our experiments showed that combining content-based video retrieval methods with search on manual catalog annotations from the archive resulted in significant increases in video retrieval performance. For content queries, performance more than doubled as compared to using the catalog annotations on their own.

12.2 Main Contributions

We group the main contributions of the thesis into two areas that correspond to the two parts of this thesis: understanding searchers in audiovisual broadcast archives, and improving search for content in audiovisual broadcast archives. The contributions are further structured per chapter.

Understanding Searchers

- Our first contribution in understanding searchers is a large-scale transaction log analysis of the electronic traces left behind by media professionals at a national audiovisual broadcast archive. As part of our contribution, we proposed a novel method for categorizing query terms, one which is applicable in environments where users search for fielded documents; query terms can be categorized by matching them to the field in which they appear in the clicked results. This allows a categorization of the long tail of query terms that are missed by typical frequency-based analyses.
- Our second contribution in this area is a framework for simulating logged user queries and purchases. Simulators developed within this framework can be used to create testbeds of artificial queries and purchases for evaluating different retrieval systems, and require only a collection of documents as input. The validation approach that we presented determines how well the rankings of retrieval systems on the simulator output correlate with the rankings of retrieval systems on a set of real logged searches and purchases. Although our experiments are performed in the setting of the audiovisual broadcast archive, the framework can be extended to other domains where users search for and purchase documents.

Improving Search for Content in the Archive

- Our first contribution for improving search for content in the archive is our study of two knowledge-intensive methods for determining which concepts from a limited vocabulary should be matched to a query for retrieval. The resulting query-to-concept mappings can be used to assess new sets of incoming concepts for the same query and predict their relative retrieval performance. The methods and insights that we have developed here can be used to help inform automatic concept selection algorithm automatic concept selection algorithmss for detector-based search. In addition, we feel they will also be of use for new problem areas where a limited number of concepts need to be matched to a query, for example in a setting where documents are tagged with terms from a small domain vocabulary.
- Our second contribution here is our investigation of the temporal mismatch and the redundancy of visual items in the video and audio signal (the latter manifested by proxy by their mention in transcript text). While this effect has been studied before, our analysis covered a greater temporal neighbourhood than other studies, and in addition was performed for hundreds of queries and concepts, in both the broadcast news and the audiovisual archive domains. Our

findings could help designers to improve the performance of transcript-based search systems for the two domains. In addition the characterizations of visual redundancy and the temporal mismatch could be insightful for researchers in the humanities who are interested in large-scale quantitative analyses of narrative structure.

- Our third contribution for this part is an analysis of ways in which automatically generated content metadata can be used to improve retrieval performance for the searches of media professionals. Our study coupled state-of-the-art content-based video retrieval methods to the practice of an audiovisual broadcast archive. The study can help inform policy makers at audiovisual archives who are facing the limitations of today’s manual annotation practices and are considering incorporating content retrieval into their work-flow. In addition, both our findings and the resources that we have created can be used by researchers who wish to explore content-based video retrieval outside of the usual laboratory benchmark setting.

12.3 Future Directions

In this thesis we analyzed the transaction logs of an audiovisual archive to gain information about what media professionals were looking for in an audiovisual archive, and found that they were looking for (and buying) video fragments. We feel a valuable direction for future research lies in applying information from video purchases fragments to improve the search for video content. For example, in Chapter 7 we illustrated a search for “shots F16”, that resulted in a purchase of several shots of material containing fighter jets from that broadcast. These keywords could be associated with the purchased shots, and returned in later searches for the same keyword. This task may not be as easy as it seems as at first glance; searchers in the audiovisual archive do not always provide clearly defined boundaries for their purchases (in other words, visual material relating to their keywords may be interspersed with non-related visual material), which brings us to another avenue for exploration.

We feel that the transaction log analysis presented in this thesis could be extended to study the potential of such logs for automatically creating large-scale sets of queries and relevance judgments for content-based video retrieval. The challenge here is automatically identifying the queries that are for visual content, and identifying which purchases and parts of purchases contain visually relevant material. This idea can be extended beyond the audiovisual broadcast archive; in online settings we are seeing the rise of *deep linking* for user-generated video. Video portals such as YouTube are enabling users to tag fragments within uploaded video. In addition, social commenting functions are enabled so that users can specify a time with

their comment. These types of annotations could also be used to create large-scale evaluation collections that approach the size of web data. Not only could such data be used for evaluation, it could also be used to learn query-specific concept detectors that are adapted to the environment in which the data is recorded.

Another area for future work lies in studying how automatically-generated metadata may be used to address queries that are not purely for visual information. In this thesis, and in content-based video retrieval studies in general, retrieval queries have so far been largely restricted to visual information needs. This is not the only dimension along which users may require audiovisual items. For example, users may require information along other information streams, such as quotations in the speech stream, or videos with specific emotional content. Or they may require iconic events, such as the marriage of a monarch. Identifying and addressing such information needs will increase the diversity of ways in which automatically generated content metadata can be applied to help searchers.

In this thesis we have seen that there is a demand in the audiovisual broadcast archive for access to video at the content level. In addition, we have demonstrated that retrieval performance for content queries can more than double when combining automatically generated content metadata with the manually created metadata already present in the archive. However, the test of acceptance will be for audiovisual broadcast archives to integrate automatically generated metadata into their video descriptions. Only then will we be able to determine the full potential for automatic methods to improve search in audiovisual broadcast archives.