
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

[http://eprints.gla.ac.uk/5739/](http://eprints.gla.ac.uk/5739/)

Deposited on: 12 April 2018

Enlighten – Research publications by members of the University of Glasgow
[http://eprints.gla.ac.uk](http://eprints.gla.ac.uk)
ABSTRACT

So far, several approaches have been studied to bridge the problem of the Semantic Gap, the bottleneck in image and video retrieval. However, no approach is successful enough to increase retrieval performances significantly. One reason is the lack of understanding the user’s interest, a major condition towards adapting results to a user. This is partly due to the lack of appropriate interfaces and the missing knowledge of how to interpret user’s actions with these interfaces. In this paper, we propose to study the importance of various implicit indicators of relevance. Furthermore, we propose to investigate how this implicit feedback can be combined with static user profiles towards an adaptive video retrieval model.

1. MOTIVATION

With the increasing availability of new tools and applications to record, broadcast and stream videos, there is a need to create new retrieval engines to assist the users in searching and finding scenes they would like to see within different video files. Research to date have a particular emphasis on the system side, resulting in the design of retrieval tools that assist the users in performing search sessions. However, since the effectiveness of current video retrieval systems is everything but satisfying for the users, more sophisticated research is needed to increase the retrieval performance to a similar level as their textual counterparts.

Unlike text retrieval systems, retrieval on digital video libraries is facing a challenging problem: The Semantic Gap. This is the difference between the low-level data representation of videos and the higher level concepts a user associates with video. In 2005, the panel members of the International Workshop on Multimedia Information Retrieval identified this gap as one of the main technical problems in multimedia retrieval [12], carrying the potential to dominate the research efforts in multimedia retrieval for the next few years. Retrieval information such as textual sources of video clips, i.e. speech transcripts, are often not reliable enough to describe the actual content of a clip. Moreover, the approaches of using visual features and automatically detecting high level concepts, as mainly studied within TRECVID [19], turned out to be not efficient enough to bridge the semantic gap.

One approach to bridge the semantic gap is to improve the interfaces of video retrieval systems, enabling the users to state their information demand appropriately. However, as the performance of state-of-the-art systems indicates, interface designs are, so far, not advanced enough to provide the users with such facilities. A promising approach to solve this problem is to incorporate an adaptive retrieval model, which automatically adapts retrieval results based on the user’s preferences. An adaptive retrieval model can be useful to significantly reduce the number of steps the user has to perform before he retrieves satisfying search results.

Identifying user preferences is a re-appearing research question and hence has been studied intensively in various areas. A common approach is to create individual (static) user profiles, where users have to provide personal information such as demographics, preferences or ratings, i.e. when they register for a service [17]. The user’s general interests can then be inferred by analysing these profiles. Arezki et al. [2] provide an example in which a computer scientist enters the text query “java course”. He will expect other retrieval results than a non expert who formulates the same query.

In a retrieval context, profiles can be used to set the user’s search query into his interest context and to re-rank retrieval results. This approach is based on the assumption that the user’s information interest is static, which is, however, not appropriate in this context. Campbell and van Rijsbergen [3] argue, that the users’ information need can change within different retrieval sessions and sometimes even within the same session. One approach to capture this sudden change of interest is to analyse the records of interactions with a retrieval system. There are different types of interaction feedback, usually divided into two categories: explicit and implicit feedback. Explicit feedback is given when a user actively informs a system what it has to do on purpose, such as selecting something and marking it as relevant. Implicit feedback is given unconsciously. An example is printing out the query “java course”. He will expect other retrieval results than a non expert who formulates the same query.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permission from the publisher, ACM.

VLDB '08, August 24-30, 2008, Auckland, New Zealand
Copyright 2008 VLDB Endowment, ACM 000-0-00000-000-0/00/00.

Studying Interaction Methodologies in Video Retrieval

Frank Hopfgartner*

Department of Computing Science
University of Glasgow
Glasgow, G12 8RZ, United Kingdom
hopfgarf@dcs.gla.ac.uk

*supervised by Dr. Joemon Jose
issues of implicit feedback can be addressed in video retrieval since digital video libraries facilitate more interaction and are hence amenable to implicit feedback.

Different questions arise from the use of implicit relevance feedback and exploiting static user profiles. How can these two approaches be set into an adaptive retrieval context? So far, rarely anything is known about which interface features can be important implicit indicators of relevance. In this research, we aim to study user interactions with different video retrieval interfaces to shed light on implicit relevance feedback, a necessary step towards further studies on combining static user profiles and user interactions to an adaptive retrieval model. In the scope of my research, I will apply a simulated user evaluation. The approach of simulating users to fine tune retrieval systems has been studied before (i.e. [9, 21, 22]), the results being promising to follow the methodology. The different approaches and methodologies will be introduced in the following.

2. BACKGROUND

This section provides an overview of methodologies to gather the user’s information need. It starts with an introduction into the idea of retrieval personalisation by incorporating user profiles. Further, research of exploiting implicit user interactions to adapt retrieval results will be introduced. Finally, an alternative evaluation framework based on user simulation is presented which can be applied to evaluate implicit relevance feedback in video retrieval.

2.1 Personalisation

Incorporating personal data provided by the users is a popular approach to customise web services. Manber et al. [15] showed that customers of online web services appreciate such customisation services. Cranor [5] provide an overview of general types of personalisation systems, one being the user-initiated personalisation. In this scenario, users can provide feedback to customise web sites. Users might provide information about their home region to see the weather forecast or news related to this region. Furthermore, they might define and re-arrange the page layout of a web site or adapt it in accordance to their display or connection bandwidth. User profiles can be used to create a simplified model of the user which represents his interests on general topics.

In the retrieval context, these user models can then be used to adapt retrieval results on concepts, i.e. “politics”, “sports” or “science”. However, as stated before, the user’s interest is not static and can change within retrieval sessions. Hence, relying on static user profiles cannot compensate this evolution. User profiles can be used to create a simplified model of the user which represents his interests on general topics.

2.2 Evaluation Framework

A common approach to study the users’ behaviour of interacting with a computer system is to perform a user study, to monitor the users’ interactions and to analyse the resulting logfiles. Such an analysis shall help to identify good implicit indicators of relevance, as it can help to answer basic questions: What did the user do to find the information he/she wanted? Can the user behaviour be used to improve retrieval results?

To get an adequate impression of the users’ behaviour
when interacting with a video retrieval system, two main criteria can be stressed out: A large quantity of different users interacting with the system is necessary to draw generalisable conclusions from this study, i.e. by analysing user clicklogs. Besides, non-expert users should be interacting with the system, as they will interact in a more intuitive way than expert users. However, such a methodology is inadequate to evaluate interactive retrieval systems. Most interactive video retrieval systems are evaluated in laboratory based user experiments. There are many issues with such evaluation methodologies such as the lack of repeatability. In addition, to achieve a robust measurement, we need a large user population, which is very expensive. Besides, it is hardly possible to benchmark different parameter combinations of features for effectiveness using user-centred evaluations.

An alternative way of evaluating such user feedback is the use of simulated interactions. In such an approach, a set of possible steps are assumed when a user is performing a given task with the evaluated system. Finin [6] introduced one of the first user simulation modelling approaches. This “General User Modelling System” (GUMS) allowed software developers to test their systems in feeding them with simple stereotype user behaviour. White et al. [22] proposed a simulation-based approach to evaluate the performance of implicit indicators in textual retrieval. They simulated user actions as viewing relevant documents, which were expected to improve the retrieval effectiveness. In the simulation-based evaluation methodology, actions that a real user may take are assumed and used to influence further retrieval results. Hopfgartner et al. [11] introduced a simulation framework to evaluate adaptive multimedia retrieval systems. In order to develop a retrieval method, they employed a simulated evaluation methodology which simulated users giving implicit relevance feedback. Hopfgartner and Jose [9] extended this simulation framework and simulated users interacting with state-of-the-art video retrieval systems. They argue that a simulation can be seen as a pre-implementation method which will give further opportunity to develop appropriate systems and subsequent user-centred evaluations. Vallet et al. [21] use the concept of simulated actions and try to mimic the interaction of past users by simulating user actions based on the past history and behaviour of users who use an interactive video retrieval system. Their study has proven to facilitate the analysis of the diverse types of implicit actions that a video retrieval system can provide.

Analysing these research efforts lead to the conclusion that even though simulation based studies should be confirmed by user studies, they can be a cheap and repeatable methodology to fine tune video retrieval systems. Hence, user simulation is a promising approach to further study adaptive video retrieval, at least as a preliminary step.

3. RESEARCH METHODOLOGY

The scope of this research is to develop an adaptive video retrieval model, which automatically adapts retrieval results to the users information need. As introduced in Section 2.1, a promising methodology to adapt retrieval results to the user’s need is the use of static user profiles and the interpretation of implicitly given relevance feedback. However, the range of implicit indicators in an video retrieval application is unclear. Within this study, we will therefore investigate the following research questions: Which implicit feedback a user provides can be considered as a positive indicator of relevance and can hence be used to adapt retrieval results? The second question is how these features have to be weighted to increase retrieval performance. It is not clear which features are stronger and which are weaker indicators of relevance, respectively. Furthermore, we will investigate how both static user profiles and implicit relevance feedback should be combined to adapt to the users need. Once the users’ intentions and information demand is clear, systems can be built that take advantage of such knowledge and optimise the retrieval output for each user by implementing an adaptive video retrieval model.

To further investigate these questions, we aim to develop an exemplary retrieval system. Therefore, we propose [10] a framework for recording, analysing, indexing and retrieving news videos such as the BBC One O’Clock News. The idea of this scenario is to automatically identify news stories which are of interest for the user and to recommend them to him.

Within the proposed scenario, users will be provided with different interface approaches for different interaction environments such as desktop PCs of iTV boxes. Hence, users are required to interact differently with the interfaces. Thus, the difference has a strong influence on the user’s behaviour, making the importance of implicit indicators of relevance application-dependent. Comparing user interactions with different applications should help to identify common positive indicators though. The research will be conducted around two different applications where we can monitor user feedback: desktop computers and television. The specific characters of these environments are introduced in the following.

- Desktop Computers: The most familiar environment for the user to do video retrieval is probably a standard desktop computer. Most video retrieval systems have been designed to run under such environment. The interface can be displayed on the screen and users can easily interact with the system in using the keyboard or mouse. One can assume that users will take advantage of this interaction and hence give a high quantity of implicit feedback. From today’s point of view, this environment offers the highest amount of possible implicit relevance feedback.

- TV: A widely accepted medium for multimedia consumption is the television. Watching television, however, is a passive procedure. Viewers can select a programme using a remote control, changing the content is not possible though. Recently, Interactive TV, is becoming more and more popular. Using a remote control, viewers can interact directly when watching television, e.g. in participating in quiz shows. In news video retrieval, this limited interaction is a challenge. It will be more complex to enter query terms, e.g. in using the channel selection buttons. Hence, users will possibly avoid to enter key words. On the other hand, the selection keys provide a method to give explicit relevance feedback. An example: The viewer sees a video segment on television. Now, he/she uses the remote control to judge the relevance of this segment.

A well accepted research methodology in the information retrieval community to evaluate different approaches is to perform user studies. In the video retrieval domain, the
proposed a simple model of adapting retrieval results based on this interaction. The model seemed to enhance retrieval results. In [9], we extended the simulation framework and simulated users interacting with state-of-the-art video retrieval systems. The introduced simulation framework can be seen as a pre-implementation method to further study implicit relevance feedback. In [21], we exploited the log files of a user study and simulated users interacting with an interface. The study has proven to facilitate the analysis of the diverse types of implicit actions. In this work, we used community based implicit feedback mined from the interactions of previous users of our video search system, to aid users in their search tasks. The results of our evaluation indicate that we achieved our goals, the performance of the users in retrieving relevant videos improved, and users were able to explore the collection to a greater extent.

Consequently, the in this paper introduced research proposal aims to shed further light on the use of implicit relevance feedback in video retrieval, a necessary study towards the development of an adaptive video retrieval model.

5. ACKNOWLEDGMENTS

This research is supported by the European Commission under the contract FP6-027026-K-SPACE. It is the view of the author but not necessarily the view of the community.

6. REFERENCES


