# 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

## 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 availablility 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. Retrievable 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 indicate, 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 infered 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 a web page, which may indicate an interest in that web page. The basic assumption is that during a search, users' actions are aimed to maximise the retrieval of relevant information. Implicit indicators have been used and analysed in other domains, such as the WWW [4] and text retrieval [23, 14], but rarely in the multimedia domain. However, traditional

*supervised by Dr. Joemon Jose


*VLDB '08,* August 24-30, 2008, Auckland, New Zealand

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 change of retrieval focus.

Analysing the content of relevant rated documents, i.e. by extracting key terms of these documents, can be used to expand the users' original search queries or to re-rank retrieval results. In the text retrieval domain, a common approach to adapt to the user is to explicitly ask the user about the relevance of retrieved results. Giving explicit feedback, users are forced to update their need, which can be problematic when their information need is vague [20] or when they are unfamiliar with the data collection [18]. Furthermore, users tend to provide not enough feedback on which to base an adaptive retrieval algorithm [7]. In addition, they are uncertain on how exactly such feedback will be used by the underlying retrieval system.

Deviating from the method of explicitly asking the user to rate the relevance of retrieval results, the use of implicit feedback techniques helps learning user interests unobtrusively. Kelly and Teevan [14] provide a literature overview of the research which has been done in the text retrieval field. The main advantage is that users are relieved from providing feedback. A disadvantage is that information gathered using implicit techniques are less accurate than information based on explicit feedback [16]. Agichtein et al. [1] evaluated the effect of user feedback on web retrieval. They show that implicit relevance feedback can improve retrieval performance by much as 31% relative to systems that do not incorporate any feedback. Furthermore, both implicit and explicit measures can be combined to provide an accurate representation of the users' interests.

In the video domain, however, implicit relevance feedback is a nearly untouched research topic. Hopfgartner and Jose [9] identified various implicit indicators of relevance in video retrieval when comparing the interfaces of state-of-the-art video retrieval tools. The most common features they identified were: clicking on a keyframe to start playing a video, browsing through a result list, sliding through a video, highlighting additional metadata and playing a video for a certain amount of time. However, which of these implicit measures are useful to infer relevance has rarely been analysed in detail. While Claypool et al. [4] identified time spend on a web site as being a valid implicit indicator of relevance in the text domain, Kelly and Belkin [13] criticise the time factor as indicator in the video domain. They assume that information-seeking behaviour is not influenced by contextual factors such as topic, task and collection. Therefore, they performed a study to investigate the relationship between the information-seeking task and the display time. Their results cast doubt on the straightforward interpretation of dwell time as an indicator of interest or relevance.

A big challenge in the analysis of implicit indicators of relevance in the multimedia domain is to draw generalisable conclusions from system-dependent studies. Hopfgartner [8] compared different state-of-the-art video retrieval systems, concluding that both backend and frontend of all systems differ significantly from each other. This includes the design of different interfaces which incorporate various combinations of features which impedes comparative studies on implicit relecance feedback in video retrieval. Moreover, varying feedback posibilities directly influence the importance of these features, making a general judgement of features a challenging research task. In our study, we aim to identify various implicit indicators of relevance by studying users' interaction with video retrieval interfaces. The findings are an important step towards the development of an adaptive video retrieval model. Once more is known about this research domain, personalisation systems can be enhanced accordingly.

## 2.2 Evaluation Framework

A common approach to study the users' behaviour of interacting with a computer system is to perfom 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 logfiles. 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 with 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 retrievel 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 probable 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

TRECVID effort is the most important platform which provides a common data collection, pre-defined search topics and relevance judgements, the main ingredients to perform user experiments. Within this study, we aim to apply this evaluation methodology and to analyse the resulting user interaction logfiles. This analysis should help to understand how users interacted with this application, and will lead to further knowledge which interface features are generalisable indicators of relevance. Furthermore, we will use the in Section 2.2 introduced simulation methodology by exploiting the user log files and analysing the effect of different feature weighting schemes on retrieval performance. This study should lead to a better understanding of user behaviour in video retrieval.

# 4. DISCUSSION

When comparing video and text retrieval systems, one notices a large difference in retrieval performance. The Semantic Gap has been identified to be the main reason for this problem. While humans can easily understand the content of images or videos, computers are not capable of doing so. Different approaches are currently studied to bridge this gap, the most prominent being the automatic detection of high level concepts in a video. As retrieval performances gathered within TRECVID show, however, this approach has not been efficient enough.

A second approach is to improve the user's facilities to formulate his information desire by enhancing the interface of retrieval systems. However, as the performance of state-of-the-art systems indicate, interface designs are not advanced enough to provide the users with facilities to enter their information need.

In this paper, we aim to study how the users' information need can be captured and be used to adapt retrieval results, accordingly. We introduced two approaches, discussed their weaknesses and argued to combine them.

The first approach is to rely on static user profiles. Profiling requires the user to provide information about his interests, i.e. when registering for a service. These information can than be used to adapt the retrieval results. When a user stated for instance that he is interested in football, triggering a sport related search query such as "goal" could result in a football dominated result list which is hence adapted to the user's interest. This approach is, however, based on the assumption that the user's information need is static. As we argued before, this cannot be applied in a retrieval context where the user's information need can change even within one search session, though.

Therefore, we propose to further adapt the retrieval model based on the user's interaction with the retrieval system's interface. In the text retrieval domain, the approach of interpreting the user's action as implicit indicator of relevance turned out to be an effective method to increase retrieval performance. In the video retrieval domain, however, rarely anything is known about which implicit feedback can be used as implicit indicators of relevance. Hence, we aim to further study implicit relevance feedback to shed light on the interpretation of the user's interests.

Our previous research seems to indicate that implicit relevance feedback can also be applied in the video domain. In [11], we simulate users providing implicit relevance feedback by interacting with a novel video retrieval interface and 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

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR '06: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 19–26, New York, NY, USA, 2006. ACM Press.

[2] R. Arezki, P. Poncelet, G. Dray, and D. W. Pearson. *Adaptive Hypermedia and Adaptive Web-Based Systems*, chapter Web Information Retrieval Based on User Profiles, pages 275–278. Springer Berlin / Heidelberg, 2004.

[3] I. Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. In *Proc. of CoLIS-96, 2nd Int. Conf. on Conceptions of Library Science*, pages 251–268, 1996.

[4] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Intelligent User Interfaces*, pages 33–40, 2001.

[5] L. F. Cranor. 'I didn't buy it for myself' privacy and ecommerce personalization. In *WPES '03: Proceedings of the 2003 ACM workshop on Privacy in the electronic society*, pages 111–117, New York, NY, USA, 2003. ACM.

[6] T. W. Finin. GUMS: A General User Modeling Shell. *User Models in Dialog Systems*, pages 411–430, 1989.

[7] M. Hancock-Beaulieu and S. Walker. An evaluation of automatic query expansion in an online library catalogue. *J. Doc.*, 48(4):406–421, 1992.

[8] F. Hopfgartner. *Understanding Video Retrieval*. VDM Verlag, Saarbruecken, Germany, 2007.

[9] F. Hopfgartner and J. Jose. Evaluating the Implicit Feedback Models for Adaptive Video Retrieval. In *ACM MIR '07 - Proceedings of the 9th ACM SIGMM International Workshop on Multimedia Information Retrieval, Augsburg, Germany*, pages 323–332, 09 2007.

[10] F. Hopfgartner and J. Jose. A news video retrieval framework for the study of implicit relevance. In *SMAP '07 - 2nd International Workshop on Semantic Media Adaptation and Personalization, London, United Kingdom*, pages 233–236, 12 2007.

[11] F. Hopfgartner, J. Urban, R. Villa, and J. Jose. Simulated Testing of an Adaptive Multimedia Information Retrieval System. In *CBMI'07 - Proceedings of the Fifth International Workshop on Content-Based Multimedia Indexing, Bordeaux, France*, pages 328–335, 06 2007.

[12] A. Jaimes, M. Christel, S. Gilles, S. Ramesh, and W.-Y. Ma. Multimedia Information Retrieval: What is it, and why isn't anyone using it? In *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pages 3–8, New York, NY, USA, 2005. ACM Press.

[13] D. Kelly and N. J. Belkin. Display time as implicit feedback: understanding task effects. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 377–384, New York, NY, USA, 2004. ACM Press.

[14] D. Kelly and J. Teevan. Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum*, 32(2), 2003.

[15] U. Manber, A. Patel, and J. Robison. Experience with personalization of Yahoo! *Communications of the ACM*, 43(8):35–39, 2000.

[16] D. M. Nichols. Implicit rating and filtering. In *Proceedings of 5th DELOS Workshop on Filtering and Collaborative Filtering*, pages 31–36. ERCIM, 1998.

[17] D. Poo, B. Chng, and J.-M. Goh. A hybrid approach for user profiling. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, page 9, 2003.

[18] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Readings in information retrieval*, pages 355–364, 1997.

[19] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[20] A. Spink, H. Greisdorf, and J. Bateman. From highly relevant to not relevant: examining different regions of relevance. *Inf. Process. Manage.*, 34(5):599–621, 1998.

[21] D. Vallet, F. Hopfgartner, and J. Jose. Use of Implicit Graph for Recommending Relevant Videos: A Simulated Evaluation. In *ECIR'08 - Proceedings of the 30th European Conference on Information Retrieval, Glasgow, United Kingdom*, 03 2008.

[22] R. White, M. Bilenko, and S. Cucerzan. Studying the use of popular destinations to enhance web search interaction. In *ACM SIGIR '07 - Proceedings of the 30th International ACM SIGIR Conference, Amsterdam, The Netherlands*, pages 159–166, 07 2007.

[23] R. White, J. Jose, C. van Rijsbergen, and I. Ruthven. A Simulated Study of Implicit Feedback Models. In *Proceedings of the 26th European Conference on Information Retrieval Research (ECIR '04). Lecture Notes in Computer Science*, 2004.