

A News Video Retrieval Framework for the Study of Implicit Relevance Feedback

Frank Hopfgartner
Department of Computing Science
University of Glasgow
Glasgow, United Kingdom
hopfgarf@dcs.gla.ac.uk

Joemon Jose
Department of Computing Science
University of Glasgow
Glasgow, United Kingdom
jj@dcs.gla.ac.uk

Abstract

In this paper, we propose a framework for recording, analysing, indexing and retrieving news videos such as the BBC One O'Clock News. We believe that such a framework will be useful to identify implicit indicators of relevance, a nearly untouched area in adaptive multimedia retrieval. Due to its advantages as a web application and its up-to-date content, it can be a promising approach to motivate a broad quantity of users to interact with the system.

1. Introduction

With the improving capabilities of current hardware systems, there are ever growing possibilities to store and manipulate videos in a digital format, leading to a growing number of video archives. People build their own digital libraries from materials created through digital cameras and camcorders, and use systems such as YouTube¹ and Google Video² to place this material on the web. Unfortunately, this data creation prowess is not matched by any comparable tools to organise and retrieve video information.

There is a need to create new retrieval engines to assist the user in searching and finding video scenes they would like to see from many different video files. Unlike text retrieval systems, retrieval on digital video libraries is facing a serious 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.

One way to address this problem is to use the interaction between users and the system. There are different types of interactions, usually divided into two categories: explicit and implicit feedback. Explicit feedback is given when a

user 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 [1] and text retrieval [10, 4], but as far as we know, not in the multimedia domain. However, traditional issues of implicit feedback can be addressed in video retrieval since digital video libraries facilitate more interaction and are hence amenable to implicit feedback. And as [3] indicate, implicit feedback can improve retrieval in digital video library retrieval systems. Therefore, we propose to develop a web based retrieval system for news video data which can be used to investigate further the relevance of implicit indicators. Our aim is to identify those implicit features which can positively influence video retrieval. To motivate a large number of different users to interact with the system on a long-term basis, it is necessary to provide a system including up-to-date news. Hence, we propose a framework to record, analyse and index local broadcasting news such as the BBC News. This paper is organised as follows: A brief introduction into web based news video retrieval is given in Section 2. The proposed framework will be introduced in Section 3: Section 3.1 proposes a framework of capturing, analysing and indexing broadcast news and Section 3.2 will introduce the retrieval backend, which shall also include user profiles to improve retrieval. We propose to implement a web application which will enable us to create different interfaces for different environments which will be introduced in Section 3.3. Section 4 will summarise the proposal.

¹<http://www.youtube.com/>

²<http://video.google.com/>

2. Background

In recent years, news videos have been the main focus of research in the field of interactive video retrieval. A main reason for this concentration on this domain is the international TRECVID [9] workshop, which provided a large corpus of news videos in the last few years.

One of the most successful competitor of this workshop is the Físchlár Digital Video System. The main interest of the developers is to study techniques to provide content-based navigation through digital video collection. Therefore, Smeaton et al. [8] record the daily evening news from the national's main broadcasting station and automatically segment news story boundaries. Their system provides story based news searching for the users. The web based interface of this system can be accessed by using a web browser, either Internet Explorer or Netscape, with an ORACLE plug-in for video streaming.

Pickering et al. [7] focus their research on content-based access to news videos in extracting keyframes and summarising shots to stories based on the similarity of the captured teletext of each shot. Their web based interface provides browsing and seeking facilities. Videos can be played using the RealPlayer.

Morrison and Jose [5] introduce the web based news video retrieval system VideoSqueak. They record the BBC One evening news and use the captured subtitles of the broadcast as retrieval source. In [5], they evaluate different presentation strategies for multimedia retrieval. A further investigation in user behaviour has been ignored.

3. A News Video Retrieval Framework

A common approach to study the user's behaviour of interacting with a computer system is the analysis of transaction 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 unsophisticated impression of 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 generalise the results of a logfile analysis. Besides, non-expert users should be interacting with the system, as they will interact in a more intuitive way than expert users.

To investigate further into the influence of implicit indicators of relevance, we propose to tackle these two criteria in implementing a web based video retrieval system. Web technology will enable us to spread the system to a large population, e.g. all students on the University campus. However, for offering an attractive news video retrieval system to the general public, the system should be based on a

up-to-date news video corpus. Hence, we propose to record, analyse and index daily broadcasting news such as the BBC One O'Clock News. The proposed system architecture will be introduced in the following.

3.1. Capturing and Indexing Process

Figure 1. System architecture of the capturing and indexing process

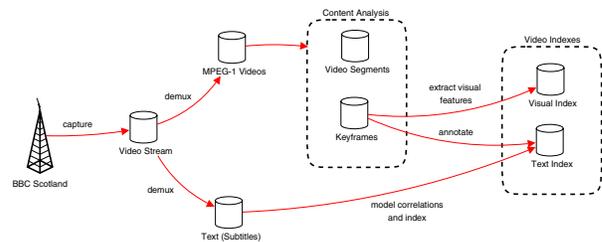


Figure 1 illustrates the proposed architecture for recording, analysing and indexing news video broadcasted by BBC Scotland. Later on, the architecture could be expanded to other news broadcasters.

First of all, the aerial broadcasting signal has to be captured, digitalised and stored on a server. The captured video stream consists of audio, visual and textual (subtitle) data which has to be demultiplexed. Due to technical conditions, we have to capture the analogue signal and store both audio and visual data in MPEG-1 format. The next step will be the content analysis, which includes the separation of the news video into segments and the extraction of appropriate keyframes. These keyframes will be the retrieval source for query-by-example retrieval. Hence, the visual features of these images have to be extracted and stored on a server. The standardised MPEG-7 format lends itself to it. In a next step, the keyframes will be annotated based on their visual features.

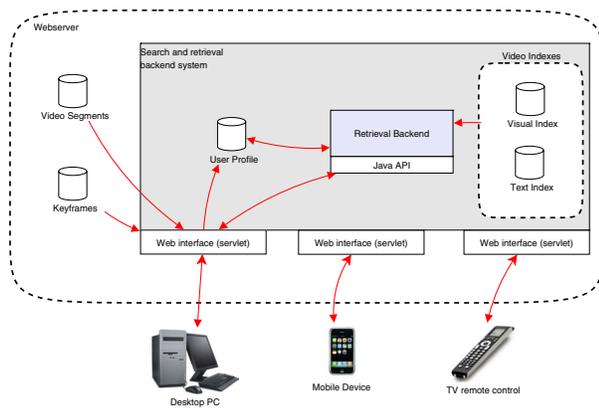
The subtitles need to be indexed. Therefore, they need to be segmented based on the same time code of subtitle and detected video shot. Besides, external sources such as WordNet [2] can be used to model correlation of the text. The Terrier Retrieval System [6] can be used for indexing the subtitles. Terrier is a Java based framework for the rapid development of large-scale information retrieval applications and provides indexing and retrieval functionalities.

3.2. The Adaptive Retrieval System Backend

As we need many users interacting with our retrieval system to study the effect of implicit relevance feedback, we

propose to implement a web based application. Using the AJAX technology will enable us to develop rich interfaces which will run on conventional web browsers. This takes away the burden of installing additional software on each client. Besides, thanks to the Web 2.0 hype, users get used to interacting with complex applications using their browser only. This might motivate them to use the system on a regular basis to retrieve broadcasting news. Figure 2 provides an overview of the proposed backend architecture. The retrieval system backend and the interface frontend are separated from each other. This approach enables us to develop and study different interface designs that provide different implicit relevance feedback techniques but share the same retrieval backend. Examples are given in Section 3.3.

Figure 2. System architecture of the adaptive retrieval system



The retrieval backend will run as a Java application on the web server and will access the video indexes. It will communicate with the interfaces, which will run as servlets. Both retrieval engine and interface will create and include user profiles to improve the retrieval performances. Each client will access the system using a web browser.

3.3. Environments

The importance of implicit indicators of relevance becomes apparent when interfaces which are designed for different user interaction environments shall be designed. In the following, we introduce three different interaction environments and discuss their restrictions and advantages for the user:

- **Desktop Computers:** The most familiar environment for the user to do video retrieval is probably a standard desktop computer. Most adaptive 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, such as query reformulation, browsing results and sliding through a video.

- **Mobile Devices:** In recent years, mobile devices have evolved from simple telephones to complex devices which are fully capable of handling multimedia content. Devices provide facilities to process multimedia data or browse the internet. Such devices, however, face two major problems: They are rather small and more complicate to handle than a desktop computer. The size of the screen does not provide much space for more complex interfaces. Besides, the lack of a real keyboard can be seen as a burden for the user to provide further keywords for retrieval. One can conclude that such an environment implies the use of a simple structured interface which does not force the user to enter many search terms. Hence, a user will have less possibilities to give implicit relevance feedback, for example query reformulation might be less important due to complicated handling of the devices.
- **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 as it is common for mobile telephones. Hence, users will possibly avoid to enter key words. On the other hand, the selection keys and a display on the remote control 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.

4. Summary

Implicit relevance feedback has been a successful approach in the textual domain to improve retrieval results. In the video data domain, however, this approach has rarely been studied. Therefore, we propose to develop a retrieval system for news videos which shall help to shed light on the possibility of implicit factors for retrieval. The news video

retrieval framework will include recording, analysing and indexing of news broadcasts such as the BBC One O'Clock News. Furthermore, we propose to develop web-based retrieval interfaces. A transaction logfile analysis of the users' interaction with the interface shall help to identify implicit indicators of relevance. To generalise the results of such an analysis, a large scale user population is needed. Using web technology will help us to address a broad number of users as web browsers can be found on nearly all computers and are more and more used for complex applications.

5 Acknowledgments

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

References

- [1] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Intelligent User Interfaces*, pages 33–40, 2001.
- [2] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA, 1998.
- [3] F. Hopfgartner and J. Jose. Evaluating the Implicit Feedback Models for Adaptive Video Retrieval. In *Proceedings of the International Workshop on Multimedia Information Retrieval (MIR 2007)*, Augsburg, Germany, 2007.
- [4] D. Kelly and J. Teevan. Implicit feedback for inferring user preference: A bibliography. *SIGIR Forum*, 32(2), 2003.
- [5] S. Morrison and J. Jose. A comparative study of online news retrieval and presentation strategies. In *ISMSE '04: Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04)*, pages 403–409, Washington, DC, USA, 2004. IEEE Computer Society.
- [6] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and D. Johnson. Terrier Information Retrieval Platform. In *Proceedings of the 27th European Conference on Information Retrieval (ECIR 05)*, Santiago de Compostela, Spain, pages 517–519, 2005.
- [7] M. J. Pickering, L. Wong, and S. Ruger. ANSES: Summarisation of news video. *Image and Video Retrieval*, 2788:481–486, 2003.
- [8] A. F. Smeaton. The Fischlar Digital Library: Networked Access to a Video Archive of TV News. In *TERENA Networking Conference 2002, Limerick, Ireland, 3-6 June 2002*, 2002.
- [9] 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.
- [10] 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*, pages 311–326, 2004.