Supporting Mobile Access to Digital Video Archives Without User Queries

Cathal Gurrin[1,2], Lars Brenna[2], Dmitrii Zagorodnov[2], Hyowon Lee[1], Alan F. Smeaton[3], and Dag Johansen[2]

[1]Centre for Digital Video Processing, [3]Adaptive Information Cluster, Dublin City University, Ireland
[2]Department of Computer Science, University of Tromsø, Norway

{cgurrin, hlee, asmeaton}@computing.dcu.ie {larsb, dmitrii, dag}@cs.uit.no

ABSTRACT

In this paper we present a technique for supporting mobile access to digital video archives without requiring explicit user queries. The idea is to infer the interests and needs of users from their WWW browsing history and represent those needs as persistent queries to the archive. An experiment, which we present here, suggests that this technique is effective for recommending video content to users on mobile devices. We also describe how to apply these findings to a mobile interface for a digital video archive.

Categories and Subject Descriptors

H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Search and Retrieval – *Information filtering, Query formulation.*

General Terms

Experimentation, Human Factors.

1. INTRODUCTION

Video delivery to mobile devices, such as mobile phones, is fast becoming an everyday occurrence due to recent advancements in network technology. For example, it is possible to view television content or pay-per-view content such as the goals from a football match on a mobile device. Expressing what to view, however, is an enduring problem with mobile devices, given the well documented limitations of user interaction with mobile devices.

In this paper we present initial results of a user experiment in which a persistent query is used as an indicator of a user's interest for content recommendation purposes on mobile devices. These persistent queries are generated from a user's WWW browsing history. In this work, we are not concerned with optimising the user interface or developing the optimal recommendation algorithm, rather we are interested in examining the possibilities that lie in utilising persistent queries for recommendation of content on mobile devices.

2. BACKGROUND

Small display size, awkward methods of data input and distractive environments have been noted as major constraints in designing systems for mobile platforms [1, 2]. For example, typical PDAs have a 3.8" TFT screen (240 x 320) in 16-bit colour; mobile phones – far more ubiquitous mobile devices – have even smaller screens and more limited navigation facilities. To overcome such limitations, mobile service providers rely heavily on personalisation techniques for optimising data presentation. They exploit the records of past user behaviour [3] to increase the usability of mobile portals, sometimes dramatically, thereby inciting their customers to use the services more often.

In addition to the optimisation of mobile portals, the utilisation of mobile devices to support content access to multimedia archives has been receiving increasing attention in recent years. For example the Físchlár-News [4] mobile video retrieval system supported personalisation and recommendation of new video content on mobile devices, where this recommendation is closely integrated with a user's browsing and playback history on a desktop version of the Físchlár-News application.

2.1 Content Access on Mobile Devices

Given the inherent limitations of mobile devices, research on interaction paradigms for the mobile environment suggests that simply following the conventional direct manipulation interfaces used successfully in desktop platforms is not sufficient [5, 6]. The general consensus is that a mobile interface requires a different interaction style from that of the desktop interface, and that attempts to replicate all the functionality of an equivalent desktop system into a mobile device are a mistake [1, 5]. In this research, we follow a number of general design guidelines for interacting with mobile devices, to limit complex user interactions:

- Minimise user input where applicable: provide simple user selections such as yes/no options or simple hyperlinking by tapping, instead of asking the user to articulate query formulation or use visually demanding browsing.
- Filter out information so that only a small amount of the most important information is presented.
- Proactively search and collect potentially useful pieces of information for a user and point these out, rather than trying to provide full coverage of all information via an elaborate searching/browsing interface.

In terms of developing any system for a mobile device which is to support searching and information retrieval tasks, all these guidelines suggest more pre-processing on the system's side in order to determine what information a particular user will most likely want to see. This encourages the development of systems that proactively recommend a particular piece of information (or pointers) to the user, and consequently lessen the demand for complex user interactions.

MobileHCI'06, September 12–15, 2006, Helsinki, Finland.

2.2 Recording Attention Data

Attention data is the record of a user's interactions with a system. We view interactions as a form of feedback, which can be either implicit or explicit in nature. Recording feedback from users is an increasingly common activity for online services, for some of which this data is critical. Much feedback is explicit, submitted voluntarily through comment forms that rate a product or an organisation (e.g. Amazon book reviews and eBay merchant ratings) or through application-specific interfaces (e.g. thumbs up/down buttons of Fischlar-News [4]). While such explicit feedback can be precise [7], not all users bother to submit it and it is not equally useful in all domains. One solution to this problem is to capture implicit feedback from a user's interactions with a system, thereby relieving the user of the overhead of providing explicit feedback. Morita & Shinoda [8] achieved improved content filtering through implicit feedback by measuring the time spent reading individual news stories. Their findings were largely confirmed in the evaluation of the Curious Browser [9].

Once collected, attention data can be utilised to improve the quality of search results or recommendations by employing a variety of information retrieval techniques. Much work in the area has examined the analysis of a user's interaction with hypermedia systems in order to personalise content and link presentation. It has been shown experimentally [3, 10] that utilising user context can increase the speed of user navigation of a hypertext system.

In the area of WWW search engines, the use of attention data is coming under increasing focus. It has been suggested [11] that conventional approaches to web search are limited by the fact that they do not utilise the previous browsing history of users, as user search behavior is (largely) repetitive and regular with similar queries and results being selected by users. As an alternative to conventional search engine ranking, group-based profiling is employed by the I-SPY experimental search engine [11], which groups users into communities and utilises the past searching behavior (using implicit feedback) of these communities to personalise the output of a conventional search engine.

3. CONTENT RECOMMENDATION

To evaluate whether attention data is useful on mobile devices for content recommendation from a video archive, we conducted a user study to capture attention data and use it to generate persistent queries for recommending video content. Five subjects took part in this initial study. Two users (A and B) captured attention data for a period of six weeks while the other three users captured data for a single week. Users were aware that this implicit monitoring of their browsing behaviour was taking place.

3.1 News Data Employed

We used the news video data from the TREC Video Retrieval Evaluation (TRECVID) [12] test collection from 2004. The main goal of TRECVID is to promote progress in content-based retrieval from digital video via open, metrics-based evaluation by providing test collections and evaluation facilities. The TRECVID test collection that we used consisted of 124 English language news video programmes (CNN and ABC) recorded daily over a three-month period. From these thirty-minute news broadcasts we segmented 1,757 individual news stories, with each news story being represented by the actual video itself, a set of keyframes (video stills) extracted from the video data, as well as a text transcript of the spoken words from the video content. It is this text transcript for each story that forms the basis of the search and recommendation described in this paper.

The individual news stories were grouped into 124 sets of stories, one set for each day's news. To each user, a fifth of these sets (about 350 stories each) were allocated for interest judgment, where each user had to apply a simple binary judgment ('interesting' or 'not interesting') to each news story in their allocated set. This judgment reflected whether the user would be likely to playback the video content of that news story if recommended it on a mobile device. As can be seen from the baseline MAP figures of the 'base' row of Table 1, each user chose different numbers of stories as being interesting to them. The more interesting stories they chose, the less the opportunity for the persistent query to influence ranking to a useful degree.

3.2 Generating the Persistent Query

For each user, every web page visited during this experiment was automatically logged to a database and, with the help of a download process executed every five minutes and indexed into the user interest model. In this way, multiple visits to the same website (e.g. CNN.com) would result in multiple entries in the user interest model. Only English language pages were indexed in the user interest model. This is an important factor as typically a user will browse web pages (especially of news sites) in their own language and news sites would be considered an important source of evidence for the user interest model. The user interest model had stopwords removed using the SMART stopword list and stemmed using Porter's algorithm, both of which are standard document pre-indexing procedures from information retrieval.

From this user interest profile, a weighted set of important terms was generated. These important terms were calculated using the following term weighting formula, which incorporated a log-nomalised TF weight. Having (*r=R=0*) where *TF* is the frequency of occurrence of a term, *N* is the size of the collection and *n* is the number of news stories that term *i* occurs in, the formula is:

$$TFRW_i = \log(TF_i) \times \log((0.5/(N-n+0.5))/((n+0.5)\times 0.5)$$

Previous work on TV news story retrieval has shown the above formula to be effective in term selection for relevance feedback purposes [13]. These important terms were then used to generate a persistent query for each user. The number of important terms chosen to generate the query affects query processing time and was the subject of experimentation and is discussed below.

3.3 Recommendation of Content

Given the ranked list of important terms for each user, a top-ranked subset of these terms was chosen to generate the persistent query. The video archive supports search and retrieval over the video news story transcripts using the BM25 [14] text retrieval model with parameters optimised for news video data from previous work [13]. The result of such a query was a ranked list of news stories for any particular day's news, which was based on the similarity of the stories to the persistent query for that user. By examining the effectiveness of this list, compared to a list generated by the natural ordering of the news stories within the news programme, the benefit of using attention data can be evaluated. A second experiment to evaluate how effective the persistent query is at generating a ranked list of stories from the whole news archive for a user is also described.

3.3.1 Daily Content Recommendations

To evaluate how effective the persistent queries are at recommending news stories to users based on that day's news, we generated persistent queries of between 10 and 1,000 terms (see Table 1), and compared these results to a baseline system that ranked content only in the original broadcast order. The effectiveness values, represented by the Mean Average Precision measures, suggest the usefulness of recommendations (proportion of recommendations correct) for that user (see Table 1).

Table 1. Effectiveness of using the persistent query for news story recommendation

	User A	User B	User C	User D	User E
Size	7,573	2,672	321	441	439
Base	0.3726	0.5041	0.7162	0.4750	0.3998
10	0.2632	0.0000	0.0417	0.0202	0.0000
20	0.2562	0.0994	0.0575	0.0688	0.0000
30	0.2307	0.1043	0.0986	0.0797	0.1413
40	0.2172	0.1162	0.0986	0.1894	0.2730
75	**0.4456**	0.1706	0.1548	0.2474	0.3512
100	**0.4763**	0.2863	0.1702	0.2780	0.3725
150	**0.5583**	0.3315	0.1602	0.3299	0.3724
250	**0.5877**	0.4647	0.2558	0.4549	0.3952
500	**0.5888**	0.4825	0.3208	0.4509	**0.4375**
1000	**0.6034**	**0.5087**	0.3455	**0.5067**	**0.4758**

It can be seen that large queries are required to improve the quality of recommendations over the baseline. That said, for User A, who had the largest attention data profile (7,573), as few as 75 terms were enough to outperform the baseline. The relative performance of persistent queries of different size for User A is presented in Figure 1 below. The optimal number of terms for the persistent query from user A is about 300, with no improvement in performance evident with additional terms beyond that number.

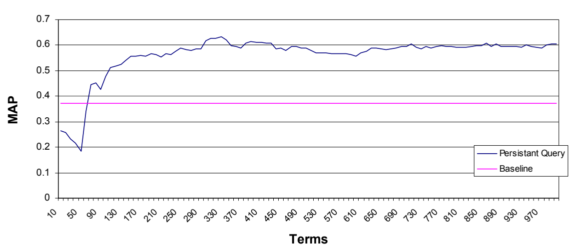

Figure 1. The Effect of Persistent query size on the quality of recommendation for User A.

For User B (see Table 1), the benefit is only noticeable when the top 1,000 terms are chosen for the persistent query. One likely explanation for the significant difference from User A is that not only did User B have a higher baseline MAP (0.5041) and smaller profile size, but also Users B, C & D generally browsed news websites in their native language, which is not English. Recall that only English-language web pages were added to the user interest model, thereby reducing the number of news-related useful terms from the user profile. However, it is interesting to note, that even with such a language mismatch issue, the use of attention data improves over baseline for two of the three Users (B & D). (In the failing case (User C) only 321 web pages were in the user profile, resulting in a base MAP of 0.7162, which gives little scope for the persistent query to have any effect.) User E represents a user who maintains a small user profile but browses web pages in the English language, and still 500 (or more terms) from the profile is shown to produce a clear improvement in recommendation effectiveness.

3.3.2 Whole Archive Recommendation

While daily recommendations of content are likely to be useful, we have extended this experiment to evaluate how effective the persistent query is at recommending content from a larger archive of pre-existing video data. For each user, this archive contains about 350 news stories.

Table 2. Precision values for 1,000 term persistent queries

	User A	User B	User C	User D	User E
P @ 5	1.0000	0.8000	0.8000	0.6000	0.6000
P @ 10	0.9000	0.8000	0.7000	0.5000	0.3000
P @ 15	0.8667	0.7333	0.6000	0.5333	0.3333
P @ 20	0.8500	0.6500	0.7000	0.4500	0.3500

As can be seen in Table 2, representing a user's interest by the persistent query is valuable when recommending content from the whole archive. Precision @ 5 (the number of interesting news stories ranked in the top 5) shows that employing a persistent query produces a ranked list of mostly interesting stories. Comparing these figures to a baseline for this experiment is not possible because a baseline run could only be time ordered and not rank ordered.

4. APPLICATION TO MOBILE DEVICES

We identify three access types for the mobile devices based on this work and our guidelines described earlier. They are content recommendation, archive search and content browsing.

4.1 Content Recommendation

The research presented in this paper illustrates one approach to recommending video content, the strength of which is that it relieves the user of having to enter queries into the mobile device, yet still presents a list of interesting news stories for any given day. While this would not replace a user's desire to browse a news story archive on a desktop device, the inherent interaction limitations of mobile devices are better suited to such a form of content recommendation. One result of this research was an implementation of a system as described above. Figure 2 illustrates an example recommendation for one user on a given day. The top news stories are visible in Figure 2 and Figure 3 shows a typical playback window for video content.

Figure 2. Personalised news story recommendations.

Figure 3. Playback of video on the mobile device.

4.2 Archive Search

Keyword search of the archive could be supported by allowing simple keyword query input on the mobile device. The focus of this research, however, was to generate recommendations automatically, not to address user's querying issues.

4.3 Content Browsing

Our previous research [4] illustrates how automatic linking between news story content can support mobile users in navigating by point-and-click through the hyperlinked archive of TV news stories. Additional research [13] illustrates the optimal feedback techniques for relevance feedback and story-story linkage within news story video archives.

5. CONCLUSIONS AND FUTURE WORK

We have shown the benefits of using implicit attention data mined from a user's web browsing activities to support effective recommendation of video content from a digital archive of TV news story videos. This relieves the user of the need to compose and submit a query for the interface to the mobile system. A typical usage scenario for this system would be where the digital library identifies content that would be of interest to a user and pushes this content across the network to the mobile device of the user, ideally at a time when the user is in a position to view the recommended content without being disrupted.

One likely criticism of our approach of maintaining persistent queries of up to 1,000 terms is the expense of processing such a large query, especially when a typical WWW search engine query is about 2 terms in length. However, this is not as expensive as one might think. Firstly, the full 1,000 terms are not processed against the text of the news stories, rather only the terms that occur in the news stories are processed. Secondly, for daily recommendation of news stories, the processing is usually done on a small number of TV news story content from the news programme on a given day. The whole archive is queried only when recommending content from all previous days' news.

We have presented our initial work in this area. Analysing the WWW browser activities of a user is only the first step in gathering user attention data. User emails, photos, documents, and other context data could be employed to build a far more complex persistent query for a user and could also be used to recommend much more than just news story content. Our future work involves a larger user study of more detailed attention data gathered over a longer period of time.

Also, we plan to study how to infer an automatic threshold for the number of news stories to push (again, relying on implicit user feedback) and the correct context for interrupting a user with the recommended news stories. Finally, privacy of user attention data is not an issue that we have addressed in this research, but we note that this is an extremely important aspect of any such system that utilised attention data.

6. ACKNOWLEDGEMENTS

This work is part-sponsored by the Norwegian Research Council Programmes No. 162349 and IKT-2010 and also supported by Science Foundation Ireland under Grant No. 03/IN3/I361.

7. REFERENCES

[1] Longoria, R. Designing mobile applications: challenges, methodologies, and lessons learned. In *Proceedings of HCI-2001*. New Orleans, Louisiana, 5-10 August 2001.

[2] Sacher, H., and Loudon, G. Uncovering the new wireless interaction paradigm. *ACM Interactions Magazine*, 9 (1), January 2002.

[3] Smyth, B. and Cotter, P. (2002) "Personalized Adaptive Navigation for Mobile Portals". In *Proceedings of the 15th European Conference on Artificial Intelligence* - Prestigious Applications of Intelligent Systems. Lyons, France. 21-26 July, 2002.

[4] Gurrin C, Smeaton A, Lee H, Mc Donald K, Murphy N, O'Connor N and Marlow S. Mobile Access to the Físchlár-News Archive. *Mobile HCI 2003, Workshop on Mobile and Ubiquitous Information Access*, Udine, Italy, 8 Sept. 2003.

[5] Marcus, A., Ferrante, J., Kinnunen, T., Kuutti, K., and Sparre, E. Baby faces: user-interface design for small displays. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems "Making the Impossible Possible"*, Los Angeles, CA, April 18-23, 1998.

[6] Rist, T. A perspective on intelligent information interfaces for mobile users. In *Proceedings of the International Conference on Human Computer Interaction (HCI 2001)*, New Orleans, LA, 5-10 August, 2001.

[7] T. Yan and H. Garcia-Molina. SIFT—A tool for wide-area information dissemination. In *Proceedings of 1995 USENIX Technical Conference*, New Orleans, LA, 16-20 Jan, 1995.

[8] M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of ACM SIGIR '94*, New York, NY, USA, 1994.

[9] M. Claypool, P. Le, M.Wased, and D. Brown. Implicit interest indicators. In *Proceedings of International Conterence on Intelligent User Interfaces '01*. New York, NY, 2001.

[10] Kaplan, C., Fenwick, J., and Chen, J. Adaptive hypertext navigation based on user goals and context. *User Modeling and User-Adapted Interaction*, 3 (3), 1993.

[11] Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M. and Boydell, O. Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. In *User Modeling and User-Adapted Interaction*, 14 (4), 2004.

[12] Over P, Kraaij W and Smeaton A.F. TRECVID 2005 - An Introduction. TRECVid 2005 – In *Proceedings of the Text REtrieval Conference TRECVID Workshop*, Gaithersburg, MD, 14-15 November 2005.

[13] Gurrin C, Johansen D and Smeaton A.F. Supporting Relevance Feedback in Video Search. In *Proceedings of ECIR 2006 - European Conference on Information Retrieval*, London, U.K., 10-12 April 2006.

[14] Robertson, S. E., Sparck Jones, K. Simple, proven approaches to text retrieval. *Tech. Rep.TR246*, University of Cambridge, 1997.