On User Modelling for Personalised News Video Recommendation

Frank Hopfgartner, and Joemon M. Jose

Department of Computing Science
University of Glasgow
Glasgow, United Kingdom
{hopfgarf,jj}@dcs.gla.ac.uk

Abstract. In this paper, we introduce a novel approach for modelling user interests. Our approach captures users evolving information needs, identifies aspects of their need and recommends relevant news items to the users. We introduce our approach within the context of personalised news video retrieval. A news video data set is used for experimentation. We employ a simulated user evaluation.

1 Introduction

Consumption of data has a centered influence on the development of our society, leading to the transformation from the industrial to the information age. Newspapers, television report broadcasts, the WWW and other media supply the society with a huge allowance of data, an expanding percentage of which is in digital format. However, facing this excessive resource of information sources might overwhelm information consumers. Hence, there is a need for providing personalised access to these data sources.

Sebe and Tian [15] discuss that for supplying personalised data from multimedia content, sophisticated study in diverse directions is required, encompassing the acquisition of user preferences and how to filter data by exploiting the user's profile. User profiles can be utilised to conceive a simplified form of the user which comprises his interests on general topics. Commercial retrieval systems incorporate such profiles, the most prominent being Google proposing iGoogle and Yahoo! proposing MyYahoo!. Query expansion is utilised to accumulate the user's interest and retrieval results are re-ranked to match their interests. These services depend on users specifically identifying preferences, a widespread approach in the text retrieval domain. By giving explicit feedback, users are forced to revise their need, which can be awkward when their data need is vague [19]. Furthermore, users are inclined to supply not sufficient feedback on which to base an adaptive retrieval algorithm [7]. Deviating from the procedure of specifically inquiring the user to rate the relevance of retrieval outcomes, the use of implicit feedback methods assists by learning user concerns unobtrusively. The major benefit is that users are not required to provide feedback.

A difficulty in user profiling is the users' evolving interest in various topics. What a client finds intriguing on day A might be absolutely uninteresting on day B, or even on the identical day. The following example shows the problem: Jane Doe is rarely interested in sports. However, throughout the 2008 Summer Olympics, she is enthralled by the euphoria exuded by the hunt for medals and pursues all reports associated to the event. After the games, her interest gradually abates again. How to capture and comprise this dynamic user interest is an unsolved problem. Moreover, a user can be involved in multiple topics, which might evolve over time. Instead of being interested in only one theme at one time, users can seek for diverse unaligned topics for example government, sports, entertainment or business.

In this paper, we present a personalised video recommender system which is designed to capture the user's evolving interest in multiple facets of news events. The system automatically processes the daily news bulletin from two national broadcasters and recommends news stories by unobtrusively profiling the user based on his interactions with the system. The evolving user interest is captured by incorporating the ostensive model of developing information need [4] and news aspects are identified by clustering the content of the profile. Parameters needed to fine tune the personalisation model are determined using a simulation-based evaluation scheme.

The paper is structured as follows: Section 2 provides an overview of related work. In Section 3, we introduce the research questions which arise from the related work. Section 4 introduces the architecture of the application. In Section 5, we present a simulated evaluation and discuss the outcome of the evaluation in Section 6.

2 Related Work

Our work builds on a number of research areas, including news video retrieval, personalised news delivery and techniques to capture evolving user needs. In the following, we introduce the state-of-the-art of these areas.

2.1 News Video Retrieval

Nowadays, almost every television channel has its own news bulletin, indicating that television is a widely accepted mass media to provide consumers with the latest news. Consequently, processing television news has been an important research area and much recent work, such as that represented by the TRECVID [17] research effort, aims to tackle the difficult problems of content based video retrieval. While some approaches have a particular emphasis on the system side, other research efforts are looking towards improving state-of-the-art video retrieval techniques from the user's point of view, such as the Open Video Project[1].

A number of conclusions can be drawn from these efforts: first of all, video retrieval is not as sophisticated as its textual counter part. The reason for this is the so-called "semantic gap" [18], the difference between the low-level representation of video and audio data, and the high-level semantics which the user

[1] http://www.open-video.org

would ideally like to associate with retrieved data. Furthermore, segmenting and indexing video is a challenge. Considering a news broadcast as a unit of retrieval will generate a result list containing whole video documents. A user must watch or browse through the whole video to finally find the information he wants, a demanding approach. Hence, it is necessary to split videos into smaller, semantically related, *segments* which should ease the access of the video data. In text retrieval, techniques have been developed to identify relevant sections of the text, e.g. [14] and to segment documents based on these sections. Hence, users can easily browse through short results to satisfy their information need. Boreczky et al. [3] argue that television news consists of a collection of *story units* which represent the different events being relevant for the day of the broadcast. An example story unit from the broadcasting news domain is a report on yesterday's football match, followed by another story unit about the weather forecast.

Indexing these segments, e.g. based on textual annotations or visual representations of the segments provides an easy access to the data collection. A challenging approach, however, is to identify those stories a user is really interested in. The problem will be introduced in the following section.

2.2 Personalised News Delivery

A common approach to capture a user's interests is user profiling. User profiling is the process of learning the user's interest over a long period of time. Using user profiles to create personalised online newspapers has been studied for a long time.

Chen and Sycara [5] join internet users during their information seeking task and explicitly ask them to judge the relevance of the pages they visit. Exploiting the created user profile of interest, they generate a personalised newspaper containing daily news. However, providing explicit relevance feedback is a cognitively demanding task and users tend not to provide much feedback [7].

Bharat et al. [2] create a personalised online newspaper by unobtrusively observing the user's web-browsing behaviour. Although their system is a promising approach to release the user from providing feedback, their main research focus is on developing user interface aspects, ignoring the sophisticated retrieval issues.

Smeaton et al. [16] introduced Físchlár-News, a news video recommender system that captured the daily evening news from the national broadcaster's main TV channel. The web-based interface of their system provides a facility to retrieve and to recommend news stories to the user based on his interest. According to Lee et al. [11], the recommendation of Físchlár-News is based on personal and collaborative explicit relevance feedback. The use of implicit relevance feedback as input has not been incorporated.

Though innovative, the above works suffers from some issues: user modelling is done through explicit means, which is not user friendly. Moreover, the users' evolving information need is not modelled properly. In the next section, we introduce open research questions which we tackle in this work.

3 Research Questions

As we have shown in the previous section, a number of research work have been performed in the fields of video retrieval, personalisation and adaptation of news content. In this work, we aim at addressing some of the shortcomings of the introduced work. Our main research questions are:

1. How can the user's evolving interest in various news topics be captured in a user profile?
2. Is it possible to identify and group stories in the profile based on similarities?
3. How can a user profile be exploited to recommend other news stories in the collection which match a user's interest?

We approach these questions by simulating users interacting with a news video recommender system. The system and its components are introduced in the following section.

4 System Architecture

Our video recommender system can be split into three conceptual parts: A data collection module, a retrieval backend, a profile module and an interface. The first two modules will be introduced in Section 4.1, the profiling module will be introduced in Section 4.2. Since we are not performing an interactive user study, we will not introduce a graphical user interface. For further details regarding state-of-the-art video retrieval/recommender interfaces, the reader is referred to [8].

4.1 Data Pre-Processing

We focus on daily news bulletins from two different broadcast channels (BBC One and ITV) which can be used for research purposes. Their programmes cover international, national (UK) and regional (Scotland) topics, which are usually presented by a single newsreader. Each bulletin has a running time of 30 minutes and is broadcasted every day from Monday till Friday. Broadcasting stations in the UK enrich their television broadcast with a closed caption (teletext) signal that provides televisual subtitles for the deaf. We use a colour histogram-based method to detect shot boundaries in our video files and detect example keyframes by calculating the average colour histogram for each shot and extract the frames within the shot which are closest to the average. These keyframes are then used to identify story boundaries as proposed by Goyal et al. [6]. Further, we capture the teletext by decoding the transmission signal of the broadcasting stations. Since teletext transcripts are manually created, the semantic quality of the text is reliable. We use the freely available LingPipe toolkit[2], at default settings (trained on the MUC-6 English corpus) to extract named entities (persons, organisations

[2] http://alias-i.com/lingpipe

and places) from each transcript and the POS tagger of the toolkit trained on the Brown corpus to extract nouns and foreign words from the transcript. Finally, we use OpenCalais[3], a web service provided by Thomson Reuters, to identify the category of each story. This service categories each story transcript into one or more of the following categories: Business & Finance, Entertainment & Culture, Health, Medical & Pharma, Politics, Sports, Technology & Internet and Other. We use the open-source retrieval engine MG4J[4] to index the processed data corpus with the transcripts, named entities and nouns being fields in the index. BM25 [13] is used to rank retrieval results.

4.2 Profiling

In this section, we introduce our approach of capturing the users' evolving interest and introduce the representation of this interest in the profile. Furthermore, we introduce our approaches of identifying semantically related topics in this profile and of exploiting the profile to recommend other news stories that match the users' interest.

Capturing user interest Our first research question in this work is to study how a user's evolving interest can be captured in a profile. As argued before, user profiling relies on users providing relevance feedback on documents in a collection. Since users tend not to provide such feedback, we aim at exploiting implicit user actions while interacting with the recommender system as an indicator of user's interest. When a user interacts with a result, he leaves a "semantic fingerprint" that he is interested in the content of this item to a certain degree. Several approaches have been studied to capture this fingerprint in a profile, the most prominent being the weighted keyword vector approach. In this approach, interests are represented as a vector of weighted terms extracted from the documents users were interested in. Each dimension of the vector space represents a term aligned with a weighting, i.e. the confidence weighting that this document is relevant. We believe however that this approach does not capture the users' interest appropriately, as a user is not interested in a single term from a story, but more in the content of the story. Therefore, we introduce a *weighted story vector* approach. The weighting of the story will be updated when the system submits a new weighted story to the profile starting a new iteration j. Hence, we represent the interaction I of a user i at iteration j as a vector of weights

$$\boldsymbol{I}_{ij} = \{W_{ij1}...W_{ijs}\}$$

where s indexes the story in the whole collection $|C|$. The weighting W of each story expresses the evidence that the content of this story matches the user's interest. In an ideal case, the higher the value of W, the closer this match is. In this work, we define a static value for each possible implicit feedback feature

[3] http://www.opencalais.com/
[4] http://mg4j.dsi.unimi.it/

provided a typical video recommender system as evaluated within TRECVID [17]:

$$W = \begin{cases} 0.1, & \text{when a user uses the highlighting feature} \\ 0.2, & \text{when a user starts playing a video} \\ 0.3, & \text{when a user browses through the keyframes} \\ 0.5, & \text{when a user expands a result} \end{cases}$$

Note that some of these features are independent, while others depend on a previous action (e.g. a video cannot be played without being clicked on). Furthermore, we represent the profile $\boldsymbol{P}_i$ of user i as a vector containing the story weight SW of each story s of the collection:

$$\boldsymbol{P}_i = \{SW_{i1} ... SW_{is}\}$$

The simplest approach to create a weighting for each story in the profile is to combine the weighting of the stories over all iterations. This approach is based on the assumption that the user's information interest is static, which is, however, not appropriate in a retrieval context. The user's information need can change within different retrieval sessions. Campbell and van Rijsbergen [4] propose in their ostensive model that the time factor has to be taken into account, i.e. by modifying the weighting of stories based on the iteration they were added to the user profile. They argue that more recent feedback is a stronger indicator of the user's interest than older feedback. In our profile, the story weight for each user i is the combination of the weighted stories s over different iterations j: $SW_{is} = \sum_j a_j W_{ijs}$. We include the ostensive evidence

$$a_j = \frac{1 - C^{-j+1}}{\sum_{k=1}^{j_{max}} 1 - C^{-k+1}} \tag{1}$$

to introduce an inverse exponential weighting which will give a higher weighting to stories which have been added more recently to the profile, compared to stories which were added in an earlier stage.

Identifying similar stories Tackling our second research question, we are interested in identifying similarities between stories that users interacted with. Therefore, we rely on hierarchical agglomerative clustering of stories with the highest story weight at the current iteration. Following Bagga and Baldwin [1], we treat the transcripts extracted from these stories as term vectors and compare them by cosine. Unlike their approach, we use the whole transcript rather than sentences linked by coreferences and use the square root of raw counts as our term frequencies rather than the raw counts. We use complete-link clustering since this approach results in more compact clusters. Moreover, we do not use inverse-document frequency normalisation since this value can be important for discremination. For tokenisation, we use standard filters (conversion to lower case, stop word removal and stemming). The numbers of clusters k is a parameter. Since each cluster should contain stories associated with an aspect of the user's interest, k should be equal to the number of different interests that a user has.

Recommending Related Stories Aiming at our third research question, we investigate how to recommend related stories by exploiting the user's profile. Assuming that each of the k clusters contains stories that cover one or more (similar) aspects of a user's interest, the content of each cluster can be exploited to recommend more documents belonging to that cluster. The simplest method is to create a search query based on the content of each cluster and to retrieve stories using this query. We identified three different sources that can be used to create a search query:

– *The most frequent terms from all stories within one cluster.* Using the most frequent terms, excluding stop words, is a common means in Information Retrieval systems to expand search queries.
– *The most frequent named entities from all stories within one cluster.* Named Entities such as "Barack Obama" or "The White House" can be used to avoid noise while automatically expanding search queries, since some of them are very specific.
– *The most frequent nouns and foreign words from all stories within one cluster.* Some words bear more information than other words, hence providing a higher "content load" than other terms. According to Lioma and Ounis [12], nouns and foreign words provide most information about the content of a document.

Extending the third research question, we focus on two different issues: *Which* of these sources provides the user with better recommendations and *what* is the optimal query length? We approach these questions using a simulation-based evaluation scheme, which will be introduced in the following section.

5 Evaluation

Most interactive video retrieval systems are evaluated in laboratory based user experiments. This methodology, based on the Cranfield evaluation methodology, is very helpful in getting valuable data on the behaviour of interactive search systems. However, it is almost impossible to test all the variables involved in an interaction and hence compromises are needed on many aspects of testing. Furthermore, such a methodology is inadequate for benchmarking various underlying adaptive retrieval algorithms. An alternative method for evaluating such systems is the use of simulations. In this section, we first introduce a usage scenario that illustrates the functions of the introduced personalised video recommender application. Then, we introduce our approach for mimicking user interactions, which is based on exploiting the log files of an interactive video retrieval experiment. Finally, we introduce assumptions and conditions underlying our simulation.

5.1 Usage Scenario

Imagine a user who is interested in four different topics: Entertainment, Business, Politics and Other. He registered with a new recommender system and provided the system with the information that he is interested in these four topics. For a period of two months, starting at the beginning of November 2008, he logs into the system, which provides him access to the latest news video stories of the day. On the system's graphical interface, he has a list of stories and four different buttons, each representing one of his interests. Clicking on these buttons will show him all of the stories of that day, which are related to this interest. He now interacts with the presented results related to each topic and logs off again. On each subsequent day, he logs in again and continues the above process.

5.2 Interaction Patterns

As we do not rely on real users in our user evaluation, we aimed to mimic the actions that potential users of our recommender system could perform. Ivory and Hearst [10] proposed exploiting real log files coming from users interacting with a similar interface. We therefore analysed the log files of a user study [9] performed on a video collection in order to identify statistical user behaviour patterns while interacting with a similar interface. Table 1 shows the probability values obtained from this study.

Independent Action type	**Probability**	**Dependent Action type**	μ	σ
$P(Click\|R); P(Click\|\neg R)$	0.8/0.3	Play Interval (3 sec interval)	2	3
$P(Tooltip\|R); P(Tooltip\|\neg R)$	0.8/0.4	Browsing through keyframes	0.25	1

Table 1. Probability and normal distribution measures for observed action types

The left column of the table shows the probabilities associated with standalone actions, i.e. actions that can be triggered independently from others. The action types shown in the right column are dependent on the actions listed in the left column, e.g. as a video cannot be played or navigated if the story has not been expanded by clicking on it in the result list. The most important action type is clicking on a story in the result list, as most of the other actions, except tooltiping, cannot be performed without previously clicking on the story. Once a story was expanded, the user can browse through the shots (represented by their keyframes). The video play duration was also monitored, by triggering a play interval action every three seconds of video playing. The navigation, browsing and play interval actions can be characterised by a Normal Gaussian distribution, with a mean value of μ and a typical deviation of σ. The click action is defined by the probability that a search result is expanded, conditioned to the result document being relevant or not to the task at hand, denoted as $P(Click|R)$ and $P(Click|\neg R)$, respectively.

5.3 User Simulation

The idea of user simulation is to mimic the possible actions of a real user. This in effect allows us to conduct various combinations of experiments and benchmark different algorithms. We applied this concept on the two month of news broadcasts from BBC One and ITV recorded between November 2008 and January 2009. These videos were segmented into more than 1600 single story units. 1000 of these stories were categorised into one or more of the six categories introduced in Section 4.1.

Following the given scenario, we simulate a user who is interested in four of these categories. We therefore set $k = 4$, resulting in four clusters which contain the stories associated with the same category. In a first step, we identified four example stories which were broadcast on the first day of the data collection and which belong to one of the given categories of the user's interest and stored them in the user profile. These four stories, one for each cluster, are used as the source for the first query expansion.

In a next step, we simulated a user logging in to the system. Since we want to evaluate how effective the clustering approach is fir identifying semantically related stories (our second research question), we had to assume that a user shows an equal interest to all categories over the whole duration of the study. Therefore, we simulated that a user clicked on *each* button in the interface which results in displaying recommendations from each cluster. In Section 5.2 we determined the probabilities of how users can interact with a video recommender system, therefore we used these actions and their probabilities to mimic users interacting with each recommendation list as illustrated in the scenario in Section 5.1. Finally, we simulated the user logging out again. This step was repeated for all days of the two months.

As argued in Section 4.2, we aimed to evaluate various parameters of our recommendation approach. We therefore repeated the whole simulation several times; for each of the three query sources and for different query lengths. The results of the simulations are presented in the next section.

6 Results

As we did not use a well-defined collection as is common in standard video evaluation campaigns like TRECVID, evaluating our simulation based on standard evaluation measures such as precision and recall or mean average precision is not possible. An appropriate measure to answer our first research question is to analyse the coherence of the produced clusters over all iterations. The more coherent each cluster is, the better the performance of the recommendation model for this run. Since we have classified each story into one of six broad categories using OpenCalais, we can use these categories as ground truth to evaluate the coherence of each cluster. For each iteration, we computed how many stories s (in percentage) in each cluster belonged to the same category C as the initial story within this cluster, hence forth referred to as C_s. This can give us an insight into the effectiveness of the clustering approach.

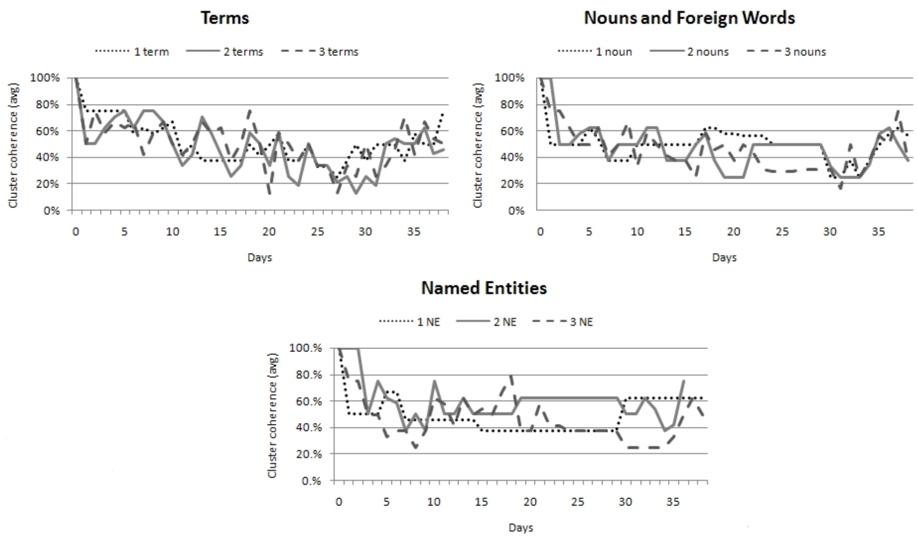

Fig. 1. Simulation results

Figure 1 shows the average of C_s across all four categories used in the simulation for different query length. In all cases, we observe a fast drop from the initial 100% cluster coherence. This is expected since our simulation does not include any judgements about relevance: Our simulated user interacts with every result, even with those which are obviously wrong, most of which a real user would not consider clicking on. This results in very noisy data, a drawback in our simulation. The figure shows similar peaks and tops for various runs, illustrating the effect of changing topics within our collection. The introduced profiling approach is hence sensitive towards evolving topics, answering our first research question. The results from the most frequent terms and most frequent nouns/foreign words seem to be rather random. The runs using named entity for recommendation however provide a more stable coherence, suggesting that named entities are a more promising source on which to base a recommendation technique on.

	1	**2**	**3**	**avg**
Terms	0.52	0.48	0.50	0.50
Nouns	0.50	0.48	0.46	0.48
NE	0.49	0.60	0.46	0.52

Table 2. Average coherence over all days

Table 2 shows the average coherence for each run over all days. The runs using two named entities provide the most coherent clusters, again suggesting that this feature is the most suitable for recommending similar results. This would answer our third research question, what is the best way to exploit the content of the clusters to recommend similar results.

The low percentage of coherent clusters can be partly explained by the high number of unclassified stories within the collection, which results in low values in our evaluation. Moreover, we considered only four out of six possible categories in order to achieve a more realistic user behaviour. This increases the chance to get noisy data though. Therefore, conclusions about the quality of the clusters are not possible, answering the second research question is hence difficult under a simulation-based evaluation scheme. Nevertheless, since all runs are effected by the same problem, the results indicate that our clustering approach can be used to identify similar stories. These findings should be confirmed by a succeeding user experiment where participants are asked to judge the quality of each cluster.

7 Conclusion

In this work, we investigated three different research questions: First of all, we wanted to analyse how a user's interests in various topics can be represented in a profile. Moreover, we were interested in identifying different aspects within this profile. Finally, we wanted to study if these aspects can be exploited to recommend similar stories to the user. We investigated these research questions using a *simulation-based* evaluation methodology: We analysed the log files of a previous user study and used the usage patterns to mimic users providing feedback while interacting with retrieval results of a video recommender system. This simulated implicit relevance feedback was used to identify stories of interest, which were then clustered accordingly. Each cluster served as a source for recommending similar results to the user.

Under an interactive evaluation scheme, we would have to rely on a large numer of participants interacting over a long period of time with our system to evaluate such settings. It is expensive and it would not be possible to benchmark various components. Therefore, we relied on a simulation-based evaluation. This scheme can be used to experiment with various parameters to identify the optimal settings for a personalised video recommender system. Nevertheless, our analysis also highlighted some limitations of a user simulation and argue that some findings should be confirmed by a real user study.

References

[1] A. Bagga and B. Baldwin. Entity-based cross-document coreferencing using the vector space model. In *Proceedings of the 17th international conference on Computational linguistics*, pages 79–85, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

[2] K. Bharat, T. Kamba, and M. Albers. Personalized, interactive news on the web. *Multimedia Systems*, 6(5):349–358, 1998.

[3] J. S. Boreczky and L. A. Rowe. Comparison of Video Shot Boundary Detection Techniques. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 170–179, 1996.

[4] 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.

[5] L. Chen and K. Sycara. WebMate: A personal agent for browsing and searching. In K. P. Sycara and M. Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pages 132–139, New York, 9–13, 1998. ACM Press.

[6] A. Goyal, P. Punitha, F. Hopfgartner, and J. M. Jose. Split and merge based story segmentation in news videos. In *ECIR'09 - 31st European Conference on Information Retrieval, Toulouse, France*, 4 2009. to appear.

[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, D. Vallet, M. Halvey, and J. M. Jose. Search trails using user feedback to improve video search. In *MM'08 - Proceedings of the ACM International Conference on Multimedia, Vancouver, Canada*, pages 339–348, 10 2008.

[10] M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, 33(4):470–516, 2001.

[11] H. Lee, A. F. Smeaton, N. E. O'Connor, and B. Smyth. User evaluation of Físchlár-News: An automatic broadcast news delivery system. *ACM Trans. Inf. Syst.*, 24(2):145–189, 2006.

[12] C. Lioma and I. Ounis. Examining the content load of part of speech blocks for information retrieval. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 531–538, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[13] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *Proceedings of the Third Text REtrieval Conference (TREC 1994), Gaithersburg, USA*, 1994.

[14] G. Salton, J. Allan, and C. Buckley. Approaches to passage retrieval in full text information systems. *ACM SIGIR conference on research and development in Information Retrieval*, pages 49–58, 1993.

[15] N. Sebe and Q. Tian. Personalized Multimedia Retrieval: The New Trend? In *MIR '07*, pages 299–306, New York, NY, USA, 2007. ACM.

[16] A. F. Smeaton. The Físchlár Digital Library: Networked Access to a Video Archive of TV News. In *TERENA Networking Conference 2002, Limerick, Ireland, 3-6 June 2002*, 2002.

[17] 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.

[18] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.

[19] 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.