

Interactive News Video Recommendation: An Example System

Frank Hopfgartner
International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA, 94704
fh@icsi.berkeley.edu

ABSTRACT

This position paper introduces a recommender system which has been developed to study research questions in the field of news video recommendation and personalization. The system is based on semantically enriched video data and can be seen as an example system that allows research on semantic models for adaptive interactive systems.

1. INTRODUCTION

In recent years, the amount of multimedia content available to users has increased exponentially. This phenomenon has come along with (and to much an extent is the consequence of) a rapid development of tools, devices, and social services which facilitate the creation, storage and sharing of personal multimedia content. A new landscape for business and innovation opportunities in multimedia content and technologies has naturally emerged from this evolution, at the same time that new problems and challenges arise. In particular, the hype around social services dealing with visual content, such as YouTube or Dailymotion has led to a rather scattered publishing of video data by users worldwide [8]. Due to the sheer amount of large data collections, there is a growing need to develop new methods that support the users in searching and finding videos they are interested in.

Video retrieval is a specialization of information retrieval (IR), a research domain that focuses on the effective storage and access of data. In a classical information retrieval scenario, a user aims to satisfy their information need by formulating a search query. This action triggers a retrieval process which results in a list of ranked documents, usually presented in decreasing order of relevance. The activity of performing a search is called the information seeking process. A document can be any type of data accessible by a retrieval system. In the text retrieval domain, documents can be textual documents such as emails or websites. Image documents can be photos, graphics or other types of visual illustrations. Video documents consist of a set of audio-visual

signals and accompanying metadata. The audio-visual features can be described by low-level feature descriptors, the main description standard being MPEG-7.

Retrieving videos using low-level features is, due to the Semantic Gap [18], a challenging approach. An analysis of state-of-the-art research on video retrieval indicates that content-based video retrieval performance is still far away from their textual counterparts [7]. An interesting approach to narrow this performance gap is to further enrich video documents using external data sources, called metadata. Blanken et al. [4] list three types of metadata: (1) Descriptive Data, (2) Text Annotations and (3) Semantic Annotation. All approaches aim to provide annotations in textual form that allow to bridge the Semantic Gap. Fernández et al. [9], for instance, have shown that ontology-based search models that exploit semantic annotations can outperform classical information retrieval models at a web scale. The advantage of these models is that external knowledge is used to set the content into their semantic context.

In [10], we introduced a news video recommender system which relies on such semantic annotations. The system captures daily broadcasting news, and segments the bulletins into semantically related news stories. DBpedia is exploited to set these stories into context. DBpedia is a structured representation of Wikipedia [2]. This semantic augmentation of news stories is used as the backbone of our news video recommendation. Our first hypothesis was that implicit relevance feedback can be used to create appropriate long-term user profiles. Implicit relevance feedback refers to user interactions that are performed implicitly during a search session, such as clicking a search result or spending time to read/view a document. We introduced an implicit user modeling approach which automatically captured users' evolving information needs, representing interests in a dynamic user profile. Another research question was to study whether the selection of concepts in a generic ontology can be used for accurate news video recommendations. Therefore, we introduced our approach of exploiting DBpedia to set concepts of news stories into their semantic context. As our evaluation indicates, semantic recommendations can successfully be employed to improve the recommendation quality.

While we evaluated within this work the underlying personalization technique, which takes advantage of an ontology, the impact of the adaptive presentation of the recommendations and search results, i.e. the interface design, has not

been evaluated yet. Given a well-evaluated backend which relies on Semantic Web technologies, we argue in this position paper that the introduced personalization system can be seen as an exemplar system which allows for studying the research questions that are within the scope of this workshop. After introducing the research domain in Section 2, we illustrate in Section 3 how users can use the system to receive frequent news video recommendations that match their personal interests. In Section 4, we introduce the interface of prior mentioned system, which is required to visualize semantically enriched video data. Section 5 discusses how this system can be used as an example to study semantic models for adaptive interactive systems.

2. SEMANTIC NEWS VIDEO RECOMMENDATION

When interacting with a video retrieval system, users express their information need in search queries. The underlying retrieval engine then retrieves relevant results to the given queries. A necessary requisite for this IR scenario is to correctly interpret the users' information need. As Spink et al. [19] indicate though, users very often are not sure about their information need. One problem they face is that they are often unfamiliar with the data collection, thus they do not exactly know what information they can expect from the corpus [17]. Further, Jansen et al. [12] have shown that video search queries are rather short, usually consisting of approximately three terms. Considering these observations, it is hence challenging to satisfy users' information needs, especially when dealing with ambiguous queries. Triggering the short search query "Victoria", for example, a user might be interested in videos about cities called Victoria (e.g. in Canada, United States or Malta), landmarks (e.g. Victoria Park in Glasgow or London), famous persons (e.g. Queen Victoria or Victoria Beckham) or other entities called Victoria. Without further knowledge, it is a demanding task to understand the users' intentions. Interactive information retrieval aims at improving the classic information retrieval model by studying how to further engage users in the retrieval process, in a way that the system can have a more complete understanding of their information need. Thus, aiming to minimize the users' efforts to fulfill their information seeking task, there is a need to personalize search. In a web search scenario, Mobasher et al. [14] define personalization as "any action that tailors the Web experience to a particular user, or a set of users". Another popular name is adaptive information retrieval, which was coined by Belew [3] to describe the approach of adapting, over time, retrieval results based on users' interests.

Most of the approaches that follow the interactive information retrieval model are based on relevance feedback techniques [17]. Relevance feedback (RF) is one of the most important techniques within the IR community. An overview of the large amount of research focusing on exploiting relevance feedback is given by Ruthven and Lalmas [16]. The principle of relevance feedback is to identify the user's information need and then, exploiting this knowledge, adapting search results. Rocchio [15] defines relevance feedback as follows: The retrieval system displays search results, users provide feedback by specifying keywords or judging the relevance of retrieved documents and the system updates the results by incorporating this feedback. The main benefit

of this approach is that it simplifies the information seeking process, e.g. by releasing the user from manually reformulating the search query, which might be problematic especially when the user is not exactly sure what they are looking for or does not know how to formulate their information need. Two types of relevance feedback exist: explicit and implicit feedback. While explicit RF models rely on users permanently providing relevance information about documents they retrieved, implicit RF models rely on automatically mining user interaction data. The main advantage is that this approach delivers the user from providing explicit feedback.

Most personalization services rely on users explicitly specifying preferences. However, users tend not to provide constant explicit feedback on what they are interested in. In a long-term user profiling scenario, this lack of feedback is critical, since feedback is essential for the creation of such profiles. Considering that each interface feature is designed to allow users to either retrieve or explore document collections, we hypothesized in [10] that the users' interactions with these features can be exploited as implicit relevance feedback. We introduced a news video recommender system which automatically generates personalized multimedia news that cover topics of the users' long-term interests.

Defining the technical conditions for such recommender systems, we argued that the creation of a private news video collection is required, consisting of up-to-date news bulletins from different broadcasting stations. Further, we argued that semantic web technology can be exploited to link concepts in the news broadcasts and suggested a categorization of stories into broad news categories. From a user profiling point of view, these links and categories can be of high value to recommend semantically related transcripts, hence creating a semantic-based user profile. For example, a user could show interest in a story about the sunset at the Greek island Santorini. The story transcript might contain the following sentence:

"This is Peter Miller, reporting live from Santorini, Greece, where we are just about to witness one of the most magnificent sunsets of the decade. [...]"

If the same user enjoys travel with emphasis on warm Mediterranean sites, he/she might also be interested in a report about the Spanish island Majorca. For example, imagine the following story:

"Just as every year, thousands of tourists enjoy their annual sun bath here in Majorca. [...]"

An interesting research question is how to identify whether this story matches the user's interests. Lioma and Ounis [13] argue that the semantic meaning of a text is mostly expressed by nouns and foreign names, since they carry the highest content load. Indeed, most adaptation approaches rely on these terms to personalize retrieval results, e.g. by performing a simple query expansion. The two example stories, however, do not share similar terms. A personalization

technique exploiting the terms only would hence not be able to recommend the second story. However, linking the concepts of the transcripts using DBpedia reveals the semantic context of both stories. It becomes evident that both stories are about two islands in the Mediterranean Sea. Exploiting this link could hence satisfy the user's interest in warm Mediterranean Sites. We therefore proposed to set news broadcasts into their semantic context by exploiting the large pool of linked concepts provided by DBpedia.

Having established a semantically annotated data collection, the recommender system can be operated on a regular basis to retrieve news stories that match the user's interests. In the next section, we illustrate a typical use-case that illustrates the use of the exemplar system.

3. USE-CASE SCENARIO

In the previous section, we provided a brief summary of the research challenges that have been tackled in [10]. Users can interact with this system on a regular basis, e.g. over several weeks, to satisfy their information need, allowing for longitudinal user studies where the system can be evaluated. The following example depicts a typical use-case scenario:

“Imagine a user who is interested in multiple news topics. They registered with a news recommender system with a unique identifier. For a period of several months, they log into the system, which provides them access to the latest news video stories of the day. On the system's graphical interface, they have a list of the latest stories which have been broadcast on two national television channels. They now interact with the presented results and logs off again. On each subsequent day, they log in again and continue the above process.”

In this scenario, a user frequently uses the system to gather latest news. The interface has been designed to adapt its content based on users' personal interests by employing the semantic context of the data collection. Each time, he/she interacts with the video documents which have been displayed by the graphical user interface, he/she leaves a “semantic fingerprint” of their interests. Based on this fingerprint, more video documents are identified by exploiting the semantic link between the video documents in the collection. Hence, each time the user interacts with retrieval results, other related videos are identified and displayed. A long-term user study focusing on evaluating the performance of different recommendation techniques has been introduced in [11].

While this evaluation is focused on the recommendation techniques, a thorough evaluation of the interface has not been done yet. An overview over the interface is given in the next section.

4. INTERFACE DESIGN

Figure 1 shows a screenshot of the adaptive news video retrieval interface which was used within the study. It can be split into three main areas: Search queries can be entered

in the search panel on top, results are listed on the right side and a navigation panel is placed on the left side of the interface. When logging in, the latest news will be listed in the results panel. Search results are listed based on their relevance to the query. Since we are using a news corpus, however, users can re-arrange the results in chronological order with latest news listed first. Each entry in the result list is visualized by an example key frame and a text snippet of the story's transcript. Keywords from the search query are highlighted to ease the access to the results. Moving the mouse over one of the key frames shows a tool tip providing additional information about the story. A user can get additional information about the result by clicking on either the text or the key frame. This will expand the result and present additional information including the full text transcript, broadcasting date, time and channel and a list of extracted named entities. In the example screenshot, the third search result has been expanded. The shots forming the news story are represented by animated key frames of each shot. Users can browse through these animations either by clicking on the key frame or by using the mouse wheel. This action will center the selected key frame and surround it by its neighboring key frames. The user's interactions with the interface are exploited to identify multiple topics of interests. On the left hand side of the interface, these interests are presented by different categories, i.e. those news categories that the user showed interest in during previous search sessions.

Summarizing, the interface provides access to different news categories in which the user showed interest in. These interests can adapt over time, i.e. when a user shows interest in a certain news aspect right now, this aspect might already be irrelevant in a few days. Imagine, for example, a user who has shown high interest in any news regarding the FIFA Soccer World Cup. Just a few days after the end of the tournament, the user's interest might drop to a minimum again. Our interface serves this evolving need by automatically updating the categories in which the user showed the most interest in during the last sessions. The evolving interest is modeled by applying the Ostensive Model [6], which provides a decay function that aligns a higher weighting to more recent user interests.

5. DISCUSSION AND CONCLUSION

Above description reveals that the interface has been designed to visualize news videos that match users' interests. The categorization of these interests is highly user-centric. The interface adapts its content, i.e. both categories on the left hand side and news videos on the right hand side based on the users' previous interactions. Even though the recommendation technique relies on interlinked data, the interface itself does not support filtering or browsing the data accordingly.

As mentioned before, this constraint is due to the different focus of the research, which was aiming at studying recommendation techniques rather than adaptive interface designs. Nevertheless, given the support of semantically enriched video data, we argue that the system can be seen as an example framework which enables to study such interface features. Example improvements include visualizing story interlinking by using a hyperbolic tree, as has been

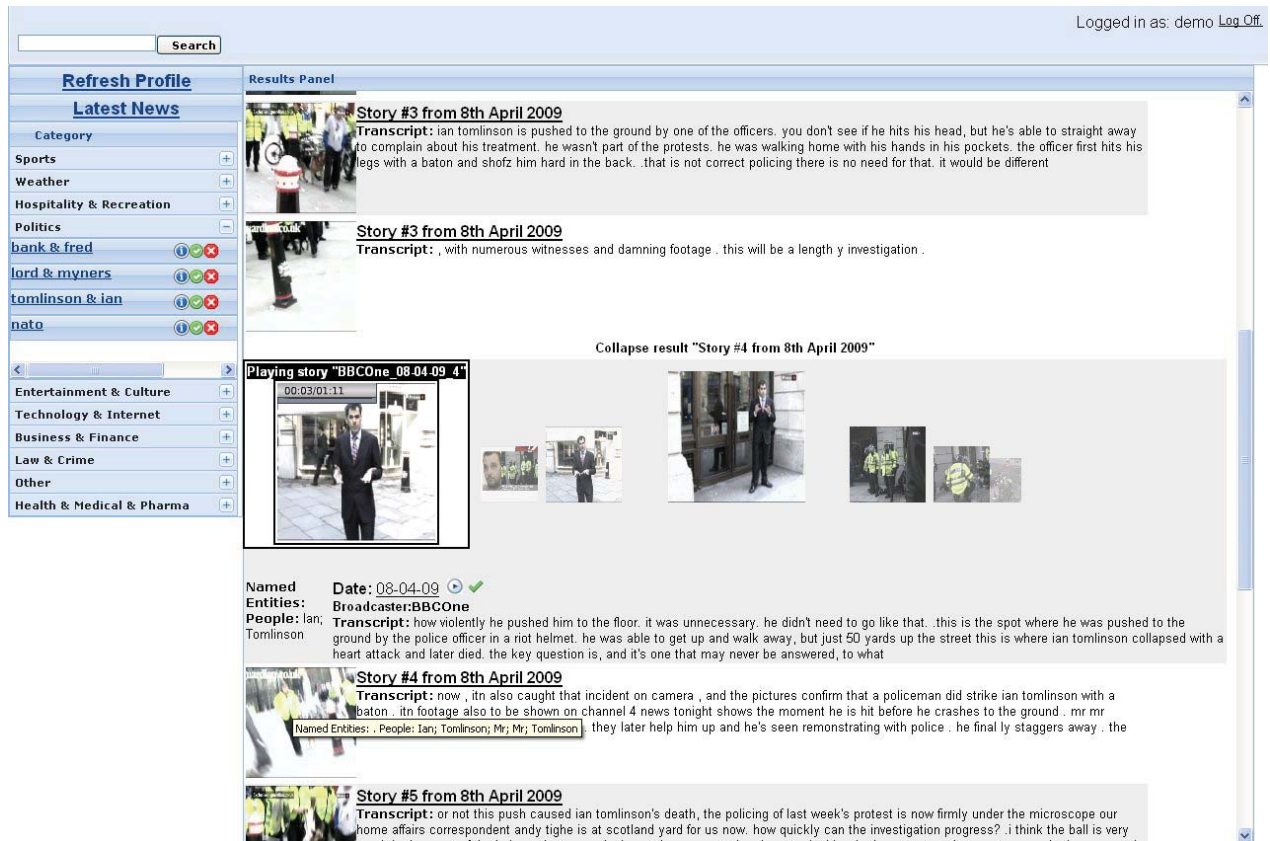


Figure 1: News Video Recommender Interface

introduced by Bürger et al. [5]. In their Smart Content Factory, each document in the index has been enriched with semantic information, i.e. places mentioned in the transcript are matched with a generic geography thesaurus. Such tree would allow users to browse the video collection based on the semantic content of each video. Another improvement could be to provide thesaurus supported query auto-completion features as shown by Amin et al. [1]. This would allow users to get an idea about the collection based on the query suggestions.

Acknowledgment

The author was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD).

6. REFERENCES

- [1] A. Amin, M. Hildebrand, J. van Ossenbruggen, V. Evers, and L. Hardman. Organizing suggestions in autocompletion interfaces. In *ECIR'09: Proceedings of the 31st European Conference on IR Research, ECIR 2009, Toulouse, France*, pages 521–529, 2009.
- [2] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. G. Ives. DBpedia: A Nucleus for a Web of Open Data. In *Proc. 6th Int. Semantic Web Conf.*, pages 722–735. Springer Berlin / Heidelberg, 11 2007.

- [3] R. K. Belew. Adaptive information retrieval: using a connectionist representation to retrieve and learn about documents. *SIGIR Forum*, 23(SI):11–20, 1989.
- [4] H. M. Blanken, A. P. de Vries, H. E. Bok, and L. Feng. *Multimedia Retrieval*. Springer Verlag, Heidelberg, Germany, 1 edition, 2007.
- [5] T. Bürger, E. Gams, and G. Güntner. Smart content factory: assisting search for digital objects by generic linking concepts to multimedia content. In *Proc. HT*, pages 286–287. ACM, 2005.
- [6] I. Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. In *Proc. Library Science*, pages 251–268, 1996.
- [7] M. G. Christel. Establishing the utility of non-text search for news video retrieval with real world users. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 707–716, New York, NY, USA, 2007. ACM.
- [8] S. J. Cunningham and D. M. Nichols. How people find videos. In *Proc. 8th ACM/IEEE-CS Joint Conference on Digital libraries*, pages 201–210, New York, NY, USA, 2008. ACM.
- [9] M. Fernández, V. López, M. Sabou, V. Uren, D. Vallet, E. Motta, and P. Castells. Using TREC for cross-comparison between classic IR and ontology-based search models at a Web scale. In

- SemSearch'09*, 4 2009.
- [10] F. Hopfgartner and J. M. Jose. Semantic user modelling for personal news video retrieval. In *MMM*, pages 336–346, 2010.
 - [11] F. Hopfgartner and J. M. Jose. Semantic user profiling techniques for personalised multimedia recommendation. *Multimedia Systems*, 16(4):255–274, 2010.
 - [12] B. J. Jansen, A. Goodrum, and A. Spink. Searching for multimedia: analysis of audio, video and image web queries. *World Wide Web*, 3(4):249–254, 2000.
 - [13] C. Lioma and I. Ounis. Examining the Content Load of Part of Speech Blocks for Information Retrieval. In *ACL'06: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia*, 2006.
 - [14] B. Mobasher, R. Cooley, and J. Srivastava. Automatic personalization based on web usage mining. *Communications of the ACM*, 43(8):142–151, 2000.
 - [15] J. J. Rocchio. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART retrieval system: experiments in automatic document processing*, pages 313–323, Englewood Cliffs, USA, 1971. Prentice-Hall.
 - [16] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2):95–145, 2003.
 - [17] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Readings in information retrieval*, pages 355–364, 1997.
 - [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. on Pattern Analysis and Machine Intelligence*, 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.