Frank Hopfgartner

Multimedia Retrieval and Adaptive Systems

Citation details
Hopfgartner, Frank (2013): Multimedia Retrieval and Adaptive Systems. In: Zander, J. & Mosterman, P. (Eds.). Computation for Humanity. CRC Press, pp. 205–234. https://doi.org/10.1201/9781315216751.

This is an Accepted Manuscript of a book chapter published by Routledge/CRC Press in Computation for Humanity: Information Technology to Advance Society on 13 November 2013, available online: http://www.routledge.com/9781439883273 or http://www.crcpress.com/9781439883273.

Chapter

Multimedia Retrieval and Adaptive Systems[1]

Frank Hopfgartner
TU Berlin, TEL 14, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

1.1 Introduction

Recently, The Guardian, one of Britain's most popular daily newspapers published an online article[2], recognising the fifth anniversary of the video sharing portal YouTube[3]. YouTube is at the forefront of a recent development that, in 2006, convinced the renowned Time magazine to dedicate their Person of the Year award to "You". "You" represent the millions of people that started to voluntarily generate (user-generated) content, e.g. in Wikipedia, Facebook and, of course, YouTube. More and more people do not only actively consume content, they have also started to create their own content. Thus, we are observing a paradigm change

[1]This is an Accepted Manuscript of a book chapter published by Routledge/CRC Press in Computation for Humanity: Information Technology to Advance Society on 13 November 2013, available online: http://www.routledge.com/9781439883273 or http://www.crcpress.com/9781439883273

[2]`http://www.guardian.co.uk/technology/2010/apr/23/youtube-five-years-on`, accessed on: 7 September 2011

[3]`http://www.youtube.com/`, accessed on: 7 September 2011

from the rather passive information consumption habit to a more active information search. Tim Berners-Lee, credited for inventing the World Wide Web, is convinced that eventually this development will completely change the way in which we engage information. During a discussion following his keynote speech "The Web at 20" at the BBC documentary "Digital Revolution", he envisioned that:

> *"As a consumer, if I have an internet connection I should be able to get at, and pay for if necessary, anything that has ever been broadcast. So Im looking forward to when the BBC, for example, [offers] a complete random access library so that I can follow a link, so that somebody can tweet about some really, really cool thing or some fun show, or some otherwise boring show, and I can follow a link directly to that. Whether its pay or free, its per view, and I get it by following a link, one way or another. I wont be searching channels. I think the concept of a channel is going to be history very quickly on the internet. It's not relevant."*

In this chapter, we illustrate how users can be assisted in exploring such digital video libraries. We first provide a brief introduction of basic concepts of video retrieval, a research area which has drawn more and more attention since the rise of YouTube. Arguing that personalisation techniques can be employed that assist the users in identifying videos they are interested in, we then survey different approaches to gather user information and introduce techniques to exploit this information to recommend video documents that match users' personal interests.

1.2 Basic Concepts of Video Retrieval

Video retrieval is a specialisation 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 accessable 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 are any type of moving images. In Section 1.2.1, we introduce the structure of a typical video document. A repository of documents that is managed by an IR system is referred to as a *document collection*. In Section 1.2.2, we argue that video documents should be analysed and processed first in order to ease access to their content. Preferably, retrieved documents are ranked in accordance to their relevance to the user's information need. In Section 1.2.3, we discuss this retrieval and ranking process further. Aiming to visualise retrieval results for the user, graphical user interfaces are required that allow the user to input their information need and to inspect the retrieved results, thus to access the document collection. Section 1.2.4 introduces challenges of graphical user interfaces in the video retrieval domain.

1.2.1 Structure of Video Documents

Computers serving multimedia and other devices are going to change the handling of videos completely. Films are consistently broadcast, recorded and stored in *digital* form. A video is a sequence of still images, accompanied by an audio stream. Classical digital video standards are the MPEG-1 and MPEG-2 formats. They were released by the Motion Pictures Expert Group (MPEG). Videos in MPEG-1 format have a resolution of 352×240 pixels and a bit rate of up to 1.5Mbps with a framerate of 25 frames per second. MPEG-1 achieves a high compression rate by the use of motion estimation and its compensation between frames. Considering the resulting low quality of the encoded videos, MPEG-1 videos are often compared to old fashioned VCR recordings. The newer MPEG-2 video format is used to encode videos in DVD quality. It is the standard used for digital television (DVB-T, DVB-S, DVB-C) [122].

Another ISO standard that has been defined by MPEG is MPEG-7. Its purpose is to provide a unified standard for the description of multimedia data using meta information. Within this standard, various descriptors have been defined to describe visual content, including colour descriptor, shape descriptor, motion descriptor, face descriptor and textual descriptor [78]. An overview of MPEG-7 descriptors is given by Manjunath et al. [78].

Besides the video's own text and audio-visual data streams, video documents can be enriched with additional data, the so-called metadata. Blanken et al. [13] survey various types of metadata: (1) A description of the video document, (2) textual annotation and (3) semantic annotations. All approaches aim to provide annotations in textual form that allow to bridge the Semantic Gap. The remainder of this section provides a brief overview over these metadata types. For a more detailed description, the reader is referred to Blanken et al. [13].

Descriptive Data Descriptive data provides valuable information about the video document. Examples are the creation date, director or editor, length of the video and so on. A standard format for descriptive data is called Dublin Core [123]. It is a list of data elements designed to describe resources of any kind. Descriptive metadata can be very useful when documents within the video collection shall be filtered based on certain document facets. Think, for instance, of a user who wants to retrieve all video documents that have been created within the last month, or all videos from one specific director.

Text Annotations Text annotations are textual descriptions of the content of video documents. More recent state-of-the-art online systems, such as YouTube and Dailymotion[4], rely on using annotations provided by users to provide descriptions of videos. However, quite often users can have very different perceptions about the same video and annotate that video differently. This can result in synonymy, polysemy and homonymy, which makes it difficult for other users to retrieve the same video. It has also been found that users are reluctant to provide an abundance of annotations unless there is some benefit to the user [45]. Van Zwol et al. [117] approach this problem by transferring video annotation into an online gaming scenario.

Apart from manually created content, an important source for textual annotations are speech transcripts. Huang [56] argues that speech contains most of the semantic information that can be extracted from audio features. Further, Chang et al. [25] argue that text from speech data plays an important role in video analysis. In literature, the most common

[4] `http://www.dailymotion.com/`, accessed on: 7 September 2011

text sources are teletext (also called closed-caption), automatic speech recognition (ASR) transcripts and optical character recognition (OCR) output. Considering that textual annotations can be a valuable source for IR systems aiming to retrieve the video documents, various approaches have been studied to automatically determine textual annotations. A survey of state-of-the-art approaches is given by Magalhães and Rüger [76]. More recent examples include [110, 73, 72, 91, 121].

Semantic Annotations Another type of annotations are semantic annotations. The idea is here to identify concepts and define their relationship between each other and the video document. Concepts can hence set the content of video documents into a semantic context. This is especially useful for semantic retrieval approaches. We will survey such approaches in Section 1.2.3. The previously mentioned MPEG-7 standard allows for describing multimedia documents and their semantic descriptions. Promising extensions include COMM (Core Ontology for Multimedia), an ontology introduced by Arndt et al. [6]. Ontologies are "content specific agreements" on vocabulary usage and sharing of knowledge [43]. Other metadata models include [39, 115, 11], who aim to enrich interactive television broadcast data with additional information by combining existing standards. All approaches build hence upon similar ideas.

Semantic annotations can either be derived from textual annotations or from the videos' low-level features, i.e. by identifying high-level concepts. Magalhães and Rüger [76] provide a survey on state-of-the-art methodologies to create semantic annotations for multimedia content. They distinguish between three semantic annotation types: (1) hierarchical models, (2) network models and (3) knowledge-based models. *Hierarchical* models aim to identify hierarchical relations or interdependencies between elements in an image or key frame. Examples include [8, 14]. *Network models* aim to infer concepts given the existence of other concepts. Surveyed approaches are [67, 52]. The third approach, *knowledge-based* models relies on prior knowledge to infer the existence of concepts. Bürger et al. [18], for example, enrich news video data with a thesaurus of geographic names. Therefore, they determine location names within the news reports' transcripts and map these with their thesaurus. Further, they identify thematic categories by mapping terms in the transcript with a controlled vocabulary. A similar approach is introduced by Neo et al. [85], who use the WordNet lexical database [41] to semantically enrich news video transcripts. Even though their approaches allow linking of related news videos, the main problem of their approaches is text ambiguity. Other examples include [112, 103].

1.2.2 Video Segmentation

As we will show in Section 1.2.3, IR systems index documents and retrieve these documents based on their relevance to a search query. This approach is problematic, however, if a document contains short paragraphs which are highly relevant to the query, while the majority of the document is not relevant at all. Classical ranking algorithms, e.g. based on the document's term frequency, will result in a low ranking of this document. A promising approach to tackle this problem in the text domain is to split documents into shorter *passages* (e.g. [97]). Various potential advantages arise when considering these passages as unit of retrieval results. First of all, individual passages will be ranked higher than documents which contain the corresponding passages. Consequently, retrieval performance increases. Second, ranking problems due to variable document lengths is minimised, assuming that passages have a similar length. Third, short passages are easier to assess for the user than long documents. Users can easily browse through short results to search for their information need. The same problems apply to videos in the news video domain. However, due to the different nature of news video documents, successful passage retrieval approaches cannot easily be adopted. This section introduces typical segmentation of news videos.

Shot Segmentation

The atomic unit of access to video content is often considered to be the video *shot*. Monaco [81] defines a shot as a part of the video that results from one continuous recording by a single camera. It hence represents a continuous action in time and space in the video. Especially in the context of professional video editing, this segmentation is very useful. Consider for example a journalist who has to find shots in a video archive that visualise the context of a news event. Shot segmentation infers shot boundary detection, since each shot is delimited by two consecutive shot boundaries. Hanjalic [48] provide a comprehensive overview on issues and problems involved in automatic shot boundary detection. A more recent survey is given by Smeaton et al. [104].

News Story Segmentation

A more consumer-oriented approach is to segment videos into semantically coherent sequences. For instance, sports fans want to watch specific highlights of a game rather than short shots depicting only parts of this highlight.

In the news video domain, such coherent sequences are news stories. News stories are commonly seen as segments of a news broadcast with a coherent news focus which contain at least two independent declarative clauses. News bulletins consists of various continuous news stories, such as reports about political meetings, natural disasters or sports events. Chaisorn and Chua [24] argue that the internal structure of news stories depends on the producer's style. While some stories consist of anchor person shots only, often with a changing background image, other stories can consist of multiple different shots, e.g. other anchor persons, graphics or animations, interview scenes or shots of meetings.

News story segmentation is essentially finding the boundaries where one story ends and the other begins. Various text-based, audiovisual-based and combinations of all features have been studied to segment news videos accordingly. Detailed surveys are given by Arlandis et al. [5] and Chua et al. [27].

1.2.3 Document Representation

Video retrieval systems aim to retrieve relevant video documents that match the users' information need. Various conditions need to be fulfilled to enable this process. Snoek et al. [107] sketched a common framework that applies for most state-of-the-art video retrieval engines. Their framework can be divided into an indexing engine and a retrieval engine.

The first component involves the indexing of the video data, so that documents can be retrieved that match the users' information need. With respect to current systems, this indexing can be incorporated on a visual, textual and semantic level. Most video retrieval engines store their indexed data collection in a database. Techniques are required to match both information need and the video collection. These tasks are fulfilled by the retrieval engine. In the remainder of this section, we briefly introduce state-of-the-art approaches that address these issues.

Video Indexing

Video indexing is the backbone of all video retrieval engines. Indexing approaches aim to develop effective and efficient methodologies for storing, organising and accessing video contents. As we have shown in Section 1.2.1, a video document consists of several modalities, e.g. a video document is made up of audio tracks, visual streams and different types of annotations. Thus, video indexing has to take numerous modality features into consideration. Moreover, these features are of various nature. Video indexing techniques can be split into three main categories: content-based indexing, text-based indexing and semantic indexing. Note that we will not focus on video indexing within this chapter, since it is out of scope of this work. Hopfgartner et al. [55] present two multimedia indexing approaches that aim to address the two main challenges in content-based indexing: (1) the high dimensional feature space of multimedia data and (2) the variable character of feature dimensions, i.e. boolean and multi-value features. A survey on content-based video indexing is given by Baeza Yates and Ribeiro-Neto [7], semantic indexing is surveyed by Snoek and Worring [106].

Retrieval

As discussed before, video data consists of multimodal features, including text, audio visual features and semantic anntotations. Consequently, there are a number of different ways in which a user can query a video retrieval system. As Snoek et al. [107] pointed out, three query formulation paradigms exist in the video retrieval domain: query-by-textual-keyword, query-by-visual-example and query-by-concept.

Query-By-Textual-Keyword Query-by-textual-keyword is one of the most popular methods of searching for video [50]. It is simple and users are familiar with this paradigm from text-based searches. Query-by-text relies on the availability of sufficient textual descriptions, including descriptive data, transcripts and annotations.

Query-By-Visual-Example Query-by-visual-example has its roots in content-based image retrieval. It allows the users to provide sample images or video clips as examples to retrieve more results. This approach uses the low-level features that are available in images and videos, such as colour, texture and shape to retrieve results. The basic idea is that visual similarity can be used to identify relevant documents. Content-based image retrieval has been well studied, a survey is given by Aigrain et al. [2].

Query-By-Concept In an attempt to bridge the Semantic Gap, a great deal of interest in the multimedia search community has been invested in query-by-concept, also referred to as concept-based, conceptual or semantic retrieval. A survey on concept-based retrieval is given by Snoek and Worring [105]. Conceptual retrieval relies on semantic annotations, i.e. high level concepts which have been associated with the video data. A well known set of high-level semantic concepts has been explored by the Large Scale Ontology for Multimedia (LSCOM) initiative [84], a subset of which is used within TRECVid[5] to study concept-based retrieval. Considering semantic concepts as additional textual annotation, documents can be retrieved by triggering textual search queries. Hildebrand et al. [54] analysed state-of-the-art semantic retrieval systems, concluding that semantic concepts are often used to filter retrieval results. Query-by-concept is an extension to both query-by-textual-keyword and query-by-visual-example, narrowing down corresponding results. Indeed, the most successful video retrieval systems that have been evaluated within TRECVid (e.g. [108, 49]) employ these two approaches to improve their retrieval results.

[5]TRECVid is an annual evaluation effort, targeting research in multimedia information retrieval.

1.2.4 Interface Designs

According to Spink et al. [109], users are often uncertain of their information need and hence have problems finding a starting point for their information seeking task. And even if users know exactly what they are intending to retrieve, formulating a "good" search query can be a challenging task. This problem even exacerbates when dealing with multimedia data. Graphical user interfaces serve here as a mediator between the available data corpus and the user. It is the retrieval systems' interface which will provide users facilities to formulate search queries and/or to dig into the available data. Hearst [53] outlines various conditions that dominate the design of state-of-the-art search interfaces. First of all, the process of searching is a means toward satisfying an information need. Interfaces should therefore avoid being intrusive, since this could disturb the users in their seeking process. Moreover, satisfying an information need is already a mentally intensive task. Consequently, the interface should not distract the users, but rather support them in their assessment of the search results. Especially in the WWW domain, search interfaces are not used by high expertise librarians only but also by the general public. Therefore, user interfaces have to be intuitive to use by a diverse group of potential users.

Schoeffmann et al. [100] survey representative video browsing and exploration interfaces. They distinguish between interfaces that support video browsing that rely on interaction similar to classical video players, video exploration interfaces and unconventional video visualisation interfaces. Another survey is given by Snoek et al. [105], who focus on concept-based video retrieval interfaces. For further reading, the reader is referred to these two publications.

1.2.5 Summary

In this section, we introduced basic principles of video retrieval. We first presented the general structure of video documents. As we have shown, videos consist of audio-visual data streams and are often accompanied with metadata. Metadata is mainly of textual nature. Further, we argued that most retrieval scenarios require videos to be split into smaller units of retrieval. We surveyed two different segmentation units and their application:

video shots and (news) video stories. Moreover, we introduced document representation and matching techniques, including indexing, retrieving and ranking of video documents. Finally, we introduced different graphical user interfaces that support users in their information seeking task.

The techniques and methods we have introduced combine all required parts of a video IR system. In the next section, we will survey how these approaches can be applied to adaptively support users' information needs.

1.3 Personalised Video Search and Recommendation

In the previous section, we surveyed basic concepts of video retrieval. We argued that when interacting with a video retrieval system, users express their information need in search queries. The underlying retrieval engine then retrieves relevant retrieval results to the given queries. A necessary requisite for this IR scenario is to correctly interpret the users' information need. As Spink et al. [109] 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 know what information they can expect from the corpus [98]. Further, Jansen et al. [59] 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. Without further knowledge, it is a demanding task to understand the users' intentions. Interactive information retrieval aims at improving the classic information retrieval model that we introduced in the previous section 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 minimise the users' efforts to fulfill their information seeking task, there is a need to personalise search.

In this section, we will introduce basic concepts of personalised IR. In Section 1.3.1, we first survey different sources that are used to gather users' interests, an important requisite for any type of personalisation. Section 1.3.2 introduces different application techniques. In

Section 1.3.3, we introduce state-of-the-art personalisation approaches with a main focus on the video domain. Section 1.3.4 summarises the personalisation section.

1.3.1 Gathering and Representing Interest

Most of the approaches that follow the interactive information retrieval model are based on relevance feedback techniques [98]. Relevance feedback 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 [95]. The principle of relevance feedback is to identify the user's information need and then, exploiting this knowledge, adapting search results. Rocchio [92] 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. There are three types of relevance feedback in interactive information retrieval which will be introduced in the remainder of this section: explicit, implicit and pseudo relevance feedback. Further, we introduce the Ostensive Model of Developing Information Need, that addresses the problem of non-static interest and discuss approaches to represent user interests in personal profiles.

Explicit Relevance Feedback

A simple approach to identify users' interests is to explicitly ask them about their opinion. In a retrieval context, they can express their opinion by providing *explicit relevance feedback*. Hence, the user is asked during their retrieval process to actively indicate which documents are relevant in the result set. This relevance judgement can be given on a binary or graded relevance scale. A binary feedback indicates that the rated document is either relevant or non-relevant for the user's current information need. Considering that binary relevance requires a rather strong judgement, a relevance scale allows the user to define different grades of relevance such as “highly relevant”, “relevant”, “maybe relevant” or “somewhat relevant”. As of December 2011, the commercial video sharing platform YouTube supports binary feedback by providing a “Thumbs up” button in their interface. The Dailymotion platform opted for the graded relevance scale scheme. Registered users of their service can express their interest on a Five “star” scale. Explicit relevance feedback is very reliable. Although the impact of explicit relevance feedback in above systems remains unclear, it has been shown in text retrieval that giving explicit relevance feedback is a cognitively demanding task and can affect the search process. Also, previous evaluations have found that users of explicit feedback systems often do not provide sufficient levels of feedback in order for adaptive retrieval techniques to work [47, 10].

Implicit Relevance Feedback

Deviating from the method of *explicitly* asking the user to rate results, the use of *implicit* feedback techniques helps learning users’ interest unobtrusively. The main advantage is that this approach delivers the user from providing explicit feedback. As a large quantity of implicit data can be gathered without disturbing the users’ workflow, the implicit approach is an attractive alternative. According to Nichols [86], however, information gathered using implicit techniques are less accurate than information based on explicit feedback. Agrichtein et al. [1] evaluated the effect of user feedback on web retrieval using over 3000 queries and 12 million user interactions. They show that implicit relevance feedback can improve retrieval performance by much as 31% relative to systems that do not incorporate any feedback. Furthermore, both implicit and explicit measures can be combined to provide an accurate representation of the users’ interests. Kelly and Teevan [65] provide a literature overview of the research which has been done in the field.

Not all implicit measures are useful to infer relevance. Thus, various research has been done to detect those features which promise to be valid indicators of interest. From the psychological point of view, a promising indicator of interest is viewing time. People look at objects or things they find interesting for a longer time than on uninteresting things. For instance, Faw and Nunnally [40] showed a positive correlation between "pleasant ratings" and viewing time and Day [34] found that most people look longer on images they liked than on images they disliked. According to Oostendorp and Berlyne [87], people look longer at objects evoking pleasurable emotions. Transferring these findings into an information retrieval context, users are expected to spend more time in viewing relevant documents than non-relevant documents. Claypool et al. [28] introduce a categorisation of both explicit and implicit interest indicators in web retrieval. They conclude that time spend on a page, the amount of scrolling on a page and the combination of these two features are valid implicit indicators for interest. Furthermore, they found that individual scrolling measures and the number of mouse clicks are ineffective indicators. Morita and Shinoda [82] evaluated if user behaviour while reading newsgroup articles could be used as implicit indicator for interest. They measured the copying, saving or following-up of an entry and the time spend for reading the entries. They reveal that the reading time for documents rated as interesting was longer than for uninteresting documents. A relation between interest and following-up, saving or copying was not found. White et al. [126] consider reading time as a technique to automatically re-rank sentence-based summaries. Their results, however, were inconclusive. Kelly [64] criticises the study approaches that focus on display time as relevance indicator, as she assumes that information-seeking behaviour is not influenced by contextual factors such as topic, task and collection. Therefore, she performed a study to investigate the relationship between information-seeking task and the display time. Her results cast doubt on the straightforward interpretation of dwell time as an indicator of interest or relevance.

Another indicator of interest which has been analysed is the users' browsing behaviour. Seo and Zhang [101] introduce a method to learn users' preferences from inobtrusively observing their web-browsing behaviour. They conclude that their approach can improve retrieval performance. However, the adaptation of users' interest over a longer period of time has not been taken into account as their search sessions were set up for a short period only. Maglio et al. [77] suggest to infer attention from observing the eye movements. In the HCI community, this has become a common technique.

Pseudo Relevance Feedback

A third relevance feedback approach is called *pseudo*, *blind* or *ad-hoc relevance feedback*. It was first introduced by Croft and Harper [30]. Differing from the previous two approaches, pseudo relevance feedback does not require users providing relevance assessments; the top ranked retrieval results are considered being relevant and used to adapt the initial search query. Considering the lack of manual input, its usage as source for personalisation techniques is questionable.

Evolving User Interest

In a retrieval context, profiles can be used to contextualise the user's search queries within his or her interests and to rerank retrieval results. This approach is based on the assumption that the user's information interest is static, which is, however, not appropriate in a retrieval context. Campbell [20] argues that the user's information need can change within different retrieval sessions and sometimes even within the same session. He states that the user's search direction is directly influenced by the documents retrieved. The following example illustrates this observation:

> Imagine a user who is interested in red cars and uses a video retrieval system to find videos depicting such cars. Their first search query returns several video clips, including videos of red Ferraris. Watching these video, he or she wants to find more Ferraris and adapts the search query accordingly. The new result list now consists of video clips showing red and green Ferraris. Fascinated by the rare colour for this type of car, he/she again reformulates the search query to find more green Ferraris. Within one session, the user's information need evolved from red cars to green Ferraris.

Based on this observation, Campbell and van Rijsbergen [22] introduce the Ostensive Model of Developing Information Need that incorporates this change of interest by considering, *when* a user provided relevance feedback. In the Ostensive Model, providing feedback on a document is seen as ostensive evidence that this document is relevant for the user's current interest. The combination of this feedback over several search iterations provides ostensive evidence about the user's changing interest. The model considers the user's changing focus of interest by granting the most recent feedback a higher impact over the combined evidence.

User Profiling

Considering the large amount of personal data that can be captured, most personalisation systems rely on user profiles to manage this data. User profiling is the process of learning user interests over an extended period of time. User profiles may contain demographic information and user feedback, that they expressed either explicitly or implicitly. In this section, we will highlight basic principles of user profiling. A state-of-the-art survey is given by Gausch et al. [42], who distinguish between two types of user profiles: *short-term* and *long-term* profiles. Short-term user profiles are used for personalisation within one session, i.e. any feedback that the user provides during their current information seeking task is used to adapt the results. Long-term user profiles, on the other hand, aim to keep track of users' long-term interests. Personalisation services based upon such profiles can hence adapt results considering user feedback which was given over multiple sessions.

In literature, three types of user profile representations exist: Weighted keywords or concepts, semantic networks and association rules. Association rules are mainly applied in the field of web log mining. By identifying relations between variables, it is possible to identify popular variable combinations. Association rules rely on large amount of data, often provided by different users. Considering this requirement, we therefore focus on weighting-based profiles and semantic network-based profiles, neglecting association rules. A survey on association rules is given by Mobasher [80].

Weighted Keywords or Concepts The most popular representation of user interests is the weighted keyword or concept approach. Interests are represented as a vector of weighted terms that have either been extracted from those documents that users showed interest in or that have been provided manually by the users. The weighting indicates the importance of the corresponding term in the user profile. The main disadvantage of this approach is the so-called polysemy problem, hence the multiple meanings that each word can have.

An early example includes the *Amalthaea* system [83] where keywords, extracted from websites are assigned with a weighting based on TF.IDF [7]. The terms-weighting combination is represented in the profile as a vector. Similar approaches are studied by Sakagami and Kamba [96], who introduce personalised online newspapers, Lieberman [70], introducing a browsing assistant and Pazzani et al. [89], who propose a recommender system that exploits weighted keyword profiles.

Even though weighted keyword profiling has been well studied in the text domain, hardly any work has been done on studying similar approaches in the video domain. Few exceptions include Weiß et al. [124], who, however, generate user profiles exploiting video metadata rather than audio-visual features.

Semantic Networks In the semantic network approach, keywords are replaced with concepts. User interests are represented as weighted nodes of a graph where each node is a concept in which the user showed interest in. A similar approach is referred to as concept profiles. Differing from semantic networks, however, concept profiles consider abstract topics rather than specific words to represent user interests.

The advantage of semantic networks and concept profiles is that concepts can be organised in a hierarchical structure. In a web search scenario, for example, a user might have shown interest in a website that has been categorised in the Open Directory Project (ODP)[6] as being a website about "Travel and Tourism in Scotland". Within ODP, "Travel and Tourism in Scotland" is a subcategory of "Travel and Tourism in the United Kingdom". Bloedorn et al. [15] argue to exploit such hierarchies, since they allow generalisation of concepts. Personalisation services could hence exploit this relationship, e.g. by recommending other websites belonging to the same or more general categories. Various approaches have been studied to exploit such public hierarchies:

[6] `http://dmoz.org/` accessed on: 7 September 2011. The ODP is a manually edited catalog of websites. It is organised as a tree, where websites are leaf nodes and categories are internal nodes.

Daoud et al. [32, 33] analyse the documents that users' provided implicit relevance feedback on, map the concepts of these documents with the Open Directory Project (ODP) ontology and store them in a hierarchical, graph-based user profile at the end of each search session. Other personalisation techniques based on ODP include [26, 102, 23, 113], who show that incorporating this taxonomy can significantly outperform unpersonalised search techniques.

Dudev et al. [38] propose the creation of user profiles by creating knowledge graphs that model the relationship between different concepts in the Linked Open Data Cloud[7]. Different from the ODP ontology, important parts of the Linked Open Data cloud have been created automatically, e.g. by converting Wikipedia pages into an ontological representation. Consequently, the available data is of immense size, but rather un-uniform.

Gauch et al. [42] argue that when exploiting such public directories, various pre-processing steps have to be performed, including transforming the content into a concept hierarchy or dealing with situations where some concepts might have multiple entries while other concepts are less important. Moreover, they argue that the more levels of the hierarchy are used, the more general the profile representation might become.

Discussion

In this section, we surveyed issues regarding gathering user interests. We first introduced relevance feedback, the most common technique used to identify user interest. As we have shown, relevance feedback techniques can be split into three main categories: explicit, implicit and pseudo relevance feedback. Rui et al. [93] propose interactive relevance feedback as a method to bridge the Semantic Gap, assuming that high-level concepts can be identified using low-level features. In their approach, users have to rate images according to their relevance for the information need. The features are weighted automatically to model high-level concepts based on user's feedback. The results of their study promise a reduction of query formulation efforts, as the relevance feedback technique seems to gather the user's

[7] `http://linkeddata.org/` Last time accessed on: 5 May 2010. The Linked Open Data collection of ontologies unites information about many different freely available concepts, ODP being one of them.

information need effectively. According to Huang and Zhou [57], relevance feedback appears to be more promising in the image field than in the classical text field, as it is easier and faster to rank images according to their relevance than ranking text documents. Various research has been done to optimise parameters and algorithms [132, 36, 57, 90]. Karthik and Jawahar [63] introduce a framework to evaluate different relevance feedback algorithms. They conclude that statistical models are the most promising algorithms. These models try to cluster images into relevant and non-relevant images. Further, we introduced the Ostensive Model of Developing Information Need, which emphasises the time when users provided relevance feedback. Various forms of this model have been developed and applied in image retrieval [21, 116, 69] and Web search scenarios [62]. Finally, we introduced different approaches of user profiling. User profiling is one of the key challenges in adaptive search and recommendation. As we discussed, two types of user profiling exist: short-term and long-term profiling.

1.3.2 Personalisation Techniques

After introducing approaches to gather users' interests, this section introduces personalisation techniques that exploit this feedback. Jameson [58] lists various adaptation paradigms, including ability-based user interfaces, learning personal assistants, recommender systems, adaptation to situational impairments, personalised search and user interfaces that adapt to the current task. Note that within this work, we will focus on recommender systems and personalised search, neglecting the other paradigms. Both paradigms will be introduced in the remainder of this section.

Personalised Search

In 2008, Marissa Mayer, the Vice President of Search and User Experience of Google Inc. predicted in an interview held at the LeWeb conference that "in the future personalised search will be one of the traits of leading search engines" [79]. This statement reflects the increasing attention that personalised search draws from both academia and industry sides.

Teevan et al. [114] argue that there is a big performance difference between personalised and non-personalised search engines. They hence argue that there is a big potential for personalisation.

As discussed before, one of the main problems in IR is that users express their information need using short queries only. Matching these short queries with the document collection will return only a relative small amount of relevant results. Providing more search terms could improve the retrieval performance, as then, more documents can be retrieved. Dependent on how much influence the user shall have, the expansion terms can either be added by the system – *automatic query expansion* or by the user – *interactive query expansion*. According to Ruthven [94], query expansion terms which have been provided in an interactive process are less useful than automatically identified term. We therefore neglect interactive query expansion and focus on automatic query expansion techniques.

Automatic query expansion based on relevance feedback is a common technique to refine search queries (e.g. [92, 98]). The reason for its success can be found by the users: they tend *not* to give relevance feedback or to formulate their search queries appropriate. The first query expansion approach was introduced by Rochio [92]. In an retrieval experiment, Rocchio added term vectors for all retrieved relevant documents and subtracted the term vectors for all irrelevant documents to refine search queries. Hence, terms are aligned with a weighting which can increase and decrease during the process. Järvelin [60] has shown that concept-based query expansion, i.e. exploiting ontologies, is helpful to improve retrieval performance. Multiple other studies show the effectiveness of ontology-based expansion [12].

Video retrieval based query expansion approaches include Volkmer and Natsev [119], who rely on textual annotation (video transcripts) to expand search queries. Within their experiment, they significantly outperform a baseline run without any query expansion, hence indicating the potentials of query modification in video search. Similar results are reported by Porkaew and Chakrabarti [90] and Zhai et al. [131], who both expand search queries using content-based visual features.

Document Recommendation

Another personalisation technique is document recommendation. Chris Anderson, editor-in-chief of the Wired Magazine, claims in his book [4] that "we are leaving the Information Age and entering the Recommendation Age." Differing from personalised search, where search results are adapted based on user interests, recommender systems provide additional items (documents). The main idea of this technique is to provide users faster and more information they might be interested in. Further, in an e-commerce scenario, recommendations shall influence the users. For example, Amazon.com[8] provides recommendations on other products that their customers might be interested in. Recommender systems hence inform users about things they might not be aware of and have not been actively searching for. They can be distinguished into two main categories: *content-based recommender systems* and *collaborative filtering systems*. In the remainder of this section, we will briefly introduce both aproaches and discuss user profiling issues.

Content-Based Recommender Systems Content-based recommender systems determine the relevance of an item (e.g. a video document, website or a product) based on the user's interest in other, similar items. Items in the data collection are evaluated in accordance to users' previous feedback and the most similar items are recommended to the user. A survey is given by Pazzani and Billsus [88]. User interest is usually stored in personal user profiles.

Collaborative Filtering Collaborative filtering systems aim to exploit the opinion of people with similar interests. Thus, items are recommended when other users of the recommender system showed interest in it. Differing from content-based recommendation, where the content of the documents has to be analysed, the challenge in collaborative filtering is in identifying users with similar interests. A survey is given by Schafer et al. [99].

[8] `http://www.amazon.com/` accessed on: 7 September 2011

1.3.3 State-of-the-art Personalisation Approaches

Implicit feedback techniques have been successfully applied to retrieval systems in the past. For instance, White [125] and Joachims [61] defined and evaluated several implicit feedback models on a text-based retrieval system. Their results indicated that their implicit models were able to obtain a comparable performance to that obtained with explicit feedback models. However, their techniques were based on textual information, and applied individually at runtime during the user's search session. As stated previously, the lack of textual annotation on video digital libraries prevents the adoption of this approach in video retrieval systems. One solution to this problem is to collect the implicit information from a large number of past users, following a collaborative recommendation strategy.

The exploitation of usage information from a community of past users is a widely researched approach to improve information retrieval systems. However, most of past and present studies focus on the text retrieval domain. The main hypothesis of such systems is that when a user enters a query, the system can exploit the behaviour of past users that were performing a similar task. For instance, Bauer and Leake [9] build up a task representation based on the user's sequence of accessed documents. This task representation is used by an information agent, which proactively suggests documents to the user.

A commonly exploited past usage information structure is clickthrough data. Clickthrough data is limited to the query that the user executed into the system, the returned documents, and the subsequent documents that the user opened to view. Sun et al. [111] and Dou et al. [35] mine query log clickthrough information to perform a collaborative personalisation of search results, giving preference to documents that similar users had clicked previously for similar queries. Sun et al. [111] apply a dimensional reduction pre-processing step on the clickthrough data in order to find latent semantic links between users, queries and documents. Dou et al. [35] complement these latent relationships with user-topic and document-topic similarity measures. Craswell and Szummer [29] use a bipartite graph to represent the clickthrough data of an image retrieval system, where queries and documents are the nodes and links are the ones directly captured in clickthrough data. A random walk is then applied in order to recommend images based on the user's last query.

Following this line of works, White et al. [127] introduced the concept of query and search session trails, where the interaction between the user and the retrieval system is seen as a trail that leads from the user query to the last accessed document of the query session or the search session, which is constituted by multiple query sessions. They argue that the user's need of information was most likely satisfied with the last document of these trails, i.e. the last document that the user accessed in the query or search session.

Above approaches rely on text to base personalisation techniques on. As we discussed, however, multimedia documents consist of multiple modalities, including text and audio-visual features. A comprehensive survey on relevance feedback in the image domain is given by Zhou et al. [133]. The main problem when incorporating content-based features is to find out which feature represents the image best. Further, which approach should be followed to build an adaptive retrieval model. This problem even increases when dealing with video content, since additional features can be considered.

An early work focusing on relevance feedback in the video domain is presented by Rui et al. [130]. They illustrate that the content rich nature of multimedia document require a more precise feedback. When a user provides relevance feedback on a video document, it is not clear, which feature of this document should be exploited to identify similar documents. It could be, for example, the colour feature of the video or the context of the video that is relevant to the user. In their work, they propose a relevance feedback framework for multimedia databases where search queries are adapted based on user's relevance feedback. Their framework, however, is focusing on explicit relevance feedback only, neglecting the possibility to exploit implicit indicators of relevance. Another study focusing on explicit relevance feedback in the video domain is provided by Hauptmann et al. [51], who asks users to manually provide labels for video shots. After retrieving similar video documents for the label, users are then asked to explicitly mark those videos that match to the corresponding label. Doulamis et al. [37] require explicit relevance feedback, given during the information seeking process, to update video retrieval results. Hence, in all three studies which have been mentioned above, explicit relevance feedback is used to personalise search queries.

Luan et al. [74] consider text, high-level features and multiple other modalities to base their relevance feedback algorithms on. In their approach, however, they do not focus on personalising techniques but rather on improving video annotation. An extension of their work is presented in [75], where they propose multiple feedback strategies that support interactive video retrieval. They argue that the more complex nature of video data, when compared with textual data, require different types of feedback strategies. Therefore, they distinguish between three categories of feedback types: recall-driven, precision-driven and temporal locality-driven feedback. The first type focuses on analysing the correlation of video features that can be extracted from positive and negative rated video documents. It aims to result in higher recall values. The second type employs active learning techniques and aims to constantly re-rank retrieval results. Its purpose is to increase precision. The third type, temporal locality-driven feedback, exploits the temporal coherence among neighboured shots. In their study, they show that giving the user the choice to chose between these different feedback types can effectively improve video retrieval performance.

Yan et al. [129] study video recommendations based on multimodal fusion and relevance feedback. They exploit viewing time to identify positive recommendations in their video recommender system. They interpret a very short video viewing time as negative relevance feedback, arguing that the user is not interested in the video's content. Further, they argue that a longer viewing time indicates a higher interest in the video. Another study focusing on negative relevance feedback is introduced by Yan et al. [128], who consider the *lowest* ranked documents of a search query to be not relevant. Hence, they suggest to re-formulate the initial search query by using these documents as negative example. They approach, however, does not require any specific user input, it is based on pseudo-relevance feedback only. Nevertheless, their findings indicate that pseudo-relevance feedback can successfully be employed to adapt retrieval results.

Vrochidis et al. [120] aim to improve interactive video retrieval by incorporating additional implicit indicators of relevance. They consider the following user actions as implicit relevance feedback: (1) Text-based search queries. Given a search query, they assume that the keywords that form the search query are relevant. (2) Visual queries. If a user provides a key frame as visual query, they assume this key frame to be relevant. (3) Side-shot and video-shot queries. When a user performs an action on a shot, they assume this shot to be relevant. Arguing that each indicator can be interpreted to a different degree to be relevant, they suggest to assign each feedback type a different weighting.

A purely content-based personalisation approach is introduced by Aksoy and Cavus [3] who extract low-level visual from explicitly relevant rated key frames. The suggest to extract different feature vectors from those key frames and to assign a relevance weighting for each vector.

A different understanding of implicit relevance feedback is introduced by Villa et al. [118]. Within their study, they aimed to elaborate whether the awareness of other users' search activities can help users in their information seeking task. They introduce a search scenario where two users remotely search for the same topic at the same time. Using their interface, each user is able to observe the other user's search activities. They conclude that awareness can have a direct influence in the information seeking process, since user's can learn from other users' search results and/or search queries. In case users copy the other user's search activity, this action can be interpreted as implicit relevance feedback, that the action has been performed on (or resulted in) relevant documents. A similar study is performed by Halvey et al. [46], who, however, study the impact of awareness in an asynchronous scenario. Within their experiment, users can interact with other user's previous search sessions. Again, continuing other user's search sessions can be seen as implicit indication that the retrieved search results are partially relevant.

1.3.4 Summary

In this section, we surveyed personalised video search and recommendation. An important prequisite for any kind of personalisation service is to identify users' personal interests. As we have shown, the most popular technique to gather this interest is relevance feedback. After introducing different feedback techniques and challenges such as evolving interest and user profiling, we introduced different personalisation services, namely personalised search and recommender systems. Finally, we surveyed state-of-the-art personalisation and recommendation systems.

1.4 Summary

In recent years, 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 uncoordinated publishing of video data by users worldwide [31]. 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. In this chapter, we first provided an overview on multimedia retrieval systems. We introduced the typical structure of video documents and argued for the segmentation and representation of these documents. Then, we illustrated that personalization techniques are required in order to identify those video documents that match users' personal interests. We therefore surveyed different techniques to gather user interests, presented different personalization techniques and discussed state-of-the-art personalisation approaches.

References

[1] Eugene Agichtein, Eric Brill, and Susan T. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR 2006*, pages 19–26, 2006.

[2] Philippe Aigrain, HongJiang Zhang, and Dragutin Petkovic. Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. *Multimedia Tools and Applications*, 3(3):179–202, 1996.

[3] Selim Aksoy and Özge Çavuş. A Relevance Feedback Technique for Multimodal Retrieval of News Videos. In Ljiljana Milic, editor, *EUROCON'05: Proceedings of the International Conference on Computer as a Tool, Belgrade, Serbia & Montenegro*, pages 139–142. IEEE, 11 2005.

[4] Chris Anderson. *The Long Tail: Why the Future of Business is Selling Less of More.* Hyperion, 7 2006.

[5] Joaquim Arlandis, Paul Over, and Wessel Kraaij. Boundary error analysis and categorization in the trecvid news story segmentation task. In *CIVR'05: Proceedings of the 4th International Conference on Image and Video Retrieval, Singapore, July 20-22, 2005*, pages 103–112, 2005.

[6] Richard Arndt, Raphaël Troncy, Steffen Staab, Lynda Hardman, and Miroslav Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In Karl Aberer, Key-Sun Choi, Natasha Fridman Noy, Dean Allemang, Kyung-Il Lee, Lyndon J. B. Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux, editors, *ISWC'07: Proceedings of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, Busan, Korea*, volume 4825 of *Lecture Notes in Computer Science*, pages 30–43. Springer, 11 2007.

[7] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval.* Addison Wesley, Harlow, 1st edition, 1999.

[8] Kobus Barnard and David A. Forsyth. Learning the Semantics of Words and Pictures. In Bob Werner, editor, *ICCV'01: Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, British Columbia, Canada*, pages 408–415. IEEE, 2001.

[9] Travis Bauer and David B. Leake. Real Time User Context Modeling for Information Retrieval Agents. In Henrique Paques, Ling Liu, David Grossman, and Calton Pu, editors, *CIKM'01: Proceedings of the ACM CIKM International Conference on Information and Knowledge Management, Atlanta, Georgia, USA*, pages 568–570. ACM, 2001.

[10] Nicholas J. Belkin, Colleen Cool, Diane Kelly, Shin jeng Lin, Soyeon Park, Jose Perez Carballo, and C. Sikora. Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing and Management*, 37(3):403–434, 2001.

[11] Marco Bertini, Alberto Del Bimbo, and Carlo Torniai. Multimedia enriched ontologies for video digital libraries. *International Journal of Parallel, Emergent and Distributed Systems*, 22(6):407–416, 2007.

[12] J. Bhogal, Andy Macfarlane, and P. Smith. A review of ontology based query expansion. *Information Processing & Management: An International Journal*, 43(4):866–886, 2007.

[13] Henk M. Blanken, Arjen P. de Vries, Henk Ernst Bok, and Ling Feng. *Multimedia Retrieval.* Springer Verlag, Heidelberg, Germany, 1st edition, 2007.

[14] David M. Blei and Michael I. Jordan. Modeling annotated data. In Callan et al. [19], pages 127–134.

[15] Eric Bloedorn, Inderjeet Mani, and T. Richard MacMillan. Machine Learning of User

Profiles: Representational Issues. In *AAAI/IAAI, Vol. 1*, pages 433–438. AAAI, 1996.

[16] Mohand Boughanem, Catherine Berrut, Josiane Mothe, and Chantal Soulé-Dupuy, editors. *ECIR'09: Proceedings of the 31th European Conference on IR Research, Toulouse, France*, volume 5478 of *Lecture Notes in Computer Science*. Springer Verlag, 4 2009.

[17] Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors. *The Adaptive Web*. Springer Verlag, 2007.

[18] Tobias Bürger, Erich Gams, and Georg Güntner. Smart content factory: assisting search for digital objects by generic linking concepts to multimedia content. In Siegfried Reich and Manolis Tzagarakis, editors, *HT'05: Proceedings of the 16th ACM Conference on Hypertext and Hypermedia, Salzburg, Austria*, pages 286–287. ACM, 2005.

[19] Jamie Callan, Gordon Cormack, Charles Clarke, David Hawking, and Alan F. Smeaton, editors. *SIGIR'03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada*. ACM, 2003.

[20] Iain Campbell. Supporting information needs by ostensive definition in an adaptive information space. In Ian Ruthven, editor, *MIRO'95: Proceedings of the Final Workshop on Multimedia Information Retrieval, Glasgow, Scotland, UK*, Workshops in Computing. BCS, 9 1995.

[21] Iain Campbell. Interactive Evaluation of the Ostensive Model Using a New Test Collection of Images with Multiple Relevance Assessments. *Information Retrieval*, 2(1):85–112, 2000.

[22] Iain Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. In Peter Ingwersen and Niels Ole Pors, editors, *CoLIS'06: Proceedings of the 6th International Conference on Conceptions of Library and Information Sciences, Copenhagen, Denmark*, pages 251–268. The Royal School of Librarianship, 10 1996.

[23] Jason Chaffee and Susan Gauch. Personal Ontologies for Web Navigation. In Arvin Agah, Jamie Callan, Elke Rundensteiner, and Susan Gauch, editors, *CIKM'00: Proceedings of the 2000 ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA*, pages 227–234. ACM, 11 2000.

[24] Lekha Chaisorn and Tat-Seng Chua. The segmentation and classification of story boundaries in news video. In Xiaofang Zhou and Pearl Pu, editors, *VDB'02: Proceedings of the IFIP TC2/WG2.6 Sixth Working Conference on Visual Database Systems, Brisbane, Australia*, volume 216 of *IFIP Conference Proceedings*, pages 95–109. Kluwer, 2002.

[25] Shih-Fu Chang, R. Manmatha, and Tat-Seng Chua. Combining Text and Audio-Visual Features in video Indexing. In Billene Mercer, editor, *ICASSP'05: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1005–1008. IEEE, 3 2005.

[26] Paul-Alexandru Chirita, Wolfgang Nejdl, Raluca Paiu, and Christian Kohlschütter. Using odp metadata to personalize search. In *SIGIR 2005*, pages 178–185, 2005.

[27] Tat-Seng Chua, Shih-Fu Chang, Lekha Chaisorn, and Winston H. Hsu. Story boundary detection in large broadcast news video archives: techniques, experience and trends. In Henning Schulzrinne, Nevenka Dimitrova, Martina Angela Sasse, Sue B. Moon, and Rainer Lienhart, editors, *ACM MM'04: Proceedings of the 12th ACM International Conference on Multimedia, October 10-16, 2004, New York, NY, USA*, pages 656–659. ACM, 2004.

[28] Mark Claypool, Phong Le, Makoto Wased, and David Brown. Implicit interest indicators. In Candy Sidner and Johanna Moore, editors, *IUI'01: Proceedings of the International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, USA*, pages 33–40. ACM, 1 2001.

[29] Nick Craswell and Martin Szummer. Random walks on the click graph. In Kraaij et al. [66], pages 239–246.

[30] W. Bruce Croft and David J. Harper. Using probabilistic models of document retrieval without relevance information. *Readings in information retrieval*, pages 339–344, 1997.

[31] Sally Jo Cunningham and David M. Nichols. How people find videos. In Larsen et al. [68], pages 201–210.

[32] Mariam Daoud, Lynda Tamine-Lechani, and Mohand Boughanem. Using a graph-based ontological user profile for personalizing search. In James G. Shanahan, Sihem Amer-Yahia, Ioana Manolescu, Yi Zhang, David A. Evans, Aleksander Kolcz, Key-Sun Choi, and Abdur Chowdhury, editors, *CIKM'08: Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, California, USA*, pages 1495–1496. ACM, 2008.

[33] Mariam Daoud, Lynda Tamine-Lechani, Mohand Boughanem, and Bilal Chebaro. A session based personalized search using an ontological user profile. In Sung Y. Shin and Sascha Ossowski, editors, *SAC'09: Proceedings of the 2009 ACM Symposium on Applied Computing, Honolulu, Hawaii*, pages 1732–1736. ACM, 2009.

[34] H. Day. Looking time as a function of stimulus variables and individual differences. *Perceptual & Motor Skills*, 22(2):423–428, 1966.

[35] Zhicheng Dou, Ruihua Song, and Ji-Rong Wen. A large-scale evaluation and analysis of personalized search strategies. In Carey L. Williamson, Mary Ellen Zurko, Peter F. Patel-Schneider, and Prashant J. Shenoy, editors, *WWW'07: Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada*, pages 581–590. ACM, 2007.

[36] A. Doulamis and N. Doulamis. Performance evaluation of Euclidean/correlation-based relevance feedback algorithms in content-based image retrieval systems. In Luis Torres, editor, *ICIP'03: Proceedings of the International Conference on Image Processing, Barcelona, Spain*, volume 4, pages 737–740. IEEE, 09 2003.

[37] Anastasios D. Doulamis, Yannis S. Avrithis, Nikolaos D. Doulamis, and Stefanos D. Kollias. Interactive content-based retrieval in video databases using fuzzy classification and relevance feedback. In *ICMCS, Vol. 2*, pages 954–958, 1999.

[38] Minko Dudev, Shady Elbassuoni, Julia Luxenburger, Maya Ramanath, and Gerhard Weikum. Personalizing the Search for Knowledge. In Vassilis Christophides and Georgia Koutrika, editors, *PersDB'08: Proceedings of the Second International Workshop on Personalized Access, Profile Management, and Context Awareness: Databases*, pages 1–8, 2008.

[39] Gwenaël Durand, Gabriella Kazai, Mounia Lalmas, Uwe Rauschenbach, and Patrick Wolf. A Metadata Model Supporting Scalable Interactive TV Services. In Yi-Ping Phoebe Chen, editor, *MMM'05: Proceedings of the 11th International Conference on Multi Media Modeling, Melbourne, Australia*, pages 386–391. IEEE Computer Society, 1 2005.

[40] T. Faw and J. Nunnally. The Effects on Eye Movements of Complexity, Novelty, and Affective Tone. *Perception & Psychophysics*, 2(7):263–267, 1967.

[41] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, May 1998.

[42] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. *User Profiles for Personalized Information Access*, chapter 2, pages 54–89. Volume 1 of Brusilovsky et al. [17], 2007.

[43] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.

[44] Ling Guan and Hong-Jiang Zhang, editors. *ICME'06: Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, Toronto, Ontario, Canada*. IEEE, 2006.

[45] Martin Halvey and Mark T. Keane. Analysis of online video search and sharing. In Simon Harper, Helen Ashman, Mark Bernstein, Alexandra I. Cristea, Hugh C. Davis,

Paul De Bra, Vicki L. Hanson, and David E. Millard, editors, *HT'07: Proceedings of the 18th ACM Conference on Hypertext and Hypermedia, Manchester, UK*, pages 217–226. ACM, 9 2007.

[46] Martin Halvey, David Vallet, David Hannah, and Joemon M. Jose. Vigor: a grouping oriented interface for search and retrieval in video libraries. In Fred Heath, Mary Lynn Rice-Lively, and Richard Furuta, editors, *JCDL'09: Proceedings of the 2009 Joint International Conference on Digital Libraries, Austin, TX, USA*, pages 87–96. ACM, 2009.

[47] Micheline Hancock-Beaulieu and Stephen Walker. An evaluation of automatic query expansion in an online library catalogue. *Journal of Documentation*, 48(4):406–421, 1992.

[48] Alan Hanjalic. Shot-Boundary Detection: Unraveled and Resolved? *IEEE Transactions on Circuits and Systems for Video Technology*, 12(2):90–105, 2 2002.

[49] Alex Hauptmann, Mike Christel, R. Concescu, J. Gao, Q. Jin, J.Y. Pan, S.M. Stevens, R. Yan, J. Yang, and Y. Zhang. CMU Informedia's TRECVID 2005 Skirmishes. In *TRECVid 2005*, 2005.

[50] Alexander G. Hauptmann. Lessons for the future from a decade of informedia video analysis research. In *CIVR 2005*, pages 1–10, 2005.

[51] Alexander G. Hauptmann, Wei-Hao Lin, Rong Yan, Jun Yang 0003, and Ming yu Chen. Extreme video retrieval: joint maximization of human and computer performance. In Klara Nahrstedt, Matthew Turk, Yong Rui, Wolfgang Klas, and Ketan Mayer-Patel, editors, *MM'06: Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA*, pages 385–394. ACM, 2006.

[52] Xuming He, Richard S. Zemel, and Miguel Á. Carreira-Perpiñán. Multiscale conditional random fields for image labeling. In Frances Titsworth, editor, *CVPR'04: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2*, pages 695–702. IEEE, 2004.

[53] Marti Hearst. *Search User Interfaces*. Cambridge University Press, 2009.

[54] Michiel Hildebrand, Jacco van Ossenbruggen, and Lynda Hardman. An Analysis of Search-based User Interaction on the Semantic Web. Technical Report INS-E0706, Centrum voor Wiskunde en Informatica, 5 2007.

[55] Frank Hopfgartner, Reede Ren, Thierry Urruty, and Joemon M. Jose. *Information Organisation Issues in Multimedia Retrieval using Low-Level Features*, chapter 15. Multimedia Semantics: Metadata, Analysis and Interaction. Wiley, 1 edition, 2011.

[56] Chih-Wei Huang. Automatic Closed Caption Alignment Based on Speech Recognition Transcripts. Technical report, University of Columbia, 2003. ADVENT Technical Report.

[57] Thomas S. Huang and Xiang Sean Zhou. Image retrieval with relevance feedback: From heuristic weight adjustment to optimal learning methods. In Ioannis Pitas, Venetsanopoulos, and Thrasos Pappas, editors, *ICIP'01: Proceedings of International Conference on Image Processing, Thessaloniki, Greece*, volume 3, pages 2–5. IEEE, 2001.

[58] Anthony Jameson. Adaptive Interfaces and Agents. *The human-computer interaction handbook: evolving technologies and emerging applications*, pages 433–458, 2008.

[59] Bernard J. Jansen, Abby Goodrum, and Amanda Spink. Searching for multimedia: analysis of audio, video and image web queries. *World Wide Web*, 3(4):249–254, 2000.

[60] Kalervo Järvelin, Jaana Kekäläinen, and Timo Niemi. Expansiontool: Concept-based query expansion and construction. *Information Retrieval*, 4(3-4):231–255, 2001.

[61] Thorsten Joachims, Laura A. Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In *SIGIR 2005*, pages 154–161, 2005.

[62] Hideo Joho, D. Birbeck, and Joemon M. Jose. An ostensive browsing and searching on the web. In Bich-Lien Doan, Massimo Melucci, and Joemon M. Jose, editors, *CIR'07: Proceedings of the Second International Workshop on Context-Based Information Retrieval*, pages 81–92, 2007.

[63] P. Suman Karthik and C. V. Jawahar. Analysis of Relevance Feedback in Content Based Image Retrieval. *ICARCV'06: Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision*, 2006.

[64] Diane Kelly. *Understanding implicit feedback and document preference: A naturalistic user study*. PhD thesis, Rutgers University, 2004.

[65] Diane Kelly and Jaime Teevan. Implicit feedback for inferring user preference: A bibliography. *SIGIR Forum*, 37(2):18–28, 2003.

[66] Wessel Kraaij, Arjen P. de Vries, Charles L. A. Clarke, Norbert Fuhr, and Noriko Kando, editors. *SIGIR'07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands*. ACM, 2007.

[67] Sanjiv Kumar and Martial Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In Katsushi Ikeuchi, Olivier Faugeras, and Jitendra Malik, editors, *ICCV'03: 9th IEEE International Conference on Computer Vision, Nice, France*, pages 1150–1159. IEEE Computer Society, 2003.

[68] Ronald L. Larsen, Andreas Paepcke, José Luis Borbinha, and Mor Naaman, editors. *JCDL'08: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Pittsburgh, PA, USA*. ACM, 2008.

[69] Teerapong Leelanupab, Frank Hopfgartner, and Joemon M. Jose. User centred evaluation of a recommendation based image browsing system. In Bhanu Prasad, Pawan Lingras, and Ashwin Ram, editors, *IICAI'09: Proceedings of the 4th Indian International Conference on Artificial Intelligence, Tumkur, Karnataka, India*, pages 558–573. IICAI, 2009.

[70] Henry Lieberman. Letizia: An Agent That Assists Web Browsing. In *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Canada*, pages 924–929. Morgan Kaufmann, 8 1995.

[71] Rainer Lienhart, Anand R. Prasad, Alan Hanjalic, Sunghyun Choi, Brian P. Bailey, and Nicu Sebe, editors. *MM'07: Proceedings of the 15th International Conference on Multimedia, Augsburg, Germany*. ACM, 2007.

[72] Ainhoa Llorente, Simon E. Overell, Haiming Liu 0002, Rui Hu, Adam Rae, Jianhan Zhu, Dawei Song, and Stefan M. Rüger. Exploiting Term Co-occurrence for Enhancing Automated Image Annotation. In Carol Peters, Thomas Deselaers, Nicola Ferro, Julio Gonzalo, Gareth J. F. Jones, Mikko Kurimo, Thomas Mandl, Anselmo Peñas, and Vivien Petras, editors, *CLEF'08: Proceedings of the 9th Workshop of the Cross-Language Evaluation Forum, Aarhus, Denmark, Revised Selected Papers*, volume 5706 of *Lecture Notes in Computer Science*, pages 632–639. Springer, 2008.

[73] Ainhoa Llorente and Stefan M. Rüger. Using second order statistics to enhance automated image annotation. In Boughanem et al. [16], pages 570–577.

[74] Huan-Bo Luan, Shi-Yong Neo, Hai-Kiat Goh, Yong-Dong Zhang, Shouxun Lin, and Tat-Seng Chua. Segregated feedback with performance-based adaptive sampling for interactive news video retrieval. In Lienhart et al. [71], pages 293–296.

[75] Huan-Bo Luan, Yantao Zheng, Shi-Yong Neo, Yongdong Zhang, Shouxun Lin, and Tat-Seng Chua. Adaptive multiple feedback strategies for interactive video search. In Jiebo Luo, Ling Guan, Alan Hanjalic, Mohan S. Kankanhalli, and Ivan Lee, editors, *CIVR'08: Proceedings of the 7th ACM International Conference on Image and Video Retrieval, Niagara Falls, Canada*, pages 457–464. ACM, 2008.

[76] João Magalhães and Stefan M. Rüger. *Semantic Multimedia Information Analysis for Retrieval Applications*, pages 334–354. IDEA group publishing, 2006.

[77] Paul P. Maglio, Rob Barrett, Christopher S. Campbell, and Ted Selker. Suitor: an attentive information system. In Doug Riecken, David Benyon, and Henry Liebermann, editors, *IUI'00: Proceedings of the Fifth International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, USA*, pages 169–176, New York, NY, USA, 1 2000. ACM Press.

[78] B. S. Manjunath, Philipe Salembier, and Thomas Sikora, editors. *Introduction to MPEG 7: Multimedia Content Description Language*. Wiley, 6 2002.

[79] Marissa Mayer. Interview at the "LeWeb 2008" Conference in Paris, France, 2008.

[80] Bamshad Mobasher. *Data Mining for Web Personalization*, chapter 3, pages 90–135. Volume 1 of Brusilovsky et al. [17], 2007.

[81] James Monaco. *How to Read a Film*. Oxford Press, London, 4 edition, 2009.

[82] Masahiro Morita and Yoichi Shinoda. Information filtering based on user behaviour analysis and best match text retrieval. In W. Bruce Croft and C. J. van Rijsbergen, editors, *SIGIR'94: Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, (Special Issue of the SIGIR Forum)*, pages 272–281. ACM/Springer, 1994.

[83] Alexandros Moukas and Pattie Maes. Amalthaea: An Evolving Multi-Agent Information Filtering and Discovery System for the WWW. *Autonomous Agents and Multi-Agent Systems*, 1(1):59–88, 1998.

[84] Milind R. Naphade, John R. Smith, Jelena Tesic, Shih-Fu Chang, Winston H. Hsu, Lyndon S. Kennedy, Alexander G. Hauptmann, and Jon Curtis. Large-Scale Concept Ontology for Multimedia. *IEEE MultiMedia*, 13(3):86–91, 2006.

[85] Shi-Yong Neo, Jin Zhao, Min-Yen Kan, and Tat-Seng Chua. Video Retrieval Using High Level Features: Exploiting Query Matching and Confidence-Based Weighting. In Hari Sundaram, Milind R. Naphade, John R. Smith, and Yong Rui, editors, *CIVR'06: Proceedings of the 5th International Conference on Image and Video Retrieval, Tempe, AZ, USA*, volume 4071 of *Lecture Notes in Computer Science*, pages 143–152. Springer, 2006.

[86] David M. Nichols. Implicit rating and filtering. In Laszlo Kovacs, editor, *Proceedings of 5th DELOS Workshop on Filtering and Collaborative Filtering, Budapest, Hungary*, pages 31–36. ERCIM, 11 1998.

[87] A. Oostendorp and D. E. Berlyne. Dimensions in the Perception of Architecture II: Measures of Exploratory Behaviour. *Scandinavian Journal of Psychology*, 19(1):83–89, 1978.

[88] Michael J. Pazzani and Daniel Billsus. *Content-Based Recommender Systems*, chapter 10, pages 325–341. Volume 1 of Brusilovsky et al. [17], 2007.

[89] Michael J. Pazzani, Jack Muramatsu, and Daniel Billsus. Syskill & webert: Identifying interesting web sites. In Dan Weld and Bill Clancey, editors, *AAAI/IAAI'96: Proceedings of the 13th National Conference on Artificial Intelligence, Portland, Oregon, USA*, pages 54–61. AAAI, 8 1996.

[90] Kriengkrai Porkaew and Kaushik Chakrabarti. Query refinement for multimedia similarity retrieval in MARS. In John Buford and Scott Stevens, editors, *MM'99: Proceedings of the 7th ACM International Conference on Multimedia, Orlando, Florida, USA*, pages 235–238. ACM, 10 1999.

[91] Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang, Tao Mei, and Hong-Jiang Zhang. Correlative multi-label video annotation. In Lienhart et al. [71], pages 17–26.

[92] J. J. Rocchio. Relevance feedback in information retrieval. In Gerard Salton, editor, *The SMART retrieval system: experiments in automatic document processing*, pages 313–323, Englewood Cliffs, USA, 1971. Prentice-Hall.

[93] Yong Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrotra. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions*

on Circuits and Systems for Video Technology, 8(5):644–655, 1998.

[94] Ian Ruthven. Re-examining the potential effectiveness of interactive query expansion. In Callan et al. [19], pages 213–220.

[95] Ian Ruthven and Mounia Lalmas. A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2):95–145, 2003.

[96] Hidekazu Sakagami and Tomonari Kamba. Learning personal preferences on online newspaper articles from user behaviors. *Computer Networks*, 29(8-13):1447–1455, 1997.

[97] Gerard Salton, James Allan, and Chris Buckley. Approaches to Passage Retrieval in Full Text Information Systems. In Robert Korfhage, Edie M. Rasmussen, and Peter Willett, editors, *SIGIR'93: Proceedings of the 16th Annual International ACM Conference on Research and Development in Information Retrieval. Pittsburgh, PA, USA, June 27 - July 1, 1993*, pages 49–58. ACM, 1993.

[98] Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. *Readings in information retrieval*, pages 355–364, 1997.

[99] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. *Collaborative Filtering Recommender Systems*, chapter 9, pages 291–324. Volume 1 of Brusilovsky et al. [17], 2007.

[100] Klaus Schoeffmann, Frank Hopfgartner, Oge Marques, Laszlo Boeszoermenyi, and Joemon M. Jose. Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1):018004–1–018004–35, 2010.

[101] Young-Woo Seo and Byoung-Tak Zhang. Learning user's preferences by analyzing web-browsing behaviors. In *AGENTS'00: Proceedings of the Fourth International Conference on Autonomous Agents, Barcelona, Catalonia, Spain*, pages 381–387, New York, NY, USA, 2000. ACM Press.

[102] Ahu Sieg, Bamshad Mobasher, and Robin D. Burke. Web search personalization with ontological user profiles. In Mário J. Silva, Alberto H. F. Laender, Ricardo A. Baeza-Yates, Deborah L. McGuinness, Bjørn Olstad, Øystein Haug Olsen, and André O. Falcão, editors, *CIKM'07: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, Lisbon, Portugal*, pages 525–534. ACM, 2007.

[103] Nikos Simou, Carsten Saathoff, Stamatia Dasiopoulou, Vaggelis Spyrou, N. Voisine, Vassilis Tzouvaras, Yiannis Kompatsiaris, Yannis Avrithis, and Steffen Staab. An Ontology Infrastructure for Multimedia Reasoning. In Luigi Atzori, Daniele D. Giusto, Riccardo Leonardi, and Fernando Pereira, editors, *VLBV'05: Proceedings of the 9th International Workshop on Visual Content Processing and Representation, Sardinia, Italy*, volume 3893 of *Lecture Notes in Computer Science*. Springer, 9 2005.

[104] Alan F. Smeaton, Paul Over, and Aiden R. Doherty. Video shot boundary detection: seven years of TRECVid activity. *Computer Vision and Image Understanding*, 4(114):411–418, 2010.

[105] Cees G. Snoek and Marcel Worring. Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, 2(4):215–322, 2009.

[106] Cees G. M. Snoek and Marcel Worring. Multimodal Video Indexing: A Review of the State-of-the-art. *Multimedia Tools & Applications*, 25(1):5–35, 2005.

[107] Cees G. M. Snoek, Marcel Worring, Dennis C. Koelma, and Arnold W. M. Smeulders. A Learned Lexicon-Driven Paradigm for Interactive Video Retrieval. *IEEE Transactions on Multimedia*, 9(2):280–292, 2 2007.

[108] C.G.M. Snoek, K.E.A. van de Sande, O. de Rooij, B. Huurnink, J.C. van Gemert, J.R.R. Uijlings, J. He, X. Li, I. Everts, V. Nedovic, M. van Liempt, R. van Balen, F. Yan, M.A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J.M. Geusebroek, Th. Gevers, M. Worring, A.W.M. Smeulders, and D.C. Koelma. The mediamill trecvid 2008 semantic video search engine. In Paul Over, Wessel Kraaij, and Alan F. Smeaon, editors, *TRECVid'08: Notebook Papers and Results*, pages 314–327, Gaithersburg, Maryland, USA, 11 2008. National Institute of Standards and Technology.

[109] Amanda Spink, Howard Greisdorf, and Judy Bateman. From highly relevant to not relevant: examining different regions of relevance. *Information Processing & Management: An International Journal*, 34(5):599–621, 1998.

[110] Vassilios Stathopoulos and Joemon M. Jose. Bayesian mixture hierarchies for automatic image annotation. In Boughanem et al. [16], pages 138–149.

[111] Jian-Tao Sun, Hua-Jun Zeng, Huan Liu, Yuchang Lu, and Zheng Chen. CubeSVD: a novel approach to personalized Web search. In Allan Ellis and Tatsuya Hagino, editors, *WWW'05: Proceedings of the 14th international conference on World Wide Web, Chiba, Japan*, pages 382–390. ACM, 2005.

[112] H. Tansley. *The multimedia thesaurus: Adding a semantic layer to multimedia information*. PhD thesis, University of Southhampton, 2000.

[113] Francisco Tanudjaja and Lik Mui. Persona: A Contextualized and Personalized Web Search. *HICSS'02: Proceedings of the 35th Hawaii International Conference on System Sciences, Big Island, Hawaii, USA*, 3:67, 2002.

[114] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. Potential for personalization. *ACM Transactions on Computer-Human Interaction*, 17(1):1–31, 2010.

[115] Chrisa Tsinaraki, Panagiotis Polydoros, Fotis G. Kazasis, and Stavros Christodoulakis. Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content. *Multimedia Tools and Applications*, 26(3):299–325, 2005.

[116] Jana Urban, Joemon M. Jose, and C. J. van Rijsbergen. An adaptive technique for content-based image retrieval. *Multimedia Tools & Applications*, 31(1):1–28, 2006.

[117] Roelof van Zwol, Lluis Garcia Pueyo, Georgina Ramírez, Börkur Sigurbjörnsson, and M Labad. Video Tag Game. In *WWW'08: Proceedings of the 17th International World Wide Web Conference (WWW developer track)*. ACM Press, 2008.

[118] Robert Villa, Nicholas Gildea, and Joemon M. Jose. A study of awareness in multimedia search. In Larsen et al. [68], pages 221–230.

[119] Timo Volkmer and Apostol Natsev. Exploring automatic query refinement for text-based video retrieval. In Guan and Zhang [44], pages 765–768.

[120] Stefanos Vrochidis, Ioannis Kompatsiaris, and Ioannis Patras. Optimizing Visual Search with Implicit User Feedback in Interactive Video Retrieval. In Shipeng Li, editor, *CIVR'10: Proceedings of the 9th ACM International Conference on Image and Video Retrieval, Xi'an, China*. ACM, 2010. to appear.

[121] Meng Wang, Xian-Sheng Hua, Xun Yuan, Yan Song, and Li-Rong Dai. Optimizing multi-graph learning: towards a unified video annotation scheme. In Lienhart et al. [71], pages 862–871.

[122] John Watkinson. *The MPEG Handbook*. Focal Press, 2 edition, 2001.

[123] Stuart Weibel. The Dublin Core: A Simple Content Description Model for Electronic Resources. *Bulletin of the American Society for Information Science and Technology*, 24(1):9–11, 2005.

[124] Diana Weiß, Johannes Scheuerer, Michael Wenleder, Alexander Erk, Mark Gülbahar, and Claudia Linnhoff-Popien. A user profile-based personalization system for digital multimedia content. In Sofia Tsekeridou, Adrian David Cheok, Konstantinos Giannakis, and John Karigiannis, editors, *DIMEA'08: Proceedings of the Third International Conference on Digital Interactive Media in Entertainment and Arts, Athens, Greece*, volume 349 of *ACM International Conference Proceeding Series*, pages 281–288. ACM, 2008.

[125] Ryen White. *Implicit Feedback for Interactive Information Retrieval*. PhD thesis, University of Glasgow, 2004.

[126] Ryen White, Ian Ruthven, and Joemon M. Jose. Finding relevant documents using top ranking sentences: an evaluation of two alternative schemes. In Kalvero Järvelin, Micheline Beaulieu, Ricardo Baeza-Yates, and Sung Hyon Myaeng, editors, *SIGIR'02: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and*

Development in Information Retrieval, Tampere, Finland, pages 57–64. ACM, 2002.

[127] Ryen W. White, Mikhail Bilenko, and Silviu Cucerzan. Studying the use of popular destinations to enhance web search interaction. In Kraaij et al. [66], pages 159–166.

[128] Rong Yan, Alexander G. Hauptmann, and Rong Jin. Multimedia search with pseudo-relevance feedback. In Erwin M. Bakker, Thomas S. Huang, Michael S. Lew, Nicu Sebe, and Xiang Sean Zhou, editors, *CIVR'03: Proceedings of the Second International Conference on Image and Video Retrieval, Urbana-Champaign, IL, USA*, volume 2728 of *Lecture Notes in Computer Science*, pages 238–247. Springer, 2003.

[129] Bo Yang, Tao Mei, Xian-Sheng Hua, Linjun Yang, Shi-Qiang Yang, and Mingjing Li. Online Video Recommendation Based on Multimodal Fusion and Relevance Feedback. In Nicu Sebe and Marcel Worring, editors, *CIVR'07: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands*, pages 73–80. ACM, 2007.

[130] Rui Yong, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega. A relevance feedback architecture for content-based multimedia information systems. In Penny Storms, editor, *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, San Juan, Puerto Rico*, 1997.

[131] Yun Zhai, Jingen Liu, and Mubarak Shah. Automatic Query Expansion for News Video Retrieval. In Guan and Zhang [44], pages 965–968.

[132] Xiang Sean Zhou and Thomas S. Huang. Exploring the Nature and Variants of Relevance Feedback. In Lorretta Palagi, editor, *CBAIVL '01: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, Fort Collins, Colorado, USA*, pages 94–100, Washington, DC, USA, 6 2001. IEEE Computer Society.

[133] Xiang Sean Zhou and Thomas S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, 2003.