Search in audiovisual broadcast archives
Huurnink, B.

Citation for published version (APA):
Huurnink, B. (2010). Search in audiovisual broadcast archives

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In this chapter we review studies of the “human in the loop” of the video search process, and outline how they have led to and motivated the user studies that we perform in the first part of the thesis. We start with a review of some relevant aspects of user studies in the discipline of information retrieval in Section 2.1. In Section 2.2 we zoom in on studies of people searching for video material “in the wild.” This is followed by a review of studies of searchers in controlled experiments in Section 2.3. Finally, we discuss the implications of the studies of audiovisual searchers for the remainder of the thesis in Section 2.4.

2.1 User Studies in Information Retrieval

Studies of information seeking go back as far as 1948, with a survey of sources used by scientists for obtaining scientific literature [138]. Following on from this, a large body of work on studies of searching behavior in general has emerged, which we do not aim to comprehensively review here. Rather, we will focus on aspects of user studies that are relevant to the thesis. Specifically we will review query typologies and transaction log analysis, which are two building blocks of the studies in Chapters 3 and 4. For more comprehensive reviews of the literature on user studies in information retrieval we refer the reader to the overviews provided by, for example, Marchionini [103] on information seeking in electronic environments and Case [17] on methodological approaches.

Query typologies An important result to come out of the increasing number of detailed studies of users’ search behavior is multiple typologies of queries and searches.
Broder [10] describes three query types in the context of web search: informational ("I need to know about a topic"), navigational ("Take me to a specific item or site") and transactional ("I need to purchase or download a product or service"). This typology has served as the basis for other query classification schemes, including those by Rose and Levinson [130], Kellar et al. [90], and Jansen et al. [83]. Smeulders et al. [147] provide a categorization specific to content-based image retrieval queries: target (or known-item) search (when the user has a specific image in mind), category search (retrieving an arbitrary image representative of a specific class), and search by association (search starts with no aim other than to find interesting things). In query typologies, queries are categorized such that the searcher is expected to be satisfied by very different needs for each type. By understanding what types of queries are commonly issued in a search environment, search engines can be designed to ensure they satisfy the different types of needs.

**Transaction log analysis** Users can be hard to access in situ, and as a result the study of user interactions recorded by retrieval systems—transaction log analysis—has become a common method of studying retrieval behavior. Information science has a long history of transaction log analysis, from early studies of the logs created by users of library online public access catalog (OPAC) systems [120] to later studies of the logs of Web search engines [79]. This was followed by the analysis of more specialized search engines and their transaction logs. For instance, Mishne and de Rijke [108] study the use of a blog search engine through a log file analysis and Carman et al. [14] examine the difference between the vocabularies of queries, social bookmarking tags, and online documents. Three frequently used units of analysis have emerged from the body of work: the session, the query, and the term [79]. These three units will also play a central role in our transaction log analysis in Chapter 3.

Over the years, transaction log analysis has proved an apt method for the characterization of user behavior. Its strengths include its non-intrusive nature — the logs are collected without questioning or otherwise interacting with the user — and the large amounts of data that can be used to generalize over the cumulative actions taken by large numbers of users [78]. It is important to note that transaction log analysis faces limitations: not all aspects of the search can be monitored by this method, for example, the underlying information need [127]. Thus, we cannot say with any certainty what a user is looking for on basis of what appears in the transaction logs alone; for this, qualitative methods are required. It can also be difficult to compare across transaction log studies of different systems due to system dependencies and varying implementations of analytical methods. Comparability can be improved to some extent by providing clear descriptions of the system under investigation and the variables used [79]. When two systems are comparable, then we can study differences and similarities between the searchers using the systems.
2.2 Field Studies of Audiovisual Retrieval

In this section we review studies of audiovisual searchers “in the wild.” We divide the review according to two locations in which search for audiovisual material often takes place: on the Web, and in audiovisual archives. There are important differences between web-based multimedia search engines and audiovisual archives: web-based search engines serve the general online public rather than a group of specialist users; they offer access to amateur video of varying quality as well as professionally produced material of the type curated by archives; and the search engines allow users to search on text obtained from web pages or user tags, rather than the manually created catalog descriptions that are maintained by archives.

2.2.1 Web Searchers

Searching for video has become part of daily life for many computer users. We have continuous access-on-demand to unlimited quantities of amateur and professional footage, enabled by the Web.

In one of the earliest studies of online audiovisual search, Ozmutlu et al. [118] performed a transaction log analysis of an Internet search engine. They compared multimedia queries from 2001 to queries from 1997–1999, and found that queries were changing rapidly as web content and searching techniques evolved. In particular, search sessions were becoming shorter with fewer query modifications, while queries themselves contained increasing numbers of terms. Jansen et al. [81] found that multimedia web searching was relatively complex as compared to general web search, with a longer average query length and higher use of boolean operators, in a study of transaction logs gathered in 2002. In a later study of transaction logs gathered in 2006, Tjondronegoro et al. [167] found that multimedia searches used relatively few search terms. In addition, they found multimedia searches to be generally short in duration, with more than 50% of searching sessions being less than one minute in duration. The authors used an open-source classification tool to categorize approximately 2% of queries and found that 64% of video searches are for information about people. In a comparison with logs from earlier years they found that online multimedia search had begun to shift from entertainment to other categories such as medical, sports, and technology. This shows that video search is not a static field; searcher needs are changing, though whether that change is due to changing demographics of online searchers or other factors such as changes in search engine technology is not clear.

Cunningham and Nichols [28] performed one of the few non log-based studies of web searchers for audiovisual material. They studied 48 undergraduate students and their friends. The study included both ethnographic (self)-reports and observa-
tions of search sessions. The authors used these means of analysis to investigate the motivations of their subjects when searching for audiovisual material. The authors found that, in their search for video, these users were often driven by their mood or emotional state, for example being bored and wanting to find videos to pass the time. Other motivations included locating instructional videos, and social reasons (for example, locating videos recommended by friends). They also found video search to be deeply embedded in the day-to-day lives of the users, and to typically be interspersed with a variety of daily activities.

2.2.2 Archive Users

Now we move on from the studies of audiovisual search in the completely digitalized Web environment to studies in incompletely digitized audiovisual archives. All but one of the studies reviewed here are performed in a search environment where an archivist acts as an intermediary between the archive user and the archive content. Though the communication between the archivist and the archive user was, in many cases, digital, the audiovisual material the user was searching for was not. Audiovisual archives have, in general, only recently started to digitize their content on a large scale [187], and as a result large-scale transaction log analyses such as those described for Web users have not been performed. However, a number of specialized audiovisual archives have performed analyses of their users without the aid of transaction logs, instead analyzing by hand requests made to the archives. These requests are generally natural language texts in the form of e-mails etc. The advantage of this approach is that the underlying information need is more explicitly identified than in transaction logs, where searches are limited to short keyword queries.

Audiovisual archives primarily provide access for searchers with specialized information needs, such as for example historians or television producers. Sandom and Enser [135] analyzed 1,270 information requests to 11 (mostly commercial) film archives. They found that approximately 10% of the information requests were for known-items, specifying video titles, films by particular directors, or starring particular actors. These requests were excluded from the analysis. The remaining 90% of information requests specified the desired content or subject of the footage. These were categorized according to the Panofsky-Shatford matrix [119], which classifies different types of visual content. They found that overall 68% of requests included specific content, 56% included generic content, and 2% included abstract content, where a request could contain multiple kinds of content. This was further broken down according to client type. Their final conclusion was that, in the majority of cases, the moving image information seeker was looking for audiovisual material that illustrated specific events and showed named individuals or groups of people, in particular places or on unique dates.
2.3 Laboratory Experiments

Hertzum [59] manually examined 275 e-mail requests to the Deutsche Film Institut, a non-commercial archive of European films. The text of the e-mail requests, which contained on average 111 words, was used to categorize them into search types: known-item retrieval, fact retrieval, subject retrieval, exploratory search, and other. He found that most requests could be classified as known-item retrieval (43%) or subject retrieval (32%) requests. A further analysis of the e-mail requests showed that 75% of the requests specified production-related attributes such as the title, director, or production year of a film. Only 38% of requests specified content, subject, or context attributes. This is in contrast to the study by Sandom and Enser [135], where 90% of the requests studied desired content or subject matter. Hertzum speculated that such discrepancies reflect differences in the requester’s tasks.

Jørgensen and Jørgensen [88] analyzed 685 searches obtained from the logs of a commercial image archive. The archive did not include any audio media, and is therefore not strictly an audiovisual archive. Nevertheless, we include this study in our overview, as it is one of very few studies of commercial archives of non-textual media. The broad goal of their research was to “fill the knowledge gap concerning ‘typical’ search sessions […] and to provide a general picture of searching by a particular group of image professionals.” They found that unique term searches (searches for proper nouns, which refer to a specific concept, in other words, a type of navigational query) were less frequent than other studies in image archives [5, 40] had found. They also found a large amount of query modifications (62% of queries), indicating that often the information need of the searcher was not completely fulfilled by the results of the first query to the archive.

2.3 Laboratory Experiments

Now we turn from studies of searchers in their natural environment to studies of users in controlled laboratory environments. In laboratory experiments aimed at eliciting requirements for experimental video search engines, (potential) searchers are brought into a controlled setting. The impact of external influences on the experiment is minimized by keeping, as much as is possible, all variables constant, and then altering a single experimental variable. In the majority of cases such experiments are aimed at examining the performance and perception of searchers on experimental retrieval systems, rather than commercially available systems.

In order to enable performance oriented evaluations, researchers gather explicitly requested user relevance data — specifically, selected relevant videos for a set of queries [20, 32, 64, 65, 168, 184]. In addition, implicit user data in the form of logged actions may be gathered [20, 48]. It is worth noting that an important source of queries and judgments of relevant videos is the TREC Video Retrieval Evaluation
16

Chapter 2. Audiovisual Searchers

(TRECVID) [146], which is described in detail in Section 6.1.

User perceptions are taken into account by requesting that the searcher describe their experiences interacting with the retrieval system(s). Such subjective data may be gathered by means of questionnaires, interviews, and discussions [20, 32, 64, 65, 168, 184].

A problem when performing laboratory experiments is gaining access to test subjects. Presumably as a consequence of this, there is a clear divide as to the background of users participating in such experiments. The majority utilize students and research staff of academic institutions. In addition, there have been some studies of professional video searchers, and studies employing simulated users. Accordingly, we discuss each in turn below.

2.3.1 Experiments Involving Students and Researchers

When students and researchers are used as test searchers, as part of a study of audiovisual search, the research goal is generally to evaluate the effect of changing variables in experimental audiovisual retrieval systems. This is measured by examining which retrieval system helps the user find the maximum number of relevant pieces of video within a given period of time. Thus the focus here is on improving system design in terms of attained retrieval performance, rather than examining users. Such experiments are valuable for creating effective search systems, but a criticism is that performance is bound to system design, and it is difficult to draw conclusions about what techniques work outside the context of a particular system [146].

In terms of retrieval interfaces, multiple studies have demonstrated that displaying temporally related video segments, as well as video segments related to a query, improves retrieval results (e.g., [33, 56]). De Rooij and Worring [32] found that displaying even more related segments on the screen—for example, according to visual similarity—can improve video retrieval performance, though performance degrades as the number of lists on the screen becomes too large. This was confirmed by feedback from questionnaires. An example of an interface incorporating both temporally related video segments, and other related segments, is shown in Figure 2.1. Hauptmann et al. [56] investigated the effect of presenting users with results very rapidly. Though this mode of video retrieval requires absolute attention from the user, they found that the approach improved video retrieval performance. The searchers’ qualitative perception of this mode of presentation was not reported.

Turning to retrieval algorithms, Hopfgartner et al. [64] found that their searchers preferred a system that provided recommendations on the basis of feedback from earlier searchers, to a system that did not provide such recommendations.

In an experiment aimed at identifying differences between types of users, Halvey and Jose [48] investigated the search behavior and perceptions of expert and novice
users on an experimental retrieval system. They found that researchers with prior experience of experimental (content-based) video retrieval systems tended to perform better in terms of retrieval performance than students with no experience with the systems. In addition, they found that expert users tended to give more negative and critical feedback than novice users, an important factor to take into account when soliciting feedback for new retrieval systems. In a large, cross-site survey, Wilkins et al. [184] found high levels of variation across users in terms of attained retrieval performance. This indicates the difficulty of quantitatively comparing search approaches on the basis of performance when there is a “user in the loop”.

2.3.2 Professionals in the Lab

Christel [20] performed a laboratory experiment with a group of six government analysts, in order to analyze their search preferences. The participants were asked to perform predefined retrieval tasks using different experimental content-based video retrieval systems. These professional users preferred to use a video retrieval system with multiple query capabilities (such as query-by-example and query-by-concept)
Chapter 2. Audiovisual Searchers

over a video retrieval system that only allowed search on the text of transcripts. This is in contrast to their previous studies with non-professional users, described in the same paper. With these non-professional users, a single query capability (query-by-transcript) dominated. In addition, the professional users expressed disdain for tasks that did not correlate well with their daily practice; this disdain appeared to negatively influence retrieval performance. This is an interesting observation given that, in laboratory experiments, retrieval tasks are often taken from benchmarking activities and not defined on the basis of real-world use [33, 48, 56, 64, 184]. In contrast, a study of professional users in a laboratory setting that did derive search tasks from on the basis of real-world use is that of Van den Heuvel [170]. She studied transaction logs of an audiovisual archive and used them to define search tasks for eight broadcast professionals, and found that the search strategies of users were consistent across different tasks.

In Part II of this thesis we will perform quantitative retrieval experiments with experimental retrieval systems, with as our ultimate goal of improving video retrieval in the audiovisual broadcast archive setting. To cross the perceived gap between tasks taken from benchmarking activities and professional practice, we will develop retrieval tasks based on information about the daily practice within the audiovisual broadcast archive. As it is difficult to quantitatively compare search approaches with “users in the loop,” we will separate our retrieval experiments from the searchers, first investigating the searchers and their information needs, and then performing retrieval experiments automatically, i.e., with no user involved.

2.3.3 Simulated Searchers

As laboratory studies with human searchers are expensive to perform, some researchers have explored simulation of users as an alternative. In a simulation, a digitized model of the searcher is used as a proxy for a real user. The advantage of such an approach is that simulated users have infinite patience, and experiments can be repeated as necessary. In addition, the researcher has near complete control over all experimental variables. However, it is not always clear whether the simulated user performs the same way as a real user, and it is not possible to gather subjective information about how a retrieval system is experienced. Here we summarize how users have been modelled in simulation experiments.

De Rooij and Worring [32] simulate an “ideal” user who issues a single query to the system, navigates through multiple result lists that are visible on the computer screen, and selects every relevant result that is encountered. This simulation is used to evaluate the potential benefit of displaying multiple result lists on the screen over displaying a single result list. The authors find that including extra result lists on the screen would give improved retrieval performance, however, in a separate eval-
uation with a group of human users, described in the same paper, the display of many result lists was indicated to be confusing, and including more result lists did not result in increased performance. This highlights a difficulty of simulation experiments; by definition a simulation model reduces the complexity of the real world [136], and there cannot incorporate all aspects of a real-world retrieval system.

Hopfgartner and Jose [63] use simulation to investigate the effectiveness of using implicitly gathered user interaction data such as video playback behavior to inform retrieval algorithms. In the simulation, different types of feedback are assigned to relevant results that are displayed in the simulated interface. Different types of feedback are shown to give different retrieval performance, but there is no relationship back to real users. In later work, Hopfgartner et al. [65] create simulated users to evaluate the performance of a faceted video retrieval system. This simulation goes beyond result browsing, and includes the formulation of new queries by adding frequently occurring keywords in clusters of results. Some simulations are informed by the logged behavior of users from a separate laboratory experiment, which results in improved retrieval performance, but no attempt is made to validate the realism of the simulation model itself.

2.4 Discussion

In this chapter we have reviewed studies of people searching for audiovisual material. Outside of the laboratory, studies of searchers on the Web have been dominated by the analysis of the electronic traces of user behavior recorded by search engines. Studies of searchers in audiovisual archives, on the other hand, have been restricted to small-scale studies of information requests given to archivists. Within the laboratory environment, experiments are performed under controlled conditions, and are often aimed at evaluating user performance on, and perception of, novel audiovisual search systems. When real users are not available or too expensive, simulated searchers may be employed as proxies.

Transaction log analysis allows us to quantify searcher behavior on a large scale, although it is difficult to determine the underlying information intention of the user on the basis of their search words alone. Studies in audiovisual archives have been limited to manual analyses of a requests issued to archivists. In Chapter 3 we will perform a large-scale transaction log analysis of an audiovisual broadcast archive. In this analysis we will take a step towards determining the underlying intention of the user by connecting terms in queries to terms in clicked results, and then using an audiovisual thesaurus to categorize the query terms.

Simulation allows us to study users even when we don't have users; a simulated
searcher has unlimited time and unlimited patience. However, it can be difficult to address the usefulness of simulators when they are not validated against real-world tasks. In Chapter 4 we will create simulators that attempt to reproduce, for retrieval experimentation purposes, searches and the corresponding purchases on the basis of audiovisual broadcast archive catalog data. In order to determine the validity of these simulated searches and purchases, we will compare them to the actual searches and purchases that we witness in the transaction logs.

In the review of related work in this chapter we have placed the emphasis on searchers rather than on search systems, which will be the focus of Part II. However, to ensure that evaluated retrieval tasks reflect real world tasks, observations of searchers are vital to informing task design. Therefore, in Chapter 7 we will define two evaluation collections, one of which is based on observations of searchers in the audiovisual broadcast archive. This collection will then be used later on in the thesis to perform experiments studying the possible impact of new content-based video retrieval techniques on the real world search task.