



SEARCHING FOR VIDEOS

THE STRUCTURE OF
VIDEO INTERACTION
IN THE FRAMEWORK OF
INFORMATION FORAGING THEORY

YNZE VAN HOUTEN

Telematica Instituut
On top of technology.

This publication is part of the Telematica Instituut's Fundamental Research Series. Telematica Instituut is a unique, co-operative venture involving businesses, scientific research institutions and the Dutch government, which carries out research into telematics, a field also known as information and communication technology (ICT).

The institute focuses on the rapid translation of fundamental knowledge into market-oriented applications such as electronic business practice, electronic co-operation and electronic data retrieval. We are also involved in the development of middleware.

In a financial, administrative and substantive alliance, the parties involved work on a joint mission, namely the development and application of high-quality telematics knowledge. By combining forces in an international network of this kind, we will be in a better position to deploy the opportunities of telematics so that ICT can make a contribution to strengthening the innovative capacity of business and industry.

In addition, telematics will fundamentally change and facilitate our way of living, working and even spending our leisure time. In all of these activities, technology remains the servant of man: On top of technology. The Dutch government also recognizes the importance of information and communication technology, and awarded Telematica Instituut the title 'leading technological institute'.

SEARCHING FOR VIDEOS

Telematica Instituut Fundamental Research Series

- 001 G. H. ter Hofte, *Working Apart Together: Foundations for Component Groupware*
- 002 P. J.H. Hinssen, *What Difference Does It Make? The Use of Groupware in Small Groups*
- 003 D.D. Velthausz, *Cost Effective Network Based Multimedia Information Retrieval*
- 004 L. van de Wijngaert, *Matching Media: Information Need and New Media Choice*
- 005 R.H.J. Demkes, *COMET: A Comprehensive Methodology for Supporting Telematics Investment Decisions*
- 006 O. Tettero, *Intrinsic Information Security: Embedding Information Security in the Design Process of Telematics Systems*
- 007 M. Hettinga, *Understanding Evolutionary Use of Groupware*
- 008 A. van Halteren, *Towards an Adaptable QoS Aware Middleware for Distributed Objects*
- 009 M. Wegdam, *Dynamic Reconfiguration and Load Distribution in Component Middleware*
- 010 I. Mulder, *Understanding Designers, Designing for Understanding*
- 011 R. Slagter, *Dynamic Groupware Services – Modular Design of Tailorable Groupware*
- 012 N.K. Diakov, *Monitoring Distributed Object and Component Communication*
- 013 C.N. Chong, *Experiments in Rights Control: Expression and Enforcement*
- 014 C. Hesselman, *Distribution of Multimedia Streams to Mobile Internet Users*
- 015 G. Guizzardi, *Ontological Foundations for Structural Conceptual Models*
- 016 M. van Setten, *Supporting People in Finding Information: Hybrid Recommender Systems and Goal-Based Structuring*
- 017 R. Dijkman, *Consistency in Multi-viewpoint Architectural Design*
- 018 J.P.A. Almeida, *Model-Driven Design of Distributed Applications*
- 019 M.C.M. Biemans, *Cognition in Context: The effect of information and communication support on task performance of distributed professionals*
- 020 E. Fiel, *Designing for Acceptance: Exchange Design for Electronic Intermediaries*
- 021 P. Dockhorn Costa, *Architectural Support for Context-Aware Applications: From Context Models to Services Platforms*
- 022 T. Broens, *Dynamic Context Bindings: Infrastructural Support for Context-Aware Applications*

Searching for Videos

The Structure of Video Interaction in the Framework of
Information Foraging Theory

Ynze van Houten



Telematica
Instituut

Enschede, The Netherlands, 2009

Telematica Instituut Fundamental Research Series, No. 023 (TI/FRS/023)

Cover Design: Studio Oude Vrielink, Losser and Jos Hendrix, Groningen

Cover Photo: Marcel ter Bekke

Book Design: Lidwien van de Wijngaert and Henri ter Hofte

Printing: Universal Press, Veenendaal, The Netherlands

Samenstelling promotiecommissie:

Voorzitter, secretaris: prof.dr. J.M. Pieters (Universiteit Twente)

Promotor: prof.dr.ir. P.W. Verhagen (Universiteit Twente)

Co-promotor: prof.dr. J.G. Schuurman (University of Lincoln, UK)

Referent: dr. P.L.T. Pirolli (Palo Alto Research Center, USA)

Leden: prof.dr. G. Marchionini (University of North Carolina, USA)

prof.dr. E.S.H. Tan (Universiteit van Amsterdam)

prof.dr.ir. A. Nijholt (Universiteit Twente)

prof.dr. F.M.G. de Jong (Universiteit Twente)

ISSN 1388-1795; No. 023

ISBN 978-90-75176-48-3

Copyright © 2009, Telematica Instituut, The Netherlands

All rights reserved. Subject to exceptions provided for by law, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the copyright owner. No part of this publication may be adapted in whole or in part without the prior written permission of the author.

Telematica Instituut, P.O. Box 589, 7500 AN Enschede, The Netherlands

E-mail: info@telin.nl; Internet: <http://www.telin.nl>

Telephone: +31 (0)53-4850485; Fax: +31 (0)53-4850400

SEARCHING FOR VIDEOS
THE STRUCTURE OF VIDEO INTERACTION IN THE FRAMEWORK OF
INFORMATION FORAGING THEORY

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 30 januari 2009 om 15.00 uur.

door
Ynze Abe van Houten
geboren op 16 augustus 1967
te Drachten

Dit proefschrift is goedgekeurd door de promotoren:
prof.dr.ir. P.W. Verhagen en prof.dr. J.G. Schuurman.

Preface

When my wife and I were living in Amsterdam, she applied for a job at the Telematica Instituut in Enschede. When they offered to hire her, she had to confess that there would be some practical problems, as her family was living on the other side of the country. The scientific director at the time, Chris Vissers, asked her about my background. When she answered that I was a researcher in applied psychology, his response was “Oh, we need those here too.” So, in the same week that the National Aerospace Laboratory offered me a permanent appointment after a three-year temporary position, I came home to discover that she had arranged me an interview for a job at the TI.

After coming to grips with this awkward situation, and after studying this potential new employer, I realized that this could be a valuable opportunity. By then I had done years of research at different institutes, on several subjects scattered over different projects, and I really felt a need to focus and specialize. Of course, one of the best ways to accomplish that is to start Ph.D. research, and the TI offered the inviting opportunity to work on that part-time in an applied research environment.

To make a long story short, we are now very happily living in Twente, and this book presents the results of the complex journey that is a doctoral research project. Of course, there is no way I could have done this without the support of others.

First I want to thank my Ph.D. supervisor (promotor), Pløon Verhagen, for his thorough guidance, his broad knowledge, and the enthusiasm which always kept me motivated. Moreover, he has a sense of humor that always made our meetings pleasant events.

Jan Gerrit Schuurman was my loyal, patient, and optimistic coach at the TI. I thank him for the stimulating discussions and streams of ideas, which were mostly based on his erudition.

I thank Mark van Setten, Jaap Reitsma, and Peter Fennema for developing the video editor and browser with which I did my user studies.

Guido Annokkée, Carla Verwijs, and Robert Slagter more than once put up with the role of pilot subject. I want to thank them for their flexibility.

Many thanks go to Bauke Freiburg and Karen van der Moolen of Fabchannel for making the survey related to their website succeed. I really enjoyed working with Fabchannel in the MultimediaN Persis project, in which I performed most of my research. Among other things, it gave me a fascinating inside look at a successful, rapidly growing online entertainment company.

Julie Phillips helped to improve the English in this book. I thank her for her great job and for being so flexible. Any bad English that is left is entirely my own responsibility.

My colleagues at the TI - and specifically the people in the Media Interaction group - provided a pleasant and stimulating work environment. Special thanks go to my officemates Robert Slagter and Niels Snoeck for all the fun and discussions.

As I am unable to function properly without a life alongside my work, I would like to thank all my friends and family simply for being around and making life enjoyable. The weekly sessions with the band BOB provided a lot of fun and distraction during the last tough phases of my research. I specifically want to thank Martijn van Rijn, Elly van der Sluis, and Mettina Veenstra for sharing their love for making music. Thank you to all members of the “Zweden-club,” including Douwe, Astrid, Paul, Jeanet, Hendrik-Jan, Olga, Freya, Stef, and all their children, for the sociable weekends, climaxing every year with Christmas Eve dinner. I particularly thank Douwe Douma, Astrid Oldenziel, and their children for the holidays we share. Moreover, I thank Douwe for being my “paranimf.” I also thank Marcel ter Bekke, Houkje Tamsma, and their children for sharing at least one holiday week every year. What’s more, I thank Marcel for letting me use his fabulous picture of the jumping hare for the cover of this book.

Unable to continue on in their education themselves, my parents, Sjouke and Gé van Houten, have always expressed their trust in me and motivated me to make use of my talents. It is to them I dedicate this book. I also want to thank the rest of my family for all the warm moments throughout the year, and in particular my niece Ymkje Hoekstra for being my other “paranimf.” My parents-in-law, Han and Jannie Veenstra, have been indispensable in the past years, always available to look after the children and our house in demanding times. I’m very grateful to them.

Most importantly, I am very happy and lucky that I have Mettina, Menno, and Ieme to look after me, and the other way around. They are the best and dearest family I could wish for.

Ynze van Houten
Borne, the Netherlands, December 2008

Contents

CHAPTER 1	Introduction	1
	1.1 Video interaction	1
	1.2 Theoretical background: Human information behavior	5
	1.3 Video Foraging	9
	1.4 Research questions and thesis overview	12
CHAPTER 2	Video patches: classifying video content	17
	2.1 Labeling and categorizing video content	17
	2.2 Research Questions	25
	2.3 Kenniswijk survey	27
	2.4 Fabchannel survey	37
	2.5 Conclusions and discussion	43
CHAPTER 3	Scent-following in a video database	51
	3.1 Video scent	51
	3.2 Research questions	61
	3.3 Method	62
	3.4 Results	67
	3.5 Conclusions and discussion	74

CHAPTER 4	Video interaction environments	79
	4.1 VIBES video browser	80
	4.2 YouTube	85
	4.3 Fabchannel	88
	4.4 Conclusion	92
CHAPTER 5	Video-foraging behavior	95
	5.1 Research questions	96
	5.2 Method	98
	5.3 Results	107
	5.4 Conclusions and discussion	131
CHAPTER 6	IFT-based browsing: conclusions and discussion	139
	6.1 Introduction	139
	6.2 Main conclusions	140
	6.3 IFT, browsing behavior, and bridging gaps: what did we learn?	145
	6.4 Limitations of the research	148
	6.5 Concluding remarks and directions for future research	150
APPENDIX A	Kenniswijk survey: questions (in Dutch)	153
APPENDIX B	Fabchannel survey: questions	173
APPENDIX C	Scent experiment: examples of result sets	181
APPENDIX D	Scent experiment: survey versions	187
APPENDIX E	Scent experiment: additional questions	189
APPENDIX F	Browse experiment: task order	193
APPENDIX G	Browse experiment: observation form	195
	Summary	197
	Samenvatting	203
	References	209

Introduction

Given the overabundance of video¹ available, the attention of someone searching for video needs to be allocated efficiently among the video sources. The objective of our research is to study how to support interaction with video in such a way that people can efficiently satisfy their needs. In this chapter, it is explained that we see interaction as a process of bridging gaps. We apply human-information interaction theory to study the problem of video interaction, leading to the concrete research questions described at the end of the chapter and studied in the remainder of this thesis.

1.1 Video interaction

Video plays an important role in our highly visual culture, and we are confronted with it all the time. Currently, people have access to numerous videos that are distributed via high-bandwidth cable or internet connections. According to computer networking company Cisco Systems, the sum of all forms of video (TV, VoD, Internet, and P2P) will account for close to 90% of consumer traffic by 2012. Internet video alone will account for nearly 50% of all consumer internet traffic in 2012 (Cisco Systems Inc., 2008). At the start of 2008 the internet video site YouTube had about 2.8 million user pages and contained over 70 million videos. Every second, 10 hours of video is uploaded to YouTube (Dahdah, 2008). These include home videos made by amateurs, (clips from) movies and TV programs made by professionals, and any other type of video. Another example is the Netherlands Institute for Sound and Vision, which looks after, and releases,

¹ Video here refers to all moving-image technologies created for viewing. Technically, video consists of a number of still pictures (also called *frames*) delivered at a rate giving the viewer the *impression* of seeing a moving image. Additions to the original recording – subtitles, charts, soundtrack, voice-over, photographs, credits et cetera – are part of the video.

70% of the Dutch audio-visual heritage. The collection contains some 700,000 hours of television, radio, music, and film, making it one of the largest audiovisual archives in Europe. Every year, about 10,000 hours of television programs are added to the collection (Beeld en Geluid, 2008). At home, people have hours and hours of video material stored on hard disks or DVDs, including broadcasted television programs that have been recorded either user-supervised or automatically, based on user-profiles. The convergence of TV and the internet is well underway (see, for example, Noam, Groebel & Gerbag, 2004), and new interaction modes are becoming available. Viewing behavior is no longer dictated by the broadcasting schedule. Users have the option of actively selecting content and using/viewing it whenever they want (Brown & Barkhuus, 2006). New digital technologies make it very easy for users to have an abundance of content available. They can interact with the content and personalize the information to their specific needs and preferences.

Of the various sources or channels people have access to, only a part will be relevant or interesting. Even worse, as watching video is very time-consuming, people will only be able to view a very small part of all the interesting video material available. Herbert Simon has remarked that “what information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it” (as cited in Varian, 1995). Providing people with access to more information is not the problem. The central problem is how to maximize the allocation of human attention to information that will be useful to them. The very abundance of digital data intensifies the most fundamental constraint on interaction with information: the limits of human information processing capacity. For media that are based on time – like audio and video – interaction is very cumbersome, worsening the problems. Little research has been done into how people interact with rich content – that is, content other than text. There has been almost no detailed psychological research into how people browse or navigate through hypermedia that include images, video, animations, and so forth (Pirolli, 2003).

1.1.1 Strategies for video interaction

Video interaction starts with allocating a relevant to be watched video, followed by interacting with its content. When seeking information, people can apply two types of strategies (Marchionini, 1995): (a) Formal, analytical strategies, based on planning, use of query terms, and iterative adaptations of the query based on evaluation of intermediate results, and (b) informal,

browsing strategies, heuristic, opportunistic, associated with recognizing relevant information.

The classic analytical approach to information retrieval (IR) is system- and content-driven (e.g. Robertson, 1977). The focus of that approach is to get a best match between the document representations and a user's query, trying to get high recall and precision measures. The assumptions behind this approach are that it is possible for the user to specify precisely the information that he/she requires, and that information needs (or at least expressions of them) are functionally equivalent to information objects. So, if the user is able to specify his/her information need in a query, a good system will retrieve a grand best set of information objects that will fulfill the user's need. Research within the content-centered paradigm typically focuses on the information objects rather than on the people who create, find, and use those objects (Marchionini, 2004).

The other strategy, browsing, is a prevalent form of human behavior that by its iterative and exploratory nature lacks the precision of direct, systematic searching. It has long been accorded less value than direct, precise searching due to the historical bias towards specific, direct searching in library and information science. In 2001, Rice, McCreadie, and Chang indicated that the concepts and nature of browsing had not yet been systematically studied and were thus not yet well understood. Recently, interest in browsing and exploratory search (where querying and browsing generally are combined) has been growing, acknowledging there are many search situations where the target is not well known and a single fact or document will not suffice (White, Kules, Drucker & schraefel, 2006). Researchers from diverse communities, such as information retrieval, user interface design, information visualization, and library sciences have been working on techniques to support browsing or exploratory search.

Since the end of the 1970s there has been more interest in the cognitive processes in information retrieval (for an overview, see Ingwersen, 1999) that should be understood to value browsing as a natural form of human behavior. In contrast to traditional IR research, the cognitive view does not per se regard user behavior as highly logical, well defined, and purposeful. Rather, random action and vagueness are seen as typical elements of retrieval behavior, due to uncertainties and ambiguities. This is reinforced by the fact that users' needs are often difficult to express in verbal form (Taylor, 1968). Belkin (1980) formulated the Anomalous States of Knowledge (ASK) hypothesis, which states that an information need arises from a recognized anomaly in the user's state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly. This may be even more true for non-textual information (such as, for example, video) when a user needs to add an extra translation in the query from images/sounds to text.

The expression of an information need is in general a statement of what the user does not know, so the query will represent an anomalous or in some sense inadequate or incoherent state of knowledge. Users often do not have predefined search criteria (Hildreth, 1982). Belkin, Oddy, and Brooks (1982) have asked why it is necessary for the searcher to find a way to represent the information need in a query understandable by the system. Systems based on the classic approach to IR cannot handle information from the user about doubt, uncertainty, or suspicion of inadequacy in the user's state of knowledge.

1.1.2 Browsing is dominant

Savoliainen and Kari (2006) studied tactics people use while searching the Web, and found that of all tactics used 18.3% were query-related and 81.7% browse-related. According to Belkin et al. (1982), the basic idea is that browsing is the means for users to bridge the gap created by their ASK. The anomaly (or more positively: the need), and the user's perception of the problem, can change with each instance of communication between user and system (Belkin et al., 1982). It seems that information needs are very often ill-defined and not static, but evolving. Information-seeking behavior is characterized by movement from one strategy to another in the course of a single information-seeking episode, as the searcher's problematic situation changes (Bates, 1989). Our everyday life is dominated by these kinds of ill-defined problems, such as choosing a career or finding a good school (Reitman, 1965; Simon, 1973).

Also most tasks on the Web are broad and ill-defined (Pirolli, 2003). The information need cannot be satisfied by a single final retrieved set, but only by a series of selections of individual references and bits of information at each stage of the ever-modifying search. Each new piece of information users encounter gives them new ideas and directions to follow. Furthermore, at each search stage, the user may identify and acquire useful information. This bit-at-a-time retrieval is called *berrypicking* (Bates, 1989), by analogy to picking blueberries in the forest: they do not come in bunches and one must pick them one at a time. This idea emphasizes that the search process is at least as important as the query terms result.

Marchionini (2004) heeds the consequences and speaks of a paradigm shift from information retrieval to information interaction, stressing the role of the human in the retrieval problem and emphasizing not discrete matches but the flow of representations and actions. A person with an information problem is best able to meet that need through action, perception, and reflection rather than through query statements alone. The importance of interaction is confirmed in video retrieval research (Hauptmann & Christel, 2004). Interactive search approaches in video

retrieval (relying heavily on the user's ability to refine queries and reject spurious answers) substantially outperform non-interactive approaches (in which the human merely enters the query into the system).

All in all, information interaction is about combining querying and browsing. When a search task is well-defined and a structured search system is available, analytical search (using queries) is more appropriate than browsing. However, we saw that information needs are more often ill-defined. In these situations, browsing becomes the dominant strategy. Forms of filtering – including querying and recommendations – are still needed to bring down the amount of data to a size that can be browsed. Browsing often includes querying at some phase, while querying is powerless without browsing. In some cases, it is hard to distinguish between querying and browsing, for example when links to information sources can be considered to be in some sense “pre-fab” queries (see also Golovchinsky, 1997). In those situations it might be better to speak of information interaction, or more specifically in this case: video interaction. This includes both querying and browsing, but especially for video interaction we acknowledge the importance of browsing strategies that may help provide access to the non-verbal and time-based properties of video content. A restriction of browsing is that there are physiological-psychological limits (mostly related to attention), and that browsing is only practical for a relatively small set of objects (for example, performance accuracy falls off rapidly between 100 and 200 image examinations (Marchionini, 1995).

This leads to the general question how and when to maximize the allocation of human attention to information that will be useful to the users. This is an efficiency question. For video interaction it means that the more efficiently people can get access to video content, the more people will be enabled to watch video material of interest to them per unit of time. Efficiency is thus also a prerequisite for effectiveness. It is this need for support for efficient video interaction that is the object of our research.

1.2 Theoretical background: Human information behavior

The information-seeking approach, based on a problem-solving perspective of human behavior, has been the dominant approach within the field of library and information sciences. Wilson (1999) provides an overview of models and theories in information science research. He distinguishes between models of information seeking and models of information searching (although this distinction is not consistently applied in the

literature). Models of information seeking describe the purposive seeking of information in relation to a goal, and are concerned with the variety of methods people employ to discover and gain access to information resources. Models of information seeking include Wilson's (1981; 1999) model of information-seeking behavior, Ellis's (1989; 1993) behavioral model of information-seeking strategies, and Kulthau's (1991) model of the stages of information-seeking behavior. Information-searching behavior is a subset of information-seeking behavior, one that is concerned with the interactions between the information user and (computer-based) information systems. Information-searching models include Ingwersen's (1996) cognitive model, Belkin's (1995) 'episode model', and Saracevic's (1996) 'stratified interaction model'. The traditional model – already described above – represents IR as a two-prong set (system and user) of elements and processes converging on comparison or matching. Where information-searching behavior is a subset of information-seeking behavior, the latter is a subset of a larger research area called information behavior, which is "the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking and information use"(Wilson, 2000).

Wilson's own problem-solving model (1999) proposes an integration of some of the models above. His model identifies four stages in the information-seeking process: problem identification, problem definition, problem resolution, and solution statement or presentation. The information search in this model begins with a need that is perceived by the information user that is referred to as uncertainty, a gap, or an anomalous state of knowledge. This model suggests that information-seeking behavior is goal-directed, with the resolution of the problem and, possibly, the presentation of the solution as the goal. The question, however, is how well models like these work for general goals like "I want to have fun" or "I want to relax." When people interact with content (e.g., surfing the internet or zapping through television channels), the interaction itself may be the center of interest without a solution being present. The problem-solution perspective underestimates the importance of the search process, and has problems with more non-academic and less-formal information seeking behaviors.

As an alternative to pragmatic and cognitive approaches, Dervin (1992) presents a sense-making theory based on communication theory. In the sense-making approach, humans are conceived as hard-wired theorizers about their world, but because they live in a world of continuous discontinuity they must continuously make new theories. When a gap in sense under an old theory develops in the individual's world, the individual tries to make new sense, thus creating a new theory. To bridge these gaps in our day-to-day lives we must have enough information to make sense of the

whole. The total situation of the user is considered, and that is why some of the uses the user puts into bridge construction do not involve information seeking at all but rather such things as gaining the emotional assurance and trust needed to continue making the journey through the time-space point.

The idea of bridging gaps can also be found in the work of John Searle. When a person initiates a certain action (whether spontaneously or from a prior intention), the psychological antecedents do not automatically determine what this person is going to do, or what the action is going to be. There is a gap between the "causes" of the action (desires, beliefs) and the "effect," that is, the action (Searle, 2001). The gap is that part of our conscious decision-making and acting where we sense *alternative* future decisions and actions as causally open to us. There are at least three gaps that need to be bridged by searching the information environment: 1) a gap between reasons for a decision and the decision; 2) a gap between the decision and the initiation of the action; and 3) a gap between the initiation of the action and the continuation and completion of the action.

People's interaction with the information environment can be framed as a bridging of the gaps. We can use the gaps to classify the problems people face when interacting with video. The other way around, we can evaluate solutions by checking how much support they provide for bridging the different gaps. The gaps define the problem space of the users, and can be considered to describe different types of interaction contexts. In order to bridge the gaps, people interact with their information environment.

One important factor that determines the characteristics of the interaction in the problem space is the cognitive "tools" (see also Gigerenzer & Todd, 1999) people have available for bridging the gaps: tools that people use to structure their environment and interact with that environment. We can define these tools in terms of information foraging theory or IFT (Pirulli & Card, 1999).

As we saw above, most information science theories, but also most psychological theories (e.g. problem-solving, decision theory) talk about goals and means to reach those goals. IFT describes how people adapt to their environment, respond to what they encounter during the process of searching, and form their goals along the way and on the fly. IFT states that people forage through an information environment in search of a piece of information that associates with their interests the way animals forage for food. For the user, the information environment has a patchy structure (compare berries scattered on berry bushes, or websites on the World Wide Web). Within a patch, a person can decide to forage the patch or switch to another patch. Users make navigational decisions guided by scent, which is a function of the perception of value, cost, and access path of the information with respect to the goal and interest of the user. Perceived scent is influenced by the design of "scent carriers": representational

elements in the information environment that relate to sought-for information. The forager is constantly adapting decision making and direction. People prefer information-seeking strategies that yield more useful information per unit cost, and they tend to arrange their environments (physical or virtual) to optimize this rate of gain. People prefer, and consequently select, technology designs that improve returns on information foraging (Pirolli, 2003).

Information foraging behavior typically occurs when a person is in a certain intentional state. Foraging is a concept used in evolutionary psychology but has found its way in information science, for example in the berrypicking model of Bates (1989). Ideas from IFT can be used for the design of information-searching environments, and as such it also qualifies as a human-computer interaction theory. The basis of IFT, though, lies in optimal foraging theory from anthropology and biology. Optimal foraging theory is concerned with the “searching efficiency” of cognitive systems, both human and non-human, for food and mating opportunities in the environment. Cognitive systems evolve towards stable states that maximize gains of valuable information per unit cost. The evolution toward such a stable state is constructed by the human forager through a process of constructing effective foraging patterns and continuously fine-tuning or adapting these patterns to the ever-changing environment. IFT takes an adaptationist approach. Users are viewed as complex adaptive agents who shape their strategies and actions to be more efficient and functional with respect to their information ecology (Pirolli, 2003).

As stated at the start of this chapter, in our research we acknowledge that most search problems are ill-defined, that the target is not always known, and that a single fact or document will not suffice. We acknowledge the importance of the search *process*, especially in new information environments and in the case of ‘rich’ content like video. We think IFT is the most promising theory to describe searching behavior in situations where simple queries don’t suffice. We agree that in many search situations people adapt to their information environment, respond to what they encounter during the process of searching, and form their goals along the way and on the fly. This emphasis on adaptation is a strong characteristic of IFT, and is less prevalent in other theories (S.K. Card & P. Pirolli, personal communication, October 7, 2003). Moreover, as we will see later, IFT provides very useful concepts for developing and evaluating information environments. Acknowledging the importance of the search process, we refine the outlook of Marchionini on information interaction. We define the interaction contexts in terms of bridging the three gaps as defined by Searle, allowing us a closer look at the process of searching.

Concluding, in our research, we try to render an account of information interaction by refining the outlook of Marchionini (2004). We try to

explain information interaction behavior on the basis of human search principles as described in IFT (Pirolli & Card, 1999), and define the interaction contexts in terms of gap bridging as defined by Searle (2001). We do this for the specific case of video. The object of our research is to study how to support interaction with video in such a way that people can efficiently satisfy their needs. We hope that our research will demonstrate the feasibility of our approach to studying the video interaction problem, and to designing and evaluating video interaction environments.

1.3 Video Foraging

Following the approach outlined in the IFT framework as stated above, regarding video interaction this research will focus on the preferred structure of the video environment (video patches), the way navigation through the environment is supported (video scent), and the problems people have to solve (bridging gaps).

1.3.1 Video patches

The patchy structure of the information environment can be observed at various levels. For example, the World Wide Web consists of portals, web sites, web pages, and parts of pages. Each of them can be considered patches, and people switch between patches at the same level via hyperlinks, or go up and down in the hierarchy. Search tools provide result sets which can be seen as newly created patches, where the individual result items are interrelated by having the same keywords (the ones used in the query). Video works the same way. Each individual video is a patch, often containing a narrative structure and consisting of a number of smaller segments. At a higher level, groups of videos are also patches. The best examples of these are TV-related patches: video on the same TV channel or from the same broadcasting company. But all videos with Bill Murray or from Monty Python or made in Japan are also video patches. Even the group of videos that my friend likes, or those that are all on my hard disk, are patches. The concept “patch” as a structural unit is very broad. Whenever a user has a reason to consider objects as belonging together, they form a patch. Patches provide a structure that is user-based, and as such it is a broader concept than classification, which is often more document-based. The most important thing, of course, is that – from the user’s point of view – those patches have meaning and are usable or pleasurable in some way. People structure their environment in patches, and ideally the information environment is designed in such a way that the patches in it match the patches people have in their minds.

Within a video we can also distinguish patches. Video patches can be defined as collections of video fragments sharing a certain characteristic (van Houten, van Setten & Schuurman, 2003), e.g. they contain the same meaningful elements or have a narrative relationship. From the IFT point of view a video is a construct that generates scent, with components that also generate scent. A video can be looked at as a database containing individual video fragments (Manovich, 2000). The original narrative of the video is “only” one out of many ways of organizing and relating the individual items. So, the original video - as the video maker intended - is a specific kind of video patch. Other patches are subsets of fragments from that video or from a video collection. A simple example of a within-video patch would be the highlights of a football game. On a website like YouTube.com, people upload homemade compilations, such as the highlights of a football player’s career. People often want to structure the information environment in their own way, so that the “decodings are likely to be different from the encoder’s intended meaning” (Hall, 1980).

The attributes that “glue” together segments into video patches may vary along many dimensions. The important thing is that they are useful for the end-user. Examples of patches may be all videos/fragments about a certain subject, such as politics; containing a certain person, for example Damien Hirst; related to a certain event, such as the Football World Cup; with songs in a certain language, such as Norwegian; recommended by my friend Paul; containing a large area of blue; and so on. Patches can form a heterarchic patchwork or a hierarchy, and several combinations of attributes can be combined in a patch. Selecting a patch gives the user a specific view on the content. In a video environment where patches are created, patches provide a means to filter video content, as users can browse a patch and ignore video fragments not belonging to the patch. Video fragments within the patch will at the same time probably belong to a number of different patches. Links to these patches can be shown to the user. This will allow users to switch to other patches when their evolving information need, as they browse, gives them new ideas. Such patches form a hyperlinked network above the video data that can be browsed (van Houten, Schuurman & Verhagen, 2004). For these links to be effective they need to be expressed in attributes/tags that carry the appropriate scent of the related patches.

1.3.2 Video scent

IFT describes how people make the decision to forage a patch, or leave a patch to find another one. These decisions are guided by the scent that is perceived. When there is a match between (associations with) elements in the information environment and (associations with) the user’s goals or

interests, the elements give off scent. People adapt their scent-following strategies to the flux of information in the environment. If the scent is strong, the information forager can make the appropriate choice. If there is no scent the forager can only perform a “random walk” through the environment, or quit altogether.

Scent can be found within an information source as well as in links and metadata that refer to that source. When users scan the information sources themselves (e.g., when switching TV channels), they can decide to stay at that source based on the scent in the small sample they were watching. When the scent in the source is low, users may still decide to watch the source when the scent in the link or metadata related to the source is high (it is a movie by a favorite director, or it was highly recommended by a friend).

Hyperlinks are representations or abstractions of the information sources, providing cues (see also Gigerenzer, 2000) which more or less tell the users what they will find at the destination. The scent of hyperlinks is this remote indication of an information source, which is also called *residue* (Furnas, 1997). Scent is wafted backward along hyperlinks – the reverse direction from browsing. People make navigational decisions on the basis of perceived scent: they follow links with good scent (from their point of view). The design of the links and metadata related to information sources – the *scent carriers* – can influence the perceived scent and thus the decision to watch a source or not.

Stored past experiences are retrieved, based on proximal features of the current context (with links to information sources), and then used to predict the likelihood of distal features (what can actually be found in the information sources) (Pirolli, 2003). So, the user’s cognitive task is to predict the likelihood of the desired distal information from the proximal cues available in the user interface. If we want to make perceived scent measurable, we have to measure this subjective likelihood. Scent can be measured by, for example, asking users to rate how confident they are before they click on a link.

1.3.3 Bridging gaps

Scent following and interaction with patches are subject to a series of decisions that can be looked upon as actions to bridge gaps in the way that has been described by John Searle. When a person has a reason for a certain action, this does not automatically determine what this person is going to do, or what the action is going to be. There is a gap between the “causes” of an action (desires, beliefs) and the “effect” (the action) (Searle, 2001). There are three gaps that need to be bridged in an information environment: (a) a gap between reasons for a decision and the decision, (b)

a gap between the decision and the initiation of the action, and (c) a gap between the initiation of the action and the continuation and completion of the action.

For the first gap, beliefs, desires, and other reasons are not experienced by the searcher as causally sufficient conditions for a decision. They merely determine a state of mind within the information ecology, leaving the decision to act open. The information environment should make the related action possible. In terms of IFT, the environment should contain accessible patches that allow for desired actions. If we, for instance, for some reason have the desire to look for a video to watch, appropriate video patches should be available to help us put that desire into action. TV guides and web environments such as YouTube are examples of suitable patches for this purpose.

Bridging the first gap is not a causally sufficient condition for an intentional action. For example, watching a desired video may require so much effort that it hampers the action. The scheduled broadcasting time may be difficult to meet, or it may be difficult to locate a video (fragment) in a database. This sets the conditions for the second gap, which, in the case of video databases, can be bridged by querying and browsing available video data. In terms of IFT, the interface of the browsing environment should help to bring that scent of a video item to the surface that will help the decision to start watching that one item out of all items in the concerned patch.

The third gap lies between the initiation of the action and its continuation and completion. Starting an action does not set sufficient conditions for its continuation or completion. For example, watching TV/video may continue as long as the video matches your information need. However, while you are watching you may get new ideas that trigger a decision to stop watching and do something else, such as go look for related video material. There is, thus, a gap between the actual information that is being watched and the desire for other information that meets the dynamically determined need of that moment. One way of bridging this gap is by offering links to video items that meet the requirements of the modified need of the user, who may then decide to sustain viewing. In terms of IFT, video items that meet these requirements share properties by which they form patches.

1.4 Research questions and thesis overview

In our framework of IFT, efficient video browsing takes the form of optimizing video patches and their related scent in a browsing structure that

supports decision-making in Searle's three-gap decision model. The general research question of this thesis is:

How to support interaction with videos in such a way that people can efficiently satisfy their needs?

We divide this general research question into three specific research questions as described below.

The first research question looks at how to optimize video patches. In IFT-based browsing, the purpose of organizing video patches in a browsing structure is to support users in their interaction with videos. This support will be optimal when there is a match between the structure of the environment and the psycho-semantic structures of the users. We may expect that users will be able to move around within that environment most efficiently when the way they structure or classify their environment corresponds most closely to the way the environment itself is structured or designed. This leads to the question of what categories of video comply with users' preferred way of selecting and interacting with video content:

Research question 1: What is the most useful way to classify video content?

We studied this issue by asking users about their preferences for video categories that may serve to organize patches. We conducted two exploratory survey studies to collect data on user preferences for video categories that may serve to organize patches: the Kenniswijk survey and the Fabchannel survey. An important difference between the two studies was that the Kenniswijk survey was very large and generic, asking about TV/video viewing behavior and preferences in general (it also provided data on scent and gap-bridging behavior which were useful for the following studies). The Fabchannel survey was very specific, asking a specific user-group about their preferred interaction with videos on a dedicated website with videos from one genre. We expected that exploring user preferences in this way would yield valuable insights about classifying and structuring video for patch-based browsing. The two studies are described in Chapter 2 of this thesis.

The second research question is about the character of good scent. Scent is contained in scent carriers, representational elements by which video items are made known to the potential user. Scent carriers take the form of links, metadata, video fragments, and whole videos. The question is which forms scent carriers should take to establish the most realistic expectations about video content:

Research question 2: What is the character of good video scent?

We studied this in an experiment in which we asked participants to select the most relevant link to a video from a group of links. We measured the perceived scent by asking for the subjective probability that the information that was needed could be found behind that link. We repeated this for different types of tasks and different types of scent carriers to study the influence of these factors. This experiment is described in Chapter 3 of this thesis.

The third research question examines design principles for a patch-based browsing environment that effectively bridges the three gaps and efficiently supports video data browsing:

Research question 3: How to design a video interaction environment that will optimally support its users?

Optimal support is reached when patches and scent carriers together support the bridging of all three gaps at a rate that maximizes user satisfaction over (search) time. Based on the results of the previous studies, we refined an experimental browsing application that had been in development for a number of years: the VIBES video browser. The idea of patch-based browsing was developed as a first implementation of IFT in video browsing (Van Houten, Van Setten & Schuurman, 2003). The practical development of that environment also gave rise to research questions about video patches and video scent as described in van Houten, Schuurman & Verhagen (2004). We used the results of the user studies described above to further develop the experimental video application, whose main goal was to provide a context in which to study browsing behavior within the IFT framework. This application is described in Chapter 4 of this thesis, together with two other popular video environments on the internet: YouTube and Fabchannel. These three applications are described from an IFT point of view. The description can be seen as part of the method section of the experiment described in the following chapter.

We conducted an experiment in which we asked participants to perform a number of tasks with the VIBES video browser. This resulted in a quantitative analysis of the usefulness of the elements of the application. In addition, we asked the participants to perform tasks with the Fabchannel and YouTube websites. This provided data for a qualitative analysis of the difficulties of video interaction in specific and general situations, and of which support is most wanted for interacting with video. This experiment is described in Chapter 5 of this thesis.

In the final chapter of this thesis (Chapter 6) we will summarize the conclusions of all four user studies. Next we will discuss the success (or lack thereof) of our approach. We will try to determine the usefulness of

applying the framework of IFT and gap-bridging to the problem of video interaction. Can we use it to explain human searching behavior, and can we use it to create or evaluate video interaction applications? At that point we will evaluate how well we have answered our main question: *How to support interaction with videos in such a way that people can efficiently satisfy their needs?*

Video patches: classifying video content

In this chapter, we will deal with the research question “What is the most useful way to classify video content?” We present results from two surveys on how end-users would prefer to structure the video environment into video patches: the Kenniswijk survey and the Fabchannel survey. First we will discuss the research literature on labeling and categorizing video content, and formulate more specific research questions. We will end this chapter by discussing the results from the surveys in the light of the research questions.

2.1 Labeling and categorizing video content

People use concepts to classify perceived information through the process of cognition. Concepts are a kind of mental glue, in that they tie our past experiences to our present interactions with the world (Murphy, 2002). People carve up the world into “uniformities” relating to concepts they use when extracting information from the environment (Searle, 1978). The categories humans impose on the world are dependent upon human “individuation” capacities (Devlin, 1991). The facility to individuate objects – that is, to see them as objects - is a fundamental cognitive ability. The world doesn’t come to us already sliced up into objects and experiences: what we see as an object is a function of our system of representation, and how we perceive the world is also influenced by that representational system. Objects are not self-identifying - the world divides the way we divide it. Patches as defined in a database are expected to be most relevant when they match with patches as the user would define them. The content of patches should thus be organized in forms that are meaningful to the user and that allow users to control how they select and navigate them. To this

end there is a need for a predetermined set of semantic concepts that can act as semantic filters and aid in video interaction (Naphade & Smith, 2004). Navigation through video data is affected by the new forms of video interaction that have been made possible by the advent of digital video. Interaction with units smaller than the video itself is one of the characteristics of digital video and is still a relatively new phenomenon. Semantic concepts can be used to describe videos as a whole, or to describe smaller video segments. In this study we also try to determine the preferred unit of interaction.

Essential for all classifications of video content is knowing in what kinds of semantic concepts users would like the video environment to be structured. Before going into what we learned about that in the two survey studies, we first present an overview of current ideas on labeling and categorizing video content.

2.1.1 Adding metadata to videos

Classifying videos involves adding descriptions or metadata to video material. There are two main ways to add metadata to videos: manually (or supervised) and automatically (or unsupervised). Manual addition of metadata can be performed by various different agents, such as a professional (e.g., a librarian or other content expert), the author of the video object, or the user. If professionals are used to add metadata, the disadvantages are the need for training/education, the relatively high cost in time and effort, and a large scalability problem. Still, it is feasible for a small subset of video objects, e.g., videos that represent an important part of a country's cultural heritage. It is, nevertheless, a problem which classification scheme the professional should use.

The author has the same problems regarding time-consumption, but can have special motivations to add metadata to a video he/she created, for example to increase the chance that it will be found by others. The user can implicitly add metadata to an information object by viewing it, citing it, or linking to it, behaviors that can be detected by algorithms for ranking videos (e.g., on the basis of number of views), or relating objects to each other (e.g., on the basis of how often they are watched by the same persons). Users can also explicitly add metadata to video objects by rating or reviewing the objects, or by adding descriptions or tags ("social tagging"), thus creating a "folksonomy." Folksonomy combines the words "folk" and "taxonomy" and is a type of distributed classification system. A folksonomy begins with tagging. People tag websites (e.g., del.icio.us), photos (e.g., Flickr), videos (e.g., YouTube), et cetera, to be able to find them again. When other people are tagging the same objects, the cumulative force of all

the individual tags can produce a bottom-up, self-organized system for classifying large collections of digital material (Mathes, 2004).

Unsupervised metadata creation concerns the use of algorithms for automatically detecting content characteristics. The fact that it requires no human time or effort is very advantageous, but it has the problem of the semantic gap. This is produced by the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation (Smeulders, Worring, Santini, Gupta & Jain, 2000). Clearly, higher-level semantic descriptions are often more useful than low-level properties such as color and texture, but automatic classifiers for such high-level features are much less accurate than those that detect low-level features (Sebe, Lew & Smeulders, 2003). Promising developments include multimodal analysis techniques, using data from the visual, auditory, and textual modality (Snoek & Worring, 2005). Still, unsupervised techniques have difficulties with conceptual, content-descriptive metadata. There is also the critique that the research is focusing too much on core technology and not enough on making it work in practice. There is lack of understanding of which semantics are important and what breadth and depth of the semantic space is required for enabling effective search (Smith, 2007).

Both the manual and automatic methods of metadata creation require some kind of guidance in the form of a classification system. The multimedia research community has identified the need to find a set of semantic concepts to focus on as it explores new automated tagging techniques (Naphade et al., 2006). This provides interoperability and lets the multimedia community focus ongoing research on a well-defined set of semantics. In the past years, the approach to research on multimedia semantics has been ad hoc, without larger coordination. Recently, an initiative has started to standardize the set of semantics for (unsupervised) tagging multimedia: the Large-Scale Concept Ontology for Multimedia or LSCOM (Naphade et al., 2006). The goal was to create a taxonomy of 1,000 concepts for describing broadcast news video. Preferably, such a classification system would relate to what people are actually looking for in videos, so concepts were partly chosen based on analyses of video archive query logs. However, one important criterion for inclusion of concepts was whether automated extraction considering a five-year technology horizon was feasible.

Moreover, concepts had to be observable. This way, concepts such as flying airplanes and riots were included, but concepts such as discovery and happiness were excluded as unobservable and infeasible. The current view is that, with fewer than 5,000 of these concepts, LSCOM is likely to provide high accuracy results, comparable to text retrieval on the web, in a typical broadcast news collection (Hauptmann, Yan & Lin, 2007). However, the

researchers leave unanswered the question of which specific concepts should be used.

For the description of audiovisual content, there are well-accepted standards such as MPEG-7 and TV-anytime (see, for example, Tsiriraki, Polydoros, Kazasis & Christodoulakis, 2005). The descriptors may refer to the whole multimedia content (programs, videos, et cetera) or to parts of the content (segments).

Research on automated tagging provides a good impression of possible relevant categories, but we think it is good to study what users really need without being hindered by issues such as technological feasibility. In our approach, we are aware of but choose to avoid the discussion on how metadata are added: by professionals, by machines, or by users. We try to get information about how people classify video content by asking them directly what kinds of video material they would like to watch. We think this users' point of view can be of added value to the discussion of metadata and the way video data should be classified.

2.1.2 **Classifying (images and) videos**

Metadata concern different types of information that are associated with videos (Del Bimbo, 1999). First, there are data which are not directly concerned with video content but are in some way related to it (content-independent metadata). Examples include Dublin Core elements such as creator, publisher, and date, but also how many times a video is viewed and the rating it has gotten. Second, there are data which refer directly to the content of the video. Del Bimbo (1999) distinguishes between data referring to perceptual facts like color, texture, and motion (content-dependent metadata), and data referring to content semantics (content-descriptive metadata). Classifying or grouping video can happen at each metadata level. As we stated above, any reason to group videos can be applied in order to create video patches. The question is which ways are really useful.

Currently, two methods for grouping results of a query are quite popular: clustering and faceted categorization (Hearst, 2006). Both search interfaces are applied and used primarily in domain-specific collections. Clustering refers to the grouping of items according to some measure of similarity, typically using associations and commonalities among features, where features are typically words and phrases. Advantages are that it is fully automatable, can reveal interesting trends, can clarify and sharpen a vague query, and works well for disambiguating unclear queries. For example, a query for "Ajax" on Clusty.com distinguishes results related to the mythological Greek hero, the web application technique Asynchronous JavaScript and XML, the Amsterdam football club, and so forth.

Disadvantages of clustering include lack of predictability, conflation of many different dimensions, the difficulty of labeling the groups, and the counterintuitiveness of cluster subhierarchies.

A faceted classification system allows the assignment of multiple classifications to an object, enabling the classifications to be ordered in multiple ways, rather than in a single, pre-determined taxonomic order. It makes information access useful, by providing multiple navigational paths to any one item of information. In contrast to a folksonomy, the information in each of the facets can be organized into a hierarchy (for instance, the location facets could be divided by state, then city, then neighborhood). Also, folksonomies are emergent properties of social tagging systems in which individuals apply “tags” as they please, without control or coordination; faceted systems require someone to make a decision about which facets to record in the database and, often, which values will be permitted.

Faceted categories are usually created manually, although assignment of documents to categories can be automated. Hearst (2006) advises against creating one large category hierarchy, and recommends hierarchical faceted categories (HFC): a set of category hierarchies each of which corresponds to a different facet (dimension or feature type) relevant to the collection to be navigated. Each facet has a hierarchy of terms associated with it, and each item in the collection can be assigned many labels from the hierarchies. Navigating within the hierarchy builds up a complex query. HFC-enabled interfaces are preferred over the standard keyword-and-result listing interfaces (Yee, Swearingen, Li & Hearst (2003). Disadvantages are that in most cases the category hierarchies are built by hand and that automated assignment of categories to items is only partly successful. Moreover, the categories of interest must be established in advance, which means they might not reflect important trends that arise in the data. An important question remains: which categories are really useful for people?

Many ideas about the semantic level have been developed for image classification. Panofsky (1955) distinguishes three levels of meaning of a picture: the pre-iconographical description (addressing what an image shows in generic terms - e.g., a woman, a building), the iconographical analysis (addressing the specific subject matter of an image, with people or places named or identified: Madonna, the Eiffel Tower), and the iconological interpretation (addressing the symbolism of the image: hope, salvation, love). Based on these levels, Shatford (1986) categorized the subjects of pictures as “Generic of”, “Specific of”, and “About” For further classification she defined four facets at each level, containing the answers to the questions Who? What? When? and Where? For example, a picture (or, for that matter, a video) may be of the 1974 World Cup football final between the Netherlands and West Germany (What – Specific of), while

also being of a football game (What – Generic of), or being an event with lots of excitement and emotion (What – About). This so-called Panofsky/Shatford model has become widely used in image classification. Jaimes and Chang (2000) extend this model by splitting What into What Action and What Object, and adding a Why, thus creating a 3x6 “semantic information table” where each “W-question” (who, when, et cetera) has a specific, generic, and abstract version. It is, nevertheless, sometimes difficult to handle different levels of specificity. For example, at the generic level we can say that a person is a woman, and at the specific level that she is a politician. But how can we then define a specifically named woman—say, Hillary Clinton? To handle such cases, it may be useful to introduce another level. In our study, we used a level “famous” as a special form of specific. For example, a general location is a city, a specific location is a museum, and a famous location is the Van Gogh Museum.

Of course, for video, we also have to be aware of the audio component: dialogue (human voice), direct/ambient sound, organized sound (music), and sound effects. Relationships between individuals and audio are especially important: who or what is making a sound. Much research in the area of image classification is, nevertheless, relevant for classifying videos because of the similarities between the two media types - image and video - regarding classification problems. Video descriptions often relate to content that can be found in frozen frames, while image descriptions often strongly suggest temporal characteristics: a picture often has a past, “What happened before this point in time,” and a future, “What will happen afterward”). The temporal characteristic in video is, evidently, not just suggested, but is actually present. This leads on one hand to the possibility of more accurate descriptions, but on the other hand to more cumbersome interaction. Regarding descriptions, the main difference between these visual media lies in the presence or absence of audio. In our study, we included categories related to audio.

Enser and Sandom (2001) analyzed 1,270 requests for moving image footage received by eleven representative film collections, using the Panofsky/Shatford model. Most requests (1,148) were subject requests; the remainder were requests for particular titles, directors, actors, shot types and the like. More than half of the subject requests comprized two or more facets. Of all subject requests, 1,143 included specifically named people, events, places, or times, and 852 requests were for generic subjects. Requests for abstract subjects were very unusual (22). In the conclusion and discussion section of this chapter we will take a more detailed look at their results and compare them with the results of this study.

The LSCOM taxonomy mentioned above for describing newscast video (Naphade et al., 2006) organized the (almost) 1,000 concepts into six categories on a top level: objects (e.g., flag, animal, computer),

activities/events (walk/run, explosion, natural disaster), scenes/locations (office, court, mountain), people (crowd, face, police), graphics (maps, charts), and program categories (weather, entertainment, sports). The gathering of concepts was partly driven by input from end-user communities, but was mostly driven by practical considerations for the multimedia research community.

Markkula and Sormunen (2000) observed and interviewed journalists trying to retrieve photos from a digital newspaper archive containing over 83,000 photos. In nearly half of the search topics, the main focus was on the context information (e.g., a particular news event) or on themes requiring high-level human reasoning. Low-level visual features were not expressed as the main search criteria in any of the search topics analyzed. Based on the observations, the researchers emphasize that browsing is an essential strategy in searching for photos, and advise the use of conceptual indexing based on broad categories to find browsable query sets (in other words: patches).

In an experiment with 30 participants, Hollink, Schreiber, Wielinga, and Worrying (2004) analyzed and classified textual descriptions of images and keyword queries for images. The majority of the descriptions (85%) were conceptual (as opposed to perceptual). Within the conceptual level, 74% of the descriptions were general, 16% specific, and 9% abstract. Descriptions of objects appearing in images were used twice as much as descriptions of scenes (the set of all objects and their arrangement, e.g., city, landscape, indoor, or still life).

Eakins, Briggs, and Burford (2004) tested the extent to which people were interested in different content categories of images. They used a taxonomy of image content, which was based on an extensive survey of the computer science, art history, and psychology literature (Burford, Briggs & Eakins, 2003). They distinguished the following categories: 1) Perceptual primitives – the lowest level of visual content (e.g., edges, texture, color, sharpness); 2) Geometric primitives – the simplest structures (lines, curves, recognizable geometric shapes) 3) Visual relationships – the spatial arrangement of objects in a scene in two dimensions + visual extension (the arrangement of objects in the third dimension); 4) Semantic units – the names of objects, or classes of objects, present in a scene, either general (“horse,” “sand”) or specific (“Abraham Lincoln,” “the Eiffel Tower”); 5) Abstraction – content which is not directly present in the image, but needs to be inferred from background knowledge and experience.

Within abstraction they distinguished four levels: 5a) Contextual abstraction – non-visual information derived from knowledge of the environment (e.g., day or night scene); 5b) Cultural abstraction – aspects that can be inferred only with the help of specific cultural knowledge (e.g., the religious significance of a scene); 5c) Emotional abstraction – emotional

responses evoked by an image; 5d) Technical abstraction – aspects requiring specific technical expertise to interpret (e.g., signs of injury in an X-ray). The final category was 6) (Content-independent) Metadata – terms describing the image itself (such as size, type, and creator).

When subjects were asked to rate the importance of the categories, the following order emerged: 1 semantic (specific); 2 semantic (general); 3 sharpness; 4 cultural abstraction; 5 technical abstraction; 6 metadata; 7 contextual abstraction; 8 color; 9 shape; 10 texture; 11 visual relationships; 12 visual extension; and 13 emotional abstraction. Clearly the ability to retrieve images by their semantic content is a priority for users of image databases, while lower-level issues are generally considered less important.

Research on facets of metadata for image search and browsing showed that a category-based approach is a successful way to provide access to image collections (Yee, Swearingen, Li & Hearst, 2003). In this study, 32 art history students had to perform image search tasks with a faceted category interface and a baseline interface. The baseline interface was a keyword-based image search interface, closely resembling Google Image Search. The faceted navigation interface enabled users to navigate along conceptual dimensions that described the images. The latter interface was strongly preferred by the participants. Facets were created bottom-up by converting (semi-automatically) the descriptions of art images into a set of metadata categories by comparing the words in the descriptions to their higher-level category labels in an electronic lexical database (WordNet). The facets that were thus created were high-level categories that were considered useful for students and scholars of art history. They included: media (e.g., book, costume), location (e.g., Europe, Australia), date (e.g., 14th century), themes (military, religion), nature (birds, mammals), places and spaces (bridges, buildings), people (aristocrats, children), shapes, colors, and materials (decorations, metal), and artists. When participants had to perform structured tasks (for example, “Gather materials for a scholarly essay on woodcuts created in the US”), facet usage was driven largely by the task content, causing participants to focus on date, location, media, artist, and theme. For unstructured searches (e.g., “Search for images of interest”), usage was more evenly distributed across all the facets. Most searches started with selecting Artists (17%), Date (15%) and Location (15%).

The general consensus in the empirical research on images is that people are mostly interested in meaningful, conceptual information. Different results are found regarding how specific or general the information should be. This seems to be influenced by the reasons people have for searching: a specific search task will require specific descriptions, while a general search task (“Find something that is interesting to you”) may require less specific

descriptions. Abstract descriptions seem to be less useful than both general and specific descriptions.

Most classification research was aimed at images, and the question is whether a video environment provides comparable results. As seen in the studies mentioned above (e.g., Yee et al.), the reasons why people search seem to have an influence on the preferred classification. All studies focused on the usage of a specific image collection. The second study we describe – the Fabchannel study – also explores the interaction with a specific collection, in this case a group of videos. However, we think it is also important to understand the reasons why people interact with video “in the field,” meaning outside a laboratory situation and not related to a specific collection. This was researched in the first study we describe: the Kenniswijk study.

Moreover, people already have strategies available for finding videos and interacting with them. It is important to know more about these strategies before going to the real classification question. Video interaction is more cumbersome than image interaction because of the time factor. Therefore, the question of how to structure videos may be related to the level of interaction: between videos within a database, between video segments within a video, or between video segments within a database. For the specific case of video we think it is necessary to study the classification problem in combination with these interaction issues.

2.2 Research Questions

This research is aimed at finding the best form of support for people as they interact with video material. So, the research questions were formulated in a practical way, in terms of users’ actual searching behavior. First, we want to know more about the searching process, about why people want to view certain videos, and what their efforts are to get to that point. In other words, we want to know what they do to bridge the three gaps as described in Chapter 1. Second, we take a look at user preferences on how that process of video interaction can be improved. We start with the structure of a database as a whole, how people prefer the database to be structured for easy access. Next we look at other possible ways of interaction: which parts of video do people prefer to interact with and what are the characteristics of those parts? This leads to the formulation of the following research questions:

RQ1 - What are the main reasons people start a TV/video viewing session?

For the Kenniswijk survey, we created a list of 25 reasons based on several studies of users' motives for watching TV/video (including Hirschman, 1987, and Lee & Lee, 1995). To check how this corresponds to program preferences, we created a list of 80 TV/video program types (current affairs, documentaries, soap operas, science fiction movies, etc.) and asked the respondents to check which programs they were interested in. For the Fabchannel study, we created a list of specific reasons in cooperation with employees of Fabchannel who were familiar with the motives and wishes of visitors to their website as a result of email correspondence, and so forth. In terms of the theory articulated by Searle (2001), "reasons" should be understood as "prior intentions."

RQ2 - What kinds of efforts do users invest in

- a) choosing a video to watch? (gap-1 problem)*
- b) finding and watching the video of their choice? (gap-2 problem)*
- c) continuing or completing their video viewing session? (gap-3 problem)*

For questioning viewing behavior in the Kenniswijk survey, we asked about the actions people perform to bridge the gaps as described by Searle (2001). So, several questions explored the reasons to view TV/video, how a decision to watch a program is made, how the decision is executed, and how the viewing action is continued or completed. As the experience with video on the internet was very limited for this group of respondents at the time of measurement, we will restrict the analysis of these research questions to the television situation. In the Fabchannel survey, we had no specific questions about gap-bridging behavior.

RQ3 - What is the preferred way to structure the video collection for easy access?

In the Kenniswijk survey, we asked about the preferred way to structure TV program information, and about different approaches to TV channels (mixed or theme-based). In the Fabchannel survey we asked how the respondents would like to sort the concerts in the archive.

RQ4 - What is the preferred unit of interaction?

We formulated questions to find out whether respondents were satisfied with the current unit of interaction, that is, programs as a whole. In the Kenniswijk survey, we did so for TV programs in general, and in the Fabchannel survey for the specific genres of live music concerts.

RQ5 - What is the most useful way to classify video content?

For the Kenniswijk survey, we created a list of 30 aspects of video content based on the Panofsky/Shatford model and the extension of Jaimes and Chang, in combination with ideas from related research on image and video classification. We included categories related to both images and sound. Most people know that DVDs offer the option of jumping straight to a particular chapter or scene of a movie. This category occupies a particular position because it relates to the “natural” syntax of a video. We asked the respondents to which aspects in video they would like to be able to jump instantly (bridging gap 2), and of which aspects in a video they would like to see more while already watching (bridging gap 3). In the Fabchannel survey, we looked at specific aspects of music concerts, and asked about the usefulness of these aspects in searching for similar parts within one concert or in other concerts.

If we can get good answers to the final two research questions, this will provide useful guidance to the characteristics of useful video patches. The first two research questions are a prerequisite for the latter three, as they provide the user context: what users want and what interaction problems they face.

2.3 Kenniswijk survey

The first survey we conducted was within the Dutch Kenniswijk project, a national project that was carried out in the Eindhoven area. Kenniswijk was an experimental neighbourhood around 40,000 families (84,000 inhabitants) in which consumers were given access to innovative products and services in the fields of computers, (mobile) communication, and internet. The survey was conducted at the end of the project, in 2005.

2.3.1 Participants

Inhabitants of the Kenniswijk area were approached via an advertisement in an email newsletter to participate (anonymously) in the survey. It is unknown how many of the inhabitants actually read the newsletter. The survey was presented as “research on TV and video viewing: the way you approach it now, and what you would like to change about it”. Respondents could win one of 25 tickets (five with a face value of €12.50, twenty of €2.50) in the national lottery. To enter the drawing they had to leave an email address.

2.3.2 Method

A survey was developed using an online survey tool (for the tool, see SurveyWorld, 2008). The advertisement in the newsletter contained a link to the online survey. Respondents could start participating directly. The survey contained several questions on demographics, viewing behaviour, and viewing preferences. The exact formulations of the analyzed questions are presented in the results section below. All survey questions are presented in Appendix A (in Dutch).

2.3.3 Response

A total of 215 people responded to the survey. Five were left out of the analyses because of having too many missing data (no response to a whole group of questions related to one of the research questions). The mean time it took to complete the survey was 24 minutes. The remaining 210 respondents consisted of 60% men and 40% women (figures for the Netherlands as a whole: resp. 49% and 51% [CBS, 2005]). Divided by age, 20% of the respondents were 16-35, 50% were 36-55, and 30% older than 56. (It was difficult to get comparable figures for the Netherlands because the central bureau for statistics, or census bureau, uses different categories. But of all people of 20 years and older, 35% were 20-40 years old, 46% were 40-65, and 19% were over 65 [CBS, 2005]). Divided by education level, 20% had completed (or were working on) education at the primary or lower-secondary level, 36% at the higher-secondary level, and 44% at the tertiary level (figures for the Netherlands: resp. 34%, 41%, and 25% [CBS, 2005]). So, this sample can be considered fairly representative for the Dutch population, although male, middle-aged, and highly educated people were slightly overrepresented.

On average, respondents reported watching 19.3 hours of TV/video images per week. They mostly watched television programs at the time of broadcast (71% almost daily). At least 30% watched recorded programs at least weekly. Almost 12% watched streaming video from the internet at least weekly, and 8% downloaded videos at least weekly. About 10% watched their own DVDs/videos at least weekly, and 4% rented a DVD/video at least weekly. Watching home videos or visiting the cinema was even more infrequently done.

The results of this survey were gathered in 2005, hence the modest role of video on the internet (YouTube went live while this survey was running). Of all respondents, 60% had never watched streaming video from the internet, 65% had never watched a downloaded video, and 76% had never used a search engine to find videos on the internet. In other words, interacting with video on the internet was a relatively infrequent activity for this group of respondents at that point in time.

2.3.4 Results

The results for each of the research questions were as follows.

RQ 1 - What are the main reasons people start a TV/video viewing session?

Respondents had to indicate on a 10-point scale how important a reason was for them to watch TV/video. The main reasons were “to keep oneself informed” (score 8.2) , “to relax” (7.4), “to learn, better understand something” (7.3); “to see something I had heard about (e.g., current affairs)” (6.9), and “to have food for thought; out of interest” (6.5). All other reasons scored in the lower half of the scale (below 5.5). In sum, the main reason is cognitive benefit, followed by mood improvement. This is confirmed by the TV/video programs the respondents preferred to watch. Of the top ten, eight were informative and related to the cognitive benefit reason. (The top three were news programs [96%], current affairs programs [82%], and the weather [66%.]) The other two in the top ten were movie genres (comedy and crime/thriller) relating to the mood improvement reason.

RQ 2 - What kind of efforts are invested by users in

- a) choosing a video to watch? (gap-1 problem)*
- b) finding and watching the video of their choice? (gap-2 problem)*
- c) continuing or completing their video viewing session? (gap-3 problem)*

The first gap (or interaction context) is the gap between the reasons for a decision and the actual decision to watch a certain TV program. We asked the respondents how hard it is to find a TV program that is of their interest. Regarding television programs, 56% of the respondents indicated that this was easy, 30% said it was sometimes hard, sometimes easy, and 14% said it was hard. A problem with these numbers is that most people are partly familiar with what television has to offer, and routinely view a number of programs: 67% said they often/always turn on the TV to watch a program they always watch (20% sometimes, 13% seldom/never). In these situations, people will hardly experience the first gap.

There are two main strategies for bridging the first gap: consulting an overview or guide (“metadata-first”), or browsing video data (“data-first”). These can be considered as basic heuristics (Gigerenzer & Todd, 1999). For the specific television situation, we asked people to respond to a number of propositions. To the proposition “Before I turn on the TV, I consult the TV guide to see whether there is something of interest to me”, 40% indicated often/always, 28% sometimes, and 31% seldom/never. To the proposition “I turn on the TV and start channel-surfing to see whether there is

something of interest to me,” 25% indicated often/always, 30% sometimes, and 45% seldom/never. Comparing the responses to these two propositions for each respondent, we can conclude that 50% were more guide-users than zappers; 33% were more zappers than guide-users; 8% applied both strategies equally; and 8% seldom used either strategy (but, for example, always watched the same programs). While 19% actively didn't like to consult TV guides (21% liked it, and 60% were neutral), 42% actively didn't like channel-surfing (20% liked it, and 39% were neutral).

In sum, of the two mentioned strategies, consulting a guide was the most popular. The most popular source for getting program information was a (printed) program guide, which was used by 75%, followed by the teletext² signal of a channel (43%), a newspaper (33%), and an electronic program guide (16%). When asked how successful the respondents were in finding an interesting program with the aid of a program guide, 66% said they were often/always successful (29% sometimes, 5% seldom/never). In cases where people were not successful, the main reasons were “There are too few interesting programs” (68%) and “The program guide does not provide enough information to determine whether a program is of interest to me” (24%).

The second gap (or interaction context) is between the decision and the initiation of the action: the decision is not automatically followed by an intentional action. This gap can be bridged by, for example, entering keyword queries, browsing video indexes, or browsing video data. For example, you know exactly which TV program you want to watch but can't locate it. When people are watching TV this is seldom a problem, as most people know where to locate the TV channels of their interest. We asked the respondents whether they were satisfied with the way their TV channels were programmed, and 79% were satisfied, 19% thought it could be improved but were used to it, and only 2% had difficulties finding channels. While they zap channels, viewers bridge the second gap automatically when they stumble upon a program that is of interest.

The third gap (or interaction context) is about the continuation or completion of the action. We asked people which strategies they applied to finding other interesting programs once they have been watching television for a while. A large number of people - 39% - often/always consulted a program guide, 41% sometimes; 20% seldom/never. Another popular strategy was channel-surfing: 26% did that often/always, 44% sometimes, and 31% seldom/never. Less popular strategies were asking someone else in

² Teletext information is broadcast in the vertical blanking interval between image frames in a broadcast television signal. It is closely linked to the PAL broadcast system and nearly ubiquitous across Europe. Common teletext services include TV schedules, regularly updated current affairs and sports news, simple games (such as quizzes), and subtitling for the deaf or in different languages.

the room (70% seldom/never, 26% sometimes, 4% often/always) or watching the overview channel that displays a mosaic of images from a number of channels (95% seldom/never, 5% sometimes, 0% often/always).

Summarizing, consulting guides and channel-surfing were two important strategies for finding TV programs of interest, with the “metadata-first” option (using guides) being slightly more popular than the “data-first” option (channel-surfing).

RQ3 - What is the preferred way to structure the video collection for easy access?

We asked the respondents about useful structures for a program guide. They had to give a rating between 1 (totally useless) and 10 (very useful). The most useful way to structure a program guide was by channel (score 7.0), next by theme and by time period of broadcast (both 6.2), and finally as a result of a query (5.9).

We also asked which type of information the respondents used to find programs of interest. The overview of movies (very common in TV guides) was used often/always by 36%, sometimes by 32%, and seldom/never by 33%. Other overviews based on themes or genres (less common in TV guides) were used often/always by 18%, sometimes by 29%, and seldom/never by 53%. When asked how useful it would be to place programs of the same theme/genre in a separate overview, 65% gave this a usefulness score between 6 and 10, and 35% between 1 and 5 (mean score was 6.1).

Regarding channels, we asked the respondents to give a rating for different types of channels. Channels gathered around a theme (e.g., nature, sports, or science and technology) got a mean rating of 6.5, while channels with programming in different themes/genres (like most classic TV channels) got a mean rating of 6.8. These mixed channels were watched by most respondents: 93% indicated that they regularly (at least once a month) watch the Dutch public channels, 89% the Dutch commercial channels, 61% the regional channels, 57% the Belgian, 44% the British, and 27% the German channels. Regarding theme channels, 72% regularly watched “knowledge” channels (such as Discovery and National Geographic), 31% news channels, 27% sports, 21% music, and 19% children’s channels. (All other types of channels were watched regularly by fewer than 10% of the respondents). Of all respondents, 93% indicated that they regularly watch 15 or fewer channels (58% fewer than 10).

When asked what type - in terms of specificity - of theme channels people prefer, most respondents (42%) had no preference. Of the remaining respondents, most (32%) preferred general theme channels (e.g., sports, drama). Fewer respondents (15%) preferred more specific channels (e.g., football, comedy). Even fewer respondents (7%) preferred even more

specific channels (e.g., Dutch football, American sitcoms), and still fewer (5%) preferred highly specific channels (e.g., PSV [a Dutch football club] or “Friends” [an American sitcom]).

In sum, the respondents were in some ways quite conservative in their preferences regarding patches. They preferred to watch mixed channels and program information by channel. Still, a significant part of the respondents preferred (or also preferred) program information gathered around a theme, as well as (general) theme channels. For an important part of the respondents, this apparently eases the task of bridging the gaps.

In light of the results presented above, we will now see what are good ways to structure the video environment, or in other words, to create useful patches.

RQ 4 – What is the preferred unit of interaction?

When users structure their environment, the question is what is perceived as belonging together, and what is perceived as being separable. Historically, videos (e.g., television programs) were presented as a whole, but nowadays other types of interaction are starting to emerge. The question is whether users prefer to be able to interact with units smaller than complete programs.

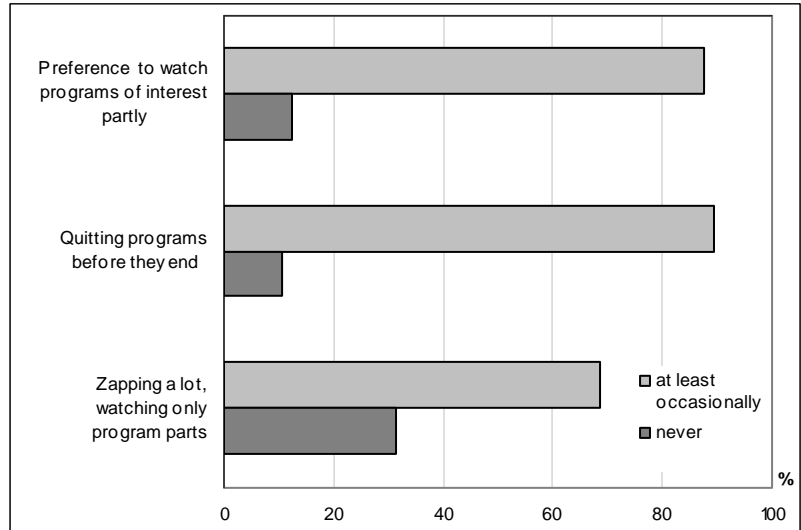
In the survey, we asked the following question: “Of all programs that have your interest, how many would you prefer to watch only certain parts of instead of viewing the whole program?” The average response to this question was 29.2%. About one-eighth (12.3%) answered 0%, indicating that a great majority (87.7%) have a number of programs they are interested in but want to watch only partly (see Figure 2-1). More than a quarter of the people (29.0%) indicated that they prefer to watch 50% or more of the TV/video programs of their interest only partly.

To the question “Once I start watching a program, I will watch it till it ends,” 10.5% stated always, 63.8% often, 20.5% sometimes, and 5.2% seldom (0% said never). In other words, 89.5% do quit in the middle of programs they were watching (see Figure 2-1).

To the question “I zap a lot, and I don’t mind watching only parts of programs,” 31.4% stated never, 32.9% seldom, 25.7% sometimes, 8.6% often, and 0.5% always. In other words, 68.6% of the respondents exhibit this viewing behavior at times (see Figure 2-1).

Because we were not referring to videos of a specific type or genre in the Kenniswijk survey, but to videos in general, the results are very general. For specific genres, other results may be found. As explained earlier, due to the time of measurement the results mostly apply to TV programs. These generally have a longer duration than most videos that can be found on, for example, YouTube.

Figure 2-1 Reported viewing behavior



In sum, there clearly is a preference for being able to select smaller units than the TV/video program as a whole. This has consequences for the way viewers interact with content and bridge the gaps. The question is which elements of TV/video programs are of interest to the viewers.

RQ5 - What is the most useful way to classify video content?

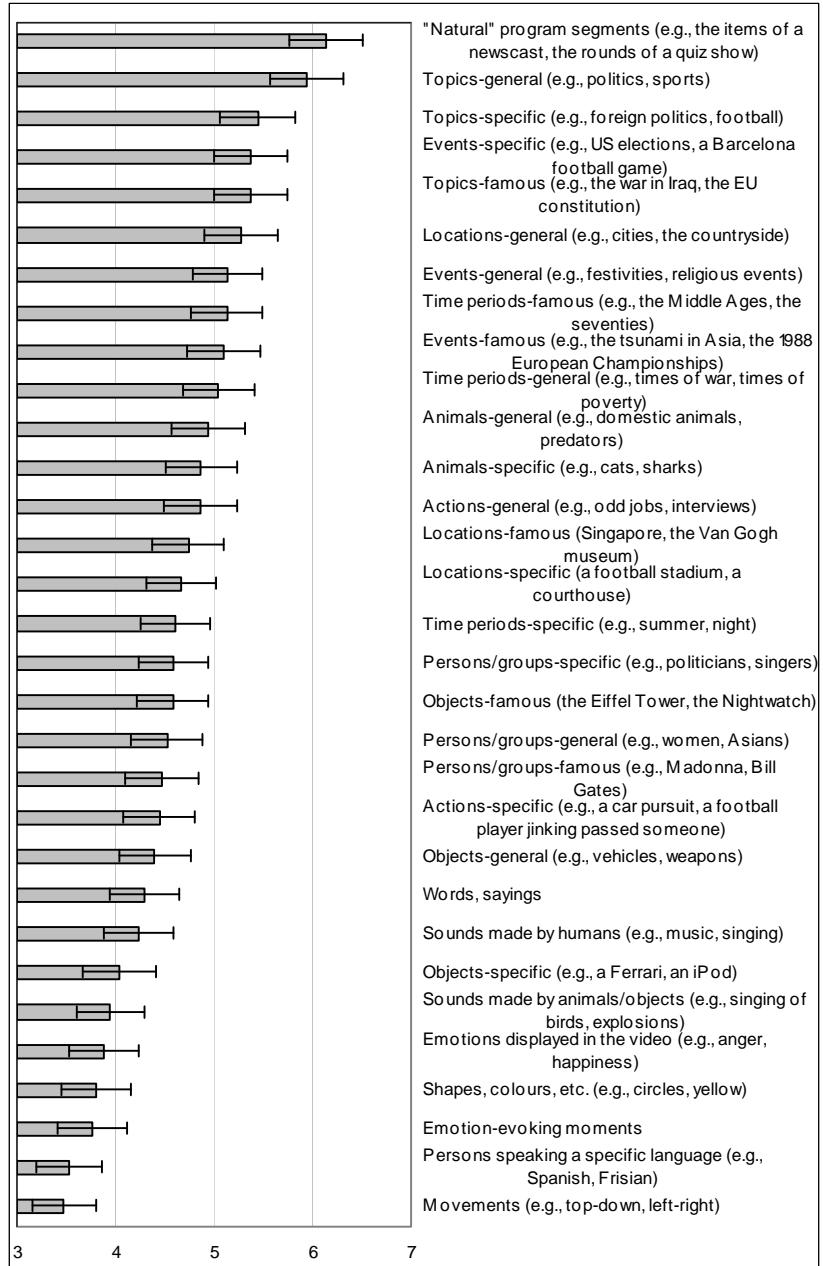
We asked the respondents to which aspects in video they would like to be able to jump to instantly (bridging gap 2), and of which aspects in a video they would like to see more while already watching (bridging gap 3). People could give ratings on a 10-point scale for the subjective usefulness of the proposed aspects.

Figure 2-2 shows the results to the question “To which aspects in the video would you like to be able to jump instantly?” where respondents had to indicate on a 10-point scale how useful it would be per aspect (1=very useless; 10=very useful).

First of all, it is notable that the mean overall rating was 4.66 on a scale of 1 to 10, and only two average scores had a pass mark (above 5.5). We don’t know the exact reason for this. It could be that this question was difficult to understand and/or difficult to answer in a general sense. The two aspects with a pass mark were “natural” program segments and topics-general. In other words, the respondents tended to agree with the semantic segmentation as applied by program makers. This indicated that this is an important unit for interaction, perhaps even more so than programs as a

whole (see research question 3). Segmenting videos into their semantically meaningful units, and first of all describing these in terms of their topic, will help people to bridge gap 2 (see research question 2b).

Figure 2-2 Aspects in videos people would like to be able to jump to, instantly showing average score (1=totally useless; 10=very useful) and 95% confidence interval



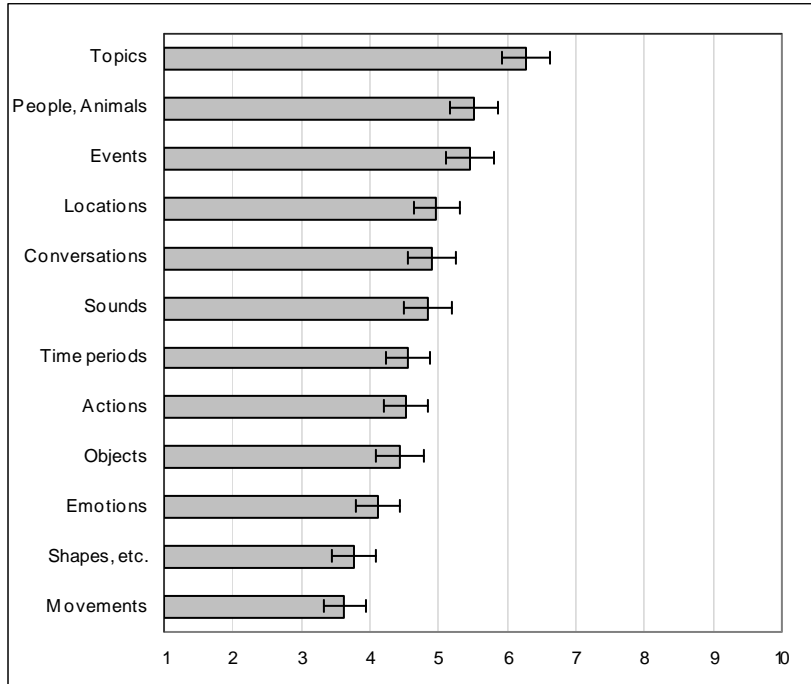
Furthermore, all events (specific, general, and famous) scored higher than 5 on a 10-point scale, as do general locations and famous and general time periods. Almost all concepts scored between 4 and 5. Typically, animals scored higher than people. There was no clear explanation for this. The typical audio concepts (speech and sounds) scored lower than most concepts that are more visual. However, this result should be interpreted with caution as with video most concepts (can) have an auditory as well as a visual component. The lowest scoring aspects included the perceptual properties of the (moving) images: shapes, colors, patterns, arrangements, and movements, a confirmation of the results of earlier studies such as Hollink et al. (2004) and Eakins et al. (2004). A result from the latter study that was also confirmed was the low position of emotions, both those that are displayed and those that are evoked.

Regarding the distinction between general, specific, and famous variants of an aspect, highest scores were found for general topics, general locations (both specific and famous locations scored lower), and general actions (specific actions scored lower – there was no category for famous actions). Further, famous time-periods scored higher than specific time-periods and famous objects scored higher than specific objects. For persons/groups, animals (with no famous category), and events, there were no differences between the categories.

The next question about classifying video content was put in the context of a situation in which a person is already watching television and wants to continue the session (so it typically was about bridging gap 3). While watching TV/videos, a viewer can get new ideas as his/her information need evolves, so he/she wants to be able to find related video material. We asked the respondents the following question: “Suppose, you are watching television, and you suddenly think: ‘I would like to know more about this.’ Which aspects would you at some time like to see more of?” Respondents gave a rating for “the same or comparable aspect, or other information about it.” Respondents gave ratings between 1 and 10, where 1 meant “I never want to see more of this” and 10 meant “I always want to see more of this.” Figure 2-3 presents the results.

Figure 2-3 largely confirms the results of the previous question, about to which aspects in the video respondents would like to be able to jump instantly. Clearly, respondents most preferred to see more of the same or comparable topics. Persons, groups, and animals had a slightly higher position than with the “jump” question (Figure 2-2), while events scored equally well. Only “topics” and “people/animals” scored a pass mark (above 5.5). Least popular – again - were emotions and the low-level, non-semantic features: shapes, colors, patterns, arrangements, and movements.

Figure 2-3 Aspects of television programs people would like to see more of (average rating [1 = “I never want to see more of this”; 10 = “I always want to see more of this”] and 95% confidence intervals)



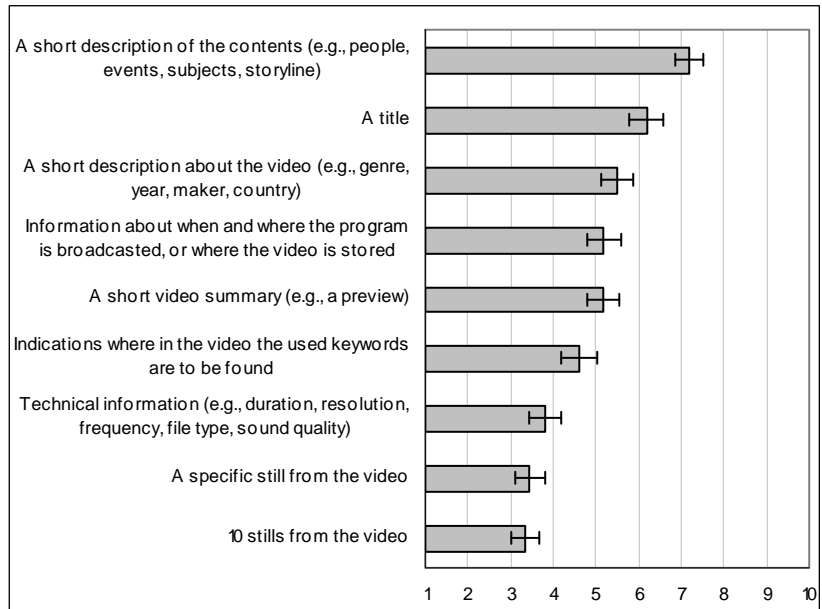
Additional question

One question referred to the usefulness of scent carriers of videos. We present the result of this question at this point, first, to keep the results of the Kenniswijk survey together (“within the Kenniswijk patch”), and, second, because the result was used in the preparation of the experiment described in the next chapter.

We asked the respondents: “Suppose you conducted a search for videos, and received over a thousand results. Which pieces of information about a video would be useful in helping you decide whether a video is interesting to you or not?” Respondents had to indicate the usefulness on a scale of 1 to 10 (1 = “totally useless”; 10 = “very useful”). Figure 2-4 shows the result of this question.

As previously stated, this result from the Kenniswijk survey is relevant to the research question dealt with in the next chapter, and will also be discussed in that chapter.

Figure 2-4 Usefulness (scored on a scale from 1 to 10) of pieces of information about a video (average rating and 95% confidence intervals)



2.4 Fabchannel survey

The second survey we conducted was within the community of visitors to the Fabchannel website (Fabchannel, 2007). Fabchannel broadcasts, on the internet, videos of concerts, festivals, competitions, and lectures from two Amsterdam concert halls (Paradiso and the Melkweg). They offer one of the biggest online concert archives in the world, and have thousands of visitors each day.

2.4.1 Participants

Respondents were gathered in two ways: First, visitors who had commented on the website in the past (for example, to add or correct data about a performer) were approached via email. Second, near every concert in the database, a link to the survey was placed with the request to help improve the Fabchannel website. No reward was offered for participating in the survey.

2.4.2 Method

As with the previous study, a survey was developed using an online survey tool (for the tool, see SurveyWorld, 2008). The survey contained several

questions on demographics, music preferences, and visiting behavior regarding the Fabchannel website, along with specific questions regarding preferred interaction with the Fabchannel concert database (specific questions will be described in the following sections). All survey questions are presented in Appendix B.

2.4.3 Response

A total of 427 people completed the survey, which took – on average – six minutes. Of the respondents, 87% were men and 17% women. Divided by age, 29% were 20 or younger, 39% were 21-30, 18% were 31-40, and 15% were 41 or older. As this concerns a music site, we also determined the respondent's music preference, using the dimensions of Rentfrow and Gosling (2003). We asked people to indicate how often they listen to music from different music-preference dimensions on a five-point scale (1 = never; 5 = always). Respondents listened most often to music that is “Intense and Rebellious” (includes the genres rock, metal, punk, reggae, and alternative), mean score 3.9; then “Reflective and Complex” (jazz, trip-hop, singer/songwriter, blues, folk, classical), mean score 3.4; then “Energetic and Rhythmic” (hip-hop, soul, dance, R&B, ska), mean score: 2.7; and finally, “Upbeat and Conventional” (pop, country, religious, sound tracks): mean score 2.6. Most respondents (56%) visited the site at least a couple of times weekly, 91% at least monthly.

We assume the sample was representative for visitors to the Fabchannel website, although a bias could be expected towards the more frequent and loyal visitors: as there was no reward, the only motivation to participate was to help Fabchannel to find out how to provide better service. The sample is not representative for the population as a whole: men and people younger than 30 were overrepresented. Moreover, respondents could be considered to be music fans with a particular preference for music in the Fabchannel database. On their website, Fabchannel describes the music it offers as “rock, hip-hop, folk, avant-garde, and everything in between”, and writes that this music is not always played on “popular radio and television” (Fabchannel, 2008). However, we do believe that the mechanisms we are interested in can also be studied with this particular group of people.

2.4.4 Results

RQ1 - What are the main reasons people start a TV/video viewing session?

The main reasons the respondents visited the site were to “find a specific concert in the archive to watch” (mean score 4.2 on a five-point scale [1 = never; 5 = always]), and to “discover new music”(3.2). These are very different reasons, requiring different support for interaction.

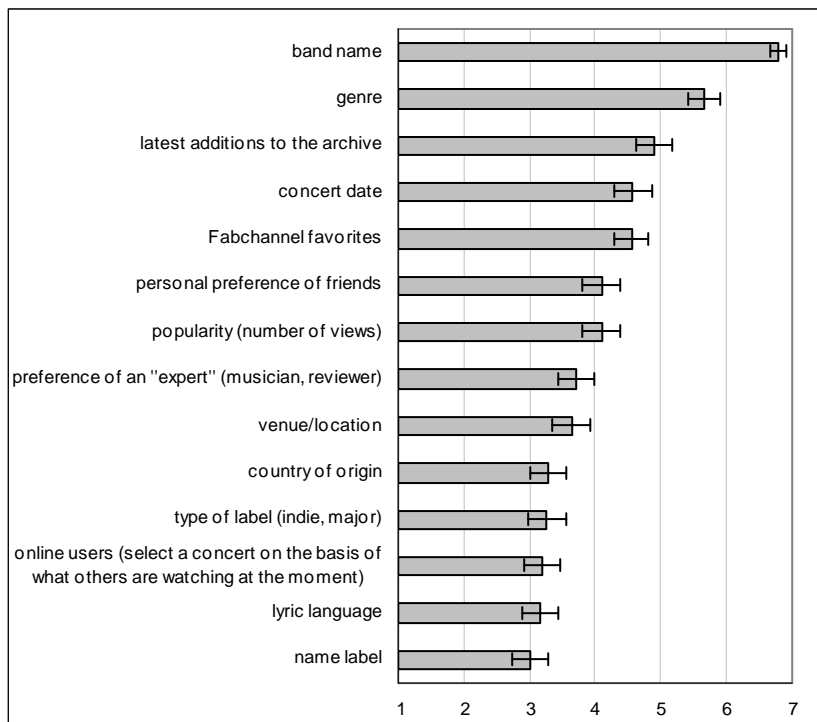
- RQ2 - What kind of efforts are invested by users in
- a) choosing a video to watch? (gap-1 problem)
 - b) finding and watching the video of their choice? (gap-2 problem)
 - c) continuing or completing their video viewing session? (gap-3 problem)

For the first reason (“find a specific concert in the archive to watch”), users know which concert to look for, so the first gap is already bridged. The problem is to locate the concert. The website offers search functionality to easily find specific concerts, thus bridging gap 2. Another way to find concerts is by browsing the database. In the Fabchannel survey, we had no specific questions about gap-bridging behavior.

RQ3 - What is the preferred way to structure the video collection for easy access?

We asked how the respondents would like to sort the concerts in the archive. Figure 2-5 shows that “band name” was the most useful option for sorting concerts (6.8 on a 7-point scale).

Figure 2-5 Usefulness of options for sorting concerts (average scores [1=totally useless; 7=very useful] and 95% confidence intervals)

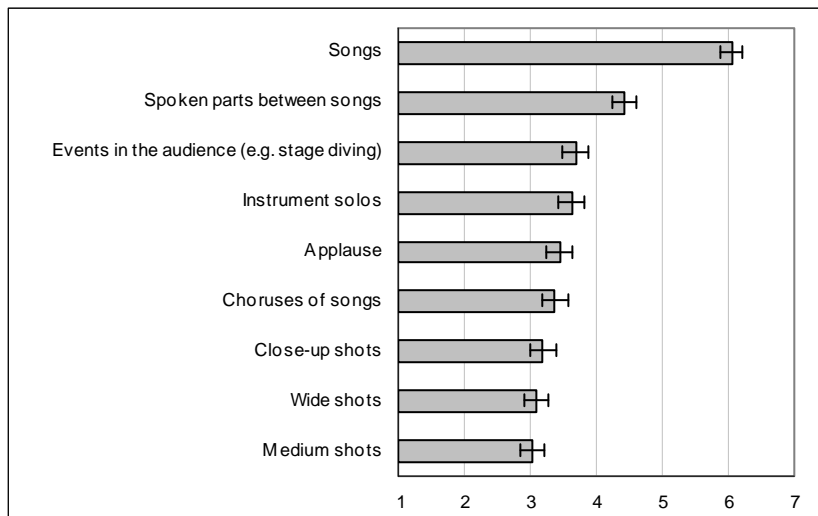


“Latest additions” (4.9) and “concert date” (4.6) are other useful ways to sort concerts and find specific performances. All other options are more closely related to the second goal: “to discover new music.” As “genre” is second in usefulness regarding sorting of concerts (5.4), respondents probably use concerts from their preferred genre to find new music. Further, several sorting options related to recommendation are considered useful: website favorites (4.6), friends’ favorites (4.1), number of views (4.1), and experts’ favorites (3.7). These sorting options are clearly helpful for finding new music, and help to bridge gap 1 and 2. Sorting by what others are currently watching was a very new option at the time of the survey and probably unknown to a lot of respondents, which may explain the relatively low score (3.2) compared to other “recommendation” options. Typical metadata options like the venue (3.6), the country of origin (3.3), type of record label: independent or major (3.3), lyric language (3.2), and name of record label (3.0) were considered less useful.

In sum, structuring the database by band name, latest additions, and date helps to bridge gap 1 and 2 for the reason “find a specific concert.” For the second reason, “discover new music,” sorting by genre and recommendations/ratings by others is useful in bridging the first gaps.

In answering the remaining two research questions on user preferences, we try to see whether there is ground for other ways of interaction to bridge the gaps.

Figure 2-6 Usefulness of concert parts to be able to skip through directly (average scores [1=totally useless; 7=very useful] and 95% confidence intervals)



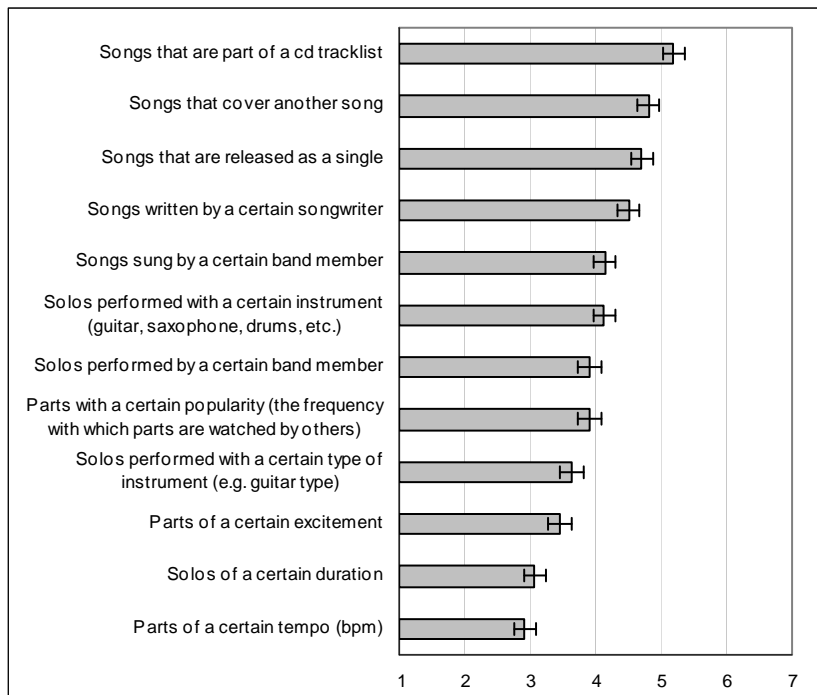
We asked the respondents the following: “In the current Fabplayer you can select and play a concert. While playing a concert you can browse through it by clicking anywhere on the timeline. We could make this interaction

‘richer’ by allowing you to directly access smaller sections within a concert. For example, you would be able to skip through all instrument solos or applause in a concert. From the options listed below, to what extent do you feel these concert parts would be useful to skip through directly?” From Figure 2-6 it is clear that songs were considered by far the most useful interaction unit. This was less so for smaller segments of songs, such as solos, choruses, and the subsequent applause. The spoken parts between songs were also considered useful units, and to a lesser extent specific events in the audience. Units related to the video registration (types of shots) were not considered very useful to interact with. Other parts respondents considered useful were interviews before and after the concert and backstage activities, but these parts are currently not recorded.

RQ4 – What is the preferred unit of interaction?

In sum, the most useful unit of interaction for this specific genre is the song. This corresponds to the result from the Kenniswijk survey that the “natural” program segments are the most useful units of interaction. In a (rock) concert, songs are the natural semantic parts.

Figure 2-7 Usefulness of characteristics of concert parts for specifying a search (average scores [1 =totally useless; 7 =very useful] and 95% confidence intervals)



We then asked the following question: “We can attach different labels, such as genre, venue, concert date, etc., to a whole concert. We are also able to label specific parts of a concert, such as songs, instrumental solos, and spoken parts. These labels or characteristics can be used to search for similar parts in one concert or in other concerts. For example, you could play all songs that have been released as a single, or all solos performed with an acoustic guitar. Another possibility would be to look for songs with similar tempo. From the options listed below, which characteristics do you feel would be useful to specify your search?” Respondents were presented with a number of characteristics of concert parts, and had to indicate on a 7-point scale how useful it would be to use those characteristics as further search criteria. Figure 2-7 shows the results of this question.

RQ5 - What is the most useful way to classify video content?

Typically, content-independent labels are considered very useful characteristics to use as further search criteria: songs from a certain CD, songs that are covers, songs that were released as a single, songs written by a certain songwriter. These metadata cannot be derived from the content directly, but need an external source such as a music database. Characteristics that can be derived from the content are considered less useful: who sings the song, which instruments are used, who performs a solo, and which specific instrument is used. Low-level features (the “excitement”, tempo [beats-per-minute], and duration [of a solo]) are even less useful. How popular certain parts are is seen as having moderate usefulness.

As a related question, we asked the following: “Imagine for instance watching one of the concert parts addressed above. To what extent do you feel it would be useful to quickly select and watch other similar concert parts 1) within a specific concert; 2) within specific parts of the archive (band, genre, singing language, etc.); 3) throughout the whole concert archive?” Respondents could answer on a 7-point scale (1 = totally useless; 7 = very useful). The average score within a concert was 4.8; within specific parts of the archive 4.6; and throughout the whole archive 4.4. This clearly relates to the important reason why people visit the site: to discover new music. It suggests that in order to bridge gap 3, there should be no strict boundaries between video files. Songs are related via their characteristics, and as such form video patches. These characteristics are not necessarily limited to a specific concert (or file), and as such patches can be formed across the borders between the concerts.

2.5 Conclusions and discussion

In the two studies presented here, we first wanted to know more about the searching process, about why people wanted to view certain videos and what their efforts were to get to that point. In other words, we wanted to know what is done to bridge the three gaps as described in Chapter 1. Second, we took a look at user preferences on how that process of video interaction could be improved. To start with we looked at the database as a whole, and at how people preferred the database to be structured for easy access. Next we looked at other possible ways of interaction: which parts of video do people prefer to interact with and what are the characteristics of those parts? In the Kenniswijk study, questions were asked in general, not referring to a specific goal or a specific content type. The Fabchannel study presented a specific case with a specific user-group and specific content. As such, the studies complement each other.

The answers to RQ 1 show that in the Kenniswijk study, the main reasons to start a viewing session were cognitive benefit, followed by mood improvement. In the Fabchannel study, the main reasons to visit the site were to find a specific concert in the archive to watch, and to discover new music. The reasons from the Fabchannel study cover both mood improvement (enjoying the music) and cognitive benefit (developing one's musical taste).

The answers to RQ 2 show that before turning on a TV, people often know what they are going to watch (for example, the 8 o'clock news), and have no problem locating the relevant channel. So in that case they do not experience the first two gaps as described by Searle. In cases where they do not know what to watch, consulting guides and channel-surfing are two important strategies for finding TV programs of interest, with the "metadata-first" option (using guides) being slightly more popular than the "data-first" option (channel-surfing). For bridging the third gap, the same strategies apply.

Most people who visit the Fabchannel website do so in order to watch a specific concert. In that case they know which concert to look for, so the first gap is already bridged. The problem is to locate the concert. The website offers search functionality to easily find specific concerts, thereby bridging the second gap. Another way to find concerts is by browsing the database, sorting the concerts by name or date. However, people also indicate that an important reason to watch videos is to learn, to discover new things. And for this they need support.

The answers to RQ 3 show that in the Kenniswijk study, the respondents were somewhat conservative in their preferences regarding patches. They preferred to watch mixed channels and program information by channel. Still, a significant part of the respondents preferred (or also

preferred) program information gathered around a theme, as well as (general) theme channels. For an important part of the respondents, this eased the task of bridging the gaps. In the Fabchannel survey, respondents indicated a need for sorting concerts by genre and getting recommendations by others.

Until now we have looked at the current behavior of respondents in both studies. In answering the remaining two research questions on user preferences, we try to see whether there are grounds for developing other ways of interaction to bridge the gaps. The answers to RQ 4 show that alongside seeing TV programs as units to be watched from beginning to end, people also approach TV/video as a set of program parts or fragments. In the Kenniswijk study, people indicated that most of the time, only part of a video is of interest, so that they were liable to quit watching a video before it was over. They clearly had a preference for being able to select smaller units than the TV/video program as a whole. This has consequences for the way viewers interact with content and bridge the gaps. People should be provided with direct access to those fragments. Designers and developers should feel encouraged to experiment with interaction modes that take video segments as the unit of interaction. Multimedia content description standards like MPEG-7 and TV-anytime are a step in this direction in that they already provide capabilities for descriptors to refer to segments.

The answers to RQ 5 show that people prefer to interact with semantically meaningful segments of video programs. In the Fabchannel study, the most useful unit of interaction for this specific genre was the song. Songs are the natural semantic parts that make up a rock concert. This corresponds to the results from the Kenniswijk survey, where the most useful aspect people would like to be able to jump to instantly was “natural” program segments. In other words, the respondents tended to agree with the semantic segmentation that is applied by program makers: the items of a news show, the items of a talk show, the different sports in a sports program, the rounds of a quiz show, and so forth. These are the semantic segments as the program maker intended, or as “naturally” provided by the characteristics of the content. This indicates that this is an important unit of interaction, perhaps even more so than programs as a whole (see RQ 3). Also, smaller units of interaction than “natural” segments are considered less useful. Segmenting videos into their semantically meaningful component parts, and first of all describing them in terms of their topic, will help people to bridge gap 2 (see RQ 2b).

Next to “natural” program segments, the only other aspect with a pass mark for being the most useful unit of interaction was topics-general. The topic is often strongly content-related: the topics of a newscast are politics, economics, art, sports, etc.; the topics of a sports program are football,

skating, track, etc. Knowing what the topic is of a video (segment) seems to provide a lot of navigational information. General topics are considered more useful than specific or famous topics. Topics can be considered to be high-level concepts. Each topic has its specific people, events, locations, etc. When we speak of topics, we often refer to videos of the informative type. Most of the time, the topics are directly related to the “natural” program segments: a change of topic often means a change of item. Likewise, events are considered useful that are also often related to “natural” program segments, especially of news and current affairs programs. When we look at which programs had the interest of the respondents, the top three were the news (96%), current affairs programs (82%) and the weather forecast (66%). This may explain this preference for “natural” program segments, topics, and events. Other content-related aspects such as time-periods, locations, animals, people, actions, and objects were considered less useful. One reason may be that a topic or an event often contains these time-periods, locations, animals/people, actions and objects: they provide the context. So, a hypothesis rising from these results may be that people prefer a level of interaction where the context is clear. They may, for example, prefer “all news items with person X” over “any video with person X”. This needs to be confirmed in future research. Typically, “people & animals” scored relatively higher in the gap 3 question (“would like to see more of”) than in the gap 2 question (“would like to jump to instantly”). This indicates the need to support this behavior in the interaction environment: people need the option to see “more about this topic” and “more about this person/animal” (the only aspects with a pass mark for the gap 3 question, whether people would like to see more of that aspect). The typical audio concepts (speech and sounds) scored relatively low. However, as with “people,” “sounds” and “conversations” seem to gain importance in gap 3, though not enough to get a pass mark. For these aspects, it seems that for a number of cases interest needs first to be aroused or stimulated by watching and listening. This emphasizes the empirical evidence that people often do not know what to look for and change their goals during the interaction process (Hildreth, 1982; Belkin et al., 1982; Bates, 1989).

The problem regarding context may also be relevant for emotions, both with regard to emotions that are displayed and emotions that are evoked. These scored very low, but they may need a context. There’s a difference between “any video with an angry person” and “a video with famous person Z being angry”. The lowest scoring aspects included the perceptual features: shapes, colors, patterns, arrangements, and movements. The results of earlier user studies on image searching (e.g., Markkula & Sormunen, 2000; Hollink et al., 2004; Eakins et al., 2004) are thus confirmed for video searching. For unsupervised metadata creation for video this emphasizes the need to bridge the semantic gap.

We can further evaluate our results by taking a more detailed look at the moving image footage requests studied by Enser and Sandom (2001). They found about as many requests for individually named persons, groups, or things (Specific Who, 373 requests) as for kinds of persons or things (Generic Who, 409 requests). In our study, we also did not find a difference between general, specific, and famous persons/groups, nor between general and specific animals. For “things” (in our study “objects”), specific objects scored below the 95% confidence interval of famous objects. So all in all the result is comparable, although the results of Enser and Sandom are a bit more detailed and show a difference for specific objects. Regarding events and actions, their study showed about three times more requests for Generic What (310) than for Specific What (100). In our study, no differences were found for events, but a difference was found for actions: the general actions scored above the 95% confidence interval of specific actions. So this result is only confirmed for actions, not for events. For locations, Enser and Sandom found three times more requests for individually named geographical locations (Specific Where, 360 requests) than for kinds of locations (Generic Where, 120 requests). In our study, general locations (like cities, the countryside) scored higher than specific and famous locations (5.28 vs. 4.74 and 4.67 on a 10-point scale). One explanation for these reversed results may be the context: users actively requesting footage vs. people interacting with TV/video to be informed or otherwise entertained. It is difficult to speculate any further as the exact reasons and backgrounds of the users in the Enser and Sandom study are not specified. Regarding time, many more requests were for linear time (dates or periods: Specific When, 310 requests) than for cyclical time (seasons or time of day: Generic When, 13 requests). Because of the context we used a different definition for the time periods, but got comparable results. Famous time periods (e.g., the seventies) and general time periods (e.g., times of poverty) scored higher than cyclical time (e.g., summer) what we called specific time periods (resp. 5.13 & 5.04 vs. 4.60 on a 10-point scale). In other words, with the exception of locations, our study largely confirms the results of Enser and Sandom.

If we compare these results with the categories in the LSCOM taxonomy for describing newscast video (Naphade et al., 2006), we see that the most important categories in our study (“natural” program segments and topics-general) are represented in the taxonomy as “program categories” (e.g., weather, entertainment, sports). Of the category “activities/events,” the events were specifically rated as useful. One reason may be that events can often be directly related to topics. The category “people” was considered especially useful in gap 3, less so in gap 2. The category “scenes/locations” was considered of average usefulness. The category “objects” (e.g., flag, animal, computer) may be very interesting

from an automated detection point of view, but with the exception of animals, objects were not considered as really useful concepts (except perhaps when they can be linked to higher-level topics). The category “graphics” seems to be quite useless without a context. Less popular categories found in our study (such as objects and actions) are often relatively easy to detect automatically. From a user interaction point of view they may hardly be of any use, but their strongest use may be the derivation of semantically richer notions like events and topics.

In the Fabchannel survey, respondents indicated they wanted to be able to interact with the songs of a concert video. Useful characteristics of concert parts to use as further search criteria were content-independent metadata: songs from a certain CD, songs that were covers, songs that were released as a single, songs written by a certain songwriter. These metadata cannot be derived from the content directly, but need an external source like a music database. Characteristics that can be derived from the content (for example, who sings the song, which instruments are used, who performs a solo, the song’s tempo) are less useful. The most important step for improving interaction is relating the video data to a database or any other source of metadata. In some cases, a central database is available (e.g., the Gracenote database for music or the Internet Movie Data Base for movies). Also, there needs to be some kind of identification process: before a relation to a database can be established, it has to be known who and what is in the video. Songs are related via their characteristics, and as such form video patches. These characteristics are not necessarily limited to a specific concert (or file), and as such patches can be formed over the boundaries between the concerts. Especially when people want to discover new music, there should be no strict borders between video files so that the third gap can easily be bridged.

Interpreting the two studies in terms of IFT, the main result is that although TV/video programs as a whole form relevant patches, there is also a need for more efficient interaction with smaller video units. The preferred units of interaction are the “natural” program segments: the semantic segments as the program maker intended (e.g., the items of a newscast), or “naturally” provided by the characteristics of the content (e.g., the songs of a concert). These segments can be considered to be good scent carriers, and are very suitable for providing navigational information. In other words, when videos are segmented this way, people are able to follow scent more efficiently and can more easily satisfy their needs.

In general, the most important aspect uniting these units in video patches is their topic. For the specific case of concerts, songs within a concert can most importantly be united into patches by being part of the same CD, by being covers, by being released as a single, or by being written by a certain songwriter. Different concerts, or individual songs from

different concerts, can also be united into patches by being from the same genre, or by being recommended by a certain party. These video patches that range beyond the borders of the original TV/video programs may efficiently stimulate the discovery of new information (the “cognitive benefit” reason that was mentioned by the Fabchannel users, and their need to “discover new music”) while such an efficient search process itself can be very enjoyable (supporting “mood improvement,” which was a second major reason for using Fabchannel). A patch-oriented database structure supporting both within-patch browsing and browsing between patches may thus lead to satisfaction.

While the Kenniswijk study was taking place, the video sharing website YouTube was founded. On this now very popular website, users can upload, view, and share video clips. Typically, YouTube is largely filled with the “natural” program parts we are speaking off. Pieces of programs are taken from their context and presented as autonomous videos people can interact with. This can be seen as a confirmation of the usefulness of interaction with smaller units. A query within YouTube creates a video patch filled with short videos related to the query term. When a video is watched, “related” videos are shown, which can be considered as belonging to the same patch. Each video itself has user-added tags, which can be considered to be prefab queries leading to other video patches. Another available patch is the collection of videos uploaded by the person who uploaded the current video. Moreover, videos fit into broad categories like “entertainment” or “music,” forming other large patches.

Given their popularity, video websites like YouTube seem to fulfill an important need regarding video interaction. However, since metadata are provided by users with varying precision and varying goals, and there are no rules or guidelines for adding metadata, metadata-based patches can have an arbitrary character. We will see in Chapter 4, where we discuss a study we performed which included a search task on YouTube, that there was general agreement among the sixteen participants that queries on YouTube yield a lot of results, many of which are irrelevant. This illustrates the importance of the question how interaction on video websites like YouTube can be improved and be more satisfying.

The research of this chapter has been aimed primarily at interaction with large TV/video programs, which used to be the dominant form of usage but now must compete with interaction with short videos. Still, our approach is relevant to collections both of long and of short videos, even though they require slightly different types of interaction. The function of the surveys presented here was to study how the video environment should best be structured in patches so that people can satisfy their needs. This research did provide some main results that were relevant in both the general and the specific situation. First, people prefer to interact with the

“natural” program segments, which can be considered useful units for carrying scent. Second, people think topics/genres are the most useful aspects for describing or grouping videos or video segments, indicating the importance of semantic descriptions of video content. In a later stage, refinement of these categories will be discussed. The studies presented here are a first step in our general approach of applying ideas from IFT to the video interaction problem. Our conclusion from these studies is that the concepts of IFT - patches and scent – and the idea of bridging gaps seem to provide a useful context for studying the problem.

In the context of IFT, the following step is to investigate how scent may guide users to these patches. This is the goal of the next chapter, Chapter 3.

Scent-following in a video database

As we saw in Chapter 1, IFT describes how people face the decision to forage a patch or leave it to find another one. These decisions are guided by the scent that is perceived. The research question we deal with in this chapter is: “What is the character of good video scent?” Scent occurs when there is a match between (associations with) elements in the information environment and (associations with) the user’s goals or interests. Therefore there are two sides to scent: the elements in the user interface, and the user. In the user interface, scent is contained in what we call “scent carriers,” representational elements that help make (video) items known to the potential user. The question is which forms scent carriers should take to establish the most realistic expectations about video content, and how this relates to the users’ tasks. We are only beginning to understand what kinds of scent carriers might be useful for videos. In this chapter we present the results of an experiment in which we asked participants to select the most relevant link to a video from a group of links, depending on the task they had to perform. Before describing the experiment and the results, we will discuss the two sides – interface and user - of video scent, with the emphasis on empirical studies when available.

3.1 Video scent

3.1.1 Scent carriers

Representations of information sources (sometimes also called *abstractions* or *surrogates*) provide cues which more or less tell users what they will find within the information source. When users navigate, their cognitive task is to predict the likelihood of finding the desired information in the source, based on the scent carriers available in the user interface. The design of

these scent carriers can influence the perceived scent and thus the decision to watch a source or not.

Scent carriers in the form of links to video sources can take several forms or shapes. They often present aggregate or diluted information about the source. However, any other data about the video that relates to the user's interests may constitute a scent carrier. For example, the fact that a video is the most popular one in a database may be a reason for a user to select it. The link to that video may carry a lot of scent for that reason, regardless of the content of the video. As such, the term *scent carrier* is broader than *video abstraction* or *video surrogate*, which refer to a condensed representation constructed to stand for a complete information object. From the users' point of view, scent carriers are elements in the information space that can point to an information source that may be of interest. Many video representation types have been developed, but only a few have been studied under experimental conditions with users. In this study, we make an attempt to get empirical results about the scent strength of a selected number of scent carriers. Several empirical studies have compared different scent carriers, or studied the optimal design of a scent carrier.

Table 3-1 displays different types of scent carriers. Groups of scent carriers are distinguished based on the type of medium: text, graphics, video, sound/speech, and combinations of media. This is relevant because – as we will see later – it relates to screen layout issues and to the way users direct their attention to the scent carriers. As such, some scent carriers may be more or less usable in different search phases.

Within the textual scent carriers, a distinction is made between scent carriers describing the video content – title, description, tags, and transcript – and scent carriers that are not related to the content but still tell something about the video. First, there are content-independent metadata such as author, date, and genre. An unknown video can have a lot of scent when it is from a director whose other movies you like, or features your favorite actress, or when it is from your home country. Online, the time that the content was added can provide scent (e.g., the “most recent” category on YouTube). A person, X, who in the past created and uploaded content interesting to a user, Y, will be associated with interesting content by Y and thus carry scent for Y. This may even go one step further: not only does the content that person X favors carry scent for user Y, in that case, but even content from the friends of person X.

Second, there are “social data,” which are related to what other users have done with the video. These influence perceived scent, either explicit (relationships formed by third parties between objects and opinion) or implicit (relationships formed by social behavior: the traces of many people acting with objects). If other users (professional reviewers, friends) claim

Table 3-1 Scent carrier types for video

Scent carrier type	Medium	Description
Content descriptions	Text (static)	A <i>title</i> is the name given to the video source. A <i>description</i> is usually more elaborate than a title, and can take the form of a table of contents. <i>Tags</i> are descriptions in the form of keywords. A <i>transcript</i> is a written version of the dialogue, directly reflecting the data in the video and very useful for search
Content-independent metadata	Text (static)	Data which are not directly concerned with video content but in some way related to it (e.g., creator, date, genre). Titles, descriptions and tags may also contain content-independent elements
Social data	Text (static)	Can take the form of <i>explicit recommendations</i> (e.g., ratings, reviews) given by others, or the display of <i>implicit behavior</i> by others (e.g., number of views, links, citations)
Spoken text	Speech (dynamic)	Any of the textual scent carriers, but in audio format
Single frame	Graphics (static)	A frame extracted from the video, based on time (e.g., first frame, middle frame) or any other selection criterion (e.g., most salient)
Multiple frames	Graphics (static)	A set of key frames usually displayed in tabular format (storyboard), sometimes with the appearance of a comic strip with differently sized frames or composed as one image. The display may involve interactive elements (e.g., scrolling)
Multiple frames	Graphics (dynamic)	A <i>slideshow</i> presenting a series of key frames one at a time for a few seconds each. This becomes a <i>fast-forward</i> when every Nth frame is selected and the frames are displayed at normal speed (30 fps). Comparable effects can be reached by using the <i>slider</i> of a player (either selecting a position on the timeline or dragging). A fast forward can also contain audio (with or without preserved audio pitch)
Video snippet	Video + sound/speech (dynamic)	A short (often representative) excerpt of a video. Can be “disguised” as a single frame and activated by, for example, a mouse-over
Video summary	Video + sound/speech (dynamic)	A <i>skim</i> is a video clip ‘abstract’ created by compacting visual and audio information while preserving the original frame rate. Comparable to a movie <i>trailer</i> , which is a pre-produced series of clips excerpted from a video
Other multimodal representations	Any combination, either static or dynamic	A key frame with a title, a storyboard with audio descriptions, etc.

that an information source is very useful, interesting, fun, et cetera (explicit recommendations), that information will probably increase the perceived scent around the link towards the information source and the chance the user will follow that path (depending, of course, on how valuable the user

thinks these claims are and how closely the user identifies with the person who is the source of the preferences, ratings, or reviews).

An example is the list of top rated videos from YouTube. Cosley, Lam, Albert, Konstan, and Riedl (2003) studied how recommendations (or other people's opinions) affect the ratings people give to movies. They found that displayed ratings influenced user ratings. This suggests that perceived scent can be manipulated by showing higher or lower ratings. In general, any recommendation, whether based on other people's opinions or on a user's personal profile, provides scent to the user. For example, in an electronic program guide (EPG), recommended TV programs are highlighted and thus propagate scent. Personalization is a way to present users with high-scented links.

Scent can also be perceived in relation to how other people act around an information source (implicit recommendations). Whenever people view, cite, or link to an information object, this behavior can be detected by algorithms and used to rank information objects (e.g., on the basis of number of views), or relate objects to each other (e.g., on the basis of how often watched by the same persons). Examples from YouTube include the categories "Most Viewed" and "Most Discussed." Behavior can also be used for creating recommendations.

For every textual scent carrier, a spoken version can be generated. A disadvantage is that while text is static, speech is dynamic and requires constant attention. It may, nevertheless, be useful in specific situations where the visual sense is restricted or burdened.

Still image representations render the video content as extracted frames. They can function as natural analogues to keywords in text. Examples of representations using multiple still frames include storyboards, slide shows, collages, key-frame-based tables of contents, and so forth. Tse, Vegh, Marchionini, and Shneiderman (1999) compared the effectiveness of two key-frame-based video representations: a static storyboard and a dynamic slide show. The storyboard had a 3x4 configuration of key frames, while the slide show "flashed" each key frame onto a screen at the rate of 3 key frames per second. Twenty participants had to perform two types of task: gist determination (goal-oriented, learn what the video is about) and object recognition (task-oriented, recognize whether a particular object or relationship exists in the key frames). Video clips (1.5-3.0 minutes) were obtained from documentaries, and per clip 12 key frames were selected (first algorithmically, then manually). No effect of type of video representation was found on task accuracy. Significant differences were found for user satisfaction: users assessed the static display as more useful and less confusing than the dynamic display for both task types. The authors argue that this may have been caused by the high display rate of the slide show (3 frames/sec) or the lack of user control for key-frame rate and/or

direction of play. Furthermore, participants had to wait for an image to loop around in order to view a particular key frame again. A slideshow obstructs the efficient perception of scent by making possibly relevant representations only partly available. This result more or less confirmed previous research (also at the University of Maryland) by Komlodi and Marchionini (1998), who concluded that static displays were better than slide shows for object identification but not for gist determination. In that study, subjective satisfaction also favored the static display.

In a study by Hughes, Wilkens, Wildemuth, and Marchionini (2003), participants selected relevant video records from results lists containing titles, descriptions, and three key frames for ten different search tasks. All participants were eye-tracked to determine where, when, and how long they looked at text and image surrogates. Participants looked at and fixated on titles and descriptions statistically reliably more than on the images. Most people used the text as an anchor from which to make judgments about the search results and the images as corroborating evidence for their selections. This study suggests that titles and descriptions are more useful in determining the relevance of a video than key frames. This is confirmed in Figure 2-4 in Chapter 2, showing that in the Kenniswijk survey, respondents thought that a short description of the contents, a title, and metadata (genre year, maker, country, etc.) were the most useful pieces of information about a video (all received a pass mark – 5.5 or higher - on a scale of 1 to 10). At the same time, one still or ten stills from the video both received very low ratings (below 4).

It does, however, matter whether frames are naively chosen or are directly related to the query. Christel, Winkler, and Taylor (1997), for instance, showed that when a frame showing the shot with the highest matching score between the query terms and the transcript was used, better results were obtained than with a frame of the video's opening shot. From a scent point of view this is a logical result, as there is (by definition) a semantic match between the query and the query-based frames, and substantial amounts of scent for those representations.

Boreczky, Girgensohn, Golovchinsky, and Uchihashi (2000) studied pictorial summarizations of video. Given an existing segmentation, they calculated shot importance for each segment based on its rarity and duration. Less important shots were pruned, leaving a concise and visually varied summary. Key frames were then displayed in different sizes according to their importance, in a style reminiscent of a comic book or Japanese manga. In one experiment, 24 participants (staff at FX Palo Alto Laboratory) were asked to browse visual summaries of three staff meetings. Three styles of summaries were created: fixed-size images sampled at regular time intervals (control), fixed-size images with selection based on importance (selected), and variable-size images with selection based on

importance (manga). Participants were asked three questions for each video, and had to find relevant video segments as fast as possible. No differences between the three conditions were found regarding time to complete the task. However, the manga version was judged by the subjects to be more effective, to provide better entry points, and to be more visually pleasing.

Multimodal representations combine textual, visual, and audio information. An example is a key frame with a title. When moving images (and audio) are incorporated, the representation resembles the original video more strongly, because the temporal aspect of video is still present. An example of this is a fast forward, either with or without the audio, and, if the audio is present, either with or without preserved pitch. Another example is a video skim, which summarizes the original video by concatenating significant subsets of video and audio data in a style resembling a movie trailer. Christel, Smith, Taylor, and Winkler (1998) measured the effects of different “video skim” techniques on comprehension, navigation, and user satisfaction. In contrast to static video representations such as titles and thumbnails, a video skim preserves the temporal dimension of video and includes audio information (of course, a skim still takes much more time to view than a short title or thumbnail). In this study, skims were 7.5 times shorter than their source video and built from segments averaging five seconds in duration. (This was a modification of a previous study, in which skims of one-tenth of the original duration were used, built from segments averaging three seconds.) The video skims differed in the rules used for selecting “important” audio and video components. Twenty-five students watched four different skims and the full video (in balanced order), and after each view had to say whether presented images and presented text phrases were part of the full source video or not. A video skim incorporating speech, language, and image processing produced the best results. However, subjects preferred the full video to any of the skim types, which mostly received poor evaluations (perhaps due to abrupt visual changes). This indicates the difficulties of creating a “good” video skim.

Wildemuth, Marchionini, Wilkens, Yang, Geisler, Fowler, Hughes, and Mu (2002) compared five video representations (or surrogates): storyboards (6x6 frames) augmented with text or audio keywords, slide shows (36 frames, displayed at 4 frames/second) augmented with text or audio keywords, and fast forward (4 times original speed). These were manually created for each of seven video segments (documentaries, educational, promotional), each 2-10 minutes long. Ten participants viewed the surrogates and performed a series of tasks: gist determination, object recognition, action recognition, and visual gist determination. Gist determination tasks included writing a brief summary of the video’s content

and selecting the video surrogate that best represented the video's content from five brief text summaries presented. Object recognition tasks included marking whether objects were present in surrogates that were first presented as text or as frame. In the action recognition task, participants viewed mini-segments (2-3 seconds long) and had to decide whether they were part of the same video that was represented in the video surrogate. In the visual gist determination task, users had to determine, for frames not used in the surrogate, whether they belonged in the video segment for which the surrogate was seen.

The main results were that for the object recognition task (with textual stimuli), the storyboard with text keywords performed significantly better than the storyboard with audio keywords. For the action recognition task, the fast forward surrogate outperformed the rest of the surrogates. For the gist tasks, there were no differences. No surrogate was universally judged "best," but the fast forward surrogate garnered the most support, particularly from experienced video users. The least support was given to the slide show with text keywords. In general, users expressed a desire for more control over the display of the surrogates (e.g., starting, stopping, speed, display time). In addition, users viewed different surrogates as being more or less useful for different types of tasks, and would have liked to be able to move from surrogate to surrogate. In follow-up research, Wildemuth, Marchionini, Yang, Geisler, Wilkens, Hughes, and Gruss (2003) studied four speed variations of one form of a video surrogate: a fast forward created by selecting every Nth frame from a video. No audio or additional metadata were provided. They recommend a fast forward default speed of 30 frames per second (normal frame speed) showing every 64th frame of the original video (meaning a fast forward of 64 times the original speed).

Song and Marchionini (2007) compared three surrogate alternatives: visual alone (a storyboard), audio alone (spoken description), and combined visual and audio (a storyboard augmented with spoken description). The combined surrogates were more effective, were strongly preferred, and did not penalize efficiency. Audio alone led to better understanding than visual alone, although people liked to have visual surrogates that they could use to confirm interpretations and add context.

In Table 3-1, we indicated the type of medium for each scent carrier and indicated whether that medium type is static or dynamic (i.e. presenting changes over time). This is an important criterion, as dynamic media require constant attention from the user if all the information they convey is to be used. Comprehension of concurrently presented dynamic media about different subjects is not feasible, as the overall level of cognitive resources required to process information about two different subjects is too great for the majority of users (Bearne, Jones & Sapsford-Francis,

1994). As such, this tells us something about the usability of these scent carriers in different stages of search: whether a choice needs to be made between a number of possibly relevant videos, or whether one of those videos is selected for closer inspection of the contents.

Some video representations, such as titles and key frames, are more suited than others for use in a scannable list of video representations. Representations with a temporal component (including moving images and audio) take time to explore, and their most useful function is probably as a substitute for viewing the whole video. Before people decide to explore such representations, they will probably make use of static representations and their scent. A typical scenario may be as follows: 1) On the basis of scent perceived in a static representation that is part of a longer list, people select the related representation with a temporal component; 2) they view/explore this representation and when the scent is high enough decide that the original video source is interesting to them; 3) they watch the video source or a specific part of it. Of course, users may skip step 2, or may skip step 3 when step 2 is enough to satisfy a (temporal) need.

We wanted to study the principles of scent-following in a video database in which users follow links to video sources, and establish which factors influence scent-following. This study focuses on step 1 as described above, so scent carriers were selected that are fit to be used in a *scannable list of video representations*. We are talking about scent that can be present in hyperlinks to video sources (not in the video data itself), in situations where people have to choose from a list of links. This includes most of the static scent carriers as presented in Table 3-1. In the experiment described below, there was no chance to further explore (or view) the video. Scent carriers that are more fit for the further exploration of a video's content were therefore excluded from this particular study. This includes dynamic scent carriers such as moving image, audio, and multimodal representations which take time to explore. In addition, representations that require a lot of space in the user interface – such as storyboards - were excluded.

In this research, we studied scent carriers only in the first stage of search, where relevant video sources have to be identified from a larger set of links to video sources. In terms of gaps, this mostly corresponds to gap 2. The experimental tasks provided the reasons to browse, and the presented result sets were the outcome of bridging the first gap. Now, the second gap needed to be bridged by choosing from the alternatives the one with the highest scent. The phase after that would be to inspect the contents of the selected video, but that phase was beyond the purpose of this study (the experiment described in Chapter 4 includes all stages of search). So video summaries that are useful for getting a quick idea of the content of a video that has already been identified as being potentially relevant fell outside the scope of the study described here. The list of scent carriers that were

selected for this study is presented in the method section of the experiment.

3.1.2 Tasks

As we saw above, scent occurs when there is a match between (associations with) elements in the information environment and (associations with) the user's goals or interests. In this section, we study different characteristics of these goals and interests.

A study by Woodruff, Faulring, Rosenholtz, Morrison, and Pirolli (2001) compared the usefulness of three different summaries of Web pages in finding several different types of information. The participants had to locate information in the following categories: a picture of a given entity, the homepage of an individual they did not know, a consumer electronics item for purchase, and three or more side effects of a given drug. They had to locate a relevant Web page in results pages that contained one of the types of Web page summaries (which also served as hyperlinks). One type was a text summary that included the page's title, an excerpted text with the search item shown in bold, and the URL. Another type was plain thumbnails (scaled versions of the original Web page). Still another was enhanced thumbnails with modified font sizes (making them more readable), with highlighted keywords (as used in the query) and reduced contrast to make the call-outs more prominent.

On average, participants had the fastest search times with enhanced thumbnails (67 seconds), then plain thumbnails (86 seconds), and then text summaries (95 seconds). These results varied strongly by question category, however. For the picture question, text summaries were clearly the slowest, while the thumbnail summaries worked best, as they allowed the user to spot the presence of a picture on a page. For the homepage question, plain thumbnails were the slowest, while text summaries and enhanced thumbnails, both of which showed the person's name, were more helpful. For the other two question categories, no significant effect of summary type was found. This study emphasizes that both task type and link design influence browsing performance.

In the Kenniswijk study described in Chapter 2, we found that the main reasons people watch video are for cognitive benefit ("to keep oneself informed," "to learn, better understand something") and mood improvement ("to relax"). Many people start to watch television or visit a video database on the internet without a specific goal. Instead they have general reasons of learning something new or getting entertained, which may become specific whenever they encounter something of interest. In this research, we make the distinction between browsing for mood improvement and browsing for cognitive benefit.

Within the latter reason, we distinguish between a general interest (“Search for something that is of interest to you”) and a specific interest (“Search for something about subject X”). A general task is open and provides the participant all freedom to explore. People probably will be guided and inspired by the alternatives that are offered and will adjust their current interest. A specific task is more restricted, and we may expect different searching behavior because people know exactly what to look for.

Moreover, within the specific situation we can distinguish easy confirmational tasks (“Is there something about X in this video?”) versus hard judgmental tasks (“Is Y true?”). In the former case, the answer to the question can relatively easy be contained in the video’s representations (for example, it is easy to use a title or a frame to indicate whether it is a video about elephants). In the latter case, even though the answer may be found in a title or description, it is more probable that additional inspection of the contents of the video is required to be sure about the answer (for example, to find out whether it is true that elephants migrate).

Furthermore, we distinguish situations where there is a good or a bad match between the users’ task and the database. Several studies indicate that this match can influence browsing effectiveness and efficiency. For example, semantic similarity between task descriptions and menu labels lets subjects perform tasks faster (Soto, 1999). Pirolli, Card, and Van Der Wege (2003) compared a hyperbolic tree browser with a conventional browser (Windows Explorer). A hyperbolic tree browser is an example of a focus + context display, where more display space is assigned to one part of the hierarchy (focus) than others (context). The browser applies distortion - “stretching” the focus and “squeezing” the context - to get all the information into the display space. The hierarchy can be mouse-dragged through the central display region to bring new parts into focus, or nodes can be mouse-clicked to bring them to the center of focus. The labels on the nodes in the hyperbolic tree provide information-scent cues, and ideally users will follow the node labels to some target node in the tree.

In an experiment, eight subjects performed search tasks using the hyperbolic tree browser and the conventional browser. The tasks were simple retrieval tasks such as “Find Lake Victoria” or “Find a hammer.” The researchers distinguished “high-scent tasks” and “low-scent tasks,” in which respectively many and few users in a preceding study correctly identified the tree node labels along the path to the target. In other words, in the “high-scent tasks” there was a clear scent trail to the target, while in the “low-scent tasks” users were more likely to choose the wrong sub-trees to browse. For each search term an “accuracy of scent” (AOS) score was calculated, which is “the proportion of participants who correctly identified the location of the task answer from looking at upper branches in the tree.” Eight tasks were selected from the extremes of high and low AOS values,

respectively between 0.00-0.10 and 0.35-0.40 (the chance of correctly guessing the location was 0.015). The result was that hyperbolic tree users visually searched more of the tree structure than the conventional browser users while performing the tasks at a faster rate. However, the performance was highly affected by information scent. Poor information-scent cues caused visual search and navigation in the hyperbolic tree to become much less efficient.

The results of these studies support the need for task differentiation when exploring the properties of scent carriers. In the experiment described below, we wanted to include a number of tasks that are representative for the situations people encounter when searching for videos. The two main reasons why people watch videos – cognitive benefit and mood improvement – were therefore included. Within the cognitive benefit task we distinguished three situations for which we had reason to believe that they affect the perception of scent: general vs. specific tasks, easy vs. difficult tasks, and good vs. bad match between task and database.

3.2 Research questions

The basic research question for this chapter is “What determines the perception of scent in a video database?” Based on the literature presented above, we expect to find effects of the type of scent carrier and the type of task. Regarding type of scent carrier, we have the general question:

RQ1 - Is there an effect of type of scent carrier on the perception of scent?

Regarding the type of task, we have the general question:

RQ2 - Is there an effect of type of task on the perceived scent of scent carriers?

The combination of type of scent carrier and type of task may further show an interaction effect:

RQ3 – Is there an interaction effect of type of scent carrier and type of task on the perception of scent?

More specifically, we have the following research questions regarding the type of task and the interaction with type of scent carrier:

- **reason to browse:** *Is there an effect of type of reason (mood improvement vs. cognitive benefit) on the perceived scent of scent carriers?*
- **specificity:** *Is there an effect of specificity of the search question on the perceived scent of scent carriers?*

- **difficulty:** *Is there an effect of difficulty of the search question on the perceived scent of scent carriers?*
- **goodness of match task/database:** *Is there an effect of the goodness of match between the task and the search results on the perceived scent of scent carriers?*

In the experiment, specific questions and links to specific videos were used. To be able to place the specific outcomes in a more general context, we added the following research question:

RQ4 – How useful (in general) are individual pieces of information about a video?

3.3 Method

3.3.1 Participants

An email request to join an experiment on video search behavior was sent to Dutch and German students of the Psychology, Communication Studies, and Educational Science and Technology programs of the University of Twente in the Netherlands. It was indicated that participants would have to fill out an online questionnaire, which would take about one hour. The reward was the chance to win one of six prizes: an iPod (value €200), or five bookstore gift certificates for €50 each.

A total of 103 students responded to the request, of whom 75 completed the experiment. The group was 67% female and 33% male. The average age was 23.2 years. The average time to complete the experiment was 56 minutes.

The online questionnaire started with questions about the respondents' normal viewing behavior. On the average, the respondents reported viewing 14.8 hours of video images in a week. They spent most of this time watching television (on the average, three times a week), followed by watching downloaded videos or streaming videos from the internet (on the average, both at least once a month). Next, they watched movies in the cinema or on their own or rented videos/DVDs (on the average, all three at least four times a year). They very infrequently watched recorded TV programs and their own home videos (both on the average about once a year).

3.3.2 Independent variables

The main manipulations were type of scent carrier and type of task. The scannable list that we discussed in 3.1.1 was created in combination with

what can be encountered in interfaces to video databases. This led to eight types of scent carriers:

- frame (the middle frame of the video)
- title
- title + frame
- title + description (truncated after three lines)
- title + tags (truncated after one line)
- title + metadata (duration, category, and date added)
- title + social data (number of views and comments, how often the video was favorited, rating, and number of ratings)
- all (all information described above combined in a single representation)

Figure 3-1 shows all scent carriers for one example video.




To create a realistic scenario, one that people searching for videos would encounter in the real world, we used data from YouTube. The scent carriers we created were not adjusted but exactly copied: even typing mistakes were adopted without rectifying.

A first important distinction is made between *title* and *frame*. These are the only simple (not compound) scent carriers, representing two separate worlds: the visual and the textual. As we saw in section 3.1.1, textual surrogates often provide better results than visual surrogates. In this experiment, we will try to replicate that result in terms of how much scent is perceived in these scent carriers for different task types.

Apart from “frame,” all scent carriers contained a title. The main reason was that when descriptions, tags, metadata, or social data are available, in most cases there is also a title present. On YouTube, for example, titles are almost always available (whether useful or not). So we did not study the scent carrying capacity of, for example, social data only, but actually measured what the added value of social data was when combined with a title. When no information is present about a video but the video data itself, a frame can still be used as a representation, so a single frame also is a realistic situation.

Of course, it would also be interesting to look at other combinations (such as “title + frame + X”). However, this would make the experiment too big and complicated, and for practical reasons we made the choices described here. The *all* scent carrier was added to estimate the perceived amount of scent in the situation where all information is available. Moreover, an extra question on the usefulness of several pieces of information in general (see RQ4) was added. This in combination with the measures for the other scent carrier types will allow us to estimate the outcome of other combinations of scent carriers that we did not study in this experiment.

Figure 3-1 All the scent carriers of one example video

title + frame		frame	
	Rare prehistoric shark appears off the coast of Japan		
	title Rare prehistoric shark appears off the coast of Japan		
	title + description Rare prehistoric shark appears off the coast of Japan <i>Description:</i> This 1.6 metre long eel-like animal, identified as a deep sea frill shark, is usually 600 metres underwater according to Awashima Marine Park officials. The sharks, while they usually turn up in deep sea trawler fishing nets...		
	title + tags Rare prehistoric shark appears off the coast of Japan <i>Tags:</i> giant shark rare prehistoric japan sea ntdtv ntd ntdnews marine frill		
	title + metadata Rare prehistoric shark appears off the coast of Japan <i>Duration:</i> 01:26 <i>Category:</i> News & Politics <i>Added:</i> 5 months ago		
	title + social data Rare prehistoric shark appears off the coast of Japan <i>Views:</i> 791,296 <i>Comments:</i> 29 <i>Favorited:</i> 2897 times ★★★★☆ 936 ratings		
	all  <i>Duration:</i> 01:26 <i>Category:</i> News & Politics <i>Added:</i> 5 months ago <i>Views:</i> 791,296 <i>Comments:</i> 29 <i>Favorited:</i> 2897 times ★★★★☆ 936 ratings Rare prehistoric shark appears off the coast of Japan <i>Description:</i> This 1.6 metre long eel-like animal, identified as a deep sea frill shark, is usually 600 metres underwater according to Awashima Marine Park officials. The sharks, while they usually turn up in deep sea trawler fishing nets... <i>Tags:</i> giant shark rare prehistoric japan sea ntdtv ntd ntdnews marine frill		

For the statistical analysis, we looked at the main effect of scent carrier, and next to that we specifically focused on the difference between *title* and *frame*. The task types and tasks are summarized in Table 3-2.

Table 3-2 Task type and tasks for the experiment on scent following

Task type	Task
1. general – mood improvement	1. “Find a video you can relax with”
2. general – cognitive benefit	2. “Find a video about a subject you would like to know more about”
3. specific – easy & good match	3. “Find a video about water pollution”
4. specific – hard & good match	4. “Find a video displaying evidence that sewage problems are a main cause of surface water pollution”
5. specific – hard & bad match	5. “Find a video displaying evidence that human behavior is unconsciously influenced by smells”

A match (see task 3, 4, and 5) is considered good when potentially relevant videos are available, and bad when they are not (see section 3.4 for a description of the video selection process).

3.3.3 Dependent variable

The amount of perceived scent of a link was measured as the *subjective probability that the information that is needed can be found behind that link*. In this experiment, it was measured on a 9-point scale (1 = very improbable; 9 = very probable).

3.3.4 Experimental materials

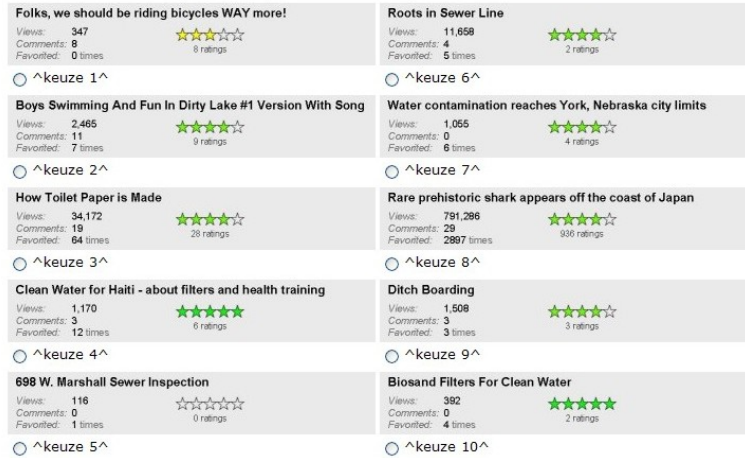
We used data from the video website YouTube.com to create eight result sets of ten videos each. Each result set had to contain some videos relevant to each of the tasks, with the exception of the task *specific – hard & bad match*. We searched for videos on the theme “water.” Our main reason was that this is a general theme everybody is familiar with, with many different – serious and less serious – sides. It is an omnipresent substance covering most of the surface of the earth; it provides a means for fun and recreation, but it is also a necessity of life of which there can be a shortage or which can be polluted.

For the mood-improvement we task added query terms such as “fun” and “humor.” For the general cognitive-benefit task, we added terms such as “interesting” and “science.” For the easy specific task we added “pollution,” and for the hard specific task we added “sewage” or “environment” to that. (The fifth task did not, of course, affect the video selection process: this task was developed afterwards as a task with no matches among the available videos.) Further, and also for use in future experiments, we searched to include videos about water with hardly any visual variation, using queries such as “lecture” and “speech.”

This way we got 80 videos, from which we created eight roughly comparable sets of 10 videos, based on their relevance to the tasks. The duration of the videos ranged from 21 s to almost 10 min. We represented

each set of ten videos by a list of ten links on a computer screen, called a “result set” (roughly resembling a result page of a query). The links were the scent carriers from which the participants had to select videos to complete a search task. Figure 3-2 shows an example of a title+ social data result set. Further examples are to be found in Appendix C.

Figure 3-2 Example of a title+ social data result set. Participants had to choose the link with the highest scent for the specific task they had to perform (the survey tool did not support the selection of pictures, hence the tick boxes)



The study of all eight types of scent carriers required that we have eight variants of each result set, one for each type of scent carrier. The ten links of a result set were thus always of the same type of scent carrier (ten titles, or ten frames, et cetera). In all there were 64 different result-set displays: eight result sets times eight scent-carrier types.

To create the scent carriers, we used the exact data that YouTube offered. We didn’t make any alterations, did not improve bad titles or descriptions; we even included typing mistakes. As such, the sets were as realistic as any query result from a popular video database.

3.3.5 Procedure

Each participant had to complete search tasks for all five task types and for all eight scent carriers. This was organized as follows. For each task type the participants performed the same task eight times, each time with a different result set for a different type of scent carrier. To complete a task, the participants had to select one link as the preferred link out of the ten links of a result set. After each link selection, they were asked to indicate the probability they would find what they were looking for on the abovementioned scale. So they only had to do that for one - the one with the highest scent - out of ten links. This way, rating problems that could result from the presence of accidentally “bad” links were avoided.

After the completion of all eight repetitions of the task for one task type, the procedure continued with the next task and the next eight result sets. After all five tasks for all five task types the participants thus performed forty (5x8) selections.

Scent-carrier and *task type* varied within subjects, so each participant was confronted with all eight scent carriers and all five task types. However, the coupling of a specific video with a specific scent carrier varied between subjects. This was necessary to avoid learning effects. So for each specific video, each participant only saw one scent carrier related to that video during the whole experiment. For this reason, we had to create eight variants of the questionnaire, so that every possible coupling between a video and a scent carrier was present in the experiment.

Moreover, of all eight questionnaire variants, two versions were created, each with a different task order. For the first task, there is a risk that the participants are contemplating a strategy for filling out the questionnaire, while for the last task there is a risk that they are tired or bored. This might influence the way they complete the tasks. However, to avoid transfer effects between tasks, the number of order variants was limited to two. Participants always started with one of the two general tasks (starting with a specific task might bias the “free” tasks). As the two specific tasks with the good match were both about the same subject (water pollution), the easy one preceded the hard one (starting with the hard one might bias the easy question). So for balance, half of the participants were presented with task order 1-2-3-4-5 (relating to the order applied in Table 3-2), and the other half with task order 2-1-3-5-4. Appendix D presents an overview of all versions of the survey.

At the end of the experiment – and in order to answer research question 4 - we asked participants to rate the usefulness of 37 different types of information about videos on a scale of 1 to 10. Ratings were asked “in general,” so not for a specific goal or situation. This differed from the experiment, where every rating was within the context of a specific task. The questions (in Dutch) are presented in Appendix E.

3.4 Results

RQ1 - Is there an effect of type of scent carrier on the perception of scent?

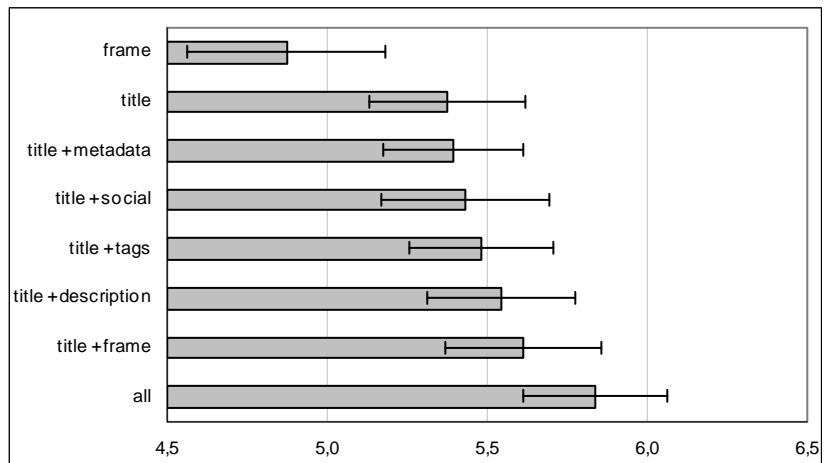
There was a significant effect of type of scent carrier (Wilks' lambda = 0.49; $F(7,68) = 10.23$; $p = .000$). Restricting the analysis to the scent carriers *title* and *frame*, the effect of scent carrier was again significant (Wilks' lambda = 0.86; $F(1,74) = 11.91$; $p = .001$).

Figure 3-3 shows the overall strength of perceived scent of all scent

carriers (averaged over all task types). Pairwise comparison shows that the “all” scent carrier has significantly higher ($p < .05$) perceived scent than all other scent carriers except title + frame and title + description. The “frame” scent carrier has a significantly lower perceived scent than all other types. When doing a Bonferroni adjustment for making multiple comparisons the results are more conservative. In that case, “all” is not significantly better than “title + description” and “title + frame” (but still better than the others), but “frame” is still significantly worse than all the others.

Of all the additions to a title, “title + frame” has the highest perceived scent, significantly higher than “title + social,” “title + metadata,” and “title” only (but with the Bonferroni adjustment this effect disappears).

Figure 3-3 Perceived strength of scent of all scent carriers (average score on a nine-point scale and 95% confidence interval)



RQ2 - Is there an effect of type of task on the perceived scent of scent carriers?

There was a significant effect of task type (Wilks' lambda = 0.22; $F(4,71) = 62.95$; $p = .000$). Restricting the analysis to the scent carriers *title* and *frame*, the effect of task type was again significant (Wilks' lambda = 0.38; $F(4,71) = 29.31$; $p = .000$).

RQ3 - Is there an interaction effect of type of scent carrier and type of task on the perception of scent?

The interaction between task type and scent carrier was significant (Wilks' lambda = 0.39; $F(28,47) = 2.58$; $p = .002$). Restricting the analysis to the scent carriers *title* and *frame*, the effect of task type was again significant (Wilks' lambda = 0.75; $F(4,71) = 5.87$; $p = .000$).

Looking at the different task types, how do the scent carriers relate to each other? The perceived scent of specific scent carriers is affected by the

type of task, but in what way? Table 3-3 shows that in all conditions except *specific – hard & bad match*, the effect of scent carrier was significant. Restricting the analysis to *title* and *frame*, there is also no effect of scent carrier for *general - mood improvement*.

Table 3-3 Effect of scent carrier per task type (separate analyses for all scent carriers and for title and frame (*t&f*))

Task type	Scent carriers	Wilks' lambda	F(7,68) F(1,74)	p
general – mood improvement	all carriers	0.77	2.89	.011
	<i>t&f</i>	0.99	0.75	.391
general – cognitive benefit	all carriers	0.57	7.44	.000
	<i>t&f</i>	0.81	17.26	.000
specific – easy + good match	all carriers	0.47	11.04	.000
	<i>t&f</i>	0.85	12.77	.001
specific – hard + good match	all carriers	0.78	2.73	.015
	<i>t&f</i>	0.91	7.49	.008
specific – hard & bad match	all carriers	0.92	0.80	.594
	<i>t&f</i>	0.98	1.52	.221

Next we look at the specific effects of task type.

Reason to browse: “Is there an effect of type of reason (mood improvement vs. cognitive benefit) on the perceived scent of scent carriers?”

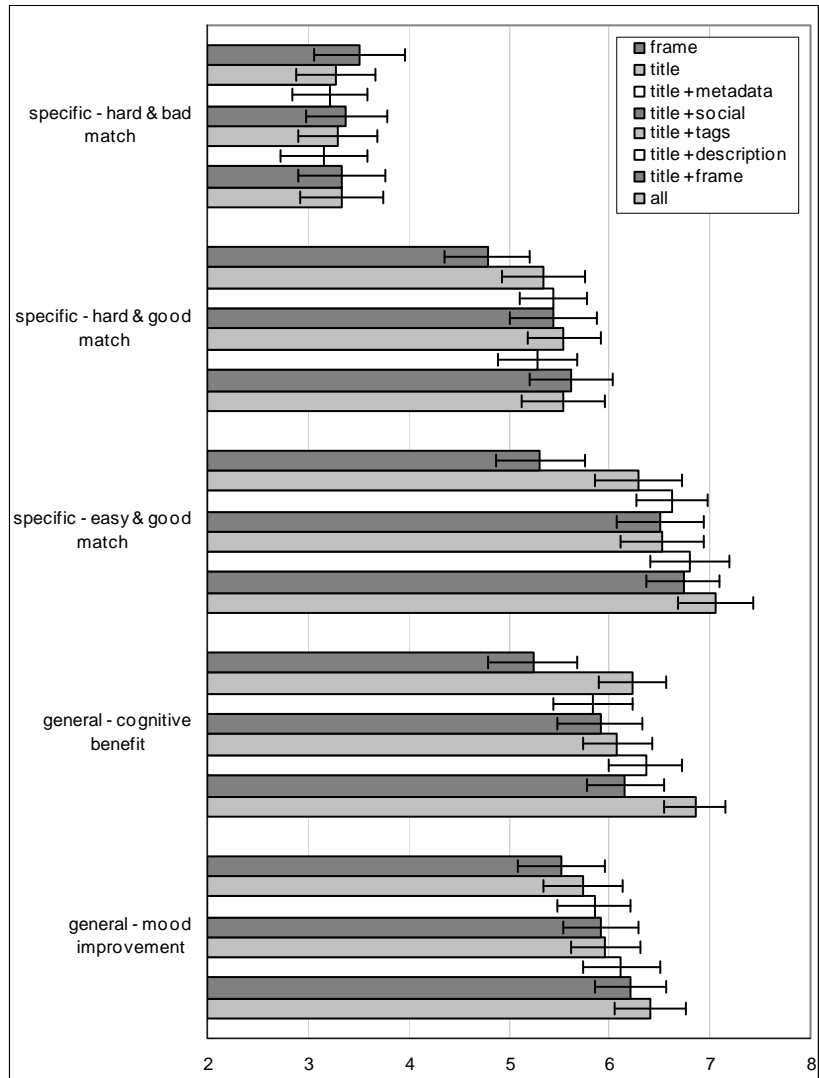
The answer results from the comparison of the tasks for the general task types: Task 1 (mood improvement): “Find a video you can relax with,” and Task 2 (cognitive benefit): “Find a video about a subject you would like to know more about.”

A significant effect of reason to browse was not found (Wilks' lambda = 0.99; $F(1,74) = 1.15$, $p = .286$), and neither was the interaction scent carrier x reason (Wilks' lambda = 0.86, $F(7,68) = 1.58$, $p = .157$). Looking only at *title* and *frame*, there is still no effect of reason to browse (Wilks' lambda = 0.99; $F(1,74) = 0.44$, $p = .508$), but the interaction with scent carrier is significant (Wilks' lambda = 0.89; $F(1,74) = 8.99$, $p = .004$). Table 3-3 and Figure 3-4 reveal that for mood improvement, there was no difference between *title* and *frame*, while for cognitive benefit there was a significant difference.

Specificity: “Is there an effect of specificity of the search question on the perceived scent of scent carriers?”

For this question we compare the cognitive benefit tasks, Task 2 (“Find a video about a subject you would like to know more about”) and Task 3 (“Find a video about water pollution”). Task 2 is a general task, while Task 3 is specific and considered to be easy given a good match with seemingly relevant videos in the result sets.

Figure 3-4 Scent carrying capacity of all scent carriers for all task types (average score on a nine-point scale and 95% confidence interval)



The main effect of specificity is significant (Wilks' lambda = 0.88; $F(1,74) = 9.82$; $p = .002$). The interaction of specificity with scent carrier is not significant (Wilks' lambda = 0.90; $F(7,68) = 1.05$; $p = .405$). Restricting the analysis to *title* and *frame*, there is no effect of specificity (Wilks' lambda = 1.00; $F(1,74) = 0.13$, $p = .724$), and neither is there for the interaction with scent carrier (Wilks' lambda = 1.00; $F(1,74) = 0.00$, $p = 1.000$). Figure 3-3 shows that, for all scent carriers in the specific but easy and

good-match situation, perceived scent is higher than in the general situation.

Difficulty: *“Is there an effect of difficulty of the search question on the perceived scent of scent carriers?”*

The answer follows from comparing Task 3 (easy+good match: “Find a video about water pollution”) and Task 4 (hard+good match: “Find a video showing proof that sewage problems are a main cause of surface water pollution”).

The main effect of difficulty is significant (Wilks’ lambda = 0.51; $F(1,74) = 71.44$; $p = .000$), and so is the interaction of difficulty with scent carrier (Wilks’ lambda = 0.76; $F(7,68) = 3.08$; $p = .007$). Restricting the analysis to *title* and *frame*, the effect of difficulty is also significant (Wilks’ lambda = 0.82; $F(1,74) = 16.57$, $p = .000$), but the interaction with scent carrier is no longer significant (Wilks’ lambda = 0.98; $F(1,74) = 1.93$, $p = .169$).

Figure 3-4 shows that, for all scent carriers, the amount of scent is less with the hard task. The interaction effect can be seen as this: the higher the perceived scent in the easy condition, the larger the decrease in the hard conditions. Or in other words, in the hard conditions there is less difference between the scores of perceived scent carrying.

Goodness of match task/database: *“Is there an effect of the goodness of match between the task and the search results on the perceived scent of scent carriers?”*

For this question we compare the following - both specific and difficult - conditions: Task 4 (good match: “Find a video showing proof that sewage problems are a main cause of surface water pollution”) and Task 5 (bad match: “Find a video showing proof that human behavior is unconsciously influenced by smells”). For Task 4, seemingly relevant videos are present in the result set. For Task 5, there are no clear targets.

The main effect of goodness of match is significant (Wilks’ lambda = 0.35; $F(1,74) = 136.20$; $p = .000$), and so is the interaction of goodness of match with scent carrier (Wilks’ lambda = 0.77; $F(7,68) = 2.89$; $p = .011$). Looking only at the scent carriers *title* and *frame*, goodness of match is still significant (Wilks’ lambda = 0.561; $F(1,74) = 57.83$, $p = .000$), and so is the interaction with scent carrier (Wilks’ lambda = 0.89; $F(1,74) = 8.93$, $p = .004$).

Figure 3-4 shows that for all scent carriers, scent carrying capacity is lower when there is a bad match between task and the search result. The interaction effect is especially visible in the “frame” scent carrier. While it

has the lowest scent-carrying capacity in the “good match” condition (as it has in all other conditions), it has the highest in the “bad match” condition.

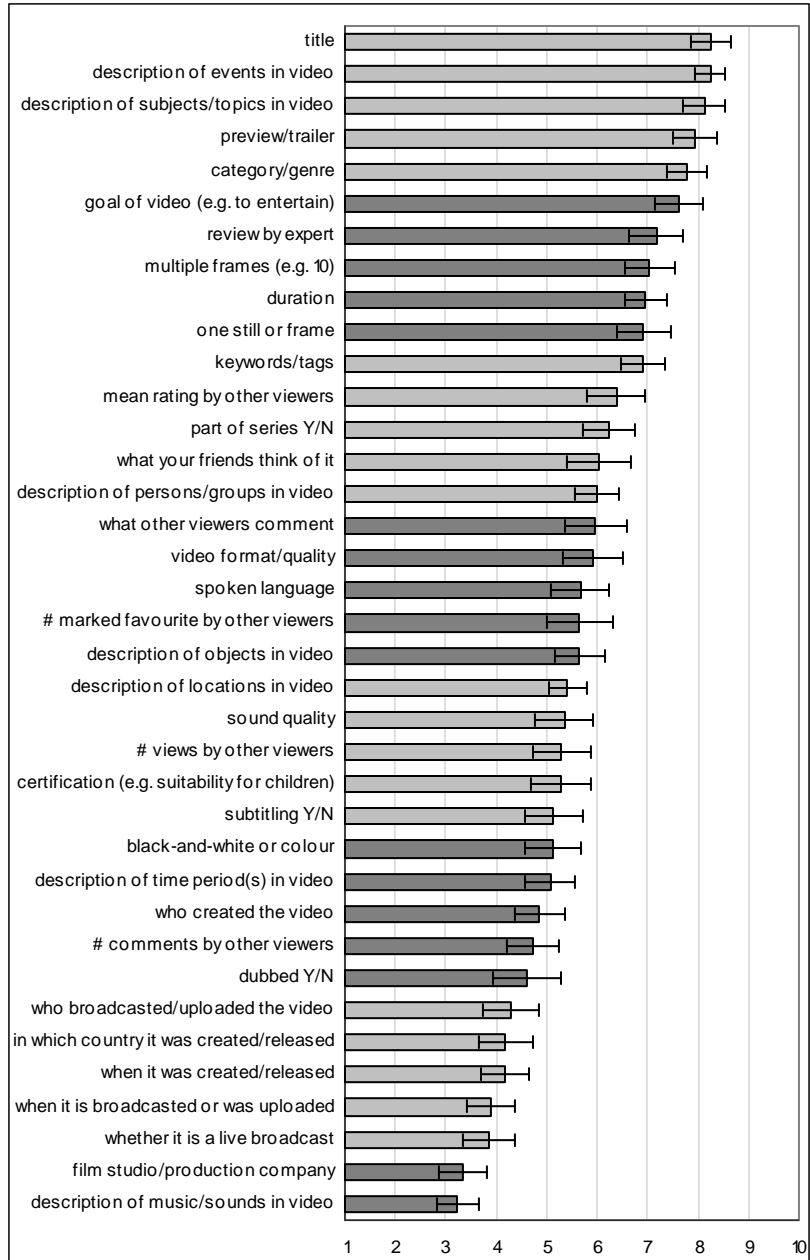
RQ4 – How useful (in general) are individual pieces of information about a video?

At the end of the experiment, we asked participants to rate the usefulness of 37 different types of information about videos on a scale of 1 to 10. Ratings were asked “in general,” so not for a specific goal or situation, unlike in the experiment where every rating was within the context of a specific task. Figure 3-5 shows that the following types of information were considered most useful in general (higher than 7.5 on a scale of 1 to 10): a title, a description of events, a description of subjects/topics, a preview/trailer, the category/genre, and the goal of the video. Other types of information with a pass mark (5.5 or higher) were: a review by an expert, multiple frames (e.g., 10), duration, one still or frame, keywords/tags, mean rating by other viewers, part of series Y/N, what your friends think about it, description of person/groups in video, other viewers’ comments, video format/quality, spoken language, how often it is marked favorite by other viewers, and description of objects in the video. The remaining 17 types of information did not get a pass mark.

In the experiment, “title” and “frame” were the scent carriers with the lowest perceived scent, with “frame” clearly being the worst of the two. In this list they are ranked 1 and 10 respectively. Other visual data representing the video - multiple frames (no. 8) and a preview/trailer (no. 4) – clearly did better than a single frame. All other scent carriers in the experiment were combinations with “title.” “Title+metadata” included category (no. 5), duration (no. 9), and when it was added or uploaded (no. 34). “Title+social” included other viewers’ rating (no. 12), number of times favorited (no. 19), number of views (no. 23), and number of comments (no. 29). The best piece of information from the social category was a review by an expert (no. 7). “Tags” were number 11 in the list.

We used several types of “description,” of which events (no. 2) and subjects/topics (no.3) did very well. Descriptions of persons/groups (no. 15), objects (no. 20), locations (no. 21), time periods (no. 27), and music/sounds (no. 37) clearly did worse. In the experiment, the general description as used on YouTube was used, which could contain any type of description. One top-10 piece of information - goal of video (no. 6) – was not used in the experiment, partly because this information was hard to establish, and partly because it is often implied or expressed in the category. A video from the category “Entertainment” is clearly meant to entertain, while a video from “News & Politics” is very probably meant to inform.

Figure 3-5 Usefulness (scored on a scale of 1 to 10) of types of information that can be used to describe a video (average score and 95% confidence interval)



This question relates to the question in the Kenniswijk survey displayed in Figure 2-4 in Chapter 2. There, respondents thought that a short description of the contents, a title, and metadata (genre year, maker, country, etc.) were the most useful pieces of information to get about a

video (all these received a pass mark – 5.5 or higher - on a scale of 1 to 10). At the same time, one still or ten still from the video both got very low ratings (below 4). These results are mostly replicated, although the usefulness of stills/frames is rated higher in this experiment.

3.5 Conclusions and discussion

The perceived scent of the scent carriers studied in this research differs significantly. The order from highest perceived scent to lowest is: 1) all; 2) title+frame; 3) title+description; 4) title+tags; 5) title+social; 6) title+metadata; 7) title; and 8) frame. The “frame” scent carrier is significantly worse than all others. The “all” scent carrier (containing all the information displayed in the other scent carriers) is significantly better than the others except for “title+frame” and “title+description.” This indicates that other combinations not used in this experiment (such as title+frame+description) would probably not have created a significantly better result than title+frame or title+description.

For the interpretation of the results we should be reminded that scent is the perceived usefulness of data in guiding users as they make navigational decisions. This usefulness appears to be a function of communication mode (verbal vs. pictorial); reason to browse (cognitive benefit vs. mood improvement); and tasks to perform (type of task, specificity, difficulty).

Verbal vs. pictorial. The results confirm the outcome of the eye-tracking study of Hughes et al. (2003), where participants looked at and fixated on titles and descriptions more than on the images. Their conclusion was that most people used the text as an anchor from which to make judgments about the search results and the images as corroborating evidence for their selections. This confirms the result of the Kenniswijk study in Chapter 2 that frames are considered to be of little use when determining the relevance of TV programs, especially compared to titles and descriptions. RQ4 on the usefulness of pieces of information about videos confirms the superiority of textual information, with “title,” “description of events,” and “description of subjects/topics” providing the most useful information (directly followed, however, by “preview/trailer”). The relative weakness of isolated frames as scent carriers may be explained by the fact that knowledge – and higher-order, more abstract knowledge in particular – is predominantly communicated through words. The combination of a frame with textual information (e.g., a title), however, is clearly strong in terms of perceived scent. The top two scent carriers in this study (“all” and “title+frame”) contained a frame.

Reason to browse. Restricting the analysis to the scent carriers *title* and *frame*, the results showed that the difference in perceived scent between

titles and frames is larger in the cognitive benefit task than in the mood improvement task. In other words, titles are clearly more important than frames when the reason to browse is cognitive benefit. This is in line with the dominance of text for knowledge acquisition that we already mentioned to explain the differences between verbal and pictorial scent. When the reason to browse is mood improvement the difference in perceived scent is much smaller, and frames can carry relatively more scent, what in this particular case means that frames are experienced as rather useful in representing relaxing videos. The explanation probably follows from the fact that video has emotion-evoking potential that is to a large extent based in the audiovisual design of the images (see Rottenberg, Ray, and Gross, 2007, for an example of research-based information about the use of movie clips for the elicitation of emotions). This analysis should be interpreted with caution: the overall effect of reason to browse was not significant.

Tasks to perform. The order of the type of tasks with the greatest effect to smallest was: 1) specific – easy + good match (Task 3); 2) general – cognitive benefit (Task 2); 3) general – mood improvement (Task 1); 4) specific – hard + good match (Task 4); 5) specific – hard + bad match (Task 5). This order is significant. Task specificity also had a significant effect: with a simple task and with relevant videos present in the result set, perceived scent is higher than with a general (cognitive benefit) task. This is a characteristic of most studies we see on this subject: task descriptions are clear, and targets are present. It is the optimal situation for a (video) search task. In many cases, however, the task is more difficult, and video representations simply cannot provide enough information to solve the problem. Perceived scent significantly drops when task difficulty increases. Our results show that task difficulty increases when the match between the task description and the videos present in the result set is bad. The perceived scent of all scent carriers is then significantly lowered: the user will find it nearly impossible to relate the information from the scent carriers to the task. The differences between scent carriers decrease when the match gets worse, causing any advantages of certain scent carriers to disappear. This is comparable to the situation in Pirolli, Card, and Van Der Wege (2003), where the performance advantages of an advanced browser disappear under conditions where scent is low. A weakness of our study is that for practical reasons there only was one query per task. There are strong indications that the task factors we studied have a significant effect on the perception of scent, but there is a need to confirm these results with additional user studies in which several queries are used for each task type. These queries should preferably be as realistic as possible and be a result of observations of the real-life types of questions people have in interaction with video databases.

The results show that the representativeness of scent carriers determines whether the scent of video targets reaches the interested user. In that respect it has to be noted that the scent carriers in the interaction between participant, task, and scent carrier in this study were taken directly from YouTube (not even the typing mistakes were changed). In other words, they represent the situation that people searching for videos will encounter in the real world. The result sets contained good and bad titles, good and bad frames, and so forth.

As far as the frames are concerned it should be mentioned that the ones used in this study were all the middle frame of a video, the same approach YouTube applies. As was demonstrated in Christel, Winkler, and Taylor (1997), frames that have a direct relation to the query can produce better results than naively chosen frames. In general it can be said that, for a person searching in videos, some frames of a given video will have more scent than others. The problem of course is to present the relevant frame for every query type. Still, as stated before, RQ4 confirms the superiority of title (usefulness score 8.3) over frame (6.9) (see Figure 3-5). This question provides us with a check on what users generally think is useful to know about a video. If the results from the experiment were influenced by the way the scent carriers were created, we would find incongruence between these results. Now, both parts of the study confirm that titles are more useful than frames.

As was stated above, the scent carriers we used were exact copies of YouTube metadata. Very probably there was room for improvement in the scent carriers. However, trying to create the best titles, best descriptions, most representative frame, and so on for each video is not a simple task. As the results of our study confirm, the representativeness of scent carriers depends on the reason to browse and type of task. A question is whether this problem might be eased when users have the opportunity to browse the video content of a patch. In this study the participants judged only the perceived scent of the one link (out of ten) that they preferred (and this for eight different result sets). They had no access to the video content behind any of the links. Would they have modified their judgment if they were able to browse through the patch that was defined by each result set?

Now how can we import the results of this user study to enhance video browsing? Combining the results of the experiment and the extra question on the usefulness of types of information about a video, the conclusion is that the most important thing to know about a video is the title. However, combining a title with other information enhances perceived scent of the represented video. We saw that “title+frame” and “title+description” did not carry significantly less scent than the “all” situation. Figure 3-5 suggests a more important role for description than frame, with the ideal description being of the events and subjects/topics in the video. A preview/trailer is

indicated to be a more useful way to represent the video data than one or multiple frames, but this scent carrier typically has a role further on in the search process (step 2 in the search scenario as described at the end of section 3.1.1). Viewing the preview or trailer for every video in a list of videos would require a lot of time and attention, so typically a user would do this only after establishing - on the basis of the available scent carriers - that the video source is worth further examination. If we were to design a composed scent carrier for the earlier stage of search - when people have to choose from a list of links to videos - it would at least contain the elements title, description (of events, subjects, or topics), the category/genre, and the goal of the video (which is often directly related to the category/genre). Preferably it would also contain a frame, especially because of the strength of the image-title combination. Other elements - an expert review, duration, ratings by others - can have importance in specific situations, but the abovementioned elements seem to be essential for supporting efficient video browsing. Additional user studies are needed to examine the design of scent carriers that are meant for exploration of the contents after a video is selected from a list.

In our approach, links to videos - as provided by video applications such as YouTube - are seen as carrying scent. This way, scent - as one of the central concepts of IFT - is made operational, allowing us to study the usefulness of IFT in approaching the problem of video interaction. The experiment in this chapter describes the relevance of the characteristics of a scent carrier (representing a video source) and the characteristics of the user's task for the perception of scent. As such, we have built further on our IFT-based framework. This study indicates that the idea of scent can - via the design of scent carriers - provide insight into the design and evaluation of an efficient video environment.

Now that we have a better concept of video patches and video scent, the next step in our research is to see whether the video interaction behavior we observe in real browsing situations can be described in terms of IFT. In the next chapter, we describe an experimental video browsing environment we developed applying ideas from IFT and using the results from the studies described in Chapters 2 and 3: the VIBES video browser. We also describe two other video browsing environments known from the internet - YouTube and Fabchannel -- from an IFT point of view.

We utilized these three environments in an experiment to observe user browsing behavior. Unlike in the study described in this chapter, users had the ability to browse or watch the video behind the link and interact with the video database. In other words, they could now evaluate the choices they had made based on the presentation of the links. The results of that study are described in Chapter 5. But first, in Chapter 4, we will describe the video environments that were used in that experiment.

Video interaction environments

In the previous two chapters, we discussed ways to classify videos (Chapter 2) and ways to present scent (Chapter 3). The next step is to look at video browsing behavior. However, browsing behavior is very much dependent on the functionalities and interaction support provided by the browsing environment, which can both hinder and facilitate browsing behavior. The objective of this chapter is to look at video browsing environments from an IFT point of view. In the next chapter, Chapter 5, we will present an experiment in which we a) study browsing behavior with the environments described in this chapter, and b) evaluate the environments. As such, this chapter can be seen as part of the method section of the experiment described in Chapter 5.

Almost as long as film and video have existed, there have been collections and more or less advanced environments for interacting with those collections. A remarkable development was the digitization of video collections, allowing new interaction modes such as direct access to specific video segments and hyperlinking. Moreover, high-bandwidth internet connections have made a large number of videos available online. Several video on demand services (Wikipedia, 2008a) and video hosting/sharing services (Wikipedia, 2008b) have set up shop on the internet. In addition, several research communities have been working on the development of video browsing environments, including Carnegie Mellon University (Informedia, 2008), the University of Amsterdam (MediaMill, 2008), Dublin City University (Fischlár, 2008), and the University of North Carolina at Chapel Hill (OpenVideo, 2008). Several other recent research projects have been aimed at designing interfaces to support exploratory search interaction (for an overview, see White, Drucker, Marchionini, Hearst, and schraefel, 2007).

In this chapter, we will concentrate on the three browsing environments we used in the experiment described in the next chapter, and which

therefore require a detailed description. First, we will present the video browsing environment we have been developing in recent years: the VIBES video browser. Then we will discuss two popular examples of video services on the internet: YouTube and Fabchannel. We chose the former because it is the most popular general video website, the latter because it is a very good example of a specific video website with interesting interaction modes.

We will provide a description mostly from an IFT point of view. This means we are interested in such questions as: “Which patches are or can be created?” “How is browsing within a patch supported?” “How is switching between patches supported?” and “In which way is scent displayed in the interface?” Also, we will see how bridging of the gaps – as described in Chapter 1 - is supported.

4.1 VIBES video browser

The VIBES video browser has been in development for several years. The idea of a patch-based browser arose as a way of implementing IFT in video browsing (van Houten, Van Setten & Schuurman, 2003). The practical development of that environment also gave rise to research questions about video patches and video scent, as described in van Houten, Schuurman, and Verhagen (2004). Using the results of the user studies described in the previous chapters, the experimental video application was further worked out. It was intended primarily to provide a context for studying browsing behavior within the IFT framework.

The first principle from IFT that was applied was the idea of a “patchy” structure. In Chapter 1, we defined a video patch as a collection of video segments sharing a certain characteristic. This can mean a number of things.

- Each individual video is a patch, often consisting of smaller segments which have some kind of relationship (for example, they are part of a narrative structure, or were recorded at a certain location and time)
- Segments from one video together form a patch (for example, the highlights of a football game, or all appearances of an actor in a movie)
- Segments from different videos form a patch (for example, highlights from an actor’s career)
- Different videos together form a patch (for example, all videos uploaded by user X, or all movies from Norway)

To make things more complicated, in the second and third cases the segments can be combined to form new individual videos. For example, a video on YouTube showing highlights from the career of Dutch football player Dennis Bergkamp contains segments taken from different matches,

all separate patches. The collection of highlights itself was probably a segment of a sports program, which is another patch.

In Chapter 2, we established that people prefer to interact with semantically meaningful segments of video programs, either the segments the program maker envisioned or those “naturally” provided by the characteristics of the content. This is an important unit for interaction, in some cases perhaps even more important than programs as a whole. Smaller units of interaction than “natural” segments, on the other hand, are considered less useful. So segmenting videos into their semantically meaningful units is the first important step in patch-based interaction.

There are two ways to do this: with or without leaving the original context intact. In the first case, a (large) video is segmented and direct access to the different segments is provided within the original video. In the second case, a video is segmented, but the segments are treated as individual videos. Often, it is only one segment that is selected from a larger video to lead a life of its own somewhere on a weblog or website such as YouTube.

The VIBES video browser can handle both cases. The VIBES video environment was developed in Java on the Eclipse platform, and consists of an editor and a browser. The editor has two main functions: it supports users in manually segmenting a video, and adding metadata to the video and its segments. Metadata are added according to the MPEG-7 standard and are stored using XML.

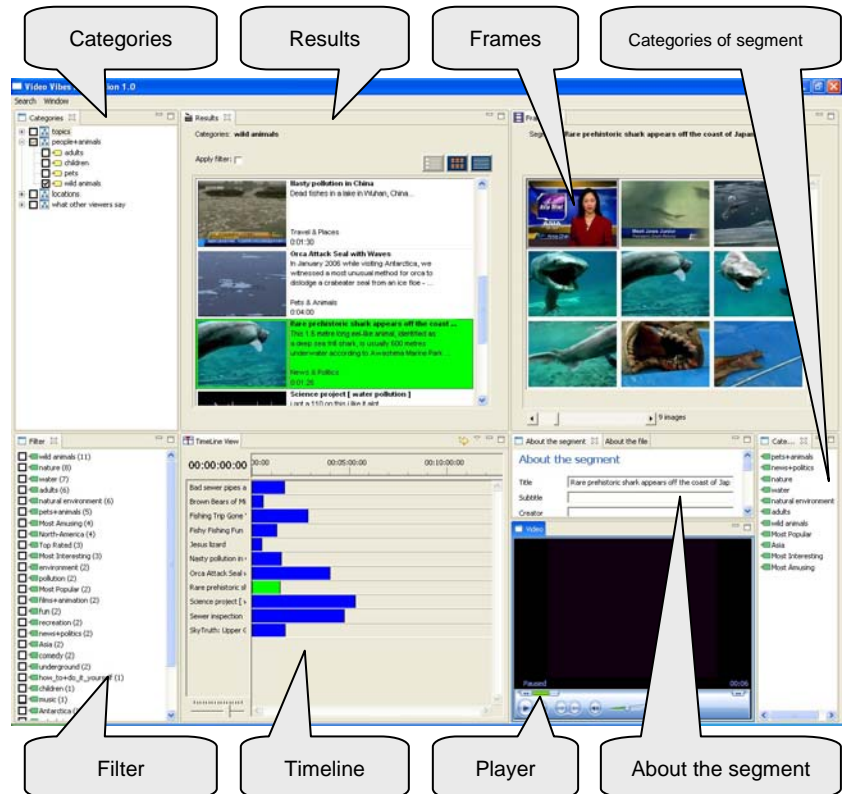
Figure 4-1 presents the interface of the VIBES video browser. The interface consists of a number of windows or modules. The size and location of these modules is flexible and can be changed by respectively dragging the outside border or the top bar with the module name. However, any change affects the display of other modules, so that a lot of consideration is needed when adjusting the interface. A default layout is presented here to which users can easily return (via the Windows menu button of the application).

The following modules are present:

- *Categories* (top left) – displays the available categories or patches. A category can be selected by checking the box in front of the category name. More than one category can be selected at the same time (which in terms of logical operators is an *OR* selection). Double-clicking a category overrules the current selection and selects only that category. All categories within a category group can be selected by checking the box in front of the category group name. The names of the categories can give off scent.
- *Filter* (bottom left) – The selection made in *Categories* contains a number of videos. All categories available in this video selection are displayed in the list in the *Filter* module (which includes the number of videos that

are present in every category). So while *Categories* displays all categories available in the database, *Filter* displays all categories that are part of the current video selection. Checking a box in front of a category in this list excludes videos from the selection made in *Categories* (which in terms of logical operators is an *AND* selection). With the exclusion of videos, some categories may be excluded from the filter (no videos are left in that category), in which case the category name is grayed out. Multiple boxes can be checked at the same time. In terms of IFT, the *OR* and *AND* selections can be seen as a type of *enrichment* (Pirolli & Card, 1999). These selections are useful for bridging gaps 2 and 3 (as described in Chapter 1): the gap between the decision and the initiation of the action, and the gap between the initiation of the action and the continuation (and completion) of the action.

Figure 4-1 Screen dump of the VIBES video browser



- *Results* (top middle) – Displays the results of the selections made in *Categories* and *Filter*. The status of the selection (which categories are chosen and which filters are applied) is presented at the top of the module. The module contains a check button for applying the filter, which facilitates the evaluation of what the effect of the filter is. Figure

4-2 shows the three different ways of displaying the results which can be chosen via the selection of an icon (at the top right of the module). The choice of these scent carriers was based on a combination of the outcome of the scent experiment described in Chapter 3, and practical user interface issues:

- the *all* scent carrier displays a thumbnail (the middle frame of the video), the title, description, main category, and duration. Although this was the scent carrier with the highest scent-carrying capacity in the previous experiment, a disadvantage is that it takes up space (about three results can be displayed at one time, depending on the chosen size of the module), requiring people to scroll a lot
- with the *title + frame* scent carrier, six results can be displayed at one time
- with the title scent carrier, twelve results can be displayed at one time

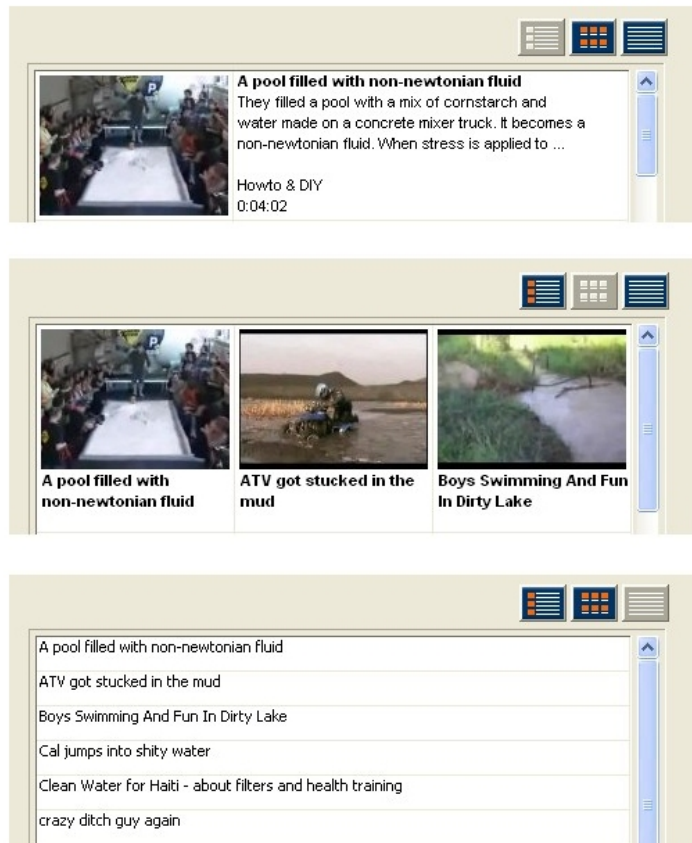
A selected video in the list is highlighted.

- *Timeline* (bottom middle) – Is actually part of the *Results* module, and displays the same videos. Shows the duration of the videos in the selection via colored bars. Selecting a video here has the same effect as selecting a video in *Results*. When a filter is applied to the results, all bars for the videos that are excluded by the filter turn white. This gives visual feedback of the application of a filter.
- *Frames* (top right) – When a video is selected in *Results* or *Timeline*, a visual summary is provided in the *Frames* module. The summary is time-based: the video is cut into equal parts, each showing the first frame of the section. By default, 9 frames are displayed, but this number can be changed to any number between 1 and 60. Displaying more than 9 frames requires scrolling (although with other module sizes this may differ). The frames are visual representations of the video and are capable of carrying scent. If a user detects scent in one of the frames, double-clicking the frame starts the video at that particular point in the *Player* module.
- *Player* (inner bottom right) – Double-clicking a video in *Results* or *Timeline* causes it to start playing. So does double-clicking a frame in *Frames*. Further, any selected video can be played by clicking the Play button. The player is an embedded Microsoft Media Player with standard functionality. The slider on the timeline and the fast-forward can be used to quickly view images from the video, which may or may not provide scent to the user.
- *About the segment* and *about the file* (above the *Player*) – Provides metadata (and thus scent) about the selected video or segment. This is the same metadata that can be seen in the *all* scent carrier in the results, although more space is available (e.g., for the transcript). If the video consists of

- only one segment (in other words, if it is not segmented), the information displayed in *about the segment* and *about the file* is the same.
- *Categories of segment* (outer bottom right) – Displays the tags of the selected video, or in other words, the categories the video (segment) fits into. These metadata are displayed in a separate module that has different functionality: double-clicking one of the categories overrules the current selection as displayed in *Results*, and selects only that category. This was implemented to ease switching between patches when one of the displayed categories contains a lot of scent for the user. It can be seen as an extra support in bridging gap 3 (aimed at the continuation of the action).

Search functionality (applying keyword queries) was not implemented at the time of the experiment. If it had been, search results would have been treated like any of the other tag-related results.

Figure 4-2 Three ways of displaying the results in the VIBES video browser: *all* (top), *title + frame* (middle), and *title* (bottom)



In sum, the VIBES video browser is an experimental video interaction environment, built around the idea of patch-based browsing. Users are

encouraged to create patches by combining categories and can easily switch between patches. According to preference, different scent carriers can be chosen to represent the video segments, encouraging users to follow links to video segments.

4.2 YouTube

YouTube is a video-sharing website where users can upload and view video clips. Currently, it is the most popular video website on the internet. Last year in the US it was used by 27% of online video consumers (Madden, 2007). Young adults are almost twice as likely to point to YouTube as a source for online video; 49% of video viewers age 18-29 say they watch YouTube videos.

YouTube supports both browsing and querying for bridging gap 2 (and gap 3 if search needs to be continued). The website’s start page shows a search bar for posing a keyword query. Browsing can be started by selecting one of three pages: Videos (=Categories), Channels, or Communities (=Groups). These pages contain patches with videos. On the Videos page, videos are grouped by category, on the Channels page by the person who uploaded the videos, and on the Community page by theme (started by users).

Figure 4-3 The “Videos” pages of YouTube, showing the video categories and different ways of arranging the videos

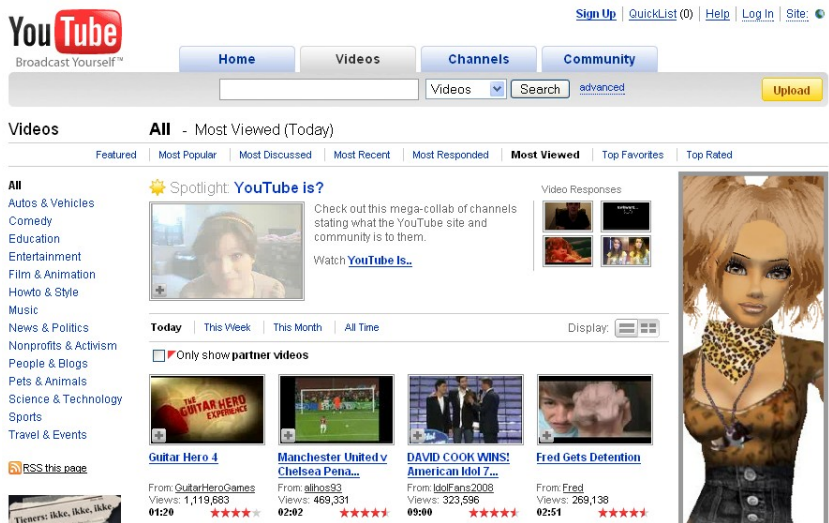


Figure 4-3 shows that the categories YouTube currently uses are Autos & Vehicles, Comedy, Education, Entertainment, Film & Animation, How to & Style, Music, News & Politics, Nonprofits & Activism, People & Blogs, Pets

& Animals, Science & Technology, Sports, and Travel & Events. The category Music is the only category with subcategories (12 music genres). Within a category, videos can be ordered by date (Most Recent) and by the following social data: Most Popular, Most Discussed, Most Responded, Most Viewed, Top Favorites, and Top Rated. Depending on the interests of the visitor, these patches or orderings can have more or less scent. A keyword query leads to a patch with results. An example is shown in Figure 4-4, displaying the results of the keyword search for “bill clinton about hillary” (an example closely related to the YouTube task which was part of the study described later on in this chapter).

Figure 4-4 Part of the YouTube interface, showing the results of the keyword query “bill clinton about hillary”

The screenshot shows the YouTube search interface. At the top, there is a search bar containing the text "bill clinton about hillary", a dropdown menu set to "Videos", and a "Search" button. Below the search bar, the text reads "“bill clinton about hillary” video results 1 - 20 of about 8,670".

The results are displayed in a list format. Each result includes a video thumbnail, a title, a description, and additional information such as the date added, the uploader's name, the number of views, an average rating (represented by stars), and the video duration. The first result is titled "President Clinton responds to Hillary's 'Children' ad." and has 172,526 views and a 5-star rating. The second result is titled "BILL CLINTON ON HILLARY CLINTON AND BARACK OBAMA 'SPAT'" and has 10,678 views. The third result is titled "Bill Clinton talks about Hillary on Larry King" and has 35,146 views.

At the bottom of each result, there is a link that says "More in News & Politics".

The patch contains videos related to the keywords. Different ways to arrange the results are presented, the default being “by relevance.” Users can explore the patch using a scroll bar. Several scent carriers are displayed for each video within the patch: a frame, a title, a description, date added, the user who uploaded the video, social data (number of views, average rating), and category. If the patch does not contain enough scent, the user can try a new search with other keywords (a form of enrichment).

The name of the uploader and the category are hyperlinked. Each of these links leads to a new patch: all videos uploaded by this user and all videos in this category, respectively. If the name of the uploader or category

carries enough scent for the user starting the search, that user may feel encouraged to explore these other types of patches.

Selecting a video from the result patch leads to a new page, of which an example is displayed in Figure 4-5.

Figure 4-5 Part of a screen dump of the YouTube interface, showing a video, its metadata, and related information

The screenshot shows a YouTube video player for the title "Bill Clinton talks about Hillary on Larry King". The video player includes a progress bar at 02:46 / 06:09. Below the player, the video has a 4-star rating from 107 users and 35,228 views. It offers sharing options for MySpace, Facebook, and Digg. The page features a "Commentary" section with "Statistics & Data" tabs, showing 0 video responses and 129 text comments. Several comments are visible, including one from "smartwheel" and another from "Cerrito1234". To the right, there is a channel profile for "victorkahn" with a "Subscribe" button, and a list of "More From" and "Related Videos" such as "Hardcore turtle", "You and I are DONE! (What?)", and "Barack Obama - Larry King Live".

The figure displays a variety of information about the selected video. The player presents the video data, which can be explored quickly using the slider on the timeline. The other data have been extended compared to the representations used on the results page. The main extension for the social data is the display of comments on the video by other users. This can involve descriptive information that can influence the scent of the video. Every commenter is a YouTube user, and a hyperlink leads directly to the patch of this user. The name of the commenter or the content of the comment may carry enough scent so that the user will feel encouraged to

follow the link. Comments can also be given in the form of a video, which may again lead the user to other patches.

The metadata are mainly extended by displaying tags. These again are hyperlinks which lead to video patches related to those tags. They also provide ideas on which keywords to use in the search.

The patch with other videos from the user who uploaded the current video is displayed prominently. The figure shows that this patch is represented with representations from five videos. If one of these representations (or their collective appearance) contains enough scent, the user may want to move to that patch.

One important addition is “Related Videos.” Although it is not made clear how these videos are related to the currently selected video, these videos by definition carry scent if the user thinks the current video is interesting. This listing may also provide new ideas to the user on how to search further, which can be seen as support for bridging gap 3.

In sum, users are encouraged to explore the hyperlinked, patchy environment that is provided by YouTube. Several scent carriers are available to lure users to other patches. They can create their own patches (enrichment) by applying different keyword searches. However, this does not mean that YouTube is necessarily the easiest place to find relevant videos. This is also a function of the quality (and quantity) of the video content present in the database, and the quality of the metadata added to the content.

4.3 Fabchannel

The Fabchannel website streams free, live, and on-demand video from the Paradiso and Melkweg venues in Amsterdam and the Roxy Theatre in Los Angeles. With more than 900 concerts (including some other activities at the abovementioned venues) Fabchannel.com has built the biggest concert video on-demand archives in the world. The website has about 10,000 visitors a day. In 2006, it won a Webby Award for best music website.

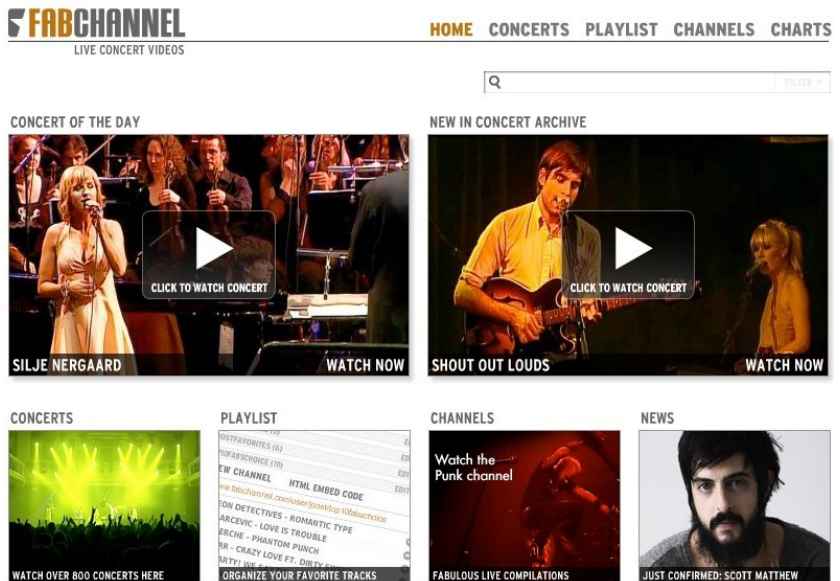
Unlike YouTube, Fabchannel has full control over the video content in its database and the metadata that are added to the content (although it is starting to experiment with user-added metadata). Another important difference from YouTube is that the database mainly contains videos from one genre: music concerts. It also has a strong relationship with the specific venues where the videos are recorded. For a certain category of visitors the website contains a lot of scent, even more so when the database contains concerts of their favorite music recorded in a venue where they were also present. As we saw in the study described in Chapter 2, the main reasons the respondents visited the site was to “find a specific concert in the archive

to watch” and to “discover new music.” These are very different reasons, requiring different support for interaction.

Fabchannel has always been eager to create an attractive video environment that also provides effective and efficient support for its visitors, and has sought ideas for this in the academic community. This led to the user study described in Chapter 2, and to an experimental video browser supporting patch-based browsing, applying the use of segmentation algorithms in a media post-processor (van Houten et al., 2005). Ideas from that browser are applied in the current user interface of the website.

Figure 4-6 shows the start page of the Fabchannel website. Visitors who want to find a specific concert can go directly to “Concerts” or go to the search bar and type the name of the performing artist, thus bridging gap 2. When the user starts typing in the search bar, the “Concerts” page (see Figure 4-7) is automatically selected. The user can also check if the concert he/she recently visited is already in the database by arranging the concerts by date and viewing the latest additions to the database. If a recent concert is not yet available, it will be in the “pending” patch on the “Concerts” page.

Figure 4-6 Part of the interface of Fabchannel, showing the start page



The visitor with a less specific goal has a number of choices at the front page for bridging the second gap. The choices are presented in a visually pleasing way which may contain scent to certain visitors and lure them to the available patches. Two concerts – the “concert of the day” and “new in the concert archive” – get a poster-like announcement with a large frame and the name of the artist. Unlike in the YouTube archive, the frames here

are carefully selected by Fabchannel employees, which means that they are able to carry more scent than when the first or middle frame of the video is used. On the bottom left, and on the menu on top of the page, visitors can choose “Concerts,” leading to the page displayed in Figure 4-7.

Figure 4-7 Part of the interface of Fabchannel, showing the page with concerts. The thumbnails on the left present a video snippet when the mouse is placed over the image. Concerts can be sorted by artist and date. Under “FILTER” there are options for selecting concerts from a certain genre or venue.

The screenshot shows the Fabchannel website interface. At the top, there is a navigation menu with links for HOME, CONCERTS, PLAYLIST, CHANNELS, and CHARTS. Below the navigation, there is a search bar and a filter dropdown menu. The main content area is divided into two columns. The left column, titled 'ARCHIVE HIGHLIGHTS', displays a grid of video thumbnails for various artists and genres, including ROCK, HIPHOP, SINGER/SONGWRITER, and METAL. The right column displays a list of concerts, sorted by artist and date, with a search bar and a filter dropdown menu.

ARCHIVE	COMING UP	PENDING	OFFLINE
SEARCH			
ARTIST	SORT DATE		
NINA KINERT	17-04-2008		
BLAUDZUN	17-04-2008		
ANE BRUN	17-04-2008		
THE IPANEMAS	15-04-2008		
THE BRUNETTES	13-04-2008		
SHOUT OUT LOUDS	13-04-2008		
TOBIAS FROBERG	11-04-2008		
THINGUMA/JIGSAW	11-04-2008		
SMOOSH	11-04-2008		
PETER MOREN	11-04-2008		
PINCHED	07-04-2008		
SILJE NERGAARD	05-04-2008		
KATE NASH	02-04-2008		
KAIZERS ORCHESTRA	01-04-2008		
GEOFF BERNER	01-04-2008		
LIGHTSPEED CHAMPION	31-03-2008		
SIMONE WHITE	30-03-2008		
PROF. DR. RON FOUCHIER	30-03-2008		
CHARLES FRAIL	30-03-2008		
THE NATIONAL BANK	26-03-2008		
LOCALS ONLY MARCH 2008	26-03-2008		
22 PISTEPIRKKO	26-03-2008		
MARCUS MUMFORD	25-03-2008		
LAURA MARLING	25-03-2008		
CANDY DJULFER & BAND	24-03-2008		

Here the site presents a list of concerts, which can be browsed by scrolling. The archive is constantly increasing, however, and the longer the list gets the less effective this simple strategy becomes. Using the “filter,” the visitor can sort the concerts by genre (for example, “punk,” “R&B”) or venue (for example, “Paradiso”), thus creating smaller patches which may have a stronger relation to the visitor’s interest. On the left, “archive highlights” are displayed, showing four highlights per genre patch. Scent can be provided by the genre, artist name, and frame, but especially by what happens when the mouse is placed over the frame: this starts a video snippet, providing a quick impression of the contents of the video. This includes the sound, which is a very important aspect of this specific genre, making this a very effective scent carrier. The snippet starts and stops immediately depending on the position of the mouse.

Another way that patches are created is with channels, which do not contain collections of videos, but collections of video *segments*, in this case songs. These patches may be created by Fabchannel and may relate to, for example, a genre (e.g., the best punk songs in the archive). They may also be collections of songs created by visitors. Under “Playlist” (see Figure 4-5), visitors can create a personal playlist with songs, which is also made possible because most concerts in the archive are segmented. These

personal playlists can be shared with other users, thus creating patches of the type “favorite songs from the archive as chosen by user X.” These playlists can also be embedded in external websites and blogs, thus creating links to the Fabchannel website. These channels may support the discovery of new music, especially when a visitor already likes a song within the channel. (“If I like a song that user X likes, than perhaps I may also like other songs – which I don’t know yet – that user X likes.”)

Another way to support the discovery of new music is by looking at popularity. If a lot of other visitors like certain patches, to a certain category of users this will mean those patches carry scent. Selecting “Charts” on the front page leads to the page displayed in Figure 4-8. The page displays the most popular concerts and channels within the database.

Figure 4-8 Part of the interface of Fabchannel, showing the top 10 of “channels” on the left and a concert top 50 on the right.

The screenshot shows the Fabchannel website interface. At the top, there is a navigation bar with the following items: HOME, CONCERTS, PLAYLIST, CHANNELS, and CHARTS. Below the navigation bar, the page is divided into two main sections. On the left, under the heading "TOP 10 CHANNELS", there are four video thumbnails. Each thumbnail has a play button and a "COPY HTML CODE TO CLIPBOARD" button below it. The thumbnails are labeled: ROCK, HIPHOP, SINGER/SONGWRITER, and PUNK. On the right, under the heading "FABCHANNEL CONCERT TOP 50", there is a search bar and a list of 50 concert entries. Each entry is numbered and includes the artist's name. The list is as follows:

Rank	Artist
1	JOSE JAMES
2	XAVIER RUDD
3	MARCUS MILLER
4	BLOC PARTY.
5	THE VIEW
6	OF MONTREAL
7	CSS
8	WOUTER HAMEL
9	APPA
10	SIN7SINS
11	OPGEZWOLLE
12	WIR SIND HELDEN
13	DAMIEN RICE
14	KULA SHAKER
15	DIO
16	KLEINE JAY
17	FLOGGING MOLLY
18	BRIGHT EYES
19	THE ARCADE FIRE
20	JOHAN
21	DO MAKE SAY THINK
22	HANNE HUKKELBERG
23	FICTION PLANE
24	SIMPLE MINDS

Once a concert or channel is chosen, the visitor can browse within this patch. Figure 4-9 shows the player with the selected video and related metadata of the video/concert. The video’s segments, the songs, can be directly accessed in two ways: via the table of contents on the right, or via the related bars on the timeline of the video below the player. Color coding relates the bar on the timeline to the name of the song.

Selecting a channel leads to a similar interface, with the exception that no timeline (with bars relating to the songs) is available. Unlike a concert patch, a channel patch consists of video segments from different videos/concert. In both cases, when the visitor recognizes the name of the

song, it may carry scent and the visitor can follow the link to the video segment or song.

The channel patch (with songs from different artists) can be browsed by selecting the songs. If a song interests the visitor, he/she may want to see the source of that song: the artist's complete concert. So, a song in a channel is a "large" scent carrier, representing a concert patch. However, a direct link to the concert – which would have eased the bridging of the third gap – is not yet implemented.

Figure 4-9 Screen dump of Fabchannel's player and the metadata of the currently selected video/concert

The screenshot shows the Fabchannel website interface. At the top, there is a navigation bar with links for HOME, CONCERTS, PLAYLIST, CHANNELS, and CHARTS. Below this, there are tabs for ARCHIVE, COMING UP, PENDING, and OFFLINE. A search bar is located on the right side of the navigation bar. The main content area is titled "CONCERTS | KAIZERS ORCHESTRA". On the left, there is a video player showing a live concert scene with a large lamp and a smaller lamp. The video player has a progress bar and an "EMBED" button. On the right, there is a section for "Kaizers Orchestra" with a play button and the date "01-04-2008" and location "Paradiso - Main Hall". Below this, there is a "TRACKS" section with a list of 18 tracks, each with a play button and a plus sign to add it to the playlist. The tracks are: 01. INTRO, 02. MASKINERI, 03. BASTARD SONN, 04. 9 MM, 05. SIGOYNERBLOD, 06. OMPA TIL DU DOR, 07. BON FRA HELVETE, 08. BLITZREGN BABY, 09. VOLVO I MEXICO, 10. KONTROLL PA KONTINENTET, 11. ENDEN AV NOVEMBER, 12. KGB, 13. RESISTANSEN, 14. APOKALYPS MEG, 15. MAESTRO, 16. KNEKKER DEG TIL SIST, 17. BAK ET HALLELUJA, 18. DIETED MEYERD INKT. Below the tracklist, there are links for "BIOGRAPHY" and "LINKS".

In sum, like YouTube, Fabchannel encourages visitors to browse the database by offering different types of patches. In particular, it has created attractive scent carriers (carefully picked frames and video snippets) to lure visitors to patches.

It should be emphasized that the Fabchannel website is under constant development, and that the description above refers to the status during the execution of the experiment described in the next chapter. Of course, this also applies to the YouTube website.

4.4 Conclusion

One goal of this chapter was to provide a description of video browsing environments, primarily from an IFT point of view. The conclusion is that this is actually relatively easy. We can indicate which patches are present or can be created, how within-patch browsing is supported, how switching

from patch to patch is supported, how scent is displayed, and so on. The question is, however, whether IFT can provide useful tools in the design and evaluation of interaction environments, or, in other words, whether our framework has practical meaning.

In the next chapter, we will describe an exploratory user study in which the VIBES video browser, YouTube, and Fabchannel are used to observe browsing behavior from the perspective of IFT. This study is aimed at evaluating the applicability of IFT and, at the same time, will provide an evaluation of these browsers, after which we will try to answer the main question that we introduced in Chapter 1: “How to design a video interaction environment that will optimally support its users?”

Video-foraging behavior

In the previous three chapters, we discussed ways to classify videos into patches (Chapter 2), ways to present video scent (Chapter 3), and environments for video interaction (Chapter 4). In this chapter, we will deal with the third research question from Chapter 1: “How to design a video interaction environment that will optimally support its users?” As explained in Chapter 1, this question is about design principles for a patch-based browsing environment that effectively bridges the three gaps for efficient support of browsing video data. Briefly, the basic question is how to create a browsing structure that will give people optimal support in finding what they need. As indicated in Chapter 1, *optimal support* is reached when patches and scent carriers together help people bridge all three gaps at a rate that maximizes user satisfaction over [search] time.) Before that question can be answered, we have to know how people search for videos, and what kind of support they need to reach their goals. We will describe an exploratory study in which we observed the browsing behavior of people interacting with the three video browsing environments described in Chapter 4.

With the VIBES video browser, we conducted an experiment in which we asked participants to perform three different tasks. This yielded two kinds of information. It allowed us to collect quantitative information on the usefulness of the elements of the application for the various task executions. At the same time, it allowed us to gather data on video interaction behavior by observing how the participants executed the tasks.

The participants also performed a task in the Fabchannel and YouTube video environments. This provided more data for a qualitative analysis of the difficulties of video interaction. For all tasks, we specifically looked at searching behavior that we could relate to patch-based browsing, scent-following, and gap-bridging. We evaluated the participants’ browsing

strategies, discussed the difficulties they encountered, and asked how the video environment could be improved for better task execution.

At the end of this chapter, we will evaluate whether we can describe interaction completely in terms of IFT, and whether we can use IFT to explain browsing behavior. Furthermore, we will evaluate whether the concepts of patches, scent, and gaps have practical meaning: can we use them to design video environments? Moreover, the results will be used to draw conclusions on what kind of support people really need when browsing videos.

First, we will formulate the specific research questions we are going to deal with in this chapter.

5.1 Research questions

In Chapter 1, we started the discussion on video interaction behavior. We saw that there are two main strategies for interacting with videos: formal, analytical ones and informal, browsing ones. Empirical studies show that there are many search situations where the target is not well known and a single fact or document will not suffice. Users can have great difficulty in articulating their information need, which often happens to be ill-defined and often evolves during interaction. We concluded that – although forms of filtering are required - for video the dominant strategy is browsing. Emphasis is on the flow of representations and actions rather than discrete matches. The problem with video is that it is time-based, and that video interaction can become very time-consuming. So what are needed are efficient solutions maximizing the allocation of human attention to useful information.

To better understand video interaction, we introduced a framework based on IFT. In that framework, people interact with videos in order to bridge gaps: gaps between reasons for a decision and the decision; gaps between the decision and the initiation of the action; and gaps between the initiation of the action and the continuation and completion of the action. IFT provides the “cognitive” tools for bridging the gaps: tools that people use to structure their environment and interact with that environment. We introduced the ideas of *video patches* - as a way to structure the environment (further worked out in Chapter 2) – and *video scent* – as a guidance principle for navigating the video environment (further worked out in Chapter 3). Now how does this all work in practice? Can we use these ideas to study video browsing behavior, or to evaluate video environments? Even better, can we use these ideas to *design* video environments?

We want to execute an experiment in which we can make manifest the structure of video interaction to be able to arrive at answers to our main

question: “How to design a video interaction environment that will optimally support its users?” For that purpose, we need to study the two sides of interaction: the user or “interactor,” and the system or interaction environment. We do so from an IFT point of view: we look at the user as a searcher trying to bridge gaps by jumping from patch to patch, navigating by following scent. We view the interaction environment as being “patchy” and providing links which can carry scent. At the same time, this experiment allows us to evaluate this approach and explore the practical value of the concepts *patches* and *scent*. This discussion on the usefulness of IFT will be part of the final chapter, Chapter 6.

The problem with studying browsing behavior is that this behavior is dependent on the tool - the browsing environment - that is provided. This is one reason why we need to study this behavior in different environments. For the same reason, we will evaluate each task emphasizing with which adjustments it would be easier to execute the task, so as to get to more general, environment-independent ideas on what is really needed to support people in video browsing.

When describing the results of the experiment, it will be hard to separate behavior and environment, so we will describe them together. We will observe how people interact with videos and use their evaluation of the tasks to answer the central question of this chapter: “How to design a video interaction environment that will optimally support its users?” As such, the research questions for this study are initially aimed at describing user behavior. Following the line of the research in this thesis, we formulate the research questions in terms of IFT and gap bridging:

RQ1 - Which patches do people choose to find relevant videos?

RQ2 - Which scent carriers do people prefer?

RQ3 - Which strategies do people apply to bridge the three gaps?

There will be overlap in answering these research questions as they typically are related: people will choose (RQ3) certain types of patches (RQ1) because they carry scent (RQ2). RQ2 describes the different ways scent is displayed (either in the links or in the data), and RQ3 describes the way people bridge gaps by following scent to patches (RQ1). Answering these questions will give us insight into the structure of interaction with video material.

When we have better understanding of how people interact with videos (by observing their behavior), *and* the participants in the experiment are sufficiently experienced (due to the task execution) to evaluate their behavior and needs, we can try to answer the following research question:

RQ4 - Which functionality do people prefer when interacting with video?

We expect that the answers to the research questions will relate to the type of task people have to perform and to characteristics of the environment, so these will get special attention in the design of the experiment and the analysis of the results.

5.2 Method

To answer our research questions, we conducted an experiment using the three video environments described in the previous chapter.

5.2.1 Participants

Students of psychology and communication studies at the University of Twente are required to participate in experiments as part of their bachelor's degree curriculum. They subscribe to experiments using an experiment management system, and get study points for completing an experiment. The experiment described here was entered into the system, and students could sign up for the available time slots. The study was listed as lasting about 2.5 hours and was described as "Browsing in videos: how do people search in videos, and what is the best way to support them?" A total of 17 students subscribed to the study, of whom the first was used as a pilot subject. (Before this first test, two other pilot subjects were used, but these were colleagues at the Telematica Instituut.)

Of the remaining 16 participants, 10 were students of psychology (62.5%) and 6 of communication studies (37.5%). Twelve were male (75%) and 4 female (25%); 10 were German (62.5%) and 6 Dutch (37.5%). Although the University of Twente is a Dutch university, it is located near the border and attracts many students from Germany. The bachelor's programs in psychology and communication studies are in Dutch. Study materials are in Dutch and English. German students must pass a Dutch exam before they can attend. The experiment was in Dutch and English.

Before the experiment we asked the participants to indicate how well they mastered the Dutch and English languages on a 7-point scale (1=very badly; 7=very well). For Dutch, 12 participants reported a 5 or higher, and four participants a 3 or 4. However, when at the end of the experiment we asked whether they had experienced any language problems during the experiment, all four in the lower group said they had not. For English, 15

participants reported a 5 or higher, while one scored a 2. But again, afterwards, this participant reported no language problems.

The average age of the participants was 22.2 years (two were 19, three were 20, four were 21, six were 22, and 1 was 41 years old). On average, they watched TV/video images 19.0 hours a week (ranging between 5 and 30 hours). This corresponds with the result in the Kenniswijk study described in Chapter 2, where the mean watching time of the 210 participants was 19.3 hours a week.

All participants indicated that they were very familiar with online video services such as YouTube and Google Video. On a scale of 1 (totally unfamiliar) to 7 (totally familiar), all scored 5 or higher (three scored 5, four scored 6, and nine scored 7). When asked how frequently they visited the YouTube website, the average answer was 4.8 days a week. The participant with the lowest frequency visited the website once every two weeks. Eight participants said they visited the website daily.

For the Fabchannel website the numbers were quite different. None of the participants had ever visited the website before the experiment.

5.2.2 Independent variables

The study had an experimental section using the VIBES video browser, and an exploratory section using the video environments of YouTube and Fabchannel. In the experimental part, the type of task was manipulated. In the scent experiment described in Chapter 3, we saw that the type of task had a significant effect on perceived scent. A main difference between that experiment and the one described here was that participants now had the ability to explore the videos, which meant that other manipulations of type of task were required. It was especially important to design tasks which would stimulate searching behavior: participants needed to be invited to interact with the video environments.

The first manipulation was task specificity. As was discussed in Chapter 3, a general task is open, leaving the participant free to explore. People probably will be guided and inspired by the alternatives that are offered and will adjust their current interest. A specific task is more restricted, and we may expect different searching behavior because people know exactly what they are looking for. We can expect people to prefer different kinds of support for these tasks. For the general task with the VIBES video browser, participants were free to search for videos they thought were nice (*mood improvement*) or interesting (*cognitive benefit*). The most important thing was for them to freely explore the database. So that the general task would not be completely without obligations, we added the restriction that participants had to find videos they would like to share with a friend.

Within the specific tasks, the dominant content from which relevant information could be found was manipulated. For video, relevant information can be present in the images, in the audio, or in both. In addition, the video representations (titles, descriptions, frames, [visual] summaries, etc.) can carry relevant information. For the VIBES video browser, we designed two specific tasks in which relevant information was predominantly available either in the images or in the audio. Metadata could indicate that a video was relevant for the task, but to be sure, the user needed to explore the video to get either visual or auditive confirmation. In the analysis, the exact role of metadata in this process is examined.

For the Fabchannel environment, we used a general task where participants could freely explore the music database for reasons of mood improvement. For finding a relevant video, auditive confirmation was required. To stimulate exploratory behavior and prevent participants from looking for familiar music, we added the restriction that they had to find a song they had never heard before. For the YouTube environment, a specific task requiring auditive confirmation was used. Exploratory behavior was stimulated by asking participants to find a number of relevant examples and then select the best.

The task types and the task definitions for each video environment are summarized in Table 5-1.

Table 5-1 Task types and tasks for each video environment

Task name	Task type	Task
VIBES-General	General - either mood improvement or cognitive benefit	"Find three videos that are very interesting or funny and that you would recommend to a friend."
VIBES-Specific(visual)	Specific - cognitive benefit + visual confirmation	"Find different videos in which people deliberately jump into the water. Indicate for each case the kind of water (e.g., sea, river, ditch)."
VIBES-Specific(auditive)	Specific - cognitive benefit + auditive confirmation	"Find different videos in which people are confronted with polluted water, whether for drinking, washing, or swimming. Indicate for each case the cause of the pollution."
Fabchannel	General - mood improvement + auditive confirmation	"Find a nice song you would like to send to a friend. It has to be a song you've never heard before."
YouTube	Specific - cognitive benefit + auditive confirmation	"Find remarks by Bill Clinton on his wife Hillary. Write them down and select your favorite."

5.2.3 Dependent variables

To answer the four research questions, we defined the following dependent variables:

- *browsing behavior* - a mostly qualitative analysis of which patches were chosen, which scent carriers were chosen, how the three gaps were bridged, and which functional parts in the interface were used. This applied to all four research questions. The variables were assessed by a human observer, backed up by video registration of the participant's activities.
- *browsing success* – measured by counting the number of relevant videos found (only for the specific task types – see Table 5-1). This quantitative variable was analyzed in relation to the previous variable – browsing behavior – to judge the success of the applied strategies.
- *subjective task difficulty* – measured by asking participants to assess on a 7-point scale how difficult the task was (1=very easy; 7=very difficult). This was also analyzed in relation to browsing behavior to judge the success of the applied strategies.
- *subjective usefulness of browser modules* – measured by letting the participants assess on a 7-point scale (1=totally useless; 7=very useful) how useful the browser modules were for the execution of each task (only for the tasks with the VIBES video browser – see Table 5-1). This especially refers to the fourth research question.
- *participants' observations and opinions* – measured by interviewing the subjects after each task and after the experiment as a whole. The results were related to the observed browsing behavior and interpreted in the context of the four research questions.

5.2.4 Experimental materials

For the tasks with the YouTube and Fabchannel websites, we used the original database and user interface as available at the time of the study. For the tasks with the VIBES video browser we had to make some specific decisions, which are described below.

The database of the VIBES video browser was filled with 80 videos downloaded from YouTube. These were the same videos that were used in the previous experiment: the scent experiment described in Chapter 3. All of them dealt with the theme “water,” either in an informative or an entertaining way. The duration of the videos ranged between 21 s and 9 min 43 s. Titles, descriptions, categories, and thumbnails were directly copied from YouTube, so no alterations were applied (typing mistakes and so on were not corrected). The only thing that was changed was the tags.

This was done to create a basis for effective and efficient patch-based browsing. In the browser, the videos are clustered via their tags into video

patches, and presented to the user in a patch menu. The goal of the browser is to encourage users to explore a patch, but also to switch between patches. Therefore we set the following criteria for the creation of patches: the total number of patches was surveyable (i.e., all categories in the list could be easily scanned), the patches were of a substantial size with overlap between the patches, and the patches were meaningful to the users.

Regarding the first two points, the original 80 videos had a total of 566 tags. Of these 425 tags were unique, that is, related to only one video. Of the remaining tags, only a few tags (including “water”) were related to more than two or three videos. There were many synonyms (e.g., “fun” and “humor”) and different versions of the same word (e.g., “fun” and “funny”). Moreover, a lot of tags were useless because, for example, the title of the video was repeated in the tags, leading to tags such as “the” and “in.” Using the original tags would lead to a very long and difficult to scan tag list, and worse, most tags would lead to a (very) small number of videos. Therefore it was decided to reduce the number of tags and create meaningful, substantial video patches that were useful for bridging all three gaps.

To start with, we looked at the outcome of the Kenniswijk study in Chapter 2. There we saw that people preferred to interact with the semantic segments as the program maker intended, or those “naturally” provided by the characteristics of the content. Typically, the videos in YouTube are autonomous videos (e.g., a shot with a webcam or a mobile phone, a video clip of a song) or semantically meaningful parts of larger videos (e.g., a news item or sports highlight). Therefore, for this specific collection of videos, we did not need to further segment the videos. So the tags were linked to the videos as a whole, not to segments.

Another outcome of the Kenniswijk study was the importance of “topics” as an aspect of a video people would like to be able to jump to instantly. So, in terms of support for navigational decisions, tags referring to the topic of the video are important. Therefore we created a topic list for this specific collection of videos. As a first step, we used the original categories the videos were placed in on YouTube. Next, we looked at the original tags and tried to identify high-level concepts with multiple occurrences in the database. For example, tags such as “school,” “university,” and “lesson” are related to the high-level concept “education.” Applying this bottom-up approach, we extended the number of topics to end up with the following list: *climate + weather, comedy, education, entertainment, environment, films + animation, fun, health + hygiene, how_to + do_it_yourself, music, natural environment, news + politics, pollution, recreation, religion, science + technology, sports, transportation, travel + places, video_weblogs, war + conflicts.*

Next on the Kenniswijk list were “events.” This proved to be a difficult category, as the database contained no clearly recognizable events (such as elections or sporting events) that were represented by more than one video. “Time periods” also proved difficult, because most videos were of recent date, and with few exceptions there were no references to time period available. So we skipped these two patch types.

We did include “people” and “animals.” Based on the tags and the video content, we made a distinction between *adults* and *children*, and *pets* and *wild animals*. Next on the Kenniswijk outcome was “locations.” Based on the tags (and sometimes the descriptions) we created a list including both geographical locations (*Africa, Antarctica, Asia, Europe, North America, and South America*) and general ones (*indoors, industrial, gardens + parks, nature, rural, urban, underground, and water*).

The category “actions” was too diverse, with hardly any overlap. The same was true for “objects.” Typically, “actions” and “objects” are often related to a topic, animal/person, or location (e.g., a player taking a penalty with a ball on a soccer field), so we figured that within the other patch categories, actions and objects would still be findable. The other categories that received low scores in the Kenniswijk study were ignored here.

So we combined a bottom-up approach (using the original metadata) and a top-down approach (applying the results from the Kenniswijk study) to create the list with patches/categories. As a final step, we watched the videos and evaluated, on the basis of each one’s content, whether it could be placed in any of the created patches. (This was especially necessary because the original metadata were not complete or accurate.) This way, still more overlap between the videos was created. So the only adjustments made to the original YouTube data and metadata were the creation of a new tag list, and an evaluation of the video content to relate the videos to that new tag list.

Finally, we created a patch category called “what other people said,” containing the following patches: *Top Rated, Most Popular, Most Amusing, and Most Interesting*. The first two were based on data from YouTube, the last two on the data from the scent experiment described in Chapter 3. Each category contained the 10 (or more when scores were equal) videos with the highest scores.

In sum, for this specific set of videos, and with the criteria for the creation of patches as stated above, we came up with this list of categories. We applied a combination of a bottom-up approach (using the information available in the tags and other metadata, and in the video’s content) and a top-down approach (using the information gathered from the studies described in Chapter 2). Categories that were hard to fill (time periods), too detailed (actions, objects), or both (events) were excluded in this specific case.

5.2.5 Setting

The experiment took place under laboratory conditions in a large furnished office room in a quiet corner of the Telematica Instituut which was normally used for demo presentations. The participants faced an empty white wall, about 5 meters away from windows on their left side. They were seated in a comfortable desk chair in front of a large table containing a keyboard, a mouse, and a 19-inch screen (all connected to a state-of-the-art computer located under the table). A pen and forms containing instructions and questions were present on the table.

To the left of the participant, the experimenter was seated at a distance which was far away enough not to be too intrusive, but close enough to be able to observe the participant's searching behavior on the screen. The experimenter had his own desk containing videotapes, forms, a chronometer, and illustrative materials for the debriefing. The experimenter used a pen and a clipboard to write his observations on a prepared form.

Between the experimenter and the participant, and slightly farther back, was a video camera on a tripod. The experimenter was able to operate the camera without leaving his chair. Recordings of the screen were made over the left shoulder of the participant.

Figure 5-1 gives an overview of the experimental setting.

Figure 5-1 The setting of the final experiment including experimenter (left) and participant



5.2.6 Procedure

Participants were required to show up at the time slot they had chosen. After they were welcomed by the experimenter and given coffee or a soft drink the goal of the experiment was explained in general terms (“we are interested in how people search in video material”). They were told they had to complete five assignments using different video browsers, and that the experiment would last about 2.5 hours in total. *How* they executed the tasks would be more important than their success in completing the assignment. Their searching behavior would be observed by the experimenter and taped with the video recorder (but without recording the participant’s face). A booklet with instructions and questions would be used to lead the participant through the experiment. Participants only had to wait for the experimenter for instructions when to turn the page. It was indicated that the instructions and questions would be in Dutch, while the browsing environments, including video data and metadata, would be in English. After these explanations, they had to fill out a form indicating that they agreed with the terms of the experiment.

Next they had to fill out a form asking for their gender, age, course of study, and mother tongue. On a 7-point scale they had to indicate how well they spoke Dutch and English (1=very badly; 7=very well). After that, they began the first of the five tasks. For all participants, the first three tasks were with the VIBES video browser. All possible orders of these three tasks were used to control for effects of transfer, learning, and/or fatigue (see Appendix F). After a coffee/tea break, participants performed the tasks with the YouTube and Fabchannel websites. The order of these tasks was also balanced. The reason the tasks with the VIBES video browser always came before the break was that this session lasted somewhat longer than the other session. We preferred a shorter post-break session because there was a little fear that the participants’ energy and motivation would decrease at the end of the study. We also preferred not to break up the three tasks with the VIBES video browser.

Before participants started the tasks with the VIBES video browser, they first went through a training session. This took the form of a walkthrough: step by step, the experimenter acquainted the participant with the browser by presenting small tasks (e.g., “select a menu item,” “try to view that video”) and asking questions (e.g., “what do you think that means?”). This way, all interface elements of the browser were brought to the attention of the participant. During the explanation, participants were allowed to try out the interface elements. The training session lasted until the participant indicated that he/she understood the complete interface.

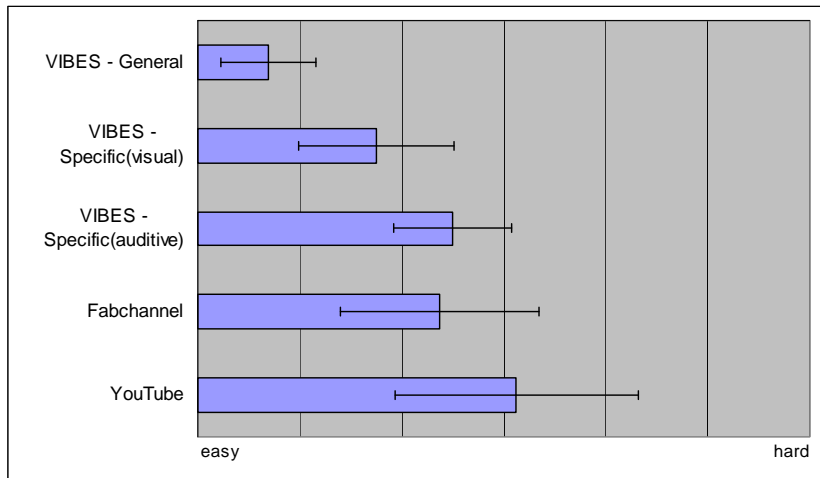
For each task, participants had to read the instructions in their booklet. Each task lasted 10 minutes. Table 5-1 presents the wording of the tasks

(originally in Dutch) and indicates what the participant had to write down in the booklet. During the task, the experimenter observed the participant's browsing choices and wrote them down – using index numbers to indicate the order of actions - on a screen dump of the browser on which all menu items were folded out (see Appendix G). After each task, the participant had to indicate on a 7-point scale how difficult the task was (1=very easy; 7=very difficult). For the tasks with the VIBES video browser, he/she had to indicate on a 7-point scale the usefulness of each of the eight browsing modules for that specific task (1=totally useless; 7=very useful). For all five tasks, the experimenter asked how the participant had approached the task (the strategy used), and which parts of the execution were easy and which were hard. Moreover, the participant was asked to suggest ideas about what kind of support could ease the execution of the task.

The first session with three tasks on the VIBES video browser ended with a general evaluation of the browser. The second session with the Fabchannel and YouTube websites ended with a general, open evaluation of video browsing, discussing what makes video browsing difficult and what kind of support is needed for efficient interaction.

The participants were then debriefed by the experimenter, who discussed the exact purpose of the experiment, explaining the approach from the IFT point of view. Finally they had to fill out a last page with questions on how many hours a week they watched TV/video, whether they had experienced language difficulties, and if they had any general remarks on the experiment. The participant was then shown to the door, after which the experimenter assigned credits to the student in the experiment management system.

Figure 5-2 Experienced difficulty of tasks indicated on a 7-point scale (average score and 95% confidence intervals)



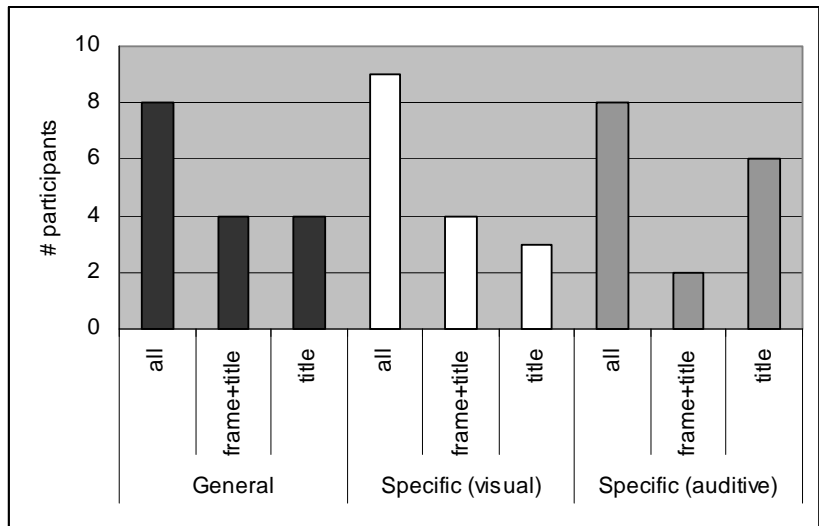
5.3 Results

Figure 5-2 shows the experienced difficulty of the five tasks. On average, the tasks were not too difficult. All tasks were on the easy side, except for the YouTube task, which was between easy and hard. The general task with the VIBES video browser was clearly experienced as the easiest of all.

5.3.1 VIBES video browser tasks

For comparison, we first show some results for all three tasks. Figure 5-3 shows the scent carriers that were chosen for the three different tasks.

Figure 5-3 The scent carriers that were chosen (in the Results module) for the three tasks with the VIBES video browser.

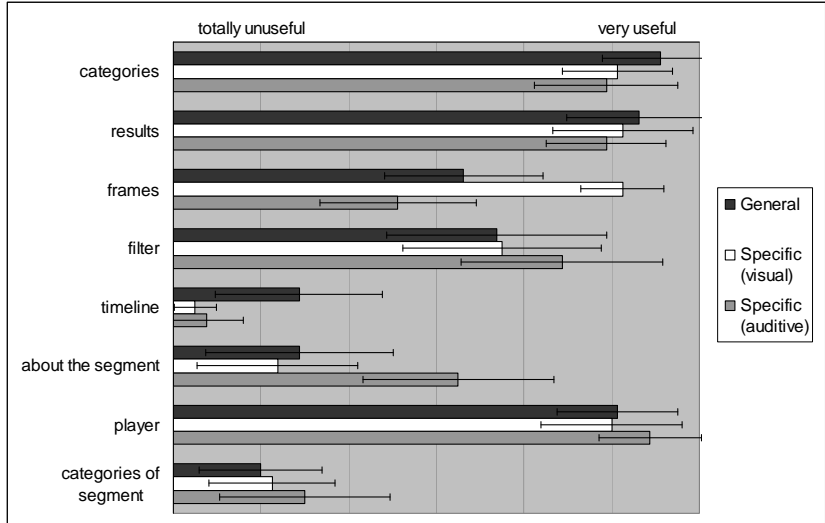


The “all” scent carrier was clearly the favorite for all tasks: for each task, at least half of the participants switched to this scent carrier. Only for the specific task with auditive confirmation was the “title” scent carrier (which was the default scent carrier) used by almost as many participants.

Figure 5-4 shows the usefulness of the different browser components for the three different tasks. Looking at differences between tasks, there seemed to be three striking results. First, the “Frames” module was more useful for the specific task with visual confirmation than for the other two tasks. Second, the “Timeline” module was more useful for the general task than for the other two tasks. Third, the “About the segment” module was more useful for the specific task with auditive confirmation.

The individual results for the three tasks will be discussed in separate sections below.

Figure 5-4 Usefulness of browser modules as indicated on a 7-point scale for the three tasks with the VIBES video browser (means and 95% confidence intervals)



Task: General

The instruction for the general task was: “Find three videos that are very interesting or funny and that you would recommend to a friend.” Participants had to indicate on a 5-point scale whether the videos were funny (1) or interesting (5), and which of the three videos they would finally choose. Eight participants chose a funny video (score 1 or 2), five an interesting video (score 4 or 5), and three a video that was both funny and interesting (score 3).

Table 5-2 shows for each participant the patch selection with which they began their search, whether or not they switched to another patch, which scent carrier was chosen in the Results module, and how difficult the participant found the task (indicated on a 7-point scale: 1=easy; 7=hard).

RQ1 - Which patches do people choose to find relevant videos?

As can be seen in Table 5-2, eight participants (50%) chose a social category (from the group “what other viewers say”) to look for a funny or interesting video. From the eight participants who did not, four included the category “comedy” or “fun” or both. Seven participants chose to start with multiple categories (OR functionality), and three applied a filter (AND functionality). Six participants started with one category.

In the evaluation, a number of people indicated that the task was made more difficult because of faulty metadata: the videos in the category “Most Interesting” were not very interesting, according to some participants’ taste.

RQ2 - Which scent carriers do people prefer?

Regarding scent in the links, Table 5-2 shows that eight participants chose the “all” scent carrier, four “frame+title,” and four “title.” Figure 5-4 shows that the “Timeline” module is more useful for the general task than for the other two tasks (although the absolute usefulness is still low). Some participants indicated that they preferred a short video: videos with short bars in the timeline view carried more scent than videos with long bars.

Regarding scent in the data, participants thought the Player module was more useful than the Frames module (see Figure 5-4). The sound and movement present in the player (and not in the frames) probably added to the perception of scent in this task.

Table 5-2 - Results for the general task using the VIBES video browser (for further explanation, see text)

P.	First patch chosen (described using logical operators)	Scent carrier	Indicated task difficulty
1	<South America> OR <North America>	all	1
2	<travel+places>	frame+title	1
3	<comedy>	title	1
4	<what other viewers said>	frame+title	3
5	<Most Amusing> OR <Most Interesting>	title	1
6	<All patch categories> AND (<Most Interesting> OR <Most Amusing>)	all	2
7	<comedy> OR <fun>	all	1
8	<music>	all	2
9	<what other viewers said> AND <fun>	title	1
10	<comedy>	all	1
11	<music> OR <comedy> OR <fun>	frame+title	2
12	<religion> OR <Most Interesting>	all	3
13	<education>	title	1
14	<Top Rated> AND (<Most Interesting> OR <Most Amusing>)	frame+title	3
15	<Top rated> OR <Most Popular> OR <Most Amusing>	all	1
16	<Top Rated> OR <Most Amusing> OR <children>	all	3

RQ3 - Which strategies do people apply to bridge the three gaps?

The reasons for interacting with videos are provided by the experimental task (“Find three videos that are very interesting or funny and that you would recommend to a friend”).

The VIBES video browser provided no other option for bridging the first gap (between reasons for a decision and the decision) than to choose a

patch from the Categories module. As we saw, most often the decision was to select a social category.

The second gap (between the decision and the initiation of the action) was easy to bridge: when a category was chosen by ticking the box in front of the category, this immediately displayed a list with results so that the action could be evaluated.

The third gap (between the initiation of the action and the continuation and completion of the action) was bridged by evaluating the results, choosing links with high scent, and exploring the selected videos. As can be seen in Figure 5-4, the Player module was considered more useful than the Frames module: people preferred to see and hear the video. When the search results did not provide enough interesting or fun videos, participants returned to the Categories (and Filter) modules to create new patches and repeat the process, thus starting again at the first gap. Five participants never switched to another patch, while four participants switched once or twice. Seven participants switched several times.

RQ4 - Which functionality do people prefer when interacting with video?

Figure 5-4 shows that for this task the most useful elements of the interface were the Categories, Results, and Player modules. Reasonably useful were the Frames and Filter modules, and not very useful were the About the Segment, Timeline, and Categories of Segment modules.

In the evaluation, we asked which improvements to the interface could ease this specific task. By far most of the suggestions were related to social data. Six participants indicated that they liked to see ratings from other people. (These were not displayed in this experiment – only the category Top Rated was created.) Links displaying high ratings can obviously carry a substantial amount of scent to some people (see also RQ2). There was also a demand for recommendations and other users' comments. Some wanted more insight into how a category such as "Most Interesting" was created.

Task: Specific (visual confirmation)

The instruction for the specific task with visual confirmation was: "Find different videos in which people deliberately jump into the water. Indicate in each case the kind of water (e.g., sea, river, ditch)."

The database contained eight more or less relevant videos of people deliberately jumping into water (10% of the database). In five cases, it was debatable whether or not it was a jump (the people slid into the water, for example, or were planning to jump but tripped instead), but the people clearly went into the water on purpose. The average duration of these videos was 2 min 45 s. The average time after the start that the relevant event appeared was 36 s (at 22% of the average duration).

On average, participants found 4.0 relevant videos in the 10 minutes they got for this task. (Three participants found 2 videos, three found 3, three found 4, five found 5, and two found 6 videos).

Table 5-3 shows for each participant the main patch selection, which scent carrier was chosen in the Results module, the performance of the participant (the number of relevant videos found), and how difficult the participant found the task (indicated on a 7-point scale – 1=easy; 7=hard).

Table 5-3 - Results for the specific task with visual confirmation using the VIBES video browser (for further explanation, see text)

P.	Main patch (described using logical operators)	Scent carrier	# videos found	Indicated task difficulty
1	(<nature> OR <gardens+parks>) AND <water>	all	6	2
2	<fun> OR <natural environment> OR <sports>	frame+title	3	2
3	(<adults> OR <children>) AND <water>	title	2	2
4	(<adults> OR <children> OR <"several other categories">) AND (<water> OR <fun>)	frame+title	5	2
5	<adults> OR <children> OR <fun> OR <health+hygiene> OR <water>	title	3	2
6	<water> AND <fun>	all	3	5
7	<comedy> AND <water>	all	2	4
8	<water> OR <fun> OR <recreation> OR <sports>	all	5	2
9	<adults> OR <children> OR <water>	all	5	2
10	<water> AND <Most Amusing>	all	4	2
11	<recreation> OR <sports>	frame+title	4	3
12	<water> AND <fun>	all	6	2
13	<fun>	title	5	1
14	(<adults> OR <children>) AND <water>	frame+title	2	5
15	(<adults> OR <children>) AND (<water> OR <fun>)	all	5	6
16	<water> OR <adults>	all	4	2
		<i>average:</i>	<i>4.0</i>	<i>2.8</i>

RQ1 - Which patches do people choose to find relevant videos?

Unlike in the general task, people stayed much longer within one patch, due to the fact that patch exploration took a lot of time. Table5-3 therefore displays the main patch, that is, the patch participants stayed the longest in. As can be seen, only one participant used one category. Six participants used multiple categories (OR functionality), four applied a filter (AND

functionality) to one category, and five applied a filter to multiple categories.

All the categories that were chosen had some kind of relation or association with the task description (“Find different videos in which people deliberately jump into the water”). Thirteen participants included the category “water.” Ten participants included a category related to funny situations: “fun,” “comedy,” or “Most Amusing.” Seven participants included people categories (“adults,” “children,” or both). Three participants chose an activity: “sports,” “recreation,” or both.

In the evaluation, participants indicated that this task would be easier if there was a category “jumping,” or if there was search functionality available so one could include “jump” or “jumping” in the query terms. In terms of IFT, these would be the category and terms with the highest scent.

RQ2 - Which scent carriers do people prefer?

Regarding scent in the links, nine participants chose the “all” scent carrier, four “frame+title,” and three “title” (see Table 5-3). The participants who chose the “all” scent carrier on average found 4.4 videos; “frame+title” choosers found 3.5 videos; and “title” choosers found 3.3 videos. These differences are not significant. However, as the latter groups are very small, these numbers are rather unreliable.

Regarding scent in the data, the “Frames” module was more useful for the specific task with visual confirmation than for the other two tasks (see Figure 5-4). For this task, the participants judged this module as important as the player. This was confirmed in the evaluation, in which participants indicated that they appreciated the Frames module (one participant said it was hardly necessary to play the video).

What made this task difficult was the fact that with one exception (“Cal jumps into shitty water”) the title was of little use. For deeper understanding of how people make choices based on the perception of scent, we will change the point of view for a while, take a look at the targets in this task, and see in what way their scent is present in the user interface.

The database contained eight relevant videos for this task. On average, each relevant video was found by 8.1 participants (51%). Five videos were found by at least half of the participants. One video was found by all but one participant: the above-mentioned “Cal jumps into shitty water.” The title clearly says it all, and the description also leaves no room for doubt about the relevance of this video: *“my mate, jumps into disgusting canal from about 25. foot.... hahaha. he realy scratched his chest dragging himself out and stung all over his face on the stingers.. oh well was worth seein him do it.[...]”* The video also has all the relevant tags: “people,” “water,” “recreation,” and “fun.” So the metadata of this video carry a lot of scent for this specific task. The

whole video is about this single event of a boy jumping from a high bridge into a canal; from the first second it is clear this is going to happen, even though the actual event only occurs after 35 seconds. In other words, the video data itself and the visual summary clearly also carry scent.

The video in second place - "Boys Swimming And Fun In Dirty Lake" – was found by 13 of the 16 participants. The title did not directly refer to someone jumping into the water, but this activity clearly can be associated with "swimming" and "fun" in a lake. The description did not add anything to the title regarding relevance for this task: "*In Weekend, Boys Fun In defile lake, water super dirty, Do you believe? [...]*" It did have all the relevant tags: "people," "water," "recreation," and "fun." So, regarding metadata, it did carry scent, but required a bit more associating than the most successful video. Regarding the video data, from the first second the video shows a compilation of boys jumping in different variations into a pond, which can also clearly be seen in the frames. As such, the video data seem to contain more scent than the previous video: the relevant action is visible from the start (not after 35 seconds), is constantly repeated (not happening once), and the people are the focus of the video (not sharing the scene with a bridge and a canal).

An example of a relevant video that was less successfully found (only once) was "Liquid Nitrogen Into A Swimming Pool." This video, which lasts 1 min 32 s, contains a short flash after 52 seconds of a person jumping into a pool. (After that event the person is mostly obscured by steam.) Although the title refers to a swimming pool, there is no mention of a person but of "liquid nitrogen" going into the pool. The brief description does not provide any additional information. It does contain the tags "people," "water," and "recreation," but the metadata clearly carry less scent than in the previous examples. The tenor of the video is to demonstrate what happens when you pour liquid nitrogen into a pool, and it shows a person with a vat containing the nitrogen walking towards a pool surrounded by people and pouring it in, after which the effect is displayed (a "steaming" pool). It can be expected that this apparent purpose of the video was distracting and generated associations that did not have any match with the task. In other words, there seems to be a lot of "negative scent." Hence, for most people, the amount of perceived scent will be limited. Only the presence of a lot of people and the swimming pool may still generate some scent. The one participant who did find the video was a successful browser (finding 6 of the 8 relevant videos), using the visual summary (the frames) to explore the video data.

RQ3 - Which strategies do people apply to bridge the three gaps?

The reasons for interacting with videos are provided by the experimental task (“Find different videos in which people deliberately jump into the water”).

The VIBES video browser provided no other option for bridging the first gap (between reasons for a decision and the decision) than to choose a patch from the Categories module. As we saw, the decision was to select categories that could be associated with the task description, such as “water,” “fun,” or “people.” Almost all patches were created by combining categories and/or applying filters.

The second gap (between the decision and the initiation of the action) was easy to bridge: when a category (or combination of categories) was chosen by ticking the box in front of the category, this immediately displayed a list with results so that the action could be evaluated.

The third gap (between the initiation of the action and the continuation and completion of the action) was bridged by evaluating the results, choosing links with high scent, and exploring the selected videos. Exploration was often done by checking the frames, then double-clicking a frame with high scent to jump to the related part of the video. Whenever a relevant video was found, the search continued by scanning the links in the result set. As the exploration of the results took some time, people switched much less between (category) patches than in the General task.

RQ4 - Which functionality do people prefer when interacting with video?

Figure 5-4 shows that for this task the most useful elements of the interface were the Categories, Results, Frames, and Player modules. Reasonably useful was the Filter module, and not very useful were the About the Segment, Timeline, and Categories of Segment modules.

In the evaluation, five participants indicated that they would have appreciated search functionality, allowing them to search for “jump.” A related suggestion was a separate category about “jumping.” In both cases a description is needed of the actions in a video. In the videos used in this experiment, only one out of eight relevant videos contained a reference to “jumping” in the title and description.

Three participants indicated that they would have liked to see some elements removed from the interface for this task, especially the Timeline and About the Segment modules. Figure 5-4 confirms that these modules were not considered useful for this task.

Task: Specific (auditive confirmation)

The instruction for the specific task with auditive confirmation was: “Find different videos in which people are confronted with polluted (drinking,

washing, swimming) water. Indicate for each case what the cause of the pollution is.”

The database contained 16 videos (20% of the database) that were more or less relevant to the task. The average duration of these videos was 3 min 46 s. The average time after the start that the cause of the pollution was mentioned was 52 s (at 25% of the average duration). If the participants found one of these videos, they had to write down the cause of the pollution or what the pollution consisted of. If they did not, the video was not counted. If they provided a cause that was trivial or incomplete, they got a half point deduction.

On average, participants found 2.9 relevant videos (ranging between 1 and 5) in the 10 minutes they got for this task.

Table 5-4 shows for each participant the main patch selection, which scent carrier was chosen in the Results module, the performance of the participant (the number of relevant videos found), and how difficult the participant found the task (indicated on a 7-point scale: 1=easy; 7=hard).

Table 5-4 - Results for the specific task with auditive confirmation using the VIBES video browser (for further explanation, see text)

P.	Main patch (described using logical operators)	Scent carrier	# videos found	Indicated task difficulty
1	<health+hygiene> OR <pollution>	all	3	3
2	<education> OR <health+hygiene> OR <pollution> OR <war+conflicts> OR <science+technology> OR <environment>	all	3,5	3
3	<water> AND <health+hygiene>	title	3	2
4	<climate+weather> AND (<environment>OR <water>)	title	3	4
5	(<adults> OR <children> OR <water>) AND <health+hygiene>	title	1	5
6	<pollution> AND <water>	title	3	4
7	<climate+weather> AND (<water> OR <health+hygiene>)	all	3	3
8	(<"several categories">) AND <pollution>	frame+title	2	4
9	<pollution> OR <water>	all	3,5	4
10	(<pollution> OR <natural environment>) AND <water>	all	2	3
11	<pollution>	all	2	1
12	(<pollution> OR <health+hygiene>) AND <water>	all	3	4
13	<environment>	title	5	3
14	(<climate+weather> OR <environment>) AND <health+hygiene>	all	2	3
15	<pollution> AND (<water> OR <health+hygiene>)	frame+title	5	5
16	<environment> OR <water> OR <nature>	title	2	5
		average:	2.9	3.5

RQ1 - Which patches do people choose to find relevant videos?

Table 5-5 displays the main patch: the one participants explored for the longest time. As can be seen, only two participants used one category. Four participants chose multiple categories (OR functionality), five applied a filter (AND functionality) to one category, and five applied a filter to multiple categories. With ten participants applying a filter, this functionality was clearly important for this task, which seems to be confirmed in Figure 5-4.

All the categories that were chosen had some kind of relation or association with the task description (“Find different videos in which people are confronted with polluted water”). Ten participants included the category “water”; nine included “pollution”; eight included “health + hygiene”; and six included “environment” (although one participant mistakenly chose “natural environment,” which has a different meaning). Typically, even though the word “people” is clearly mentioned in the task description, only one participant chose a related category (“adults” and/or “children”). One reason may be that the other categories, such as “pollution” and “health + hygiene,” imply a relation to people.

RQ2 - Which scent carriers do people prefer?

Regarding scent in the links, eight participants chose the “all” scent carrier, two “frame + title,” and six “title” (see Table 5-5). As figure 5-3 reveals, this was the task in which the most participants chose the “title” scent carrier (with four in the General task and three in the specific task with visual confirmation). The participants who chose the “all” scent carrier on average found 2.8 videos; “frame + title” choosers found 3.5 videos; and “title” choosers found 2.8 videos. These differences are not significant. However, as the groups are very small, these numbers are unreliable.

Regarding scent in the data, the “About the Segment” module was more useful for the specific task with auditive confirmation than for the other two tasks (see Figure 5-4). The player was also rated very useful, which was predictable as participants often had to listen to the audio channel of the video. For the same reason, the Frames module was not considered very useful for this task.

What really made the task difficult was that participants had to locate the point in the video where the cause of the pollution was revealed. In the evaluation, this was noted by seven participants, and three others complained that having to listen to the audio was a hard task. A lot of “impatient” behavior was observed by the experimenter during this search. Many participants started to browse a video that seemed relevant, but gave

up when they could not quickly enough find the cause of the pollution, and switched to another video. Exploring the audio channel of a video is a very time-consuming task requiring a lot of attention, and is especially irritating when there is time pressure. Based on the scent in the metadata and in the first data explored in the video, the participant had to decide whether the scent was strong enough to allow some time for extra - time-consuming - audio exploration. Again, we will change the point of view towards the targets in the database to get an idea how scent is present in the user interface.

Each of the 16 relevant videos - on average - was found by 3.0 participants. Seven videos were found by at least a quarter of the participants. Only one video was found by more than half of the participants. The title of this video was “New River - Most Polluted River in North America,” and it contained very clear references to polluted water in both title and description. It had the relevant tags, “environment,” “pollution,” “health & hygiene,” “water,” and “people.” The frame shows a vague image of a waterside and some (unreadable) text. In the scent experiment described in Chapter 3, we asked participants in this same database to “find a video about water pollution” using only links to the videos. For that question, this video was most often chosen, indicating that the information carried in the links contained a lot of scent for this task. The video is a short documentary of 1 min 42 s showing images of the river and a voice-over telling about the pollution. Right from the start, the pollution and the problems it causes are the clear object of the video. After 30 seconds, the voice-over gives the reasons for the pollution. Interestingly, the video shows almost no people, only a short, dark shot of barely distinguishable illegal immigrants swimming in the water. The task clearly asks for videos in which “people are confronted with polluted water.” The images hardly show people and pollution, but the combination of images of a river with the metadata and the voice-over saying that this polluted river is a danger to the health of the people living in the area makes this video carry scent.

Another highly relevant video was “Nigerian Families Desperate for Clean Water.” This clip tells the sad story of people in Niger who have to get their water from polluted wells. This video was found by only two participants. The title suggests relevance for the task and does carry scent (although it does not directly mention “pollution”). The description, however, starts as follows: *Like most women and many children in Niger, daily routine starts at the traditional well. For one woman, Kilma, it is a 30 minute round trip, a trip she will make 15 times today. Finding something to eat is also a challenge [...] Here the emphasis lies on a woman having to make numerous trips to the well, not on the fact that the well is polluted. Hence, the description carries very little scent, and may draw people away from the video. The key*

frame shows the face of the main character in the video being interviewed. In the scent experiment described in Chapter 3, the “frame” scent carrier of this video was never chosen for the task “find a video about water pollution,” indicating the lack of scent in the frame. The video had three relevant tags: “people,” “pollution,” and “health & hygiene.” Especially the latter two made this video appear in most relevant result sets, but it was seldom further explored. (It does contain relevant images of the well and the water in it, and of people getting water for consumption.) A suggestion might be that the scent in the title and tags was overruled by “negative scent” in the description and frame. One participant found the video at the end of the 10 minutes because she chose the tag “Africa,” having an association between water problems and that continent. There was only one video related to that tag, and she decided to explore it.

A last example is a video called “Surfers Against Sewage - Doctor Loo.” The description of that video was as follows: *Currently, the sewage from the Island of Guernsey gets pumped straight into the sea. Campaigners Surfers Against Sewage set out to change this, and sent their favourite time travelling Doctor to try and put a stop to it. . .* Clearly, this was a very relevant video. Still, it was found by none of the participants. This probably was because it did have the partly relevant tags “people” and “water,” but lacked the tags “environment,” “health & hygiene,” and “pollution” (the database including metadata was never meant to be perfect). Because of this, the video was not present in most result sets that were generated for this task. Some participants did, nevertheless, explore the video, but were probably put off by the “lightness” of the presentation. (The video started with a parody of the TV series “Doctor Who.”)

RQ3 - Which strategies do people apply to bridge the three gaps?

The reasons for interacting with videos are provided by the experimental task (“Find different videos in which people are confronted with polluted water”).

The VIBES video browser provides no other option for bridging the first gap (between reasons for a decision and the decision) than to choose a patch from the Categories module. As we saw, the decision was to select categories that could be associated with the task description: “water,” “pollution,” “health+hygiene,” and “environment.” Almost all patches were created by combining categories and/or applying filters.

The second gap (between the decision and the initiation of the action) was easy to bridge: when a category (or combination of categories) was chosen by ticking the box in front of the category, this directly displayed a list with results so that the action could be evaluated.

The third gap (between the initiation of the action and the continuation and completion of the action) was bridged by evaluating the results, choosing links with high scent, and exploring the selected videos. Exploration was done by checking the title and especially the description, either in the Results module or in the About the Segment module. When a video with high scent in the metadata was found, the video data were explored by playing the video and listening to the audio, either with or without jumping on the timeline. Whenever a relevant video was found, the search continued by scanning the links in the result set. As the exploration of the results took some time, people switched much less between (category) patches than in the General task.

RQ4 - Which functionality do people prefer when interacting with video?

Figure 5-4 shows that for this task the most useful elements of the interface were the Player, Categories, Results, and Filter modules (during the evaluation of the task, three participants explicitly mentioned that they liked using the Filter module). Reasonably useful were the About the Segment and Frames modules, and not very useful were the Categories of Segment and Timeline modules.

When asked which improvements to the interface could ease this specific task, eleven participants mentioned the need for better and more extensive descriptions. There seemed to be a preference for finding the answer in the textual description of the video rather than in the audio. Seven participants came up with solutions where metadata and segments were linked in more detail. For example, one suggested that each frame in the Frames module should get a title, description, and/or tags. Another felt that the description should contain time indications. Clearly, for this task extensive support for browsing within the video was desired.

Overall browser evaluation

As we saw in Figure 5-4, the usefulness of the different modules was rather different. Three modules were considered very useful for all tasks: Categories, Results, and the Player. A little less useful, but still scoring above 4 on a 7-point scale, were Frames and Filter. The Frames module was especially useful for the specific task with visual confirmation. It did get a lot of positive response, as can be seen in Figure 5-4.

The least useful modules were About the Segment, Categories of Segment, and Timeline. The About the Segment module was only considered moderately useful for the specific task with audio confirmation. The information in this module largely overlaps with the information in the “all” scent carrier (a lot of extra information can be added to this module, such as a transcript, but this information was not available to participants

during the experiment). So it mainly has a function when the “all” scent carrier is not chosen. And indeed, participants who had chosen the “all” scent carrier judged the usefulness of the About the Segment module as 3.8, while participants who chose the “title+frame” or the “title” scent carrier, judged its usefulness as 4.8 (on a scale of 1 to 7, with 7 meaning “very useful”).

The Categories of Segment module probably did not have enough added value. It was not used much, most likely because the categories were also available and visible in the Categories module. Perhaps the usefulness would increase if the number of categories were much larger: now it was not too difficult to pick the same category from the Category module. Moreover, the only function available was to double-click on the categories in this module, meaning that the current selection would be overruled. This was also not often done in the Categories module; the participants preferred to use the tick boxes. Some suggested using this module as a filter.

The timeline was considered the least useful of all modules. Only in the general task did it seem to have a function, namely to select relatively short videos. However, the timeline was originally developed for other purposes. For lengthy, segmented videos, it can display information on the frequency, distribution, and occurrence of certain subjects, events, persons, and so on in a video (see Van Houten et al., 2004). In this experiment, only short, non-segmented videos were used, and so the functionality of the timeline was only partly exploited.

After the three tasks with the VIBES video browser, we asked participants in general to name three negative and three positive aspects of the browser. Table 5-5 shows the results.

Table 5-5 Negative and positive aspects of the VIBES video browser as indicated by the participants (each could name three negative and three positive aspects). # indicates the number of different participants mentioning the aspect (maximum = 16). Only aspects mentioned by three or more participants are displayed.

Negative aspects	#	Positive aspects	#
<i>Lacking functionality</i>	9	<i>Frames module</i>	11
- keyword query	5		
- continue search while video plays	4		
<i>General interface design</i>	9	<i>Interaction with categories</i>	10
- too cluttered, too much information	6	- search using categories	6
- ugly, design not interesting	4	- filter functionality	4
- player too small	3		
<i>Redundant functionality</i>	7	<i>General interface design</i>	8
- timeline module	4	- well-organized, logical	7
<i>Slowness</i>	7	<i>Different display options in Results module</i>	3
<i>Method of working of filters</i>	3		

Starting with the negative aspects, the main complaints were about the lack of functionality related to search. Five participants missed an option to query the database, which is a useful tool for bridging the second gap. Four

participants would have liked to be able to search further in the database while a video is playing (multi-tasking), which is an aid to efficiently bridging the third gap. The only reason this functionality was absent was a lack of development time.

Another group of negative aspects had to do with the visual design, which was judged too cluttered (by six), too ugly/uninteresting (by four), and to have a too-small player (according to three participants). Although considered important, creating a “nice” user interface was not one of the major objectives in developing the experimental browser.

Some functionality was considered redundant, especially the timeline module (mentioned by four). As explained above, the timeline was especially developed for interaction with lengthy, segmented videos, and needs to be tested using those types of videos.

The slowness of the application was another negative aspect indicated by seven participants—justifiably so. Again, this was directly related to the amount of development time.

Finally, three participants complained about the working of the filters, which is an issue related to the IFT concept of “enrichment.” Even though the Filter module did get positive remarks, the idea of filtering was quite confusing to some. If one applies a filter to a collection in the form of a category, are all videos within that category included or excluded? This is a confusing issue in general. One group of solutions lies in providing maximum feedback on the status of the selection that is made.

All these negative issues will be on the requirements agenda when a follow-up version of the VIBES video browser is developed.

The positive aspect that was mentioned most often (by eleven participants) was the Frames module. Figure 5-4 shows that it was considered especially useful in the specific task with visual confirmation, but it clearly made a positive impression in general. The frames (nine in the default setting, but increasable to 60 if this is found convenient) provide a visual summary that can be quickly scanned for signs of scent in the video data. It delivers a fast impression of the video contents (providing a way of getting around the time-based nature of a video), and if any scent is perceived, double-clicking a frame provides direct access to the related data in the player. For the specific task with auditive confirmation, participants spontaneously suggested adding textual descriptions to the frames for easier task execution.

Next, the interaction with categories was appreciated by ten participants, with a special mention for the Filter module. The categories are prefabricated patches, which help to bridge gap 2 by providing recognizable cues on what to select. It should be pointed out that the database was small (80 videos) and so was the number of categories (43 categories in four main categories). Even with this small number of

categories, participants incidentally indicated during the evaluation that they had not seen certain categories which might have been useful for the particular task they had just executed. There clearly is a limit to the size of the list, and additional means – such as searching within the categories – are needed when the list gets longer.

Eight participants indicated that they liked the interface design, especially the logical organization of the modules. This is in contrast with the often-mentioned negative aspect that the display was too cluttered. As such, it may be a matter of taste and personal preference. A solution may be a flexible interface where users can display or hide certain modules, and determine the size of separate modules. Actually, this is already possible with the VIBES video browser, but for reasons of control participants were instructed not to change the user interface organization. It would be interesting to conduct an experiment in which participants are instructed to adjust the interface to their liking, depending on the task they have to execute.

Finally, three participants mentioned that they liked the different display options in the Results module. In other words, they liked to be able to change the type of scent carrier, depending on the type of task. Seven participants never changed their “favorite” scent carrier; four of them preferred the “all” scent carrier and three the “title” scent carrier. Nine chose to vary the scent carrier depending on the task. This indicates the usefulness of being able to choose a scent carrier.

As a final evaluation, we asked the participants whether they agreed or not with a number of remarks about the browser (to indicate on a 7-point scale, with a score of 1 meaning “agree a lot” and 7 meaning “don’t agree at all”). They mostly agreed with the following remarks (the only remarks scoring below 2): “I can understand and apply the information in the video browser,” and “the video browser’s behavior is consistent.” They agreed the least with the following remarks (the only remarks scoring above 4): “the video browser has an attractive appearance,” and “the video browser responds quickly.”

5.3.2 Fabchannel task

The instruction for the task using the Fabchannel website was: “Find a nice song you would like to send to a friend. It has to be a song you’ve never heard before.” Participants could stop as soon as they had found a song they liked, with a maximum search time of 10 minutes. This was a general task with hardly any restrictions, comparable to the general task with the VIBES video browser, but with a different database and a different browsing environment. Table 5-6 shows the outcome of this task.

Table 5-6 Outcome of the Fabchannel task, showing the song that was chosen, from which source it came, and how the participant got to the source

P.	Song	Source of song (patch)	Road to the source
1	Jaga Jazzist - All I Know Is Tonight	Concert Jaga Jazzist	Channels>Scandinavianchannel > search "Jaga Jazzist"
2	None	-	-
3	Madness - Wonderful World, Beautiful People	Concert Madness	Concerts
4	Nightwish - Crazy Train	Concert Nightwish	search "Nightwish"
5	None	-	-
6	The View - Wasted Little DJs	Concert The View	Concerts>Archive highlights (1 of the 4 rock highlights)
7	Millencollin - Fox	Punk channel	Channels
8	Fiction Plane - Two Sisters	Melkweg channel	Channels
9	Flogging Molly - Swagger	Concert Flogging Molly	Concerts>Archive highlights (1 of the 4 punk highlights)
10	Jiggy Dje - Positivisme	Hiphop channel	Channels
11	CSS - Off the Hook	Fabchannelfavorites	Channels
12	Sons and Daughters - Gilt Complex	Rock channel	Charts > Top 10 Channels
13	Sneakerfreak - Bigger Things	Concert Sneakerfreak	Concerts
14	Tim vs Baas B finale	Concert Spitt	Charts
15	Midlake - Head Home	Britishbands	Channels
16	Josh Ritter - Monster Ballad	Concert Josh Ritter	Concerts>Archive highlights (1 of the 4 singer/songwriter highlights)

RQ1 - Which patches do people choose to find relevant videos?

As can be seen in the description of the Fabchannel website in Chapter 4, songs can be part of different patches. Mainly they belong to a concert patch, but they can also be part of a channel patch or a visitor's playlist (not available in this experiment). Concerts and channels can also be part of larger patches such as a genre or chart.

Table 5-6 shows that eight songs were found in a concert and six in a channel. (Two participants could not find a song they liked.) The participants who found a song in a channel all came there via the Channels page. Participants got to the concerts in different ways. Two scanned the long list with concerts until they saw something (vaguely) familiar. Two used the search function to find a concert. One looked for a familiar band, after which he looked for a song he did not yet know. The other searched for a band he had been charmed by in a channel (no direct link is provided [yet] in Channels to the concert a channel song came from). Three participants got to a concert via the concert highlights on the Concerts

page, where a video snippet plays when the mouse is put over the concert's representation (title+frame). Finally, one participant got to a concert via the concert charts.

RQ2 - Which scent carriers do people prefer?

Different scent carriers are available in the Fabchannel interface. A concert - the main video type in the database - is first of all represented by a title: the name of the artist. As most of the artists were unknown to the participants, there was hardly any scent to be perceived in the titles. Participants scanning the long list with concerts often did not perceive any scent and changed their strategy. As described above, two participants did find a scent-carrying title in the list.

If a concert is placed in a genre or channel, it can acquire scent if the genre or channel is liked, even if the artist is unknown. Most genres and channels can evoke certain - positive or negative - associations, while the name of an unknown artist cannot. Most participants did find a song in a channel, or got to a concert via the genre-based concert highlights. In the evaluation, some participants indicated that they had had difficulties with the genres currently in the database. Some had ideas for renaming the genres. For example, genre names such as house, electronic, and alternative were missing, even though those kinds of music were present in the database, because that music was described in other terms. Other participants would have appreciated more specific genre terms. They wanted to see subgenres of house, or a genre combined with a location (e.g., "British ska"). There clearly was a need for more genres: within music, specific genres seem to carry a lot of scent to specific people.

The Charts page was visited by many participants, but here they were again confronted with a list of unknown artists. The concert that was number one in the charts was often explored: even though the artist's name was unknown, its position in the charts gave it scent.

On the homepage, two concerts in the archive were highlighted with a large photo: the "concert of the day" and a concert that was "new in concert archive." These were never selected by the participants. The prominent position on the front page did not seem to provide any scent, and they seemed to be treated like most advertisements: ignored.

In the evaluation, participants explicitly praised the video snippets that started to play (including sound) when the mouse was put over the video representation. For this genre, this clearly is a very functional scent carrier.

During the evaluation, a lot of participants remarked that they liked the appearance and look-and-feel of the Fabchannel website. They thought it was a "nice place to be." It can be said that the website as a whole carries

scent because of its appearance, which may mean that people stay around longer than if the website were less attractive (see also Norman, 2004).

RQ3 - Which strategies do people apply to bridge the three gaps?

The reasons for interacting with videos are provided by the experimental task (“Find a nice song you would like to send to a friend. It has to be a song you’ve never heard before”).

The Fabchannel website provides different ways to bridge the first gap (between reasons for a decision and the decision). Visitors can choose to query the database, or browse concerts and channels. For this specific assignment, most participants began by browsing the concerts or channels.

The second gap (between the decision and the initiation of the action) was bridged by selecting and entering the browse pages.

The third gap (between the initiation of the action and the continuation and completion of the action) was mostly bridged by selecting and exploring a channel, or by selecting and exploring a concert from a certain genre. What was explicitly appreciated by some participants was that the concerts were segmented, so they could jump from song to song. This really eased the exploration of a concert. However, participants did have difficulties interacting with individual songs: it was very hard to jump to the middle of a song. This was felt as an obstacle to exploring songs quickly.

What made the execution of the task difficult was the presence of advertisements: when a concert is selected, first a short advertisement (10-20 seconds) is shown. This may be acceptable when one is sure one is going to watch that concert. However, during the browsing process, when one wants to get an idea what a concert is like (whether it carries scent or not), this really disturbs the navigation process. To make matters worse, the volume of the advertisement was often higher than that of the concerts. Many participants were shocked by the sudden burst of sound and quickly pushed the back-button. However, instead of going back to their previous position in the website, this meant they left the Fabchannel website entirely. Participants felt like they were being “punished” for trying to explore a concert, and really thought this needed to be changed. If the scent in the link is not too strong (one is just trying things out), it may completely disappear with such an obstacle, so that one just gives up trying to follow that particular link. It was acknowledged that advertisements are inevitable, but in this setup it clearly hinders exploration. As we saw in Chapter 2, people visit the Fabchannel website for two reasons: to find a specific (known) concert, and to discover new music. Especially the second reason seems to be hampered this way.

RQ4 - Which functionality do people prefer when interacting with video?

For finding a specific concert, search functionality is becoming more and more crucial because of the expanding size of the database.

For discovering new music, useful categories which carry a lot of scent to specific users are very important, especially when the artist names are not familiar. It should be possible for an artist (or song) to fit in more than one category. In terms of patch-based browsing, this would stimulate switching from patch to patch. In order for this to happen, each concert or song should provide links to the patches it belongs to.

For this type of material, the segmentation of large videos into smaller units (songs) is very useful for exploration. It should be possible to quickly jump to the middle of a song to get a fast impression.

In the evaluation, users said that they would appreciate information about the opinions and behavior of other visitors. For example, for each concert, and for each song in a concert, the metadata should include how often it was viewed, what rating the other viewers had given, and the viewers' comments. In addition, whenever a concert/song was selected, participants would appreciate being able to see other, comparable concerts and songs ("related" or "more like this"). They would like to get recommendations such as those provided by Last.FM or Amazon ("if you like this, you might also like this," or "people who listen to this, also listen to this,").

Another improvement would be to provide more information on what is in the database, how the order in the charts is determined, why a song is in a channel, and so on. In other words, the participants wanted a mental image of the available patches: they really wanted to understand why things were the way they were.

5.3.3 YouTube task

The instruction for the task using YouTube was: "Find remarks by Bill Clinton on his wife Hillary. Write them down and select your favorite."

Figure 5-2 showed that this was considered the hardest task of the experiment (a 4.1 on a scale from 1 to 7), although there was quite some variation in the scores (with seven participants scoring 3 or lower).

As can be seen in the description of YouTube in Chapter 4, the website supports both browsing and querying. For this specific task, all participants chose to start with a query. Browsing was not an option, as no existing categories were specific enough to complete the task. Moreover, even if such a specific category existed, getting to it would require too many navigational steps. So querying was the fastest and the only available option to get to a patch with relevant videos.

The first problem now was to decide which query terms to use. Table 5-7 displays for each participant the chosen queries (each query leading to a patch with results).

Table 5-7 Outcome of the YouTube task, showing for each participant (P) the queries (leading to result patches), the number of switches to other queries/patches, the total number of remarks found, and the indicated task difficulty on a 7-point scale (1=easy; 7=difficult)

P	Queries	switches	remarks	difficulty
1	"bill clinton hillary"; "bill clinton hillary funny"; "bill clinton talks about hillary funny"; "bill clinton speech"	3	1	3
2	"hillary bill clinton"	0	3	1
3	"bill clinton about hillary"	0	5	1
4	"bill clinton private"; "bill clinton wife"; "bill clinton hillary"	2	3	5
5	"bill clinton hillary"; "bill clinton about hillary"	1	0	7
6	"bill clinton"; "bill clinton hillary"; "bill clinton about hillary"; "hillary"; "hillary bill"	4	2	6
7	"bill clinton speech hillary clinton"	0	3	3
8	"clinton campaign bill"; "clinton campaign hillary"; "bill clinton election campaign hillary"; "bill talks over hillary clinton"; "bill clinton retires"; "bill clinton supports hillary"	5	1	2
9	"bill clinton hillary"; "bill clinton hillary quotes"; "bill clinton talks about hillary"; "hillary clinton bill"; "hillary clinton bill clinton speech"; "bill clinton hillary speech"	5	3	6
10	"bill clinton"; "bill clinton wife"; "bill clinton hillary"	2	5	2
11	"bill clinton about hillary"	0	5	5
12	"bill clinton about his wife"; "bill clinton"; "bill clinton about hillary"; "bill clinton hillary clinton"; "bill clinton saying something about his wife"	4	0	7
13	"bill clinton hillary"	0	6	1
14	"bill clinton hillary statement"; "bill clinton statement on hillary"; "bill clinton hillary statement"; "bill clinton' hillary statement"; "bill clinton hillary wife"	4	1	5
15	"hillary clinton bill"; "hillary clinton bill about"; "hillary clinton bill about funny"; "hillary clinton bill talk about"; "hillary clinton bill talking about"; "clinton bill talking about hillary"; "clinton bill talks about hillary"; "bill clinton talks hillary"; "bill clinton hillary"	8	0	7
16	"bill clinton hillary"; "bill clinton' about hillary"; "bill clinton on hillary"	2	3	5
	<i>average:</i>	2.5	2.6	4.1

RQ1 - Which patches do people choose to find relevant videos?

Almost all executed queries included the terms "bill," "clinton," and "hillary," leading to thousands of resulting videos. Adding words such as "funny" or "speech" decreased the number of results, but still left hundreds

of videos, including a lot of irrelevant ones. This was indicated as a serious problem during the evaluation (by four participants). Ten participants indicated the need for some kind of filtering or clustering of the results.

In almost all cases, the “related videos” of a (potentially) relevant video were scanned. This often led to the discovery of other (potentially) relevant videos. In the evaluation, four participants indicated that the “related videos” were very useful during search, providing a kind of filter over the large patch with query results.

RQ2 - Which scent carriers do people prefer?

Participants scanned the results for relevant videos, often finding scent-carrying links on the first result page. Titles and descriptions were clearly the most useful scent carriers in this stage, although a frame showing Bill Clinton added to the scent. In the evaluation, four participants pointed to the sometimes poor quality of the metadata: titles were often misleading, and descriptions were insufficient.

Selecting a link led to the video and its metadata. After the video was started, participants went looking for remarks, either by playing the video or by jumping on the timeline. In the evaluation, the most difficult aspect of the task was finding relevant segments within a video, as seven participants indicated.

While listening to the video, participants scanned the video’s metadata (especially the description). If enough scent was perceived (“this might well be a video with a useful remark”), participants took some more time for these actions, and also began reading other metadata, such as the comments by other viewers. Either a remark was found in the video, or not.

RQ3 - Which strategies do people apply to bridge the three gaps?

The reasons for interacting with videos are provided by the experimental task (“Find remarks by Bill Clinton on his wife Hillary”).

The YouTube website provides different ways to bridge the first gap (between reasons for a decision and the decision), both by querying the database and by browsing categories. As we saw, most often the decision was to query the database.

The second gap (between the decision and the initiation of the action) was bridged by choosing query terms. The result - in almost all cases - was a very large number of videos, too large to explore each individual video.

The third gap (between the initiation of the action and the continuation and completion of the action) was bridged by selecting a video with high scent from the list and exploring its content.

If not enough scent was perceived in the video's data and metadata, participants went back to the results page to look for other scent-carrying links. If not enough scent was perceived within the results, participants started a new query.

Only five participants stuck to their initial patch, while all eleven others tried new searches. Typically, the participants who did not switch found more remarks than the group who did switch (on average 4.4 against 1.7 remarks found - independent two-sample t-test results: $t=3.2$, $df=14$, $p=.006$). Three participants (all "switchers") were unable to find any remarks. The "no-switchers" also experienced the tasks as less difficult (2.2 against 5.0 on a 7-point scale - independent two-sample t-test results: $t=2.8$, $df=14$, $p=.015$). It can be assumed that the results the two groups were presented with were about the same (their queries hardly differed). The difference seemed to be in their approach to exploring the patch. In particular, the "switchers" seemed less patient and less inclined to sit back and listen to the audio. In terms of scent: they either had more difficulties in detecting scent within the (audio channel of the) video data, or had the idea they would find more scent elsewhere. Another possibility was that they had a different scent threshold for the decision to stay within or leave a patch. Comparing the "no-switchers" and the "switchers" on the other audio task with the VIBES video browser, the "no-switchers" clearly thought that task was easier than the "switchers" did (2.4 against 4.0 on a 7-point scale - independent two-sample t-test results: $t=3.7$, $df=14$, $p=.003$). They also found more videos in that task (3.3 against 2.7) although this result was not significant. No other relevant differences were found between the two groups.

RQ4 - Which functionality do people prefer when interacting with video?

For such a large and varied database as YouTube, search functionality is indispensable. This is especially true for a specific task as described above.

However, querying the database still leaves too many results, and some kind of filtering or clustering is needed to create smaller patches that are easier to explore. Filtering out irrelevant results is considered especially important. A patch with "related" videos partly fulfills this need, although it is unclear *why* these videos are related, and this patch again contains a lot of irrelevant videos. In other words, for tasks like these, "related video" provides a less than optimal solution.

Titles, descriptions, and other metadata should represent the contents, and not be misleading. To use a Dutch expression: "the flag has to cover the cargo."

For tasks like the one described here, people need to be able to efficiently scan the contents of the video. It should be easy to jump within

the video. Four participants indicated that showing frames from the video would aid in finding relevant segments. For example, for this specific task, frames showing the face of Bill Clinton would have been relevant

When people are interested in the information in the audio channel of a video, the availability of the transcript (including search functionality) would simplify the search. Five participants indicated that this functionality would have eased the task.

5.3.4 General evaluation on required support for video browsing

After all tasks were completed, we asked in general what kind of support is really needed to facilitate video browsing. It was an open question and participants could name as many issues as they liked.

A number of participants (six) rightly commented that this depends on the situation, especially whether you are looking for specific information or just browsing for fun. Still, we thought the outcome of this question - which was posed after the participants had been faced with different types of tasks and browsing environments - would provide useful information on solutions to the video interaction problem. The results are presented in Table 5-8.

Table 5-8 Issues participants said were important when interacting with videos, ordered by how many participants mentioned the issue

# Participants	Issue
9	Accurate titles and descriptions
9	Good interaction with video content: effective timeline interaction, or segmented videos showing frames (with tags/descriptions for each segment)
9	Opinions (e.g., ratings, comments) and behavior (e.g., views) from others
8	Related videos
7	Good categories/tags
5	Order/filter results
4	Quality (of image and sound) and speed (no delays)
3	Search functionality
3	Assistance in finding the right query terms ("did you mean ...?")

Three issues were mentioned by nine participants. First, videos should have accurate titles and descriptions. From an IFT point of view, this means that if a video contains scent, this should be clear from the textual scent carriers.

Second, there is a need for good interaction with video content, meaning effective timeline interaction, or segmented videos showing frames (with tags/descriptions for each segment). From an IFT point of view, this can be considered to be support for within-patch browsing.

Third, people want information about opinions (e.g., ratings, comments) and behavior (e.g., views) from other people. Clearly, what

other people think of a video is an important scent provider. It helps users make the decision which video to watch, and thus helps to bridge the first gap.

Eight participants mentioned the need for related videos. From an IFT point of view, videos that are related together form a patch. Moreover, if a video has scent, a related video may also have scent. Finally, related videos support between-patch browsing, thus supporting bridging of the third gap.

Seven participants mentioned the need for good categories or tags. This can be interpreted as saying that patches need to be meaningful.

Five participants wanted to be able to order or filter results. This is also about the creation of meaningful patches, or what IFT calls *enrichment*.

An issue mentioned by four participants was the quality (of image and sound) and the speed (especially no delays). Although there does not seem to be a relationship with IFT, this issue may affect the perception of scent (especially in a negative way when quality and speed are low).

Finally, three participants mentioned the need for search functionality, which is about creating meaningful patches. Also mentioned by three participants was the need for assistance in finding the right query terms (“did you mean ...?”), which is also about the creation of meaningful patches. Both these functionalities would aid in bridging the second gap.

5.4 Conclusions and discussion

In this chapter, we posed the research question “How to design a video interaction environment that will optimally support its users?” We tried to find the answer by letting participants execute different tasks in different video interaction environments and observing and evaluating their behavior.

Can we describe and explain interaction behavior completely in terms of IFT? Our results indicate that the answer has to be yes. The results section of this chapter shows that practically all relevant behavior fits into the IFT framework. This was also confirmed in the debriefing of the experiment, in which the IFT framework was explained to each participant. It was easy to use examples of the participant’s own behavior in the experimental tasks to illustrate the framework: how the participant searched for or created patches; how choices were made depending on the amount of scent that was perceived; how when the amount of scent in a patch was too low, participants switched to other patches; and how when there was enough scent in a patch, participants stayed for more exploration. In other words, the framework was a useful tool for communicating about the participants’ behavior. The ideas of patches and scent were easily understood, and all participants accepted these concepts as an explanation of their own behavior. Based on our research, we conclude that the structure of

interaction with video material can be understood as foraging the way IFT describes.

Regarding the first research question, which patches people choose to find relevant videos, this clearly depends on the task description. In the general task with the VIBES video browser and the Fabchannel task, the objective was to find something – *anything* – of interest. In these kinds of situations, people prefer patches that are popular or highly rated by other people: “social” patches. General categories that include the reason or goal for browsing - such as the category “fun” when one is looking for a funny video – also contain a lot of scent to some. Otherwise, people choose categories with which they have positive associations, such as a specific music genre.

The Fabchannel task shows that a video segment (a song) can be taken out of its original context (the concert), put in another patch (a channel), and subsequently be found as an element of that newly formed patch. As long as the new video patch has a meaningful label (which means it carries scent to some of the users) this is a useful way of (re)structuring the video database.

If patches are prefabricated (as categories or genres) they are selected depending on the amount of scent they carry. In other words, categories that are highly associated with the task description are chosen, often leading to multiple selections (as in the VIBES video browser) that are or are not filtered. If patches have to be created (by applying queries), people use query terms that are associated with their task or interest, and by definition have scent.

Problems with patches arise when they are too large or contain a lot of irrelevant videos. In that case some kind of filtering or clustering is needed in order to create more meaningful patches. One way of filtering is to choose the patch “related videos” or “more like this” whenever a relevant video is found. The algorithm behind those patches, however, is often unknown.

Regarding the second research question, the preference for scent carriers also depends on the task. We distinguish between scent in the links and scent in the video data. Scent in the links is found mainly in titles and descriptions, confirming the results of the previous studies described in Chapters 2 and 3. The choice for a scent carrier in the VIBES video browser (where all scent carriers contained a title) did not affect the performance, however, but was probably related to personal preference only. The “all” scent carrier was chosen in most cases.

In some cases, the title is not sufficient, as with the Fabchannel website, which contains a lot of unknown artists. Categories or descriptions (for example the musical genre) will carry more scent in those situations. In

some cases, very specific categories (such as musical subgenres) may carry a lot of scent.

Interestingly, “featured videos” such as the ones on YouTube, or the “concert of the day” on Fabchannel, are mostly ignored. Even though these links have a salient position and appearance, they carry hardly any scent. People may treat these links just like they treat advertisements.

For tasks which need visual or audio confirmation, the links often carry very little scent: people have to go to the data to be sure. Only when the title or description happens to contain all relevant information - as in the title “Cal jumps into shity water” when looking for videos of people jumping into the water – can a lot of scent be perceived in the links. Many links are misleading. Titles, descriptions, and frames may suggest content that is not there, while videos with relevant content may be badly represented. Links may be objectively bad, with an unclear title such as “video14” or a blurred frame, or the link does not relate to the user’s goal while the actual content of the video is relevant. Video is a very rich multimedia type, containing data that may be relevant in different situations. Representations like titles often relate to only one of those situations. If that situation does not relate to the user’s goal, the representation may even send out a negative signal (“ignore me, I’m irrelevant”), or have “negative scent.”

When people search for information in the audio channel of a video, the description becomes very important. The availability of a transcript (including search functionality) may improve the search process.

Scent in the data can be detected using the slider of the player, or with other, more advanced representations. For the music genre (as in the Fabchannel database), video snippets prove to be valuable scent carriers, giving a quick impression of the character or atmosphere of the video contents. The frames module in the VIBES video browser proved very useful for visual search, to quickly detect in a video where some (visually detectable) event occurs. However, for finding audio events (such as a remark) the frames module is not sufficient. Solutions may be the extension of such a module with the transcript, tags, or audio snippets.

Regarding the third research question, people apply different search strategies to bridge the three gaps. This is related to the task, but also strongly related to the functionality of the interaction environment. To bridge the first gap (between the reasons for a decision and a decision) people can either start browsing video patches or query the database. Patches or query terms respectively are chosen that have an association with the reason to start interacting with videos. If needed, multiple patches or query terms may be selected to create a meaningful patch. This patch can be adjusted (or “enriched”) by adding a filter. The decision to start browsing or querying is – among other things - dependent on how well the

target can be described in terms of keywords (which was a problem with the Fabchannel task, for example), and the direct availability of relevant, meaningful patches (which was a problem with the YouTube task, for example). When the number of categories gets above a certain threshold, people may fail to notice certain categories. This was already observed with the VIBES video browser, which contained “only” 43 categories arranged in four main categories. Solutions may be to provide search functionality (within the list of categories), to add advanced ways of presenting the category menu items (e.g., in hyperbolic trees), or to accept that categories are useful only in relation to a selection of the database, e.g. a specific video (as with the Categories of Segment module in the VIBES video browser).

The second gap (between the decision and the initiation of the action) is often easily bridged in a hyperlinked environment: a decision is put into action by simply clicking on a hyperlink.

The third gap (between the initiation of the action and the continuation and completion of the action) can be crossed in different ways. If the reason to browse does not involve a specific target, the search can continue a long time, with the searcher picking up relevant or interesting video bits everywhere (compare berrypicking as described in Chapter 1). If the task does have a target, and the database contains a target, the search is completed when a target is found within the patch. In some cases, the metadata - most often: titles and descriptions - of a video may suffice. In other cases, the video data need to be explored. The usefulness of different exploration strategies is determined by the task, especially if the needed information can only be found in the visual or audio part of the video. When a video is relevant, after exploration one can switch to “related” videos (if available), or choose another video with scent from the current patch. When a video that has been explored proves irrelevant, one can switch to another video in the patch which has scent. If the patch as a whole has insufficient scent, one can switch to another category or query result. Individual differences in search strategies and successful execution of the tasks seem to be related to the participant’s perception of scent *and* the individual threshold at which the decision is made to stay in a patch or leave a patch.

For the continuation of the browsing process, it is important that the search be not disrupted by long delays or intrusions such as advertisements. When people are exploring the data, not sure what to look for or what to expect, such interruptions may cause them to break off the exploration. They will continue, despite disruptions, only when the scent of the target is very strong.

Regarding the fourth research question, this experiment provided a lot of data for the question of which functionality people prefer when interacting with video. As was shown in section 5.3.4 (on the required

support for video interaction), but also in the description of video interaction environments in Chapter 4, each desired type of support can easily be translated to the IFT framework. The concepts of patches and scent clearly have practical meaning, and can be used to design and evaluate video environments. Questions such as “Which patches are or can be created?” “How is browsing within a patch supported?” “How is switching between patches supported?” and “In which way is scent displayed in the interface?” prove very useful.

The results of this experiment – especially when looking at the tasks with the VIBES video browser – confirm that the kind of support that is needed depends on the task or the reason why people interact with video. This confirms the results from Wildemuth et al. (2002), in which users viewed different surrogates as being more or less useful for different types of tasks, and said they would like to be able to move from surrogate to surrogate. This suggests the need for a flexible interface, in which modules that are important for a specific task can be enlarged and other minimized, although this may give rise to usability issues.

In cases when users know exactly which video to watch (for example, a specific concert in the Fabchannel database), keyword search functionality needs to be provided in order to search for the title of the video. (Of course, for large video databases search functionality is indispensable.) If the uploading of the video is related to a specific event in time, a patch with “most recent” videos is also useful.

In other cases, the first step is to narrow down the search by selecting a patch with potentially relevant videos, that is, with high scent. This can be done by querying the database using keywords, or by selecting a category (or a combination of categories) provided by the video environment. The more aspects of a video are described, the easier it will be to create a patch that is closely related to the user’s goal. Whenever the number of videos in a patch is too large (which is often the case), it should be possible to filter or cluster the results (for example, exclude home videos on YouTube) to create smaller patches that are easier to explore.

When people are mostly unfamiliar with the videos in a database and titles contain little scent (for example, in a music database with unknown bands), creating useful categories to provide starting points is very important. In some cases it may be helpful to provide information on the database itself: what types of videos are in it, how the order of the charts is determined, why a video appears in a certain category, and so on. In other words, people want to be able to create a mental image of the available patches.

When a list of videos is presented, essential elements to include are titles and descriptions, as we saw in Chapter 3. As indicated by the participants in the study above, these must be accurate: if the video contains scent, this

should be clear from the textual scent carriers. Adding a frame can help to create more scent. It is useful to be able to select different types of scent carriers, depending on the type of task. For general tasks (“Find anything that is funny or of interest to me”), social data such as recommendations, comments, and ratings by other users are reported to be helpful. Videos can be organized using this information, thus creating patches with high scent. If people are likely to visit a given database with such a general purpose in mind, then it is advisable to structure that database according to the behavior (e.g., number of views) and opinions (e.g., ratings) of other users. Participants in the study above indicated the importance of social data. Clearly, what other people think of a video is an important scent provider. It helps to make the decision which video to watch, and thus helps to bridge the first gap.

Whenever a potentially relevant video is selected on the basis of its metadata, inspection of the contents is often needed to detect any scent in the data. The participants in the study above indicated the importance of good interaction with video content. It should be easy to jump within the video using a timeline. When people are searching for visual events, frame-based visual summaries (like the Frames module in the VIBES video browser) are very useful, especially when there is direct access to the relevant part of the video via a frame in which scent is detected. When people are looking for information in the audio channel of a video, such a frame-based summary should ideally be extended by adding a title, description and/or tags to the separate frames. Also the availability of the transcript (including search functionality) would ease the search. The results of the study described in this chapter indicate a preference for searching textual descriptions rather than the audio of a video.

Especially for larger videos, segmentation into smaller units is very useful for exploration. As we saw in Chapter 2, these units should be the semantically meaningful “natural” parts of a video, such as the items of a newscast or the songs of a concert. It should be easy to jump between segments as well as within a segment.

To support exploration of the database (that is, between-patch browsing by following scent-trails), every video should provide a starting point to other videos (“related” or “more like this”). Participants in the study above indicated the importance of this. From an IFT point of view, videos that are related together form a patch. Moreover, if a video has scent, a related video can potentially also have scent. Finally, related videos support between-patch browsing, and as such support bridging of the third gap. Relating videos can be done for example via their tags or the categories they belong to. For this kind of exploration to occur, it is very important that videos be part of multiple overlapping patches. This is also true for video *segments*.

The video they are part of can be considered just one of the patches they belong to.

For example, an item from a newscast about person X or event Y should also provide access to other videos containing person X or event Y. In the Fabchannel task described in this chapter, participants would have liked to get recommendations similar to those provided by websites such as Last.FM or Amazon (“if you like this, you might also like this,” or “people who listen to this, also listen to this”).

The functionality described above is based on observing people executing different types of tasks using different types of video environments, and interviewing them on what they thought was useful. As such, this provides a fairly extensive view on what kind of functionality is really needed in video interaction. The study was not exhaustive, but from the IFT point of view, the most important elements that need to be present to support video interaction are identified here. Regarding specific functions several developments can be expected. For example, regarding the creation of patches, there is evidence that an arsenal of tools – e.g., query-by-text, query-by-image, query-by-concept – leads to better interactive performance than any single tool (Christel & Yan, 2007). Also, regarding within-patch browsing, the availability of several types of video surrogates or representations seems to be preferred by users, so they can easily switch between them depending on the task at hand (Wildemuth et al., 2002). This is also confirmed in Li, Gupta, Sanocki, He, and Rui (2000). They implemented a browsing application providing a wide array of features, and studied browsing behavior for six different video genres. Depending on the genre, different interface features were useful for detecting scent.

The last question to answer here is: did we create a good browser? Looking at which functionality people prefer, we can say that the VIBES video browser offered much of the functionality that seems to be essential for efficient video interaction. People could select patches using the Categories and Filter modules. The interaction with categories was much appreciated, with a special mention for the Filter module. The categories are prefabricated patches, which help to bridge gap 2 by providing recognizable cues on what to select. It should be pointed out that the database was small (80 videos) and so was the number of categories (43 categories in four main categories). Even with this small number of categories, participants incidentally indicated during the evaluation that they had not seen certain categories which might have been useful for the particular task they had just executed. There clearly is a limit to the size of the list, and additional means – such as searching within the categories – are needed when the list gets longer.

People can detect scent within patches using the Results module, selecting different scent carriers depending on the type of task. Although

seven participants in the study never changed their “favorite” scent carrier (of which four preferred the “all” scent carrier and three the “title”), nine chose to vary the scent carrier depending on the task. This indicates the usefulness of being able to choose a scent carrier.

People can detect scent within a video using the Frames module and the player. The Frames module was especially useful for the specific task with visual confirmation, and got a lot of positive response.

What this version of the browser clearly lacked was the option to query the database, which is a useful tool for bridging the second gap. What was also less well supported was between-patch browsing. The Categories of Segment module had not enough added value. It was not used very much, probably because the categories were also available and visible in the Categories module. Perhaps the usefulness would increase if the number of categories were much larger; now it was not too difficult to pick the same category from the Categories module. But what really seemed to be lacking was a link to related or comparable videos.

The timeline was considered the least useful of all modules. Only in the general task did it seem to have a function, namely to select relatively short videos. However, the timeline was originally developed for displaying information on lengthy, segmented videos. In this experiment, only short not-segmented videos were used, and so the functionality of the timeline was only partly exploited.

Participants were divided on the matter of visual design of the browser. Half indicated that the design was too cluttered, ugly, and/or uninteresting, and that the player was too small. The other half said they liked the interface design, especially the logical organization of the modules. As such, it may be a matter of taste and personal preference. Still, the attractiveness of the browser is something to be taken seriously. As Donald Norman (2004) states: “Attractive things make people feel good, which in turn makes them think more creatively. How does that make something easier to use? Simply, by making it easier for people to find solutions to the problems they encounter.” When you feel good, you are better at brainstorming, at examining multiple alternatives (Isen, 1993). In other words, an attractive interface can help the process of bridging the gaps. Especially regarding gap 3, people probably will explore longer and less easy give up.

Our main conclusion based on the experiment presented in this chapter is that our IFT framework (including the concepts of patches, scent, and gaps) is useful for communicating about browsing behavior, useful for describing and evaluating video environments, and useful for deriving design principles. We can safely state that our IFT framework is applicable and fruitful, and has the potential for a complete human-computer interaction theory.

IFT-based browsing: conclusions and discussion

In this chapter we will first discuss the main conclusions of this thesis in the light of the research questions we stated in Chapter 1: “What is the most useful way to classify video content?” “What is the character of good video scent?” and “How to design a video interaction environment that will optimally support its users?” Next, we will see what we have learned from our studies regarding IFT, browsing behavior, and bridging gaps. Then we will discuss the limitations of this research, and will end with concluding remarks and directions for future research.

6.1 Introduction

The main objective of our research was to study how to support interaction with video in such a way that people can efficiently satisfy their needs. We presented an account of information interaction by refining the ideas of Marchionini (2004), stressing the role of the human in the retrieval problem with the emphasis on the flow of representations and actions rather than discrete matches. We tried to explain information interaction behavior on the basis of human search principles as described in Information Foraging Theory, or IFT (Pirolli & Card, 1999), and defined the interaction contexts in terms of gap bridging as proposed by Searle (2001). We then looked at four user studies focusing on the specific case of video. The main goal of the studies was to see whether the IFT framework is a useful approach to the problem of video interaction. We explored principles from IFT regarding how people perceive and structure their environment (in patches), and how they navigate through that environment (by following scent).

We tried to see whether the framework helps us to understand what is important in the design of interaction environment. First, in Chapter 2, we looked at the implementation of the idea of *video patches*, which structure the information environment. We described two exploratory surveys that collected data on user preferences for video categories that may serve to organize patches: the Kenniswijk-survey and the Fabchannel-survey.

Second, in Chapter 3 we studied *video scent*, especially regarding the design of *scent carriers*, which guide the user's navigational decisions. We described an experiment in which we asked participants to select the most relevant link to a video from a group of links. We measured the perceived scent by asking for the subjective probability that the information that was needed could be found behind that link. We repeated this for different types of tasks and different types of scent carriers to study the influence of these factors. Based on the IFT framework and the results of these studies, we designed an experimental video browsing environment: the VIBES video browser. We described this browser in Chapter 4, together with two other video environments: the websites of Fabchannel and YouTube.

Third, we tried to see whether the IFT framework helps us to understand *video browsing behavior*. In Chapter 5 we described a laboratory experiment in which we observed people performing tasks with the three video environments (described in Chapter 4), and evaluated whether we could describe and explain their behavior in terms of our framework. Moreover, our goal was to see whether the framework would help us to understand what kind of support is really needed to let people efficiently interact with videos. In other words, we wanted to see whether the framework has practical use for the design and evaluation of video environments.

6.2 Main conclusions

The main conclusions of this research are presented below. We present the conclusions by answering the research questions introduced in Chapter 1.

6.2.1 Research question 1: What is the most useful way to classify video content?

Which patches are useful depends on the type of task at hand, especially in cases where the user's need is specific. In that case, the easiest way to create useful patches is by querying the database using keywords that associate with the user's need. The created patches will have scent by definition (if the database contains relevant videos which are adequately described). If

patches are prefabricated - as categories - they are selected depending on the amount of scent they carry.

In our research, in general the most important classification of a video (segment) is the topic it covers or the genre it fits in. In specific cases (for example, music videos), specific categories (such as musical sub-genres) may carry a lot of scent. When people have no specific search goal, but generally want to see something that is interesting or funny, they prefer patches that are popular or highly rated by other people: “social” patches.

Although TV/video programs as a whole form relevant patches, there is also a need for more efficient interaction with smaller video units. The preferred units of interaction are the “natural” program segments: the semantic segments as the program maker intended (e.g., the items of a newscast), or those “naturally” provided by the characteristics of the content (e.g., the songs of a concert). These segments can be considered to be good scent carriers, and are very suitable for providing navigational information.

A video segment can be taken out of its original context, put in another patch, subsequently be found as an element of that newly formed patch. As long as the new video patch has a meaningful label (which means it carries scent to some of the users) this is a useful way to (re)structure the video database. A patch-oriented database structure supporting both within-patch and between-patch browsing may thus lead to user satisfaction.

Interaction with patches becomes more difficult when they are too large or contain a lot of irrelevant videos. In that case some kind of filtering or clustering is needed in order to create more meaningful patches. One way of filtering is choosing the patch “related videos” or “more like this” whenever a relevant video is found.

6.2.2 Research question 2: What is the character of good video scent?

As we saw in our studies, some elements in the interface, or scent carriers, can carry more scent than others. Regarding video scent, the perceived usefulness of data in guiding users’ navigational decisions is a function of communication mode (verbal vs. pictorial); reason to browse (cognitive benefit vs. mood improvement); and tasks to perform (type of task, specificity, difficulty).

The most useful pieces of information to get about a video in links are a title, a description of events, and a description of subjects/topics. In other words: textual information seems to be superior, especially in the earlier stages of search. As the visual channel is often dominant in video, this is a remarkable conclusion.

A preview/trailer is next in usefulness, but this scent carrier typically has a role further on in the search process. Scent carriers requiring a lot of time

and attention (which is the case when they contain audio or moving images) are a problem in the earlier stages of search. Typically, a user would only inspect such a scent carrier after establishing - on the basis of the available scent carriers - that the video source is worth further examination. A frame is considered less useful when determining the relevance of videos, especially compared to titles and descriptions. The difference in perceived scent between titles and frames is larger when the reason to browse is cognitive benefit than when it is mood improvement. The combination of a frame with textual information (e.g., a title), however, is clearly strong in terms of perceived scent. In general it can be said that combining a title with other information enhances perceived scent of the represented video. In the scent experiment it was shown that a title + frame and a title + description did not carry significantly less scent than the situation in which all information was displayed. If we were to design a composed scent carrier for the earlier stage of search - when people have to choose from a list of links to videos - it would contain at least a title, a description (of events, subjects/topics), the category/genre, and the purpose of the video (which is often directly related to the category/genre). Ideally it would also contain a frame, especially because of its strength in combination with the title. Other elements - such as an expert review, the duration, or ratings by others - can have importance in specific situations, but the abovementioned elements seem to be vital for supporting efficient video browsing.

With a simple task and with relevant videos present in the result set, perceived scent is higher than with a general (cognitive benefit) task. With increasing task difficulty, perceived scent drops. When the task becomes too difficult, video representations simply cannot provide enough information to solve the problem. Difficulty can also increase when the match between the task description and the videos present in the result set is bad. The perceived scent of the scent carriers is then significantly lowered: the user faces a difficult to impossible assignment in relating the information from the scent carriers to the task. The differences between types of scent carriers decrease when the match gets worse, causing any advantages of certain scent carriers to disappear.

Interestingly, a link having a salient position and appearance does not necessarily carry much scent. For example, “featured videos” on YouTube and the “concert of the day” on Fabchannel are mostly ignored. People treat these links just like they treat advertisements.

We distinguish scent in the links and scent in the video data. For tasks which need visual or audio confirmation, the links often do not carry a lot of scent. The description becomes very important, and the availability of a transcript (including search functionality) may improve the search process. Often, people need to inspect the data to be sure. Scent in the data can be detected using the slider of the player, or using more advanced

representations. For music (as in the Fabchannel database), video snippets prove to be valuable scent carriers, giving a fast impression of the character or atmosphere of the video contents. The frames module in the VIBES video browser proved very useful for visual search, to quickly spot the segment in a video where some (visually detectable) event occurs. However, for finding audio events (such as a remark) the frames module is not sufficient. One solution may be the extension of such a module with the transcript, tags, or audio snippets.

6.2.3 **Research question 3: How to design a video interaction environment that will optimally support its users?**

As we concluded in Chapter 5, our IFT framework is useful for designing and evaluating video browsing environments. Questions such as “Which patches are or can be created?” “How is browsing within a patch supported?” “How is switching between patches supported?” and “In which way is scent displayed in the interface?” prove to be very useful. Applying the results of our research, the issues below are important for the design of video interaction environments.

First, it should be stated that what kind of support is needed depends on the task or the reason why people interact with video. In cases where the user knows exactly which video to watch, keyword search functionality to search for the title of the video must be provided. If the uploading of the video is related to a specific event in time, a patch with “most recent” videos is also useful. In other cases, the first step is to narrow down the search by selecting a patch with potentially relevant videos, that is, with high scent. This can be done by querying the database using keywords, or by selecting a category (or a combination of categories) provided by the video environment. Categories related to the topic or genre of the video are most relevant. Whenever the number of videos in a patch is too large (which is often the case), it should be possible to filter or cluster the results to create smaller patches that are easier to explore. When people are mostly unfamiliar with the videos in a database and titles hardly contain any scent, creating useful categories to provide starting points is very important. Patches related to other people’s behavior or opinions (including recommendations) are especially useful in that case.

When a list of videos is presented, essential elements to present are titles and descriptions. These have to be good: if the video contains scent, this should be clear from the textual scent carriers. Adding a frame can help to create more scent. It is useful to be able to select different types of scent carriers, depending on the type of task. For general tasks (“find anything that is funny or of interest to me”), social data such as recommendations, comments, and ratings by other users are very important. When people are

likely to visit a given database with such a general purpose in mind, then that database can best be structured according to the behavior (e.g., views) and opinions (e.g., ratings) of other users. What other people think of a video is an important scent provider.

Whenever a potentially relevant video is selected on the basis of its metadata, inspection of the contents is needed to detect any scent in the data. Efficient interaction with video content is important. It should be easy to jump within a video using its timeline. Our research has shown that when people are searching for visual events, frame-based visual summaries (like the Frames module in the VIBES video browser) are very useful, especially when there is direct access to the relevant part of the video via a frame in which scent is detected. When people are looking for information in the audio channel of a video, they prefer to search textual descriptions rather than the audio itself. A frame-based summary then would ideally be extended by adding a title, description, and/or tags to the separate frames. Availability of a transcript (including search functionality) would also ease the search.

Especially for larger videos, segmentation into smaller units is very useful for exploration. As we saw in Chapter 2, these units should preferably be the semantically meaningful “natural” parts of a video, such as the items of a newscast or the songs of a concert. It should be easy to jump between segments as well as within a segment. Designers and developers should feel encouraged to experiment with interaction modes that take video segments as the unit of interaction.

To support exploration of the database (that is, between-patch browsing by following scent-trails), every video should provide a starting point to other videos (“related” or “more like this”), which can have the shape of a recommendation (“if you like this, you might also like this”). From an IFT point of view, videos that are related to each other form a patch. Moreover, if a video has scent, a related video can potentially also have scent. Finally, related videos support between-patch browsing, and thus help in bridging the third gap. Relating videos can be done for example via their tags or the categories they belong to. For this kind of exploration to occur, it is very important that videos be part of multiple overlapping patches. This is also true for video *segments*. The video they are part of can be considered just one of the patches they belong to.

In sum, a video interaction environment should offer tools to bridge the three gaps. It should provide useful patches (or the tools to create useful patches), support the detection of scent within a patch by providing good scent carriers, and support scent-following between patches.

6.3 IFT, browsing behavior, and bridging gaps: what did we learn?

When we look at our research as a whole, our conclusion is that the concepts of IFT - patches and scent – and the idea of bridging gaps provide a useful context for studying the video interaction problem. The last experiment in particular – in which we closely observed browsing behavior and acquired users' opinions on what is needed to efficiently browse video environments – confirmed that we can describe and explain practically all relevant interaction behavior in terms of IFT. This was also confirmed in the debriefing of the experiment, in which the IFT framework was explained to each participant. It was easy to use examples of the participant's own behavior in the experimental tasks to illustrate the framework: how the participant searched for or created patches; how choices were made depending on the amount of scent that was perceived; how when the amount of scent in a patch was too low, participants switched to other patches; and how when there was enough scent in a patch, participants stayed for more exploration. In other words, the framework was also a useful tool for communicating about the participants' behavior. The ideas of patches and scent were easily understood, and all participants accepted these concepts as an explanation of their own actions.

The way videos are classified, or better, the way the video environment is structured in patches determines users' navigational behavior. Patches (which can be separate videos, collections of videos, or collections of video segments) carry scent by the way they are defined. The way patches are represented in the links to those patches – the scent carriers - determines the navigational decisions people make. We have described the relevance of the characteristics of a scent carrier and the characteristics of the user's task for the perception of scent. Our work indicates that the idea of scent can - via the design of scent carriers - provide insight into the design and evaluation of an efficient video environment.

The last experiment has clearly indicated the essence of within-patch browsing, and, related to that, between-patch browsing for finding relevant information. Between-patch browsing occurs when a person explores a patch (most often a video), and is inspired by the data or metadata he/she finds there. At this point, the interface should provide links to other, similar patches or to patches that are related specifically to the item of interest that was discovered.

As was indicated in Chapter 1, this research confirms the dominance of browsing in video interaction. The flow of representations and actions is the central quality of the interaction, rather than discrete matches (Marchionini, 2004).

Our main conclusion is that it is very feasible to explain user behavior and formulate design principles in terms of the IFT framework, and that the framework as such is useful: IFT “works.” Based on our research, we conclude that the structure of interaction with video material can be understood as foraging the way IFT describes.

Despite its exploratory character, browsing is highly structured. Not only do the four studies reported in this thesis testify to the value of the IFT framework in thoroughly describing all browsing behaviors relevant to our research, it also seems possible to summarize all the observed browsing behaviors under three headings: the three “gaps” (Searle, 2001). In order to bridge these three gaps, we saw, people can choose different search strategies. Their choice is influenced by their goals and, in particular, by the functionality of the interaction environment.

To bridge the first gap (between the reasons for a decision and a decision), people can either start browsing video patches or query the database. In hyperlinked environments, the second gap is often easy to bridge, because the decision that is reached after the first gap is often based on the choice of a link. If that link leads to a relevant video, bridging the second gap simply means the actual selection of that link. Problems can arise when the first gap is bridged by a query, or when a link leads to a collection of videos, and the result is a very large collection, too large to explore each individual video. Especially when there are no clear filtering or clustering options, people can get overwhelmed and lost in this gap, and return to the starting point to select another patch. This can also happen when a potentially relevant video needs to be explored. If this is a time-consuming process (e.g., because the video is not segmented, or because advertising or other delays interrupt the exploration) then this can lead to “impatient” behavior.

The third gap (between the initiation of the action and the continuation and completion of the action) can be crossed in different ways. If the reason to browse does not involve a specific target, the search can continue for a long time, with the searcher picking up relevant or interesting video bits everywhere (compare berrypicking as described in Chapter 1). If it is a task with a target, and the database contains a target, the search is completed when a target is found within the patch.

In a television environment, people often know what they are going to watch before they turn on the TV (for example, the 8 o’clock news) and have no problem locating the relevant channel. In that case they hardly experience the gaps as described by Searle. In cases where they do not know what to watch, consulting guides and channel-surfing are two important strategies for finding TV programs of interest, with the “metadata-first” option (using guides) being slightly more popular than the “data-first” option (channel-surfing). For bridging the third gap, the same

strategies apply. These two strategies for bridging the first and third gaps - consulting an overview or guide (“metadata-first”) and browsing video data (“data-first”) - can be considered as basic heuristics (Gigerenzer & Todd, 1999). The main heuristic is scent-following (“select the link with the highest scent”) and there are two sub-heuristics related to finding scent in either the data or the metadata.

Most visitors to the Fabchannel website go there to watch a specific concert. In that case the first gap is already bridged: they know which concert to look for. The problem is to locate it. The website offers search functionality to help them find specific concerts and bridge the second gap. Another way to find concerts is by browsing the database, sorting the concerts by name or date. However, people also indicate that an important reason to watch videos is to learn, to discover new things. For this they need support. Especially when people want to discover new music, there should be no strict borders between video files, so that the third gap can easily be bridged. For the continuation of the browsing process, it is important that the search not be disrupted by long delays or advertisements. Especially in cases where people are exploring the data, not sure what to look for or what to expect, such interruptions may break off the exploration.

So, one simple heuristic (“select the link with the highest scent”) and two possible sub-heuristics (data-first or metadata-first) sum up users’ within-patch and between-patch switching. The concepts of patch, scent, and switching within and between patches are necessary and sufficient to do justice to the exploratory nature of browsing. It seems that good IFT-based design allows users to bridge the gaps without noticing any discontinuity.

It was clear from the start that browsing behavior is highly interactive. This is not only because video is a very attention-demanding medium with a linear and time-based character, but also because the user’s goals can evolve over time. Both aspects justify skepticism regarding the possibility for users – or computers – to guess in any fruitful manner where the search might lead, and therefore also what the next fruitful step in the search process might be. Not only does the video contain information that can only be accessed directly (by viewing the video), the users can also change their goals based on the information that is gathered through watching the time-based video data directly. Hence the importance we attach to acknowledging the “gaps.” As expressed by Searle, the gap implies indeterminacy of action. Searle wants us to take this indeterminacy of action as a fact, not as an issue to be resolved. We cannot know or predict users’ actions, but we can facilitate options for action, as well as their execution, by using our framework to come up with new ways of supporting interaction. For that purpose our approach has been to take human-computer interaction as central.

In fact, when taking into account the dominance of browsing (see Chapter 1), it was our non-arbitrary choice to pay attention to the commonness of interaction. As was stated by Marchionini (2004), action, perception, and reflection appear to be unalienable from the search process itself. In IFT, all relevant actions concern moves within a patch or between patches. Perceptions always amount to perceptions of scent, and reflections always amount to the assessment of scent in a scent carrier relative to other scent carriers. Taken together, scent and patches can explain the exploratory nature of browsing. Scent explains the variability of the impact of cues upon the direction and re-direction of the search. Patches explain the locality of searches and the switching between localities. Unlike categories (see Chapter 2), patches as localities can be overlapping, nested, neighbors, or far apart.

With this in mind, we can now comment on the meaning of optimality in the context of IFT as a framework for (video) browsing. There is some risk in using the term optimality, which is close in meaning to the term optimization (Klein, 2001). Within the framework of IFT and with regard to the three gaps, we mean by optimal the following: an IFT-based browsing environment is optimal to the extent that it allows users to bridge the three gaps with minimal experience of discontinuity. Thus, the more discontinuous the user's browsing experience, the less *IFT gap-optimal* the browsing environment. As such, the control parameter for IFT gap-optimality is the user experience of (dis)continuity.

In sum, IFT gap-optimality entails three levels: gap 1, gap 2, and gap 3. The degree of IFT gap-optimality depends on the respective measures of (dis)continuity, as experienced by users. We consider the concept of IFT gap-optimality to be tentative. Our research has covered enough ground to develop this concept further, to make it into a useful tool for the development of high-quality browsing environments, and to measure those environments' quality.

6.4 Limitations of the research

Looking at the generalizability of our research, our four studies display some weaknesses. These are mostly related to the general limitations of user research: limitations of means, time, money, support, and so on. A large problem is to create studies that keep participants motivated from beginning to end: there is a limit to the number of questions or tasks that can be imposed on the participants. On the side of the experimenter, there is a trade-off between the number of participants and how thoroughly they are questioned or observed. The development of tools to be used in experiments also depends on the available support.

Taking a closer look at video browsing behavior, Wildemuth et al. (2003) identify four main classes of variables that influence video browsing success: the user task/need, individual user characteristics, video characteristics, and characteristics of the surrogates that represent the full videos. In our research, we covered all these variables except user characteristics, partly because it would multiply the number of analyses.

The Kenniswijk study described in Chapter 2 was a survey with general questions mostly aimed at TV viewing behavior (at the time of the study, video on the internet was still a relatively new phenomenon). The research format – a survey – prevented a more in-depth analysis, keeping the results rather general. For example, it was hard to relate specific behaviors to specific reasons to watch TV. Moreover, the method relied on reported rather than observed behavior. The Fabchannel study described in Chapter 2 also lacked a task-specific and in-depth analysis due to the research format (again a survey). The specific genre used in the Fabchannel study counterbalanced the Kenniswijk study, which was aimed at video in general. Ideally, a number of other specific video genres and their users would be studied, as well as more specific reasons why people start to interact with videos (other than “cognitive benefit” and “mood improvement,”) or more specific interpretations of those reasons.

In the scent experiment described in Chapter 3, to restrict the total number of tasks, each type of task was represented by only one version. For the same reason, a limited number of scent carriers was tested, and only in the first stage of search (scanning a list of video representations). We did not study how participants could find scent in the data. The browse experiment described in Chapter 5 had the same restrictions regarding the number of tasks. Moreover, the VIBES video browser that was used lacked certain functionalities, and only one version of the user interface was used. It is hard or impossible to separate behavior from environment, so that the behavior we observed was limited by the video environments that we used. With the exception of the Fabchannel database, only interaction with short videos was observed. Both experiments were performed with students, who are younger and more educated than the average participant would be.

Regarding reproducibility, we can state that people in general are becoming more experienced in interacting with video, especially on the internet. The Kenniswijk study in particular dates from a time when only a few people had had that experience. It may be that user preferences have changed over time as a result of being confronted with new interaction modes. Also, the ideas we looked at in the Fabchannel study were relatively new at that time. As the Fabchannel website has evolved, visitors' opinions may have changed.

In previous sections, we discussed the exploratory character of our research. This and the limitations presented above influenced the general impact of this research.

6.5 Concluding remarks and directions for future research

The research presented in this thesis has a highly multidisciplinary character. A psychological theory is applied to an information science subject, and is tested on a multimedia information carrier (video) with many technical and user-interaction issues (which are studied from a computer science and a human-computer interaction point of view respectively). The problem of "searching for videos" or "video interaction" has been studied by many different research communities and discussed in related forums. Although the multidisciplinary approach that we deemed necessary carries the risk of never reaching the depth that might be achieved within a single discipline, it is less hampered by the tunnel vision that can occur in a single field. It can provide new points of view, bridge the gaps between the disciplines, and broaden existing perspectives.

Regarding the generalizability of the results as discussed in the previous section, the studies had their limitations. However, even though video databases and different interaction modes have become much more available and popular in recent years, the general principles related to human searching behavior remain no less valid. In the restricted domains we studied, our IFT framework (including the concepts of patches, scent, and gaps) proved very useful for describing searching behavior and providing tools for the design and evaluation of video interaction environments. The next step is to see whether the framework can be generalized to broader domains.

Results from our studies may be judged to have a trivial character, meaning that our findings correspond to what seems to be common practice. For example, a website such as YouTube has evolved to a state that largely corresponds to what we think is necessary support for video interaction. However, our framework based on patches, scent, and gaps gives the often implicit ideas a theoretical body. We can now start to develop tools to build and test new browsing environments more systematically, and with more confidence that new design will actually work in the appropriate contexts of use.

When we ask the question where to go from here, we see that supporting users, not predicting their actions, is the central issue. Future research may aim at developing methods and tools that help the user with

his/her start-continue-end decisions for bridging the three gaps by loops of action, perception, and reflection. The research presented here yields some starting points. Important questions to ask in the future relate to how the environment is structured into patches, and what means are available to the user to create useful patches; how scent is presented in the interface; and how the detection of scent is supported by browsing within a patch or between patches. For all these issues, techniques need to be developed to bring either humans (e.g., via social tagging) or machines into action.

To conclude: for video classification, scent carrier presentation, and video interaction the optimal solutions always have a local, temporal character. We introduced the tentative concept of IFT gap-optimality to make this explicit. The degree of optimality of any given situation can be measured by looking at the type (gap 1, 2, and/or 3) and degree of (dis)continuity that is experienced by users when trying to bridge a gap. The degree of experienced (dis)continuity is a function of the support provided by the video environment in terms of patch selection, scent-following, and within-patch and between-patch browsing. Our research has covered enough ground to develop this concept further, and to start working towards the measurement and more objective comparison of the value of existing and new browsing environments.

Appendix

A

Kenniswijk survey: questions (in Dutch)

This appendix shows the questions of the Kenniswijk study described in Chapter 2. For further information, see that chapter.

Kenniswijk

<< Vorige pagina

Algemene gegevens
 Voor het onderzoek is het van belang dat wij enkele gegevens over u weten. Deze gegevens worden vertrouwelijk behandeld.

Vraag 1
 Bent u een man of een vrouw?

man
 vrouw

Vraag 2
 Tot welke leeftijdscategorie behoort u?

16-20 jaar
 21-25 jaar
 26-35 jaar
 36-45 jaar
 46-55 jaar
 56-65 jaar
 66-75 jaar
 75 jaar en ouder

Vraag 3
 Wat is uw hoogst genoten opleiding? (indien u nog met een opleiding bezig bent, deze invullen)

Kenniswijk

Beste Kenniswijker,
Hartelijk dank voor het deelnemen aan deze enquête!
 De enquête bestaat uit 11 pagina's met 43 vragen. Invullen kost ongeveer een half uur. Als u alle vragen beantwoordt en de enquête verzendt, maakt u kans op één van de 5 hele Staatsloten L.w.v. € 12,50 of één van de 20 1/2 Staatsloten L.w.v. € 2,50.

Druk nu op de onderstaande knop om met de enquête te beginnen (dit kan enkele seconden duren)

Volgende pagina >>



<< Vorige pagina

TV/video
 Het meeste gebruikte soort van TV/video, dan bedoelen wij daarmee alle vormen van bewegende beelden die gemaakt zijn om naar te kijken (das niet foto's en games).

Vraag 4
 Hoe vaak kijkt u naar de volgende vormen van TV/video? Kies het antwoord dat het best past.

	Nooit	jaar	jaar	4x per maand	1x per maand	2x per maand	1x per week	3x per week	Minimaal dagelijks
TV-programma's tijdens de uitzending	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zelf opgenomen TV-programma's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gehuurde video's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen video's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Streaming video van het internet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Van het internet gedownloade video's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen film/video's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Films in de bioscoop	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 5
 Hoeveel uur per week kijkt u naar TV/video-beeld? (schatting)

uur per week

- basisonderwijs
- VMBO (of een vergelijkbare opleiding, bv. MAVO, MIULO, LBO)
- HAVO/VWO (ook: MMS, HBS)
- HBO
- HBO/WO
- Anders:

<< Vorige pagina

Volgende pagina >>

Vraag 6
Vindt u het gemakkelijk TV-/videoprogramma's te vinden die aansluiten bij uw interesses?
(nvt="niet van toepassing")

	Soms moeilijk,		
	Heel moeilijk	Moelijk	Gemakkelijk
TV-programma's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gehuurde video's/dvd's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Video op het internet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Films in de bioscoop	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 7
Hoe groot is uw collectie TV-/video-beelden?

	Tussen			
	Nul (of minder collectie)	10 of minder	Tussen 50 en 100	Tussen 100 en 250
Zeef opgenomen TV-programma's (aantal)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen video's/dvd's (aantal)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Van het internet gedownloade video's (aantal)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen homevideo's (in uren)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 8
Vindt u het gemakkelijk om in uw eigen collectie programma's of beelden terug te vinden?
(nvt="niet van toepassing")

	Soms moeilijk, soms		
	Heel moeilijk	Moelijk	Gemakkelijk
Zeef opgenomen TV-programma's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen video's/dvd's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Van het internet gedownloade video's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Eigen homevideo's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<< Vorige pagina

Volgende pagina >>



<< Vorige pagina

Interesse voor TV-/videoprogramma's

U heeft interesse voor een type programma als u wel eens een programma van dat type met aandacht bekijkt.

Er volgt nu een lijst met 80 verschillende typen programma's.

Vraag 9

Kruis aan welke van de volgende typen programma's uw interesse hebben.

- NIEUWS (bv. NOS Journaal, RTL Nieuws)
- ACTUALITEITEN (bv. NOVA, Netwerk, Editie NL, Twee Vandaag)
- "LUCHTIGE" ACTUALITEITEN (bv. Barend & van Dorp, Hart van Nederland, VARA laat, Man bijt hond)
- WEERBERICHT
- LIVE REGISTRATIE van een gebeurtenis (bv. uitvaart prins Bernhard, vuurkramp Enschede, nasleep moord Theo van Gogh)
- PRAATPROGRAMMA, DISCUSSIE (bv. B&W, Het lagerthuis, Rondom tien, Wat zou jij doen?)
- POLITIEK (bv. Den Haag Vandaag, Buitenhof, documentaires over politiek)
- OUDE ARCHIEFBEELDEN (bv. Polygoon-journaals)
- DOCUMENTAIRE (bv. Zembia, NCRV Dokument, Tegenlicht, VPRO's import)
- ECONOMIE & ONDERNEMERSCHAP (bv. RTL-Z, 100% ondernemer, Nederland in bedrijf, De werkplek)
- JUSTITIE, RECHT & ORDE (bv. Opsporing gezocht, De rijdende rechter, Peter R. de Vries, Misdaad en veiligheid)
- MENSEN ZOEKEN (bv. Vermist, Spoorloos)
- INFORMATIEVE JEUGDPROGRAMMA'S (bv. Klokhuis, Jeugdjournaal)
- LOKALE INFORMATIE (bv. lokaal nieuws, lokale actualiteiten, lokaal weerbericht, lokale politiek & economie, lokale documentaires)
- LOKALE ARCHIEFBEELDEN (bv. oud materiaal van de lokale omroep, amateurbeelden)

- NATUUR & MILIEU (bv. programma's op National Geographic Channel & Animal Planet, documentaires over natuur & dierenleven)
- WETENSCHAP & TECHNOLOGIE (bv. Nieuwslicht, Noordlicht, programma's op Discovery Channel, documentaires over wetenschap)
- GESCHIEDENIS (bv. Andere tijden, Polygoonfilms, de Grote Geschiedenis Quiz, documentaires over geschiedenis)
- DOCUSOAP (bv. Ingang Oost, 112 Weekend, Airport, De bevalling, Bieren op spreekuur, Missen)
- KUNST & CULTUUR (bv. R.A.M., Mijch versus Goderie, Het uur van de wolf, AVRO close-up, Museumsgasten, Tussen kunst en kitsch)
- RELIGIE (bv. Kruispunt, Brieven aan God, Studio RKK, Soeterbeek, Eucharistieviering)
- SHOWBUSINESS/ROYALTY (bv. RTL-Boulevard, Shownieuws, Rijk en beroemd)
- GEZONDHEID EN UITERLIJK (bv. Nederland in beweging, Make me beautiful, Extreme makeover, Vinger aan de pols)
- HUIS, TUIN & WONEN (bv. TV MajaJaar, Eigen huis en tuin, In Holland staat een huis, Klussen met kijkers, Woonbrigade)
- ETEN & DRINKEN (bv. Life & cooking, Jamie Oliver's twist, Gordon Ramsey: oorlog in de keuken)
- REIZEN (bv. Yorn travel, Bestemming Nederland, Rail away, films/documentaires over vakantiebestemmingen)
- AUTO'S, MOTOREN & RUGEDRAG (bv. Blik op de weg, Top Gear, Wegrijdsbruikers)
- CONSUMENTEN (bv. Kassa, Radar)
- NEDERLANDSE SOAP (bv. Goede tijden slechte tijden, Onderweg naar morgen, Het glazen huis)
- BUITENLANDSE SOAP (bv. The bold and the beautiful, As the world turns)
- NEDERLANDSE COMEDYSERIE (bv. All stars, Flodder, Toen was geluk heel gewoon, Oppassen!!!)
- BUITENLANDSE COMEDYSERIE (bv. Friends, Absolutely fabulous, The Cosby show, The office, Sex and the city)
- NEDERLANDSE POLITIESERIE (bv. Baantjer, Spangen, Russen, Dok 12)
- BUITENLANDSE POLITIESERIE (bv. A touch of Frost, Hill Street Blues, Tatort, Inspector Dajiglish)
- SCIENCE FICTION & HORROR SERIES (bv. Buffy the vampire slayer, The X-files, Sliders)
- OVERIGE NEDERLANDSE SERIES (bv. Dunya en Desie, Rozengeur en wodka lime)
- OVERIGE BUITENLANDSE SERIES (bv. ER, LA Law, Band of Brothers)

- JEUGDSERIES (bv. Zoop, De Daitons, diverse programma's op Z@ppelin, Fox Kids, Nickelodeon)
- TEKENFILMS/ANIMATIES (bv. Spongebob Squarepants, Tom & Jerry, The Simpsons, Ducktales)
- EROTIEK/SEKS (bv. Playboy-TV programma's, seksfilms op Canal+)

Vraag 10

[vervolg van de vorige vraag] Kruis aan welke van de volgende typen programma's uw interesse hebben.

- SPEELFILMS: COMEDY
- SPEELFILMS: DRAMA
- SPEELFILMS: ROMANTIEK
- SPEELFILMS: ACTIE & AVONTUUR (ook: oorlog, western, martial arts)
- SPEELFILMS: SCIENCE FICTION, FANTASY & HORROR
- SPEELFILMS: MUZIEK (ook: musical)
- SPEELFILMS: FAMILIE (inclusief animatiespeelfilms)
- SPEELFILMS: MISDAAD/THRILLER (ook: mysterie)
- SPEELFILMS: FILMHUJS (= geen populaire publieksfilms, maar vaak internationale, culturele films met kwaliteit als norm)
- KORTE SPEELFILMS (vaak te zien op internet, bv. cinema.nl, atomfilms.com, ifilm.com)
- TV-films, gebaseerd op "waargebeurde verhalen"
- SPORTREPORTAGE (bv. Nederlands elftal, WK schaatsen, Olympische spelen)
- SPORTINFORMATIE (bv. Studio Sport, Sportjournaal, Sport aan tafel, Voetbal Inside, Holland sport)
- POPMUZIEK (bv. Top of the pops, clips & concerten op MTV, TMF, The Box)
- SERIEUZE MUZIEK & BALLET (bv. klassieke concerten, ballet & dans)
- LICHTHE MUZIEK (bv. concert Marco Borsato, Frans Bauer, André Rieu, schlagers, levensliederen, registraties van musicals)
- OVERIGE MUZIEK (bv. jazz, country & western, wereldmuziek)
- INFORMATIE OVER MUZIEK (bv. Wintertijd, Popquiz a gogo, Vrije geluiden: interviews, documentaires, popnieuws op MTV, TMF)
- TALKSHOW NEDERLANDS (bv. TV Show, Jensen!, PaPaul, Kopspijkers, Villa Felderho)

- TALKSHOW BUITENLANDS (bv. The Oprah Winfrey Show, Dr. Phil)
- CABARET/SATIRE (bv. Koefnoen, Dit was het nieuws, Andermans veren, voorstelling Hans Teeuwen, Najib Amhali, Brigitte Kaandorp etc.)
- KENNISQUIZ (bv. Lingo, Get the picture, Lotto weekendend miljoeners, Postcodecloterij 1 tegen 100)
- SPELSHOW (bv. Ter land ter zee en in de lucht, Fear factor, Idols, De lekkerste)
- ENOTIE-TV (bv. All you need is love, Spoorloos, Surpriseshow, Love Letters)
- REAL-LIFE SOAP (bv. Big Brother, Expeditie Robinson, Jouw Vrouw Mijn Vrouw, The Osbournes, De Bauers)
- HOMEVIDEO'S & BEELDEN VAN BEVEILIGINGSCAMERAS (bv. De leukste thuis, World's funniest, Maximum Exposure)
- "POP"-PROGRAMMAS (bv. Bananasplit, Mijn vieze, vette, vervelende verloofde, Punk'd)
- REALITY-PROGRAMMAS MET BUITENSPORIG GEDRAG (bv. Jackass, Wildboyz)
- JEUGDAMUSEMENT-/INFORMATIE (bv. Kinderen voor kinderen, Willem Wever, Museumbende, Buya)
- LOKALE SPORT-, MUZIEK- EN AMUSEMENTSPROGRAMMAS
- LIVE WEBCAM-BEELDEN VAN PUBLIEKE LOKATIES (bv. landingsbaan Schiphol, de pier in Scheveningen, Julianaplein Amersfoort, Stadhuysplein Enschede)
- LIVE WEBCAM-BEELDEN VAN PRIVÉ LOKATIES (bv. mensen thuis of op het werk, Snowworld in Zoetermeer)
- VIDEOBOODSCHAPPEN-/MEDEDELINGEN
- TRAILERS (voor filmpjes van bioscoopfilms)
- "THE MAKING OF" filmpjes
- DEMONSTRATIE-, INSTRUCTIE-, OF VOORLICHTINGSVIDEO'S
- PRESENTATIEVIDEO'S VAN BEDRIJVEN, INSTELLINGEN OF DIENSTEN
- RECLAMES (bv. commercials STER/IP)
- TELESHOPPEN & INFOMERCIAL (bv. Tell sell)
- VIDEOKUNST (bv. getoond op kunstkanalen, in musea)


(eventueel) nog een ander type programma waar u interesse voor heeft:

<< Vorige pagina

Volgende pagina >>

(eventueel) andere redenen:

<< Vorige pagina



<< Vorige pagina

Televisiezenders

Vraag 13
Naar welke TV-zenders kijkt u regelmatig? (regelmatig = minimaal 1x per maand 3 minuten of langer)

Nederlandse publieke zenders (Nederland 1,2 & 3)

Nederlandse commerciële zenders (RTL4 & 5, Net5, SBS6, Yorin, Veronica)

Regionale zenders (bv. TV Oost, Omrop Fryslân, Omroep Brabant)

Kinderzenders (bv. Fox Kids, Nickelodeon, Cartoon Network)

Populaire muziekzenders (bv. MTV, TMF, The Box)

Kunst- en cultuurzenders (bv. Arte, Mezzo, Fashion TV)

Sportzenders (bv. Eurosport, Extreme sports, ESPN Classic Sports)

Nieuwszenders (bv. CNN, BBC World, Al Jazeera)

Kenniszenders (bv. Discovery Channel, National Geographic, Animal Planet)

Filmzenders (Cineno, MGM Movie Channel, Turner Classic Movies)

Erotiekzenders (bv. Playboy TV, Adult Channel)

Belgische zenders (bv. VRT TV1, KETNET/Canvas)

Duitse zenders (bv. ARD, ZDF)

Britse zenders (bv. BBC1 & 2)

Overige Europese zenders (bv. TV5, TVE, Rai Uno)

Zenders van buiten Europa (bv. TRT int., RTM, Zee TV, TVBS-E)

(eventueel) andere zenders waar u regelmatig naar kijkt:

(eventueel) andere zenders waar u regelmatig het programmaoverzicht van bekijkt:

Vraag 14
Naar hoeveel TV-zenders in totaal kijkt u regelmatig? (regelmatig = minimaal 1x per maand 3 minuten of langer)

- minder dan 3 ongeveer 12 ongeveer 30
 ongeveer 3 ongeveer 15 ongeveer 35
 ongeveer 6 ongeveer 20 ongeveer 40
 ongeveer 9 ongeveer 25 meer dan 40

Programmaoverzichten

Programmaoverzichten zijn in diverse vormen te vinden:
 - op papier (bv. omroepgidsen, kranten, tijdschriften)
 - op het internet (bv. tvguides.nl); soms alleen voor een genre (bv. filmsoptv.nl)
 - op de televisie zelf (bv. teletekst, elektronische programmagids, mozaïekkanaal)

Vraag 15
Van welke TV-zenders bekijkt u regelmatig het programmaoverzicht? (regelmatig = minimaal 3x per maand)

- Nederlandse publieke zenders (Nederland 1,2 & 3)
 Nederlandse commerciële zenders (RTL4 & 5, Net5, SBS6, Yorin, Veronica)
 Regionale zenders (bv. TV Oost, Omrop Fryslân, Omroep Brabant)
 Kinderzenders (bv. Fox Kids, Nickelodeon, Cartoon Network)
 Populaire muziekzenders (bv. MTV, TMF, The Box)
 Kunst- en cultuurzenders (bv. Arte, Mezzo, Fashion TV)
 Sportzenders (bv. Eurosport, Extreme sports, ESPN Classic Sports)
 Nieuwszenders (bv. CNN, BBC World, Al Jazeera)
 Kenniszenders (bv. Discovery Channel, National Geographic, Animal Planet)
 Filmzenders (Cinova, MGM Movie Channel, Turner Classic Movies)
 Erotiekzenders (bv. Playboy TV, Adult Channel)
 Belgische zenders (bv. VRT TV1, KETNET/Canvas)
 Duitse zenders (bv. ARD, ZDF)
 Britse zenders (bv. BBC1 & 2)
 Overige Europese zenders (bv. TV5, TVE, Rai Uno)
 Zenders van buiten Europa (bv. TRT int., RTM, Zee TV, TVBS-E)

Vraag 16
Van hoeveel TV-zenders in totaal bekijkt u regelmatig het programmaoverzicht? (regelmatig = minimaal 3x per maand)

- minder dan 3 ongeveer 12 ongeveer 30
 ongeveer 3 ongeveer 15 ongeveer 35
 ongeveer 6 ongeveer 20 ongeveer 40
 ongeveer 9 ongeveer 25 meer dan 40

Televisiezenders instellen

Bij het instellen van televisiezenders wordt bepaald onder welke knoppen van de afstandsbediening de verschillende zenders komen te zitten.

Vraag 17
Wie heeft bij u bepaald hoe de televisiezenders zijn ingesteld (op de televisie waar u het meest gebruik van maakt)?

- Ikzelf (eventueel in samenspraak met anderen)
 Iemand anders
 Is automatisch gebeurd

Vraag 18
Heeft u ooit de instelling van de zenders op uw TV veranderd om gemakkelijker zenders te kunnen vinden? (dus niet omdat de frequenties veranderd waren of omdat er nieuwe zenders bijkwamen, maar echt vanwege het gemak)

- Ja (zelf gedaan of door iemand anders laten doen)
 Nee

Vraag 19
Bent u tevreden met de manier waarop de zenders op uw TV zijn ingesteld?

- Ja, ik vind het een handige indeling
 De indeling is niet echt heel handig, maar ik ben er aan gewend
 Nee, ik moet regelmatig nadenken over onder welke knop een zender zit

<< Vorige pagina

Volgende pagina >>



Kenniswijk

<< Vorige pagina

Programmaoverzichten

Vraag 20
Welke bronnen gebruikt u regelmatig om gegevens te krijgen over televisieprogramma's? (regelmatig = vaker dan 1x per maand)

- een (papieren) ontroepgids
- een krant
- een elektronische programmagids (via satelliet of kabel)
- informatie over programma's in tijdschriften
- teletekst op de afzonderlijke zenders
- het mozaiek-kanaal (met beelden van de diverse zenders)
- websites over films, series etc.
- vrienden, kennissen, collega's, familie etc.

(eventueel) andere bronnen:

Vraag 21
Welke informatie vindt u MEESTAL nuttig om te krijgen over een programma?

- de naam van het programma
- waar het programma over gaat
- het tijdstip van de uitzending
- de duur van de uitzending
- het genre (of het thema)
- de taal die gesproken wordt
- in welke land het gemaakt is
- een toelichting of recensie
- of het de selectie van de dag is (bv. de speelfilm van de dag)
- een foto of illustratie van het programma
- een preview (voorfilmpje)
- het beeldformaat (bv. breedbeeld)
- de beeldkwaliteit (bv. aantal beeldpunten)
- of het een zwart-wit uitzending is

- in welk jaar het gemaakt is
- wie er in voorkomen (bv. acteurs, gasten, teams)
- door wie het gemaakt of bedacht is
- de zender of de omroep die het programma uitzendt
- of het live wordt uitgezonden
- of het programma eenmalig is of uit meerdere afleveringen bestaat
- de geluidskwaliteit (bv. Dolby Surround)
- of het Nederlands ondertiteld is
- of het ondertiteld is via Teletekst
- of het nagesynchroniseerd is (i.p.v. ondertiteld)
- wat de kijkwijzer zegt over voor welke leeftijd het geschikt is
- wat de kijkwijzer zegt over of het geweld, seks, drugs/alcohol, angst, discriminatie, of grof taalgebruik bevat

(eventueel) andere informatie die u nuttig zou vinden:

Vraag 22
Stel, u raadpleegt een programmaoverzicht op zoek naar een programma dat uw interesse heeft. Hoeveel tijd bent u gemiddeld (bij benadering) kwijt voor u een keuze hebt gemaakt voor een programma?

 minuten

Vraag 23
Lukt het u doorgaans een interessant programma te vinden met behulp van een programmaoverzicht?

- nooit
- zelden
- soms
- vaak
- altijd



<< Vorige pagina

Thema's

In programmaoverzichten worden soms alle programma's van een bepaald thema bij elkaar neergezet, bv. alle muziekprogramma's, alle sportprogramma's, etc. Het bekendste voorbeeld hiervan is het overzicht van alle films die op een dag worden uitgezonden.

Vraag 25

Maakt u bij het zoeken naar programma's gebruik van dit type informatie?

het filmoverzicht	<input type="radio"/>	nooit	<input type="radio"/>	zelden	<input type="radio"/>	soms	<input type="radio"/>	vaak	<input type="radio"/>	altijd	<input type="radio"/>
andere overzichten op basis van thema/ genre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 26

Hee nuttig vindt u het als in een programmaoverzicht programma's van hetzelfde thema of genre in een apart overzicht zijn geplaatst? (1="totaal niet nuttig"; 10="zeer nuttig")

1 2 3 4 5 6 7 8 9 10

Themazenders

Veel zenders hebben een gemengd aanbod, en laten achterelkaar verschillende soorten programma's zien (bv. nieuws, quiz, sport, film).

Op themazenders worden - meestal 24 uur per dag - allerlei programma's rondom hetzelfde thema uitgezonden.

Er bestaan diverse themazenders, zoals bijvoorbeeld voor:

- sport (bv. Eurosport)
- nieuws (bv. CNN, Journaal24)
- film (bv. CineNova)
- muziek (bv. MTV)
- kunst (bv. Arte)
- etc.

Vraag 24

Als u problemen heeft een keuze te maken met behulp van een programmaoverzicht, waar ligt dat dan aan? (meerdere antwoorden mogelijk)

- Er zijn vaak te weinig interessante programma's op televisie
- Er zijn vaak teveel interessante programma's: ik kan moeilijk kiezen
- Er zijn zoveel zenders en programma's, dat ik vaak door de bomen het bos niet meer zie
- Een programmaoverzicht geeft mij vaak niet genoeg informatie om vast te stellen of een programma mij zal interesseren
- Ik heb er vaak het geduld niet voor: ik ga liever zappen

(eventueel) een andere reden:

<< Vorige pagina

Volgende pagina >>

Vraag 27
Hoeverveel waardering heeft u voor het volgende? (geef een rapportcijfer)

	1	2	3	4	5	6	7	8	9	10
zenders met een gemengd aanbod	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
thema-zenders	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 28
Waar gaat uw voorkeur naar uit?

- algemene themazenders: bv. Sport, Informatie, Fictie
- wat specifiekere themazenders: bv. Voetbal, Vrije tijd, Comedyseries
- nog specifiekere themazenders: bv. Nederlands voetbal, Huis en tuin, Amerikaanse sitcoms
- heel specifieke themazenders: bv. PSV, Kamerinrichting, Friends
- geen van bovengenoemde

Kennislijst

Zoeken (en vinden) op de televisie

Vraag 29
In hoeverre zijn de volgende stellingen op u van toepassing?

	nooit	zelden	soms	vaak	altijd
Voor ik de TV aanzet, kijk ik in een programmaoverzicht of er een programma is dat ik wil kijken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik zet de TV aan, en zap net zo lang langs de kanalen tot ik iets interessants tegenkom	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik zet de TV aan om te kijken naar een programma waar ik altijd naar kijk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik zet de TV aan om te kijken naar een programma maar ik over getipt ben (bv. door een aankondiging die eerder op TV is geweest, door een vriend)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 30
Wat vindt u van de volgende activiteiten?

	erg vervelend	vervelend	niet leuk, niet vervelend	erg leuk
Zoeken in programmaoverzichten naar interessante programma's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zappen langs TV-zenders om interessante programma's te vinden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Kenniswijk

<< Vorige pagina

Video's zoeken

Op het internet is het mogelijk met behulp van een zoekmachine te zoeken naar video's.

Vraag 34
De volgende illustratie toont (een deel van de) zoekresultaten op de zoekterm "park" bij de AltaVista videozoekmachine.

Linkin Park ... at.mpeg
352,240 • 178 sec
[Meer info](#)

linlin_park...HDIP.mpeg
352,240 • 195 sec
[Meer info](#)

worldcup_park_700k.vmv
320,240 • 193 sec
[Meer info](#)

tajjack_park.mpg
352,240 • 64 sec
[Meer info](#)

Jurassic park ... 80k.mov
480,240 • 142 sec
[Meer info](#)

park_verspreiding.vmv
320,240 • 100 sec
[Meer info](#)

20020627_kus...fl.mpeg
352,240 • 56 sec
[Meer info](#)

2002.1.7_Gel...fl.mv
320,240 • 93 sec
[Meer info](#)

guedard_park_m180.mov
480,240 • 111 sec
[Meer info](#)

south_park.mov
320,240 • 77 sec
[Meer info](#)

Day at the Park.vmv
320,240 • 82 sec
[Meer info](#)

South Park T...d.mpeg
240,150 • 312 sec
[Meer info](#)

Vraag 31
In hoeverre zijn de volgende stellingen op u van toepassing?

	nooit zelden soms vaak altijd	
Als ik eenmaal naar een programma kijk, kijk ik het ook helemaal tot het einde uit	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Als ik de afstandsbediening heb, houd ik rekening met wat mijn kamergenoten willen zien	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Ik volg een aantal programma's tegelijkertijd	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Als er aanstootgevende of onaangename beelden zijn, zap ik snel weg	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Ik zap veel en vind het niet erg om steeds maar stukjes van programma's te zien	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	

Vraag 32
Welke reden is voor u het belangrijkste om vóór het einde van een programma te stoppen met kijken?

Ik vind het programma niet zo interessant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik denk dat andere programma's interessanter zijn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vraag 33
Hoe vaak doet u de volgende zaken tijdens het TV-kijken om nog een programma te vinden dat u interesseert?

	nooit zelden soms vaak altijd	
zappen langs de kanalen	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
een programmaoverzicht raadplegen	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
het mozaïek-kanaal (met beelden van de diverse zenders) bekijken	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
een ander vragen of die wat weet	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	

<< Vorige pagina Volgende pagina >>

nooit zelden soms vaak heel vaak
 Zoekt u wel eens met behulp van een zoekmachine naar video's?

Vraag 35

Stel, u heeft een zoekactie ondernomen, en als resultaat meer dan 1000 video's gekregen. Welke gegevens zijn voor u nuttig om te helpen bepalen of u een video interessant vindt? (1="totaal niet nuttig"; 10="zeer nuttig")

	1	2	3	4	5	6	7	8	9	10
de titel van de video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een korte beschrijving van de inhoud (bv. mensen, gebeurtenissen, onderwerpen, het verhaal etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een korte beschrijving over de video (bv. genre, jaar, maker, land etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
technische gegevens over de video (duur, aantal beeldpunten, frame-frequentie, bestandstype, geluidskwaliteit etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een specifiek stilstaand beeld uit de video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10 stilstaande beelden uit de video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een korte videosamenvatting (een soort "preview" of voorfilmpje)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegevens over wanneer en waar het programma wordt uitgezonden, of waar de video opgeslagen ligt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gegevens over waar en op welke manier de zoekterm in de video zit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Het is tegenwoordig mogelijk om met een zoekmachine te zoeken naar TV/video-programma's. U geeft dan één of meerdere woorden (zoektermen) op aan de zoekmachine, en alle programma's die die woorden bevatten worden getoond.

Vraag 36

Er zijn diverse manieren om een programmaoverzicht te tonen. Geef een cijfer tussen 1 en 10 voor de hieronder genoemde manieren, waarbij 1= totaal niet nuttig/handig, en 10= zeer nuttig/handig.


	1	2	3	4	5	6	7	8	9	10
een overzicht per zender	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een overzicht per tijdsperiode (bv. alle programma's tussen 8 en 10 uur)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een overzicht per thema	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
een overzicht als resultaat van een zoekactie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Stelt u zich voor, u bent naar TV/video aan het kijken en u denkt opeens: "hier zou ik meer van willen zien of weten".

U wilt bijvoorbeeld:

- meer van een bepaalde persoon zien
- meer van een bepaalde persoon weten
- meer over een bepaald onderwerp horen
- meer beelden uit dezelfde periode zien
- etc.

U bent als het ware geïnspireerd geraakt door wat u ziet, en wilt iets vergelijkbaars zien of meer achtergrondinformatie hebben.



<< Vorige pagina

Afsluiting

Vraag 42
Vindt u het goed als wij u (eventueel) nogmaals benaderen om op één of meerdere vragen nog even door te vragen?

Ja
 Nee

Vraag 43
Als u in aanmerking wilt komen voor één van de 5 hele Staatsloten L.w.v. € 12,50 of één van de 20 1/5 Staatsloten L.w.v. € 2,50, vul dan hieronder uw e-mailadres in.

Voor de duidelijkheid: als u heeft aangegeven niet verder meer mee te willen werken, dan wordt uw e-mailadres niet voor andere doeleinden gebruikt. U wordt dan nog wel steeds mee naar een heel of 1/5 Staatslot!

Dit waren alle vragen. Hartelijk dank voor uw medewerking!

<< Vorige pagina

Verzenden

(eventueel) andere onderdelen waar u naar zou willen kunnen springen:

<< Vorige pagina

Volgende pagina >>



Mede namens Kenniswaik BV hartelijk dank voor uw medewerking aan deze enquête.

Met vriendelijke groet,
Ynze van Houten
Telematica Instituut
Sluit venster

Appendix **B**

Fabchannel survey: questions

This appendix shows the questions of the Fabchannel study described in Chapter 2. For further information, see that chapter.



Language: English

Thank you for participating in this survey!

Completing it will only take you a couple of minutes.

When answering questions with a 5 or 7-point scale, please try and differentiate your answers relating them to the other options. That way we will be able to recognize which options you prefer over other options

First there are a couple of things we would like to know about you:

Question 1 *

Are you

- female, or
- male

Question 2 *

To which age category do you belong?

- < 15 years
- 15-20 years
- 21-25 years
- 26-30 years
- 31-35 years
- 36-40 years
- 41-45 years
- > 45 years

In the following you will be presented with 10 additional questions regarding your interest in music, your first encounter with and use of the Fabchannel website, and several options which we feel might enhance your online concert experience.

Question 3 *

From the music types listed below, which types of music do you usually listen to?

- | | | |
|--|-----------------------|-----------------------|
| Energetic & rhythmic (hiphop, soul, dance, r&b, ska) | never | always |
| Intense & rebellious (rock, metal, punk, reggae, alternative) | <input type="radio"/> | <input type="radio"/> |
| Upbeat & conventional (pop, country, religious, sound tracks) | <input type="radio"/> | <input type="radio"/> |
| Reflective & complex (jazz, trip-hop, singer/songwriter, blues, folk, classical) | <input type="radio"/> | <input type="radio"/> |

other:

Next page >>



<< Previous page

Question 4 *

When did you first visit the Fabchannel website?

- this is my first visit
- 1 to 4 weeks ago
- 1 to 6 months ago
- 6 months to 1 year ago
- 1 to 2 years ago
- more than 2 years ago

Question 5 *

How did you learn about the existence of Fabchannel?

- Via the newspapers
- Via TV
- Via a band website
- Via a community site
- Via a news site
- Via Google
- From hearsay
- Other:

Question 6 *

How often do you visit the Fabchannel website?

- Daily
- Once or a few times per week
- Once or a few times per month
- Once or a few times per year
- Less than yearly

Question 7 *

From the options listed below, in which ways do you or do you not use the Fabchannel website? I use the Fabchannel website to...

- | | never | always |
|--|-----------------------|-----------------------|
| discover new music | <input type="radio"/> | <input type="radio"/> |
| find a specific concert in the archive to watch | <input type="radio"/> | <input type="radio"/> |
| watch a concert I saw earlier in Paradise or the Melkweg | <input type="radio"/> | <input type="radio"/> |
| to watch a concert I saw earlier live in a concert hall elsewhere in the world | <input type="radio"/> | <input type="radio"/> |
| watch live webcasts | <input type="radio"/> | <input type="radio"/> |
| chat with other people during online live concerts | <input type="radio"/> | <input type="radio"/> |
| compile my own playlist from the Fabchannel archive | <input type="radio"/> | <input type="radio"/> |

something else:

<< Previous page

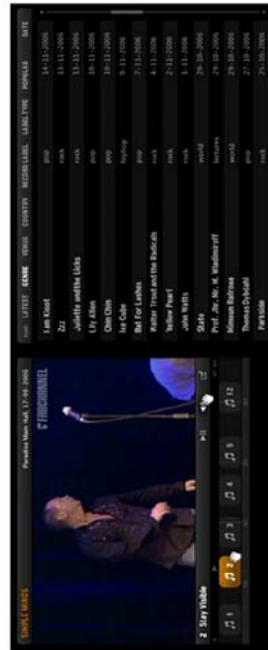
Next page >>

FABCHANNEL

<< Previous page

Sorting concerts

Currently Fabchannel has over 600 concerts available for you in the archive. These concerts can be sorted by name and genre. Of course, to assist you in finding exactly what you are looking for, there are many other ways of sorting the concerts in our archive.



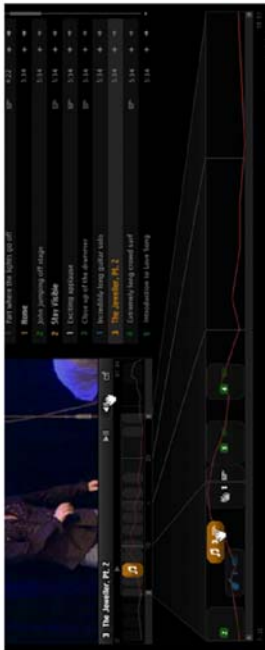
Screenshot taken from our future FabPlayer!

Question 8 *
From the options listed below, how useful do you think these options are for sorting concerts? I would like to sort the concerts by...

	totally unuseful	very useful
band name	<input type="radio"/>	<input type="radio"/>
genre	<input type="radio"/>	<input type="radio"/>
concert date	<input type="radio"/>	<input type="radio"/>
venue/location	<input type="radio"/>	<input type="radio"/>
lyric language	<input type="radio"/>	<input type="radio"/>
country of origin	<input type="radio"/>	<input type="radio"/>
name label	<input type="radio"/>	<input type="radio"/>
type of label (indie, major)	<input type="radio"/>	<input type="radio"/>
popularity (number of views)	<input type="radio"/>	<input type="radio"/>
latest additions to the archive	<input type="radio"/>	<input type="radio"/>
Fabchannel favorites	<input type="radio"/>	<input type="radio"/>
preference of an "expert" (musician, reviewer)	<input type="radio"/>	<input type="radio"/>
personal preference of friends	<input type="radio"/>	<input type="radio"/>
online users (select a concert on the basis of what others are watching at the moment)	<input type="radio"/>	<input type="radio"/>

What other ways of sorting the archive would be useful to you?:

Concert segments
 In the current FabPlayer you can select and play a concert. While playing a concert you can skip through it by clicking anywhere on the timeline. We could make this interaction "richer" by allowing you to directly access smaller parts in a concert. For example, you would be able to skip through all instrument solo's or applauds in a concert.



screenshot taken from our future FabPlayer

Question 9 *
 From the options listed below, to what extent do you feel these concert parts would be useful to skip through directly?

	totally unuseful					very useful				
Songs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Choruses of songs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instrument solos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoken parts between songs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Close-up shots	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medium shots	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wide shots	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Applause	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Events in the audience (e.g. stage diving)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What other parts of a concert would be of interest to you to select?:

<< Previous page

Next page >>

FABCHANNEL

<< Previous page

Searching for parts with certain characteristics

As addressed earlier, we can attach different labels such as genre, venue, concert date etc. to a whole concert. We are also able to label *specific parts* of a concert, like songs, instrumental solo's, or spoken parts.

These labels or characteristics can be used to search for similar parts in one concert or in other concerts. For example, you could play all songs that have been released as a single, or all solos performed with an acoustic guitar. Another possibility could be to look for songs with similar tempo (bpm).



Screenshot taken from our future FabPlayer!

Question 10 *

From the options listed below, which characteristics do you feel would be useful to specify your search?

	totally unuseful					very useful				
Songs that are part of a cd tracklist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Songs that are released as a single	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Songs that cover another song	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Songs written by a certain songwriter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Songs sung by a certain band member	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solos performed with a certain instrument (guitar, saxophone, drums, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solos performed with a certain type of instrument (e.g. guitar type)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solos performed by a certain band member	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Solos of a certain duration	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Parts of a certain tempo (bpm)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Parts of a certain excitement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Parts with a certain popularity (the frequency with which parts are watched by others)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What other characteristics of concert parts do you feel are useful to further search the archive for?:



<< Previous page

Question 12 *

Then finally, would you be interested in giving your opinion about Fabchannel or become a Fabchannel beta tester in the future?

- Yes, I would like to give my opinion about Fabchannel more often in the future. (please fill in your email address below)
- Yes, I would like to become a Fabchannel beta tester. (please fill in your email address below)
- No, rather not.

email address:

<< Previous page

Submit

Question 11 *

Imagine for instance watching one of the concert parts addressed above. From the options listed below, to what extent do you feel these would be useful to quickly select and watch other similar concert parts? Selecting similar concert parts...

	totally unuseful	very useful
within a specific concert	<input type="radio"/>	<input type="radio"/>
within specific parts of the archive (band, genre, singing language, etc.)	<input type="radio"/>	<input type="radio"/>
throughout the whole concert archive	<input type="radio"/>	<input type="radio"/>

<< Previous page

Next page >>



Thank you very much for participating in this survey. It will certainly help us to create a better Fabchannel website.

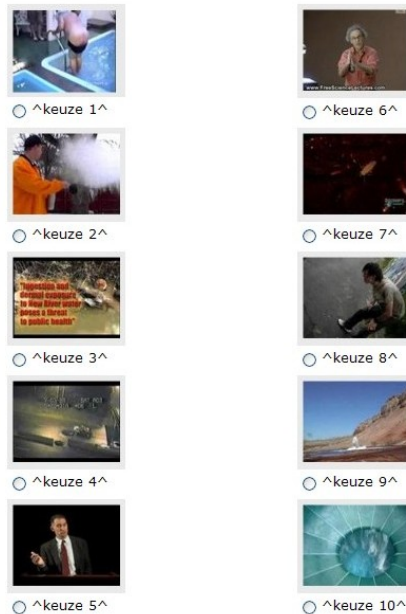
Best regards,

Fabchannel

[Close window](#)

Scent experiment: examples of result sets











This appendix shows examples of the different types of result sets as used in the scent experiment described in Chapter 3. For further information, see that chapter.



Example of a result set with the “frame” scent carrier

The Link Between Meat Eating and Climate Change	Contaminated Water for US Troops
<input type="radio"/> ^keuze 1^	<input type="radio"/> ^keuze 6^
Brown Bears of Mikvik River, Alaska	Jeep Underwater
<input type="radio"/> ^keuze 2^	<input type="radio"/> ^keuze 7^
Northern Baja Ocean Pollution	can you dig it
<input type="radio"/> ^keuze 3^	<input type="radio"/> ^keuze 8^
How to pee off your boat without falling overboard	Recycling Water
<input type="radio"/> ^keuze 4^	<input type="radio"/> ^keuze 9^
Water Pollution: The Conspiracy	Can you talk to water?
<input type="radio"/> ^keuze 5^	<input type="radio"/> ^keuze 10^

Example of a result set with the “title” scent carrier

 <p>Science project [water pollution]</p> <input type="radio"/> ^keuze 1^	 <p>Computer-generated animation of water</p> <input type="radio"/> ^keuze 6^
 <p>Bad sewer pipes across America</p> <input type="radio"/> ^keuze 2^	 <p>Super Wave</p> <input type="radio"/> ^keuze 7^
 <p>America's Most Endangered Rivers of 2007</p> <input type="radio"/> ^keuze 3^	 <p>Congress Acts to Protect Clean Water</p> <input type="radio"/> ^keuze 8^
 <p>Funny water accidents</p> <input type="radio"/> ^keuze 4^	 <p>Surfers Against Sewage - Doctor Loo</p> <input type="radio"/> ^keuze 9^
 <p>Liquid Nitrogen Into A Swimming Pool</p> <input type="radio"/> ^keuze 5^	 <p>Groundwater Contamination</p> <input type="radio"/> ^keuze 10^

Example of a result set with the “title+frame” scent carrier

A pool filled with non-newtonian fluid <i>Description:</i> They filled a pool with a mix of cornstarch and water made on a concrete mixer truck. It becomes a non-newtonian fluid. When stress is applied to the liquid it exhibits properties of a solid... <input type="radio"/> ^keuze 1^	Elway vs. the water <i>Description:</i> Elway the bulldog attacks the water <input type="radio"/> ^keuze 6^
Nasty pollution in China <i>Description:</i> Dead fishes in a lake in Wuhan, China... <input type="radio"/> ^keuze 2^	The environment and religion <i>Description:</i> If you are a person of faith listen to this... <input type="radio"/> ^keuze 7^
Flood in Grimsby! <i>Description:</i> The road between Great Coates and Healing turned into a lake 25.06.07. Watch us attempt to cross in a Vauxhall Corsa, door-high in sewage-contaminated water! <input type="radio"/> ^keuze 3^	Sewer Diver <i>Description:</i> Guy in Mexico whose job it is to scuba dive in the sewers and remove giant turds, corpses, et al. It was a free video podcast so I hope it's allowed. <input type="radio"/> ^keuze 8^
Severn Suzuki speaking at UN Earth Summit 1992 <i>Description:</i> Raised in Vancouver and Toronto, Severn Cullis-Suzuki has been camping and hiking all her life. When she was 9 she started the Environmental Children's Organization (ECO), a small group of children committed to ... <input type="radio"/> ^keuze 4^	SkyTruth: Upper Green River Valley - A View From Above <i>Description:</i> Using the latest in satellite imagery, aerial photography, and Google Earth technology, this ten minute SkyTruth video explores the environmental impacts of gas and oil drilling in the Upper Green River Valley, an ecologically... <input type="radio"/> ^keuze 9^
Water fun in Maui <i>Description:</i> Skinning, makena, big beach, little beach, kehei, Boogie Boarding, Kiting, and surfing on Maui. <input type="radio"/> ^keuze 5^	Amazing Liquid <i>Description:</i> put cornstarch, water and science together... this is what you get! <input type="radio"/> ^keuze 10^

Example of a result set with the “title+description” scent carrier

Fishy Fishing Fun <i>Tags:</i> Fishy Fishing Fun jeffery funny comedy <input type="radio"/> ^keuze 1^	the flood of horrible toxic sewage in my home. <i>Tags:</i> tim tam flood sewage yucky poo pee water everywhere blech <input type="radio"/> ^keuze 6^
The water isn't contaminated or anything... <i>Tags:</i> water seadoo <input type="radio"/> ^keuze 2^	Toriton Plus: Water Surface as Music Controller II <i>Tags:</i> sebastian tomczak little scale adelaide music sound laser water <input type="radio"/> ^keuze 7^
Environmental Health - The Role of Epidemiology <i>Tags:</i> health environment policy epidemiology TheWebVideo PolicyChannel <input type="radio"/> ^keuze 3^	Wash Your Hands <i>Tags:</i> alphacat alpha cat wash your hands hygiene germs germaphobe <input type="radio"/> ^keuze 8^
Jesus lizard <i>Tags:</i> jesus lizard water running across <input type="radio"/> ^keuze 4^	Turn on the Tap <i>Tags:</i> Turn on the Tap is a project of Samaritan's Purse Canada <input type="radio"/> ^keuze 9^
Whirlpool Sewer <i>Tags:</i> water flood tornado whirlpool pool beach diy sewer springfield ma <input type="radio"/> ^keuze 5^	Cal jumps into shity water <i>Tags:</i> funny canal jumps scream di music laugh lol tvat hurt water boat <input type="radio"/> ^keuze 10^









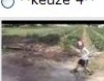

Example of a result set with the “title+tags” scent carrier

Gaza Sewage Flood ! Duration: 01:54 Category: News & Politics Added: 3 months ago ^keuze 1^	Nigerian Families Desperate for Clean Water Duration: 02:15 Category: News & Politics Added: 3 weeks ago ^keuze 6^
#8 Neuse River, North Carolina Duration: 03:27 Category: News & Politics Added: 4 months ago ^keuze 2^	Water Rocket Duration: 02:48 Category: Howto & DIY Added: 10 months ago ^keuze 7^
Gizmo Flushes Duration: 02:49 Category: Pets & Animals Added: 1 year ago ^keuze 3^	Water bed candid camera Duration: 00:30 Category: Comedy Added: 1 year ago ^keuze 8^
slow motion water balloon Duration: 00:48 Category: Howto & DIY Added: 9 months ago ^keuze 4^	Groundwater sampling Duration: 00:32 Category: People & Blogs Added: 3 weeks ago ^keuze 9^
ATV got stuck in the mud Duration: 02:08 Category: Autos & Vehicles Added: 8 months ago ^keuze 5^	drain overflowing with sewage. Welsh Water / Dwr Cymru Duration: 02:47 Category: Howto & DIY Added: 11 months ago ^keuze 10^

Example of a result set with the “title+metadata” scent carrier

Folks, we should be riding bicycles WAY more! Views: 347 Comments: 8 Favoited: 0 times 6 ratings ^keuze 1^	Roots in Sewer Line Views: 11,858 Comments: 4 Favoited: 5 times 2 ratings ^keuze 6^
Boys Swimming And Fun In Dirty Lake #1 Version With Song Views: 2,465 Comments: 11 Favoited: 7 times 9 ratings ^keuze 2^	Water contamination reaches York, Nebraska city limits Views: 1,055 Comments: 0 Favoited: 6 times 4 ratings ^keuze 7^
How Toilet Paper is Made Views: 34,172 Comments: 19 Favoited: 64 times 28 ratings ^keuze 3^	Rare prehistoric shark appears off the coast of Japan Views: 791,286 Comments: 29 Favoited: 2897 times 936 ratings ^keuze 8^
Clean Water for Haiti - about filters and health training Views: 1,170 Comments: 3 Favoited: 12 times 6 ratings ^keuze 4^	Ditch Boarding Views: 1,508 Comments: 3 Favoited: 3 times 3 ratings ^keuze 9^
698 W. Marshall Sewer Inspection Views: 116 Comments: 0 Favoited: 1 times 0 ratings ^keuze 5^	Biosand Filters For Clean Water Views: 392 Comments: 0 Favoited: 4 times 2 ratings ^keuze 10^

Example of a result set with the “title+social data” scent carrier

 <p>Duration: 09:05 Category: News & Politics Added: 2 months ago</p> <p>Views: 1,921 Comments: 12 Favorited: 4 times</p> <p>★★★★☆ 10 ratings</p> <p>Bonaire waste water</p> <p>Description: Documentary about Bonaire's waste water, made by Sean Paton and Hendrik Wuyls. Contact Sean at www.ourconfidence.com for more information</p> <p>Tags: bonaire waste water</p> <p>⊙ ^keuze 1^</p>	 <p>Duration: 02:52 Category: Comedy Added: 1 year ago</p> <p>Views: 15,824 Comments: 13 Favorited: 78 times</p> <p>★★★★☆ 31 ratings</p> <p>Fishing Trip Gone Wrong...</p> <p>Description: .</p> <p>Tags: Fishing Funny Accident</p> <p>⊙ ^keuze 6^</p>
 <p>Duration: 01:54 Category: Howto & DIY Added: 3 months ago</p> <p>Views: 17,278 Comments: 83 Favorited: 46 times</p> <p>★★★★☆ 62 ratings</p> <p>Does Drinking Water Affect Weight Loss</p> <p>Description: Will you lose more weight if you drink tons of water?</p> <p>Tags: Weight loss fitness eat obese diet anorexia exercise muscles</p> <p>⊙ ^keuze 2^</p>	 <p>Duration: 07:43 Category: People & Blogs Added: 3 weeks ago</p> <p>Views: 35 Comments: 0 Favorited: 0 times</p> <p>☆☆☆☆☆ 0 ratings</p> <p>Down the Drain: an Educational Video</p> <p>Description: This was my final project for Intro to Documentary Filmmaking. It is a light-hearted yet educational video about wastewater and the Kalamazoo water treatment plant. My goal for this piece was to deliver ordinarily mundane...</p> <p>Tags: wastewater kalamazoo Anastasia Wright</p> <p>⊙ ^keuze 7^</p>
 <p>Duration: 01:36 Category: News & Politics Added: 6 months ago</p> <p>Views: 878 Comments: 3 Favorited: 6 times</p> <p>★★★★☆ 5 ratings</p> <p>Protect and Survive - Sanitation</p> <p>Description: Another in the Protect and Survive series.</p> <p>Tags: nuclear civildence protect survive uk coldwar atomic</p> <p>⊙ ^keuze 3^</p>	 <p>Duration: 04:00 Category: Pets & Animals Added: 10 months ago</p> <p>Views: 106,768 Comments: 99 Favorited: 283 times</p> <p>★★★★☆ 139 ratings</p> <p>Orca Attack Seal with Waves</p> <p>Description: In January 2006 while visiting Antarctica, we witnessed a most unusual method for orca to dislodge a crab-eater seal from an ice floe - they made large waves to wash the seal off the relative safety of the ice. Later the...</p> <p>Tags: orca seal Antarctica waves unusual training juvenile</p> <p>⊙ ^keuze 8^</p>
 <p>Duration: 02:12 Category: News & Politics Added: 2 months ago</p> <p>Views: 233 Comments: 0 Favorited: 0 times</p> <p>★★★★☆ 4 ratings</p> <p>Waste Water Treatment - Something to Talk About</p> <p>Description: THE EUHRATES RIVER HAS BECOME A MAJOR WASTE RECEPTACLE FOR THE RESIDENTS OF FALLUJAH. THE IRAQI GOVERNMENT AND THE CORPS OF ENGINEERS HAVE A PLAN TO ...</p> <p>Tags: USACE ARMY CORPS OF ENGINEERS IRAQ RECONSTRUCTION</p> <p>⊙ ^keuze 4^</p>	 <p>Duration: 09:22 Category: Travel & Places Added: 2 months ago</p> <p>Views: 21 Comments: 0 Favorited: 0 times</p> <p>★★★★☆ 1 rating</p> <p>... Then the lake said.....</p> <p>Description: A relaxing lunch break, listening to Lake Superior talk. Very soothing.</p> <p>Tags: lake talk water nature lunch spring</p> <p>⊙ ^keuze 9^</p>
 <p>Duration: 04:08 Category: People & Blogs Added: 1 year ago</p> <p>Views: 5,770 Comments: 0 Favorited: 3 times</p> <p>★★★★☆ 5 ratings</p> <p>Mud Puddle</p> <p>Description: Watch as Michael, along with Jessica, Hannah, and Emma, enjoys a mud puddle.</p> <p>Tags: mud puddle</p> <p>⊙ ^keuze 5^</p>	 <p>Duration: 03:08 Category: Music Added: 2 months ago</p> <p>Views: 242 Comments: 2 Favorited: 2 times</p> <p>★★★★☆ 3 ratings</p> <p>Love That Dirty Water</p> <p>Description: Dirty Water by the Shandells to the Grand River, Mohawk Lake Canals, Wilkes Dam</p> <p>Tags: Mohawk Lake Canals Brantford Pollution Water Grand River</p> <p>⊙ ^keuze 10^</p>

Example of a result set with the "all" scent carrier

Scent experiment: survey versions

The following table presents an overview of all versions of the surveys used in the scent experiment described in Chapter 3. Survey versions varied in the order of the tasks (1,2,3,4 or 5) and of the scent carrier types (*All*, *Title*, *Frame*, *Title+description* [T+d], *Title+frame* [T+f], *Title+metadata* [T+m], *Title+social data* [T+s], and *Title+tags* [T+t]).

For example, a participant who completed survey version 5a started with task 1 (*general - mood improvement*) and had to make a choice from the 10 videos in result set 1, which were represented by the scent carrier *Title+frame* (see Appendix C for an example). The participant then moved on to make a choice from the 10 videos in result set 2, which were represented by the scent carrier *Title+metadata*, and so forth. After making eight choices for task 1, the participant moved on to task 2 (*general - cognitive benefit*) and again had to make eight choices from the result sets. So the eight result sets (and the related scent carriers) were the same for each of the five tasks. For further information, see Chapter 3.

Survey version	Task order	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS8
1a	1-2-3-4-5	All	Title	Frame	T+d	T+f	T+m	T+s	T+t
1b	2-1-3-5-4	All	Title	Frame	T+d	T+f	T+m	T+s	T+t
2a	1-2-3-4-5	T+t	T+s	T+m	T+f	T+d	Frame	Title	All
2b	2-1-3-5-4	T+t	T+s	T+m	T+f	T+d	Frame	Title	All
3a	1-2-3-4-5	Frame	T+d	T+f	T+m	T+s	T+t	All	Title
3b	2-1-3-5-4	Frame	T+d	T+f	T+m	T+s	T+t	All	Title
4a	1-2-3-4-5	Title	All	T+t	T+s	T+m	T+f	T+d	Frame
4b	2-1-3-5-4	Title	All	T+t	T+s	T+m	T+f	T+d	Frame
5a	1-2-3-4-5	T+f	T+m	T+s	T+t	All	Title	Frame	T+d
5b	2-1-3-5-4	T+f	T+m	T+s	T+t	All	Title	Frame	T+d
6a	1-2-3-4-5	T+d	Frame	Title	All	T+t	T+s	T+m	T+f
6b	2-1-3-5-4	T+d	Frame	Title	All	T+t	T+s	T+m	T+f
7a	1-2-3-4-5	T+s	T+t	All	Title	Frame	T+d	T+f	T+m
7b	2-1-3-5-4	T+s	T+t	All	Title	Frame	T+d	T+f	T+m
8a	1-2-3-4-5	T+m	T+f	T+d	Frame	Title	All	T+t	T+s
8b	2-1-3-5-4	T+m	T+f	T+d	Frame	Title	All	T+t	T+s

Scent experiment: additional questions

This appendix shows the questions (in Dutch) on the usefulness of types of information about videos. These were additional questions in the scent experiment described in Chapter 3. For further information, see that chapter.



<< Vorige pagina

<< Vorige pagina

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Vraag 127 *
1=totaal niet nuttig; 10=zeer nuttig

Vraag 128 *
1=totaal niet nuttig; 10=zeer nuttig

- 1 2 3 4 5 6 7 8 9 10
- een titel
- een beschrijving van wie er in voorkomen (bv. personen, groepen, teams)
- een beschrijving van waar de video zich afspeelt (lokaties)
- een beschrijving van wat er zich afspeelt (gebeurtenissen)
- een beschrijving van in welke tijdsperiode de video zich afspeelt
- een beschrijving van de muziek/geluiden die in de video voorkomen
- een beschrijving van welke onderwerpen voorkomen
- een beschrijving van welke objecten in de video voorkomen

Commentaar:

<< Vorige pagina

Volgende pagina >>

- 1 2 3 4 5 6 7 8 9 10
- een plaatje of "frame" uit de video
- meerdere frames (bv. 10) uit de video
- een preview of voorflimpje

Commentaar:

<< Vorige pagina

Volgende pagina >>



<< Vorige pagina

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Vraag 129 *
1=totaal niet nuttig; 10=zeer nuttig

- | | | | | | | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| het aantal maal dat de video is bekeken door anderen | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| het aantal commentaren gegeven door anderen | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| hoe vaak anderen de video als favoriet hebben aangemerkt | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| de inhoud van de commentaren gegeven door anderen | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| de gemiddelde beoordeling die anderen aan de video hebben gegeven | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| een recensie door een expert | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| kijkwijzerinformatie (vanaf welke leeftijd geschikt, of het geweld, seks, drugs/alcohol, angst, discriminatie, of groftaalgebruik bevat) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| wat je vrienden er van vinden | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Commentaar:

<< Vorige pagina

Volgende pagina >>



<< Vorige pagina

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Vraag 130 *
1=totaal niet nuttig; 10=zeer nuttig

- | | | | | | | | | | | |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| wie de video uitzendt of heeft geupload | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| wanneer de video wordt uitgezonden of is geupload | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| door wie de video gemaakt of bedacht is | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| wanneer de video gemaakt is | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| de taal die gesproken wordt | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| in welk land de video gemaakt is | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| van welke filmstudio de video komt | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| of de video live wordt uitgezonden | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Commentaar:

<< Vorige pagina

Volgende pagina >>



<< Vorige pagina

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Vraag 131 *
1=totaal niet nuttig; 10=zeer nuttig

de duur	1	2	3	4	5	6	7	8	9	10
of de video (Nederlands) ondertiteld is	0	0	0	0	0	0	0	0	0	0
of de video nagesynchroniseerd is (i.p.v. ondertiteld)	0	0	0	0	0	0	0	0	0	0
of de video kleur of zwart-wit is	0	0	0	0	0	0	0	0	0	0
beeldformaat en -kwaliteit (bv. aantal pixels)	0	0	0	0	0	0	0	0	0	0
geluidskwaliteit (b.v surround)	0	0	0	0	0	0	0	0	0	0

Commentaar:

[Empty text box for comment]

<< Vorige pagina

Volgende pagina >>



<< Vorige pagina

Hoe nuttig vind je de volgende stukken informatie bij het bepalen of een video voor jou interessant is of niet?

Vraag 132 *
1=totaal niet nuttig; 10=zeer nuttig

sleutelwoorden of "tags" (deze kunnen ook gebruikt worden om andere video's met dezelfde tags te vinden)	1	2	3	4	5	6	7	8	9	10
de categorie of het genre (bv. "wetenschap" of "comedy")	0	0	0	0	0	0	0	0	0	0
of de video op zichzelf staat of onderdeel is van een serie	0	0	0	0	0	0	0	0	0	0
welk doel de video dient (bv. om te amuseren, om wijzer van te worden, om een product te promoten, etc.)	0	0	0	0	0	0	0	0	0	0

<< Vorige pagina

Volgende pagina >>

Browse experiment: task order

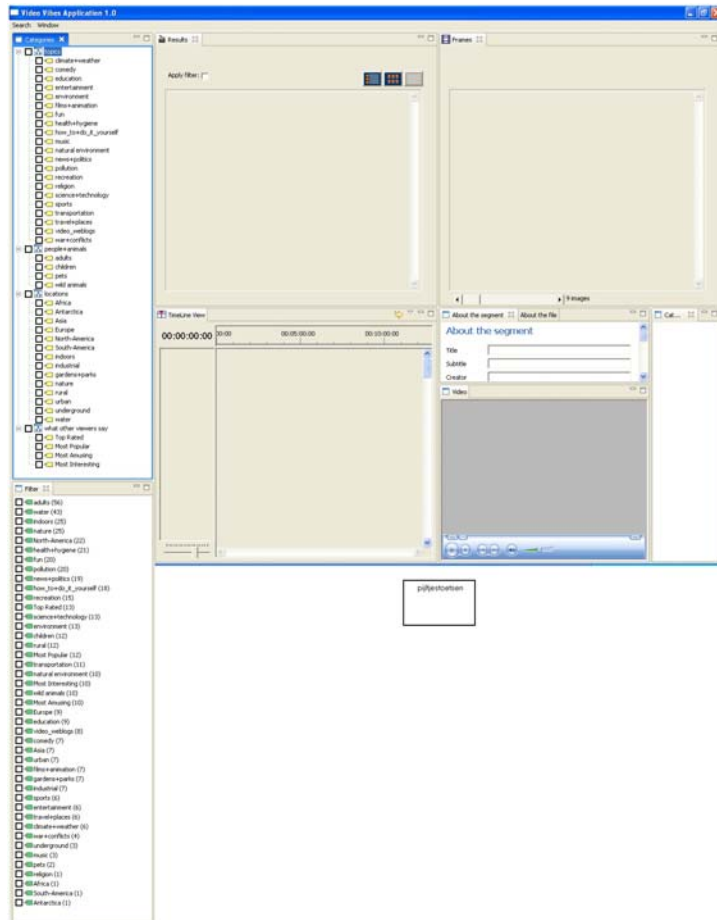
The following table presents the task order of each participant in the browse experiment described in Chapter 5. For further information, see that chapter.

Participant	Task order
1	1-2-3 <break> 4-5
2	1-2-3 <break> 4-5
3	1-2-3 <break> 4-5
4	3-2-1 <break> 5-4
5	3-2-1 <break> 5-4
6	3-2-1 <break> 5-4
7	2-3-1 <break> 4-5
8	2-3-1 <break> 4-5
9	2-3-1 <break> 4-5
10	1-3-2 <break> 5-4
11	1-3-2 <break> 5-4
12	2-1-3 <break> 4-5
13	2-1-3 <break> 4-5
14	3-1-2 <break> 5-4
15	3-1-2 <break> 5-4
16	3-1-2 <break> 5-4

Task legend:

- 1 = VIBES-General
- 2 = VIBES-Specific(visual)
- 3 = VIBES-Specific(auditive)
- 4 = Fabchannel
- 5 = YouTube

Browse experiment: observation form



Summary

Video plays an important role in our highly visual culture, and we are confronted with it constantly. Given the overabundance of video available, the attention of someone searching for video needs to be allocated efficiently among the video sources. The objective of the research described in this thesis is to study *how to support interaction with video in such a way that people can efficiently satisfy their needs.*

When talking about the search for videos, we stress the role of the human in the retrieval problem with the emphasis on the flow of representations and actions rather than discrete matches. Interaction is seen as a process of bridging gaps. There are at least three gaps that need to be bridged when searching the information environment: 1) a gap between reasons for a decision and the decision; 2) a gap between the decision and the initiation of the action; and 3) a gap between the initiation of the action and the continuation and completion of the action. People have cognitive “tools” available for bridging the gaps: tools that are used to structure their environment and interact with that environment. We can define these tools in terms of information foraging theory or IFT.

IFT consists of a number of elements. (1) IFT states that people forage through an information environment in search of a piece of information that associates with their interests the way animals forage for food. (2) For the user, the information environment has a patchy structure (compare websites on the World Wide Web). Within a patch, a person can decide to forage the patch or switch to another patch. (3) Users make navigational decisions guided by scent, which is a function of the perception of value, cost, and access path of the information with respect to the goal and interest of the user. Perceived scent is influenced by the design of “scent carriers”: representational elements in the information environment that relate to sought-for information. (4) The forager is constantly adapting decision-making and direction, preferring information-seeking strategies that yield more useful information per unit cost.

In this thesis, the framework of IFT is applied to the case of searching for videos. In the framework, efficient video browsing takes the form of optimizing video patches and their related scent in a browsing structure that supports decision-making in the three-gap decision model. The thesis describes four user studies that aim to provide an answer to the following research questions: “What is the most useful way to classify video content?” “What is the character of good video scent?” and “How to design a video interaction environment that will optimally support its users?”

We conducted two exploratory survey studies (both described in Chapter 2) to collect data on user preferences for video categories that may serve to organize patches: the Kenniswijk survey and the Fabchannel survey. An important difference between the two studies was that the Kenniswijk survey was very large and generic, asking about TV/video viewing behavior and preferences in general. The Fabchannel survey was very specific, asking one particular user-group about their preferred interaction with videos on a dedicated website with videos from one genre. The third study (described in Chapter 3) was an experiment in which we asked participants to select the most relevant link to a video from a group of links. We measured the perceived scent by asking for the subjective probability that the information that was needed could be found behind that link. This was repeated for different types of tasks and different types of scent carriers to study the influence of these factors. The fourth study (described in Chapter 5) was an experiment in which we asked participants to perform a number of tasks with an experimental video application we developed: the VIBES video browser (described in Chapter 4). This resulted in a quantitative analysis of the usefulness of the elements of the application. In addition, we asked the participants to perform tasks with the Fabchannel and YouTube websites (also described in Chapter 4). This provided data for a qualitative analysis of the difficulties of video interaction in specific and general situations, and of which support users want most in interacting with video. Taking the four studies together, the main results are as follows.

Video patches (which can be separate videos, collections of videos, or collections of video segments) structure the information environment. Which patches are useful depends on the type of task at hand. When the user’s need is specific and well defined, the easiest way to bridge the first gap and create useful patches is by querying the database using keywords that associate with the user’s need. If patches are prefabricated - as categories - they are selected depending on the amount of scent they carry. In general, the most important classification of a video (or video segment) is the *topic* it covers or the *genre* it fits in. In specific cases (for example, music videos), specific categories (such as musical sub-genres) may carry a lot of scent. Whenever the number of videos in a patch is too large (which is often the case), it should be possible to filter or cluster the results to create

smaller patches that are easier to explore. When people have no specific search goal, but generally want to see something that is interesting or funny, they prefer patches that are popular or highly rated by other people: “social” patches.

Although TV/video programs as a whole form relevant patches, there is also a need for more efficient interaction with smaller video units. The preferred units of interaction are the “natural” program segments: the semantic segments as the program maker intended (e.g., the items of a newscast), or those “naturally” provided by the characteristics of the content (e.g., the songs of a concert). These segments can be considered to be good scent carriers, and are very suitable for providing navigational information. A video segment can be taken out of its original context, put in another patch, and subsequently be found as an element of that newly formed patch. As long as the new video patch has a meaningful label (which means it carries scent to some of the users), this is a useful way to (re)structure the video database. A patch-oriented database structure supporting both within-patch and between-patch browsing may thus lead to user satisfaction.

Regarding video scent, the perceived usefulness of data in guiding users’ navigational decisions is a function of communication mode (verbal vs. pictorial); reason to browse (cognitive benefit vs. mood improvement); and tasks to perform (type of task, specificity, difficulty). With a simple task and with relevant videos present in the result set, perceived scent is higher than with a general (cognitive benefit) task. With increasing task difficulty, perceived scent drops. The most useful pieces of information to get about a video in links are a title, a description of events, and a description of subjects/topics. A frame is considered less useful when determining the relevance of videos, especially when the reason to browse is cognitive benefit rather than mood improvement. However, the combination of a frame with textual information (e.g., a title) is clearly strong in terms of perceived scent. If we were to design a composed scent carrier for the earlier stage of search - when people have to choose from a list of links to videos - it would contain at least a title, a description (of events, subjects/topics), the category/genre, and the purpose of the video (which is often directly related to the category/genre). Ideally it would also contain a frame, especially because of its strength in combination with the title. For general tasks (“find anything that is funny or of interest to me”), social data such as recommendations, comments, and ratings by other users are very important. When people are likely to visit a given database with such a general purpose in mind, then that database can best be structured according to the behavior (e.g., views) and opinions (e.g., ratings) of other users. What other people think of a video is an important scent provider.

Scent carriers requiring a lot of time and attention (which is the case when they contain audio or moving images, like a preview/trailer) are a problem in the earlier stages of search. Typically, a user would only inspect such a scent carrier after establishing - on the basis of the other available scent carriers - that the video source is worth further examination. However, for a specific genre like music (as in the Fabchannel database), video snippets prove to be valuable scent carriers, giving a quick impression of the character or atmosphere of the video contents.

We distinguish scent in the links and scent in the video data. Whenever a potentially relevant video is selected on the basis of its metadata, inspection of the contents is needed to detect any scent in the data. Efficient interaction with video content is important. It should be easy to jump within a video using its timeline. Our research has shown that when people are searching for visual events, frame-based visual summaries (like the Frames module that was part of the VIBES video browser) are very useful, especially when there is direct access to the relevant part of the video via a frame in which scent is detected. When people are looking for information in the audio channel of a video, they prefer to search textual descriptions rather than the audio itself. A frame-based summary then would ideally be extended by adding a title, description, and/or tags to the separate frames. Availability of a transcript (including search functionality) would also ease the search.

To support exploration of the database (that is, between-patch browsing by following scent-trails), every video should provide a starting point to other videos ("related" or "more like this"), which can have the shape of a recommendation ("if you like this, you might also like this"). From an IFT point of view, videos that are related to each other form a patch. Moreover, if a video has scent, a related video can potentially also have scent. Finally, related videos support between-patch browsing, and thus help in bridging the third gap. Relating videos can be done for example via their tags or the categories they belong to. For this kind of exploration to occur, it is very important that videos be part of multiple overlapping patches. This is also true for video *segments*. The video they are part of can be considered just one of the patches they belong to.

The main goal of the studies was to see whether the IFT framework is a useful approach to the problem of video interaction. Based on the user studies, our conclusion is that the IFT framework is useful for describing and explaining human searching behavior. This research confirms the dominance of browsing in video interaction. The flow of representations and actions is the central quality of the interaction, rather than discrete matches. We conclude that the structure of interaction with video material can be understood as foraging the way IFT describes. Despite its exploratory character, browsing is highly structured. Not only do the four

studies reported in this thesis testify to the value of the IFT framework in thoroughly describing all browsing behaviors relevant to our research, it also seems possible to summarize all the observed browsing behaviors under three headings: the three gaps. In order to bridge these three gaps, we saw, people can choose different search strategies. Their choice is influenced by their goals and, in particular, by the functionality of the interaction environment. One simple heuristic (“select the link with the highest scent”) and two possible sub-heuristics (data-first or metadata-first) sum up users’ within-patch and between-patch switching. The concepts of patch, scent, and switching within and between patches are necessary and sufficient to do justice to the exploratory nature of browsing.

The framework has practical use for the design and evaluation of video browsing environments. It seems that good IFT-based design allows users to bridge the gaps without noticing any discontinuity. Related to this, we introduced the idea of *IFT gap-optimality*: an IFT-based environment is optimal to the extent that it allows users to bridge the three gaps with minimal experience of discontinuity. The degree of experienced (dis)continuity is a function of the support provided by the video environment in terms of patch selection, scent-following, and within-patch and between-patch browsing. Questions such as “Which patches are or can be created?” “How is browsing within a patch supported?” “How is switching between patches supported?” and “In which way is scent displayed in the interface?” prove to be very useful. The research presented here has covered enough ground to develop the concept of IFT gap-optimality further, to make it into a useful tool for the development of high-quality browsing environments, and to measure those environments’ quality. In sum, a video interaction environment should offer tools to bridge the three gaps and become IFT gap-optimal. It should provide useful patches (or the tools to create useful patches), support the detection of scent within a patch by providing good scent carriers, and support scent-following between patches.

To conclude, in the restricted domains we studied, the IFT framework (including the concepts of patches, scent, and gaps) proved very useful for describing searching behavior and providing tools for the design and evaluation of video interaction environments.

Samenvatting

Video³ speelt een belangrijke rol in onze zeer visueel ingestelde cultuur, en we worden er voortdurend mee geconfronteerd. Aangezien er zeer veel videomateriaal beschikbaar is, is het zaak dat de aandacht van een persoon die naar video zoekt op een efficiënte manier over de diverse videobronnen wordt verdeeld, zodat vooral relevant materiaal wordt bekeken. Het doel van het in dit proefschrift beschreven onderzoek is: *Hoe moet interactie met video worden ondersteund zodat videogebruikers op een efficiënte manier in hun informatiebehoefte kunnen voorzien?*

Als we het hebben over zoeken naar video's, benadrukken we de rol van de gebruiker in het zoekproces. Het zoekproces staat hierbij centraal, waarbij er sprake is van een stroom van handelingen waarbij de zoeker voortdurend reageert op de representaties die daarbij aangetroffen worden. Interactie zien we daarbij als het overbruggen van kloven (*gaps*). Er zijn minimaal drie soorten *gaps* die overbrugd moeten worden bij het zoeken in een informatieomgeving: 1) een *gap* tussen de reden voor een beslissing en de beslissing zelf; 2) een *gap* tussen de beslissing en het beginnen van de handeling; en 3) een *gap* tussen het beginnen van de handeling en de continuering en afronding van de handeling. Voor het overbruggen van de *gaps* hebben gebruikers de beschikking over cognitieve "gereedschappen". Deze gereedschappen worden gebruikt om de omgeving te ordenen en structureren, en om met die omgeving te interacteren. We kunnen deze gereedschappen definiëren in termen van *information foraging theory* of IFT.

IFT bestaat uit een aantal elementen. (1) IFT beweert dat mensen op zoektocht gaan in een informatieomgeving om een relevant stuk informatie te vinden, vergelijkbaar met hoe dieren naar voedsel zoeken. (2) Voor een zoekende mens is de informatieomgeving als een soort van lappendeken opgedeeld in plekjes (*patches*) met informatie (zoals bijvoorbeeld de afzonderlijke websites op het wereldwijde web). In een *patch* kan een

³ Hiermee worden alle bewegende beelden – inclusief TV-programma's, dvd's, video op het internet etc. - bedoeld.

persoon er voor kiezen de *patch* te verkennen dan wel van plaats te veranderen naar een andere *patch*. (3) Bij het navigeren door een informatieomgeving laten mensen zich leiden door de waargenomen “geursporen” (*scent*). Deze *scent* is een functie van de waargenomen waarde, kosten, en toegankelijkheid van de informatie in relatie tot de doelen en interesses van de zoeker. De hoeveelheid *scent* die wordt waargenomen wordt beïnvloed door het ontwerp van de “*scent*-dragers”: de elementen in de informatieomgeving die bronnen met informatie representeren. (4) De foeragerende mens past zich voortdurend aan aan de omstandigheden door het bijstellen van beslissingen en zoekrichting. Daarbij heeft men een voorkeur voor zoekstrategieën die de meeste bruikbare informatie opleveren tegen de laagste ervaren kosten.

In dit proefschrift wordt het raamwerk van IFT toegepast op het geval van zoeken naar video’s. In het raamwerk heeft efficiënt zoeken of rondsnuffelen (“browsen”) de vorm van het optimaliseren van video *patches* en de daaraan gerelateerde *scent*, in een browsestructuur die het nemen van beslissingen voor het overbruggen van de drie *gaps* goed ondersteunt. Het proefschrift beschrijft vier gebruikersstudies die proberen een antwoord te geven op de volgende drie onderzoeksvragen: “Wat is de meest bruikbare manier om de inhoud van video’s te classificeren?”, “Wat zijn de kenmerken van goede video *scent*?”, en “Hoe moet je een video-interactieomgeving ontwerpen zodat gebruikers van die omgeving goed ondersteund worden?”.

We voerden twee exploratieve surveystudies uit (allebei beschreven in Hoofdstuk 2) om gegevens te verzamelen over de voorkeuren van gebruikers voor videocategorieën die gebruikt kunnen worden voor de organisatie van video *patches*: het Kenniswijkonderzoek en het Fabchannelonderzoek. Een belangrijk verschil tussen de twee onderzoeken was dat het Kenniswijkonderzoek uitgebreid en algemeen was, waarbij de vragen gingen over TV-/videokijkgedrag en daaraan gerelateerde voorkeuren in het algemeen. Het Fabchannelonderzoek was heel specifiek, waarbij vragen gesteld werden aan een specifieke gebruikersgroep over hun geprefereerde interactie met video's op een specifieke website met video's uit één genre. De derde gebruikersstudie (beschreven in Hoofdstuk 3) was een experiment waarin we de deelnemers vroegen uit een groep van verwijzingen (*links*) naar video's de meest relevante link te kiezen. We beoordeelden de waargenomen *scent* door te vragen hoe waarschijnlijk ze het vonden dat ze de gewenste informatie zouden vinden als ze de *link* zouden volgen. Dit werd herhaald voor verschillende soorten taken en *scent*-dragers om de invloed van deze factoren vast te kunnen stellen. De vierde gebruikersstudie (beschreven in Hoofdstuk 5) was een experiment waarbij deelnemers een aantal taken moesten uitvoeren met een door ons ontwikkelde experimentele videoapplicatie: de VIBES videobrowser

(beschreven in Hoofdstuk 4). Dit resulteerde in een kwantitatieve analyse van de bruikbaarheid van de verschillende elementen van de applicatie. Verder vroegen we de deelnemers taken uit te voeren met de websites van Fabchannel en YouTube (ook beschreven in Hoofdstuk 4). Dit leverde gegevens voor een kwalitatieve analyse van de moeilijkheden die men tegenkomt bij de interactie met videomateriaal, en welke ondersteuning het meest gewenst is bij de interactie met video. Hieronder worden de gezamenlijke resultaten uit de vier gebruikersstudies beschreven.

Video-patches (wat aparte video's, videocollecties, of collecties met fragmenten kunnen zijn) structureren de informatieomgeving. Welke *patches* bruikbaar zijn hangt af van de soort taak die uitgevoerd moet worden. Als de behoefte van een gebruiker specifiek en duidelijk gedefinieerd is, is de gemakkelijkste manier om de eerste *gap* te overbruggen en bruikbare *patches* te creëren het uitvoeren van een zoekactie waarbij zoektermen worden gebruikt die associëren met de behoefte. Als er "voorgefabriceerde" *patches* zijn – in de vorm van categorieën – wordt de *patch* die de meeste *scent* draagt gekozen. In het algemeen is de meest bruikbare manier om video(segmenten) te classificeren op onderwerp of genre. In specifieke gevallen (bijvoorbeeld, bij muziekvideo's) kunnen specifieke categorieën (zoals muzikale subgenres) veel *scent* dragen. Als het aantal video's in een *patch* te groot is (wat veelvuldig voorkomt), moet het mogelijk zijn de resultaten te filteren of te clusteren zodat er kleinere *patches* ontstaan die gemakkelijker te verkennen zijn. Als gebruikers geen specifiek zoekdoel hebben, maar in het algemeen op zoek zijn naar iets leuks of interessants, hebben ze de voorkeur voor *patches* die veel bekeken of hoog gewaardeerd worden door andere gebruiker: "sociale" *patches*.

Hoewel TV-/videoprogramma's als geheel relevante *patches* kunnen zijn, is er ook een voorkeur voor een efficiëntere interactie met kleinere videoeenheden. Er is voorkeur voor "natuurlijke" programmasegmenten als videoeenheid: de semantisch segmenten zoals de programmamaker ze al heeft aangebracht (bijvoorbeeld, de items van een journaal), of die eigen zijn aan de inhoud van de video (bijvoorbeeld de liedjes van een concert). Deze segmenten zijn goede *scent*-dragers en leveren zeer geschikte informatie voor het navigeren door de informatieomgeving. Een videosegment kan uit zijn originele context worden gehaald en (mede) onderdeel gaan uitmaken van een andere *patch*. Zolang die ander *video-patch* een betekenisvol label heeft (wat betekent dat hij voor sommige gebruikers *scent* draagt) is dit een bruikbare manier om de videodatabase te (her)structureren. Een *patch*-gebaseerde databasestructuur die zowel het browsen in een *patch* als het browsen van *patch* naar *patch* ondersteunt wordt door gebruikers gewaardeerd en leidt tot succesvolle interactie.

Wat betreft *video-scent* is de waargenomen bruikbaarheid van gegevens die bepalend zijn voor het nemen van navigatiebeslissingen een functie van

de communicatiemodus (verbaal of picturaal), de reden om te browsen (kennis opdoen of stemming verbeteren), en hoe specifiek of moeilijk de taak is. Met een simpele taak en als er veel relevante video's aanwezig zijn, is de waargenomen hoeveelheid *scent* in de links hoger dan met een algemene taak. Als de taakmoeilijkheid toeneemt, neemt de hoeveelheid waargenomen *scent* in de links af. De meest bruikbare stukjes informatie om over een video te krijgen zijn de titel, een beschrijving van de gebeurtenissen, en een beschrijving van de onderwerpen in de video. Een beeld (*frame*) uit de video is minder bruikbaar om de relevantie van een video te bepalen, vooral als de reden om te browsen is om kennis op te doen (en niet om vermaakt te worden). Echter, de combinatie van een frame met tekstuele informatie (bv. een titel) werkt weer sterk wat betreft waargenomen *scent*. Als we een samengestelde *scent*-drager zouden moeten ontwerpen voor het vroege zoekstadium – als gebruikers uit een lijst met links moeten kiezen – zou die tenminste het volgende bevatten: een titel, een beschrijving (van gebeurtenissen, onderwerpen), het genre, en het doel van de video (wat meestal direct gerelateerd is aan het genre). Idealiter zou er ook een *frame* bijzitten, vooral wegens de kracht daarvan in combinatie met een titel. Voor algemene taken (“vind iets dat leuk of interessant is”) zijn sociale gegevens zoals aanbevelingen, opmerkingen, en beoordelingen (*ratings*) door andere gebruikers zeer belangrijk. Als veel gebruikers een database bezoeken met een dergelijk algemeen doel, is het doelmatig die database te structureren volgens het gedrag (bv. het aantal bezichtigingen) en de meningen (bv. *ratings*) van andere gebruikers. Wat andere gebruikers van een video vinden is een belangrijke verschafter van *scent*. In het vroege stadium van zoeken zijn *scent*-dragers die veel tijd en aandacht kosten (omdat ze bewegende beelden of audio bevatten, bijvoorbeeld bij een voorfilmpje of *trailer*) een probleem. Iemand zal zo'n *scent*-drager typisch alleen inspecteren nadat op basis van andere *scent*-dragers is vastgesteld of de video verdere verkenning verlangd. Echter, voor een specifiek genre als muziek (als in de database van Fabchannel) kunnen korte stukjes video (*snippets*) juist heel goede *scent*-dragers zijn.

We onderscheiden *scent* in de links en *scent* in de data (de video zelf). Als een potentieel relevante video is geselecteerd op basis van zijn metadata, is vervolgens inspectie van de inhoud nodig om daar *scent* vast te stellen. Efficiënte interactie met de inhoud van een video is daarbij belangrijk. Het moet gemakkelijk zijn te springen in een video met behulp van een tijdslijn. Ons onderzoek heeft aangetoond dat als gebruikers naar visuele gebeurtenissen zoeken, op *frames* gebaseerde samenvattingen (zoals de Frames module in de VIBES video browser) zeer bruikbaar zijn, vooral als via een frame waarin *scent* is aangetroffen er direct toegang is tot de bijbehorende bewegende beelden. Als gebruikers zoeken naar informatie in het audiokanaal van de video, hebben ze voorkeur voor tekstuele

beschrijvingen boven de audio zelf. Een frame-gebaseerde samenvatting zou dan idealiter uitgebreid worden met titels, beschrijvingen en labels (*tags*) voor elk afzonderlijk frame. De beschikbaarheid van het transcript (de uitgeschreven gesproken tekst) inclusief de mogelijkheid daarin te zoeken zou het navigeren door het materiaal nog verder vergemakkelijken.

Om het exploreren van de database te ondersteunen (het springen van *patch* naar *patch* door het volgen van *scent*), zou elke video startpunten naar andere video's moeten bevatten. Dit kan bijvoorbeeld worden gepresenteerd als "gerelateerde video's", "meer video's zoals deze", of "als deze video u bevalt, zal deze video u ook bevallen". Vanuit het gezichtspunt van IFT vormen alle video's die gerelateerd zijn een *patch*. Als een video *scent* heeft voor een gebruiker, zal een gerelateerde video ook *scent* kunnen hebben voor diezelfde gebruiker. Gerelateerde video's ondersteunen het browsen van *patch* naar *patch*, en dus ook het overbruggen van de derde *gap*. Het relateren van video's kan bijvoorbeeld gedaan worden via overeenkomende *tags* of via de categorieën waartoe ze behoren. Om dit soort exploratie mogelijk te maken is het van belang dat video's onderdeel uitmaken van meerdere, overlappende *patches*. Dit geldt ook voor afzonderlijke videosegmenten. De video waar ze onderdeel van uitmaken kan worden beschouwd als slechts één van *patches* waartoe ze kunnen behoren.

Het belangrijkste doel van de gebruikersstudies was vast te stellen of het IFT-raamwerk een bruikbare benadering is van het videointeractieprobleem. Op basis van de studies kunnen we concluderen dat het IFT-raamwerk bruikbaar is voor het beschrijven en verklaren van menselijk zoekgedrag. Dit onderzoek bevestigt de dominantie van browsen in het zoekgedrag. Bij interactie gaat het vooral om de stroom van acties waarbij de zoeker voortdurend reageert op de representaties die daarbij aangetroffen worden en minder om het opgeven van zoektermen en het terugkrijgen van resultaten. De conclusie van dit proefschrift is dat de structuur van interactie met videomateriaal kan worden begrepen als foerageergedrag zoals wordt beschreven in IFT. Ondanks het exploratieve karakter is browsen zeer gestructureerd. De vier hier beschreven studies bevestigen niet alleen de waarde van het IFT voor het beschrijven van browsegedrag. Het blijkt ook mogelijk het browsegedrag onder drie kopjes onder te brengen: de drie *gaps*. We zagen dat videogebruikers verschillende strategieën kozen om de *gaps* te overbruggen. Hun keuzes worden bepaald door hun doelen, en – in het bijzonder – door de functionaliteit van de interactieomgeving. Eén simpele heuristiek ("selecteer de link met de meeste *scent*") en twee mogelijke sub-heuristieken (op basis van de data of de metadata) vatten het keuzegedrag binnen een *patch* en tussen *patches* samen. De concepten *patch*, *scent* en switchen binnen en tussen *patches* zijn

nodig en voldoende om het exploratieve karakter van browsen te verantwoorden.

Het raamwerk heeft praktisch nut voor het ontwerp en de evaluatie van video-omgevingen. Het lijkt erop dat een goed ontwerp gebaseerd op IFT gebruikers toestaat de *gaps* te overbruggen zonder dat ze discontinuïteit bemerken. Hieraan gerelateerd introduceren wij het idee van IFT-*gap*-optimaliteit: een IFT-gebaseerde omgeving is in die mate optimaal dat zij gebruikers de drie *gaps* laat overbruggen met een minimum aan ervaren discontinuïteit. De mate van ervaren (dis)continuïteit is een functie van de ondersteuning die de video-omgeving biedt in termen van het selecteren van *patches*, het volgen van *scent*, en het browsen in een *patch* en tussen *patches*. Vragen als “Welke *patches* zijn of kunnen worden gecreëerd?”, “Hoe wordt browsen in een *patch* ondersteund?”, “Hoe wordt het switchen tussen *patches* ondersteund?”, en “Op welke manier wordt *scent* weergegeven in de interface?” blijken zeer nuttig te zijn. Het onderzoek beschreven in dit proefschrift heeft genoeg basis gelegd om het idee van IFT-*gap*-optimaliteit verder uit te werken en er een bruikbaar gereedschap van te maken voor het ontwikkelen van browseomgevingen van hoge kwaliteit. Verdere operationalisatie van IFT-*gap*-optimaliteit zal daarbij ook de manier vormen om de kwaliteit van die omgevingen te meten. Samengevat moet een video-omgeving de mogelijkheden bieden de drie *gaps* te overbruggen en IFT-*gap*-optimaal te worden. Het moet nuttige *patches* bieden (dan wel de mogelijkheden bieden om nuttige *patches* te creëren), de detectie van *scent* door het leveren van goede *scent*-dragers ondersteunen, en het volgen van *scent* van *patch* naar *patch* ondersteunen.

Concluderend: in het beperkt domein dat we hebben onderzocht blijkt het IFT-raamwerk (inclusief de concepten *patches*, *scent*, en *gaps*) bijzonder nuttig te zijn voor het beschrijven van zoekgedrag en duidelijke handvatten te verschaffen voor het ontwerp en de evaluatie van videointeractieomgevingen.

References

- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the on-line search interface. *Online Review*, 13, 407-431.
- Bearne, M., Jones, S., & Sapsford-Francis, J. (1994). Towards usability guidelines for multimedia systems. In *Proceedings of ACM Multimedia '94* (pp. 105-110). New York: ACM Press.
- Beeld en Geluid (2008). *Beeld en Geluid Instituut*. Retrieved March 15, 2008 from <http://instituut.beeldengeluid.nl/index.aspx?ChapterID=8506>
- Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143
- Belkin, N.J., Cool, C., Stein, A. & Thiel, U. (1995). Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems With Applications*, Vol. 9, No. 3, pp.379-395.
- Belkin, N.J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. Part I: background and theory. *Journal of Documentation*, 38, 61-71.
- Boreczky, J., Girgensohn, A., Golovchinsky, G., & Uchihashi, S. (2000). An interactive comic book presentation for exploring video. In *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems* (pp. 185-192). New York: ACM Press.
- Brown, B. & Barkhuus, L. (2006). The Television Will Be Revolutionized: Effects of PVRs and Filesharing on Television Watching. In *Proceedings of the CHI 2006 Conference on Human Factors in Computing Systems* (pp. 663-666).
- Burford, B., Briggs, P., & Eakins, J. P. (2003). A taxonomy of the image: on the classification of content for image retrieval. *Visual Communication*, 2, 123-161.
- CBS (2005). Available from <http://www.cbs.nl>
- Christel, M. G., Smith, M. A., Taylor, C. R., & Winkler, D. B. (1998). Evolving video skims into useful multimedia abstractions. In *Proceedings of the CHI 1998 Conference on Human Factors in Computing Systems* (pp. 171-178). New York: ACM Press.
- Christel, M. G., Winkler, D. B., & Taylor, C. R. (1997). Multimedia abstractions for a digital video library. In *Proceedings of the 2nd ACM Conference on Digital Libraries (DL '97)* (pp. 21-29).

- Christel, M. & Yan, R. (2007). Merging Storyboard Strategies and Automatic Retrieval for Improving Interactive Video Search. In: *Proceedings of the International Conference on Image and Video Retrieval (CIVR)* (Amsterdam, The Netherlands, July 2007), pp. 486-493.
- Cisco Systems, Inc. (2008). *Visual Networking Index - Approaching the Zettabyte Era* (white paper). Retrieved June 17, 2008 from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374.pdf
- Cosley, D., Lam, S. K., Albert, I., Konstan, J., & Riedl, J. (2003). Is Seeing Believing? How Recommender Systems Influence Users' Opinions. In *Proceedings of CHI 2003* (pp. 585-592). ACM.
- Dahdah, H. (2008). *Schmidt describes YouTube upload rate as "disturbing"*. Retrieved March 30, 2008 from the Computerworld website: <http://www.computerworld.com.au/index.php/id;1523127763;fp;2;fpid;1>
- Del Bimbo, A. (1999). *Visual information retrieval*. San Francisco, California: Morgan Kaufmann Publishers.
- Dervin, B. (1992). From the mind's eye of the user: the sense-making qualitative-quantitative methodology. In J.D. Glazier & R.R. Powell, *Qualitative Research in Information Management* (pp.61-84). Englewood, CO: Libraries Unlimited.
- Devlin, K. (1991). *Logic and information*. Cambridge University Press.
- Eakins, J. P., Briggs, P., & Burford, B. (2004). Image retrieval interfaces: a user perspective. In P. Enser, Y. Kompatsiaris, N. O'Connor, A. F. Smeaton, & A. W. M. Smeulders (Eds.), *Lecture Notes in Computer Science 3115 - Image and Video Retrieval - Proceedings of CIVR 2004* (pp. 628-637). Springer-Verlag.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, 45(3), 171-212.
- Ellis, D., Cox, D. & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49(4), 356-369.
- Enser, P., & Sandom, C. (2001). Retrieval of archival moving imagery - a step too far for CBIR? In *Proceedings of the International Workshop on MultiMedia Content-Based Indexing and Retrieval (MMCBIR'01)*
- Fabchannel (2008). Retrieved February 29, 2008, from <http://www.fabchannel.com>
- Fischlár (2008). Retrieved May 31, 2008, from <http://www.cdpv.dcu.ie/aboutfischlar.html>
- Furnas, G. W. (1997). Effective view navigation. In *Proceedings of the conference on Human Factors in Computing Systems, CHI '97* (pp. 367-374). Atlanta, GA, USA: ACM.
- Gigerenzer, G. (2000). *Adaptive thinking: rationality in the real world*. New York: Oxford University Press.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: the adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3-36). New York: Oxford University Press..
- Golovchinsky, G. (1997). Queries? Links? Is there a difference? In *Proceedings of the Conference on Human Factors in Computing Systems, CHI '97* (pp. 407-414). Atlanta, GA, USA: ACM.

- Hall, S. (1980). Encoding/Decoding. In *Culture, Media, Language: Working Papers in Cultural Studies 1972-1979*. London: Hutchinson.
- Hauptmann, A. G., & Christel, M. G. (2004). Successful approaches in the TREC video retrieval evaluations. In *Proceedings of the 12th annual ACM international conference on Multimedia* (pp. 668-675).
- Hauptmann, A., Yan, R. & Lin, W.-H. (2007). How many high-level concepts will fill the semantic gap in news video retrieval? In *Proceedings of CIVR'07, July 9-11, Amsterdam, Netherlands* (pp. 627-634)
- Hearst, M.A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM, April 2006/Vol.49, No.4*, 59-61.
- Hildreth, C. R. (1982). Online browsing support capabilities. In *Proceedings of the Annual Meeting of the American Society for Information Science 19* (pp. 127-132).
- Hirschman, E. C. (1987). Consumer preferences in literature, motion pictures, and television programs. *Empirical Studies of the Arts*, 5, 31-46.
- Hollink, L., Schreiber, G., Wielinga, B., & Worring, M. (2004). Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61, 601-626.
- van Houten, Y., Naci, U., Freiburg, B., Eggermont, R., Schuurman, S., Hollander, D., Reitsma, J., Markslag, M., Kniest, J., Veenstra, M., & Hanjalic, A. (2005). The MultimediaN concert-video browser. In *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME 2005)*.
- van Houten, Y., Schuurman, J. G., & Verhagen, P. (2004). Video content foraging. In P.Enser, Y. Kompatsiaris, N. O'Connor, A. F. Smeaton, & A. W. M. Smeulders (Eds.), *Lecture Notes in Computer Science 3115 - Image and Video Retrieval - Proceedings of CIVR 2004* (pp. 15-23). Springer-Verlag.
- van Houten, Y., van Setten, M., & Schuurman, J. G. (2003). Patch-based video browsing. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.), *Human-Computer Interaction INTERACT '03* (pp. 932-935). IOS Press.
- Hughes, A., Wilkens, T., Wildemuth, B., & Marchionini, G. (2003). Text or pictures? An eyetracking study of how people view digital video surrogates. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR 2003)* (pp. 271-280).
- Informedia (2008). Retrieved May 31, 2008, from <http://www.informedia.cs.cmu.edu/>
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52, 3-50.
- Ingwersen, P. (1999). Cognitive information retrieval. *Annual Review of Information Science and Technology*, 34, 3-52.
- Isen, A.M. (1993). Positive affect and decision making. In M. Lewis & J.M. Haviland (Eds.), *Handbook of emotions* (pp.261-277). New York: Guilford.
- Jaimes, A., & Chang, S.-F. (2000). A conceptual framework for indexing visual information at multiple levels. In *IS&I/SPIE Internet Imaging 2000, Vol. 3964*.
- Kenniswijk (2005). *Kenniswijk- Experimenting for the Future*. Retrieved January 31, 2007, from <http://www.kenniswijk.nl>

- Klein, G. (2001). The fiction of optimization. In Gerd Gigerenzer and Reinhard Selten (eds.) *Bounded Rationality: The Adaptive Toolbox*, pp. 103-121. Cambridge, Mass.: The MIT Press.
- Komlodi, A. & Marchionini, G. (1998). Key frame preview techniques for video browsing. In *Proceedings of the third ACM conference on Digital Libraries (DL'98)* (pp. 67-75).
- Kuhlthau, C.C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42, 361-371.
- Lee, R. & Lee, B. (1995). How and why people watch TV: implications for the future of interactive television. *Journal of Advertising Research*, nov/dec 1995, 9-18.
- Li, F. C., Gupta, A., Sanocki, E., He, L., & Rui, Y. (2000). Browsing digital video. In *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems* (pp. 169-176). New York: ACM Press.
- Madden, 2007. Online video. Pew Internet project report. Retrieved December 5, 2007 from http://www.pewinternet.org/pdfs/PIP_Online_Video_2007.pdf
- Manovich, L. (2001). *The language of new media*. Cambridge, MA: The MIT Press.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge University Press.
- Marchionini, G. (2004). From information retrieval to information interaction. In S. McDonald & J. Tait (Eds.), *Advances in Information Retrieval, 26th European Conference on IR Research (ECIR 2004), Proceedings - Lecture Notes in Computer Science 2997* (pp. 1-11). Springer.
- Markkula, M. & Sormunen, E. (2000). End-user searching challenges indexing practices in the digital newspaper search archive. *Information Retrieval*, 1, 259-285.
- Mathes, A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. Retrieved 31 January 2007 from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- MediaMill (2008). Retrieved May 31, 2008, from <http://www.science.uva.nl/research/mediamill/>
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, Massachusetts: The MIT Press.
- Naphade, M. R., & Smith, J. R. (2004). On the detection of semantic concepts at TRECVID. In *Proceedings of the 12th annual ACM international conference on Multimedia* (pp. 660-667).
- Naphade, M., Smith, J.R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., Hauptmann, A. & Curtis, J. (2006). Large-scale concept ontology for multimedia. *IEEE MultiMedia*, July-September 2006, 86-91
- Noam, E., Groebel, J. and Gerbarg, D. (Eds.) (2004). *Internet Television*. London: Lawrence Erlbaum Associates, Publishers.
- Norman, D. (2004). *Emotional Design: Why we love (or hate) everyday things*. New York, Basic Books.
- Open Video (2008). Retrieved May 31, 2008, from <http://www.open-video.org/>
- Panofsky, E. (1955). *Meaning in visual arts*. London: Penguin.
- Pirolli, P. (2003). A theory of information scent. In J. Jacko & C. Stephanidis (Eds.), *Human-Computer Interaction I* (pp. 213-217). Mahwah, NJ USA: Lawrence Erlbaum.

- Pirolli, P. & Card, S. K. (1999). Information foraging. *Psychological Review*, 106, 643-675.
- Pirolli, P., Card, S. K., & Van Der Wege, M. M. (2003). The Effects of Information Scent on Visual Search in the Hyperbolic Tree Browser. *ACM Transactions on Computer-Human Interaction*, Vol.10, No.1, 20-53.
- Reitman, W. (1965). *Cognition and Thought: An information-processing approach*. New York: Wiley.
- Rentfrow, P. J. & Gosling, S. D. (2003). The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology*, 84, 1236-1256.
- Rice, R. E., McCreddie, M., & Chang, S.-J. L. (2001). *Accessing and browsing information and communication*. Cambridge, USA: MIT Press.
- Robertson, S. E. (1977). Theories and models in information retrieval. *Journal of Documentation*, 33, 126-148.
- Rottenberg, J., Ray, R.R., & Gross, J.J. (2007). Emotion elicitation using films. In J.A. Coan & J.J.B. Allen, (Eds.), *The handbook of emotion elicitation and assessment*. New York: Oxford University Press.
- Saracevic, T. (1996). Modeling interaction in information retrieval (IR): a review and proposal. In: S. Hardin (ed.), *59th Annual Meeting of the American Society for Information Science* (pp.3-9). Silver Spring, MD: American Society for Information Science.
- Savolainen, R. & Kari, J. (2006). Facing and bridging gaps in Web searching. *Information Processing and Management*, 42, 519-537.
- Searle, J. R. (1978). The philosophy of language. In B. Magee (Ed.), *Men of ideas* (pp. 153-172). Oxford University Press.
- Searle, J. R. (2001). *Rationality in action*. Cambridge, Massachusetts: The MIT Press.
- Sebe, N., Lew, M. S., & Smeulders, A. W. M. (2003). Video retrieval and summarization (Editorial introduction). *Computer Vision and Image Understanding*, 92, 141-146.
- Shatford, S. (1986). Analyzing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly*, 6, 39-62.
- Simon, H.A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4, 181-202.
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 1349-1380.
- Smith, J.R. (2007) The real problem of bridging the "semantic gap". In N. Sebe, Y. Liu & Y. Zhuang (Eds.): *Lecture Notes in Computer Science 4577, Proceedings of MCAM 2007* (pp.16-17).
- Snoek, C. G. M. & Worring, M. (2005). Multimodal video indexing: a review of the state-of-the-art. *Multimedia Tools and Applications*, 25, 5-35
- Song, Y. & Marchionini, G. (2007). Effects of audio and visual surrogates for making sense of digital video. In *Proceedings of CHI 2007, April 28-May 3, 2007, San Jose, California USA* (pp 867-876).
- Soto, R. (1999). Learning and Performing by Exploration: Label Quality Measured by Latent Semantic Analysis. In *Proceedings of CHI '99* (pp. 418-425). ACM.
- Surveyworld (2008). Retrieved May 31, 2008, from <http://www.surveyworld.net>

- Taylor, R. S. (1968). Question negotiation and information seeking in libraries. *College & Research Libraries*, 29, 178-189.
- Tse, T., Marchionini, G., Ding, W., Slaughter, L., & Komlodi, A. (1998). Dynamic key frame presentation techniques for augmenting video browsing. In *Proceedings of the Advanced Visual Interfaces International Working Conference (AVI '98)*
- Tsiniraki, C., Polydoros, P., Kazasis, F. & Christodoulakis, S. (2005). Ontology-based semantic indexing for MPEG-7 and TV-Anytime audiovisual content. *Multimedia Tools and Applications*, 26, 299-325.
- Varian, H. R. (1995). The Information Economy - How much will two bits be worth in the digital marketplace? *Scientific American*, September 1995, 200-201.
- White, R.W., Drucker, S.M., Marchionini, G., Hearst, M. & schraefel, m.c. (Eds.) (2007). *Proceedings of the ACM SIGCHI 2007 Workshop on "Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction"*, 29 April 2007, San Jose, CA, USA.
- White, R.W., Kules, B., Drucker, S.M. & schraefel, m.c. (2006). Supporting exploratory search. *Communications of the ACM*, April 2006/Vol.49, No. 4, 37-39.
- Wikipedia (2008a). *Video on demand*. Retrieved June 16, 2008 from http://en.wikipedia.org/wiki/Video_on_demand
- Wikipedia (2008b). *Comparison of video services*. Retrieved June 16, 2008 from http://en.wikipedia.org/wiki/Comparison_of_video_services
- Wildemuth, B.M., Marchionini, G., Wilkens, T., Yang, M., Geisler, G., Fowler, B., Hughes, A. & Mu, X. (2002). Alternative Surrogates for Video Objects in a Digital Library: Users' Perspectives on Their Relative Usability. *Proceedings of the European Conference on Digital Libraries (ECDL)*, 493-507.
- Wildemuth, B.M., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A., & Gruss, R. (2003). How fast is too fast? Evaluating fast forward surrogates for digital video. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Houston, May 2003*.
- Wilson, T.D. (1999) Models in information behaviour research. *Journal of Documentation*, 55(3), 249-270.
- Wilson, T.D. (2000). Human information behavior. *Informing Science*, 3(2), 49-55.
- Wilson, T.D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1), 3-15.
- Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J., & Pirolli, P. (2001). Using thumbnails to search the Web. In *Proceedings of the SIG-CHI on Human Factors in Computing Systems (CHI '01)* (pp. 198-205).
- Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. A. (2003). Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 401-408).



About the author

Ynze van Houten studied experimental psychology at the University of Groningen. His master's thesis was on the assessment of the effects of mental fatigue on selective attention using event-related brain potentials.

From 1993 to 1995 he worked at the Traffic Research Centre of the University of Groningen, studying road-user behavior and driver support systems with the goal of increasing traffic safety.

He then worked for three years at the National Aerospace Laboratory in Amsterdam. There he worked as a human factors engineer on the user interface design of cockpit displays in civil aircraft.

Since 1999 he has been a researcher at the Telematica Instituut in Enschede. In the Media Interaction group he carried out his Ph.D. research on how to support people in the process of efficiently finding relevant information in video material. His main research interest as a cognitive ergonomist is in improving the interaction between humans and information systems.

SEARCHING FOR VIDEOS

THE STRUCTURE OF VIDEO INTERACTION IN THE FRAMEWORK OF INFORMATION FORAGING THEORY

Ynze van Houten

Video plays an important role in our highly visual culture, and we are confronted with it constantly. Given the overabundance of video available, the attention of someone searching for video needs to be allocated efficiently among the video sources.

Searching for Videos studies how to support interaction with video in such a way that people can efficiently satisfy their needs. Interaction is seen as a process of bridging gaps. The cognitive tools to bridge these gaps are defined in terms of information foraging theory or IFT. This theory states that people forage through an information environment in search of a piece of information that associates with their interests the way animals forage for food. In the framework of IFT, efficient video browsing takes the form of optimizing video patches and their related scent in a browsing structure that supports decision-making in a three-gap decision model. The qualities of video patches and scent were analyzed in two survey studies and two experiments.

Within the restricted domains that were studied, the IFT framework (including the concepts of patches, scent, and gaps) proved highly useful for describing searching behavior. IFT is a valuable concept for understanding browsing: the research described here convincingly supports the theory. Moreover, the IFT framework provides useful tools for the design and evaluation of video interaction environments.

ISBN 978-90-75176-48-3



Telematica
Instituut