Christoph LAGGER

User Intent Classification for Video Retrieval

MASTERARBEIT

zur Erlangung des akademischen Grades
Diplom-Ingenieur

Studium
INFORMATIK

Alpen-Adria-Universität Klagenfurt
Fakultät für Technische Wissenschaften

Begutachter:
O.Univ.-Prof. Dr. Laszlo Böszörmenyi, Institut für Informationstechnologie

Vorbegutachter:
Dr. Mathias Lux, Institut für Informationstechnologie
Prof. Oge Marques, Florida Atlantic University, USA

January 2012

Ehrenwörtliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende wissenschaftliche Arbeit selbstständig angefertigt und die mit ihr unmittelbar verbundenen Tätigkeiten selbst erbracht habe. Ich erkläre weiters, dass ich keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle aus gedruckten, ungedruckten oder dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte sind gemäß den Regeln für wissenschaftliche Arbeiten zitiert und durch Fußnoten bzw. durch andere genaue Quellenangaben gekennzeichnet.

Die während des Arbeitsvorganges gewährte Unterstützung einschließlich signifikanter Betreuungshinweise ist vollständig angegeben.

Die wissenschaftliche Arbeit ist noch keiner anderen Prüfungsbehörde vorgelegt worden. Diese Arbeit wurde in gedruckter und elektronischer Form abgegeben. Ich bestätige, dass der Inhalt der digitalen Version vollständig mit dem der gedruckten Version übereinstimmt.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Unterschrift: ______________________________

Klagenfurt, February 6, 2012

Word of Honour

I hereby confirm on my honor that I personally prepared the present academic work and carried out myself the activities directly involved with it. I also confirm that I have used no resources other than those declared. All formulations and concepts adopted literally or in their essential content from printed, unprinted or Internet sources have been cited according to the rules for academic work and identified by means of footnotes or other precise indications of source.

The support provided during the work, including significant assistance from my supervisor has been indicated in full.

The academic work has not been submitted to any other examination authority. The work is submitted in printed and electronic form. I confirm that the content of the digital version is completely identical to that of the printed version.

I am aware that a false declaration will have legal consequences.

Signature: ________________________

Klagenfurt, February 6, 2012

Contents

List of Tables

List of Figures

Acknowledgements

It is a pleasure to finally thank all the people who made this possible. I would like start with my family, especially my mother and brother, my girlfriend and all my friends who supported me throughout my whole life, my career and my studies. You gave me the strength to cope with all the obstacles which crossed my way.

Next, my thesis would not have been possible without Prof. Laszlo Böszörmenyi, Dr. Mathias Lux, and Prof. Oge Marques. Their support, knowledge, and creative ideas were very valuable to me. Also the great cooperation, and the many challenging discussions were extremely important for the development of this thesis. I also would like to thank all the people I got in touch with at the University of Klagenfurt for the great organizational work and the great support with all the issues and questions regarding my studies and especially before and during my semester abroad at the Florida Atlantic University (FAU).

The knowledge I gathered during my stay in Florida, the people I've met, and the friends I've made, the things I've experienced and the things I've seen - it was a very valuable experience for me on an educational as well as on a personal level. Moreover, I am really thankful for the funding I received from certain institutions and parties for this venture - without your help this wouldn't have been possible for me. I would also like to thank my employer for granting me an educational leave.

Lastly, I wish to thank all my study participants who took the time to fill out my surveys, sit down with me and answer my questions during the interviews and for testing the prototype. Without all your input and constructive feedback this thesis would not have been possible.

Thank You!

Abstract

Searching and retrieving videos in a meaningful way on the web is still an open problem. The integration of a user's context into the search process is one of the most promising approaches to enhance current search interfaces and algorithms. In this paper we present the results of two exploratory studies on the topic: a qualitative study where 22 participants reported on situations when they retrieved and watched videos; and an online quantitative survey with more than 200 participants answering comparable questions. We provide a detailed analysis of the results from both studies and report on the insights that they provide in terms of video search, retrieval, watching, and sharing behaviour.

Also, we propose a software prototype that implements an adaptive video retrieval system, that utilizes the users' intentions to provide better search results and a user interface adapted to the intentions and needs of users. The prototype presented in this thesis is based on our most interesting study results and offers four predefined search intentions. Each intention prioritizes different categories and therefore adapts the search results. In order to evaluate our prototype, we performed a user test with five participants and each one had to solve the same five predefined tasks.

Our findings and the related future work can be used to enhance current video retrieval systems, search interfaces, and algorithms, in order to improve the overall user satisfaction and experience.

CHAPTER 1 Introduction

Multimedia content on the web is very popular these days and still experiences significant growth. Billions of videos are being watched and downloaded each day on various online platforms, services or content providers. Multimedia content, especially video content on the web, is growing massively. The reason for this growth is mainly the emergence of Web 2.0 video platforms, namely *YouTube*[1], *Vimeo*[2], or similar ones. These new generations of video sharing sites allow users to publish videos easily. When uploading new content, videos are automatically converted from many different formats into one uniform format (e.g. H.264/MPEG-4 AVC), making it viewable for a wide range of users and devices. Contributors can categorize videos using predefined taxonomies, enrich them with useful and meaningful titles, descriptions and tags so that others in the community can search, view and share them. Nowadays videos are accessible on various devices, e.g. laptops, TV, mobile phones, which also have a huge effect on accessibility and popularity of online video content. Sharing of videos is also made easier through embedded features on video platforms where users can recommend links to other users, make a list of favourite videos, share them on social networks, or embed them on web pages and blogs. Additionally the capability to rate videos, or engage in a conversation through comments on the video topic brought new social aspects to the whole online video consumption behaviour. The social networking aspect of *YouTube* further enables the formation of communities and groups and it allows to connect videos to users and videos to other related videos. Videos and users are no longer independent from each other and this factor has substantially contributed to the success of *YouTube* and similar sites.

On *YouTube* for example, users generate 60 hours of video content every minute or one hour per second and *YouTube* does have more than four billion video views a day[3, 4]. In the overall traffic of the internet, CISCO forecasts in [5] that video traffic

[1] YouTube: `http://www.youtube.com` (accessed January 27, 2012)
[2] Vimeo: `http://www.vimeo.com` (accessed January 27, 2012)

will surpass global peer-to-peer (P2P) traffic and that the global video community will surpass one billion users, both events should have happened by the end of 2010 already. In 2014, the sum of all forms of video (TV, video on demand, Internet, and P2P) will exceed 91% while internet video (e.g. *YouTube*, *Vimeo*) alone still accounts for 57% of all that traffic. Also in mobile networks the majority of traffic will be video content. Predictions for 2014 are that 66% of all mobile data traffic will be video content as well [6]. A graphical representation of the predicted video traffic is given in figure 1.1, where one can clearly see that internet video constitutes for the majority of traffic all over these years and that it is still growing. This is a strong indicator that video content plays a significant role in future. Video content is growing massively each day and that

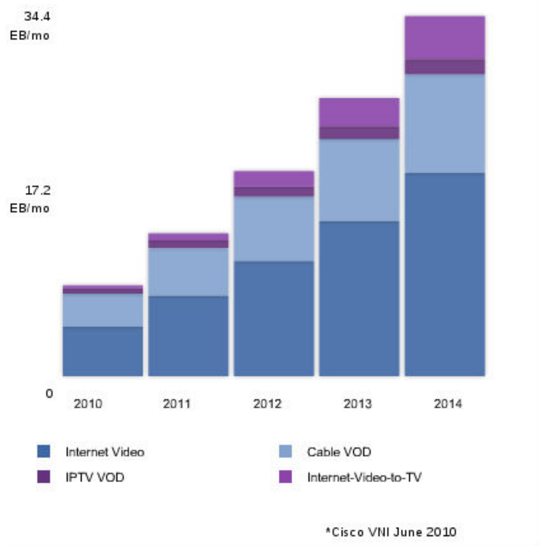

Figure 1.1: Forecast of video internet traffic in exabytes per month from 2010 to 2015[3]

demands research for new methods to organize videos as well as new search techniques. Some features are implemented already to enhance user experience and to tweak search results within video search. When users upload videos to e.g. *YouTube*, they have to add a title and a description for the video and they have to choose between one of the 15 predefined categories[4] *YouTube* provides. A screenshot of the video upload page on *YouTube* can be seen in figure 1.2. Additionally, users can add 'tags' to describe their

[3]Graphic is created via Cisco VNI Forecast Widget, accessible at `http://www.ciscovni.com` (accessed March 19, 2011)

[4]YouTube predefined categories: Autos & Vehicles, Comedy, Education, Entertainment, Film & Animation, Gaming, Howto & Style, Music, News & Politics, Nonprofits & Activism, People & Blogs, Pets & Animals, Science & Technology, Sports, Travel & Events

video with the most important and meaningful terms. It is also possible to add a date and the exact location when and where the video was recorded.

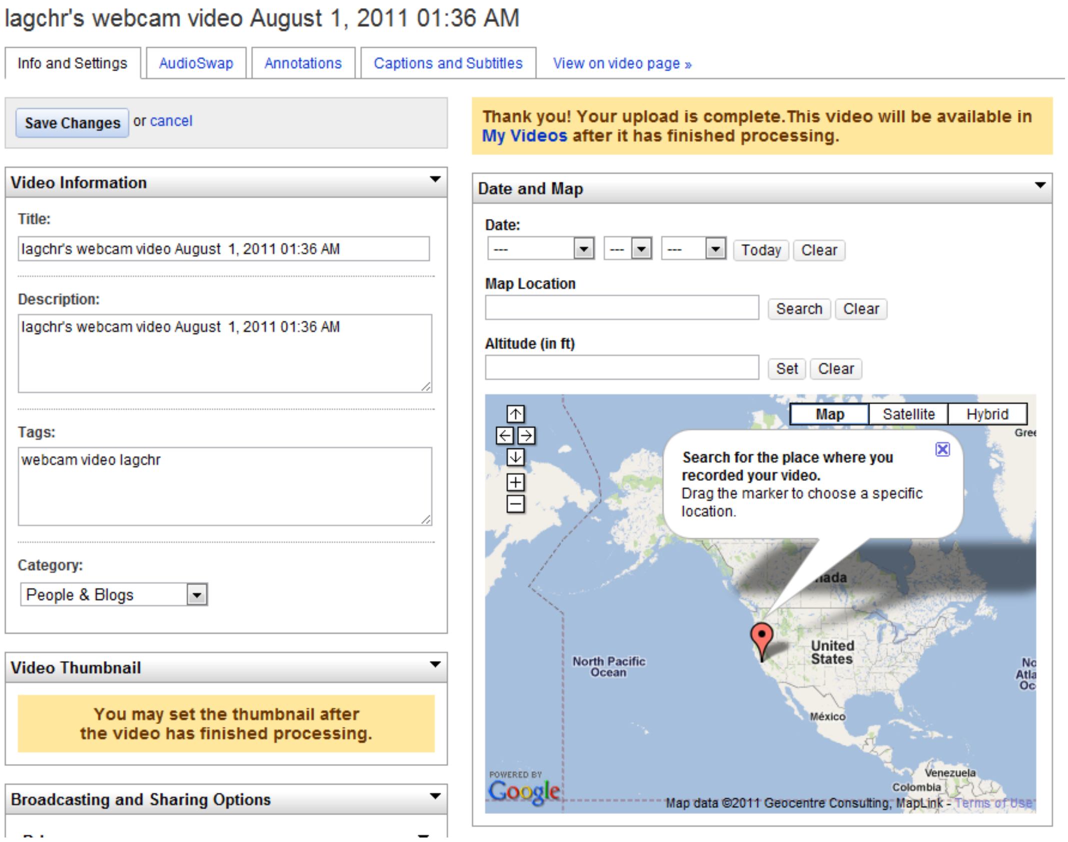

Figure 1.2: *YouTube*'s video upload[5]

If a video is published, other users in the community can *like* or *dislike* a video which results in an overall rating. *YouTube*'s search algorithm takes the following meta-data into account: title, description, tags, number of views and the rating of other users [7]. User's can further constrain their search results when using the 'Filter & Explore' feature. In figure 1.3 a screenshot of the 'Filter & Explore' feature is shown where users can sort and filter the result set or further *explore* a topic through *YouTube* query recommendations.

Results can be filtered on *result types* (Videos, Channels, Playlist), *upload date* (Anytime, Today, This week, This month), *categories* (to the categories in which a video is found based on your search term), *durations* (Short (∼4 minutes), Long (∼20 minutes)),

[5]Screenshot from YouTube's video upload site taken on August 1, 2011

[6]Screenshot from YouTube's 'Filter & Explore' feature taken on August 1, 2011

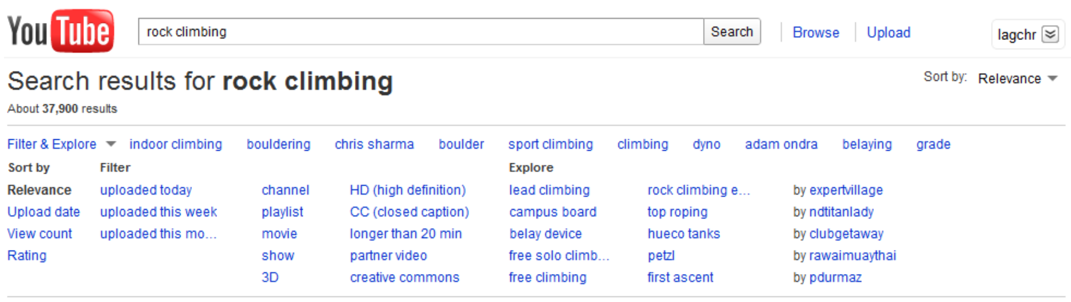

Figure 1.3: *Youtube*'s 'Filter & Explore' feature[6]

features (Closed Caption[7], HD, Partner videos, Rental, WebM[8]), and *ordered by certain attributes* (Relevance, Upload date, View count, Rating).

Since 60 hours of video content gets published every minute[3], this meta-data may not be sufficient any more. These masses of content demand new models, algorithms, approaches and techniques to organize, search, and filter this huge amount of videos. In this work we try to exploit user intentions in addition to other descriptors to make results more meaningful and precise.

We postulate that the integration of a users context into the search process is one of the most promising approaches to enhance current video search interfaces and routines. We anticipate that it is possible to achieve better search results, i.e., higher precision relative to a users context based on intentions, e.g., by dynamically adapting relevance functions used by search algorithms. In general, intentions are an aim that guides a certain action[9]. Almost every action of a person is caused by an intention to accomplish a certain goal. Therefore also almost every search performed on the internet, either on general search engines or more specified search engines like image and video search, are caused by an intention. It is reasonable to take a users intention into account to deliver more precise and meaningful results. Furthermore, a better understanding of queries through intentions driving information need including query reformulation, query extension, and relevance feedback is feasible. A better user experience can also be achieved through dynamic hiding of options and view adaptation based on the users intention.

[7]Subtitles either added by the user manually or added by the Auto-caps feature of *YouTube*

[8]WebM is an open, royalty-free, media file format designed for the web. `http://www.webmproject.org/` (accessed January 27, 2012)

[9]Dictionary: `http://www.thefreedictionary.com/intentions` (accessed February 12, 2011)

In order to better understand users and their behaviour, we conducted two different exploratory studies between October 2010 and February 2011:

- Semi-structured interviews with 22 participants, where we asked questions about general platform usage, online video consuming and sharing behaviour, as well as specific scenarios illustrating why, where and when people watched a video online.
- Online survey with 270 participants, in order to confirm the findings from the semi-structured interviews and to gain further insight into user goals for watching videos online.

1.1 Research objectives

Little is known about the motivations why people watch videos online. In our work we aim to fill the gap between multimedia information systems design and development and user context and behaviour, especially user goals. We do not conduct research in a system-centric way, but rather in a holistic approach integrating relevant actions before and after the video consumption on a specific platform. Questions motivating research in this thesis include:

1. How and why are people using video platforms?
2. What is the actual user's behaviour and motivation to retrieve videos?
3. How do users share their video experience?
4. Are events that lead to video consumption related to video genres?
5. Is it beneficial to translate these findings into a hands-on prototype?

1.2 Thesis overview

Chapter 1 gives an introduction and a general overview on topic of interest. Chapter 2 of this thesis reviews related work from the viewpoint of text-based web search, image retrieval and finally about video retrieval. Regarding video retrieval, particularly of interest are those publications in literature that focus on user surveys as well as literature focusing on the analysis of e.g. the *YouTube* platform and its video metadata. In Section 3 we elaborate on the design and results of our exploratory study. Chapter 4 presents our implementation of the prototype which is based on the findings in our exploratory study. In Chapter 5 we conclude the present work, summarize our findings and discuss limitations. Furthermore we present research challenges, open research questions and possible future work based on the outcome, limitations, restrictions, and scope of this Master's thesis.

CHAPTER 2 Related Work

A person's motivation to use a information retrieval system is often based on a task which needs to be solved. Users need information in order to accomplish a certain goal. They construct a query in natural language and submit it to an information retrieval system. The system then selects a collection of matching documents or artifacts based on the query and displays the result to the user. Research on information systems (text, multimedia), digital libraries, hypertext and information retrieval began early because researchers were one of the first who had access to computer based search tools. Even before the time of computers and the digital age, librarians were observed in their work environment to learn from their manual tools, reference catalogs, indexes and bibliographies when browsed through masses of data, searching for information and answers. Later, these methods got adapted and computerized and were one of the first applications within information retrieval.

Research on information retrieval in the digital age dates back to 1945 when Vannevar Bush's article *'As We May Think'* appeared in the Atlantic Monthly [8]. In this publication, he expressed demand for a machine, some sort of collective memory, that organizes knowledge and therefore makes information more accessible. His vision can be projected on all sorts of technologies we have today like the World Wide Web, online encyclopedias and similar services. The term *information retrieval* was then coined by Calvin Mooers [9] in the 1950s. Since then, researchers and academia began to see information retrieval systems as a solution to many problems regarding knowledge management, knowledge retrieval and organization. In [10] a good summarization and overview on the history of information retrieval, divided in seven ages ranging from *schoolboy phase* to the *retirement phase*, is given.

Our focus within this work lies on user intent research in video retrieval, which can be explored with the following methods [11]:

- Empirical studies and surveys,
- manual analysis of query logs, and
- automatic classification of web searches.

Going deeper into the topic of information retrieval (IR), we analyse basic literature on text based web-search, especially focusing on user intentions. Understanding *why* a search is performed is essential to satisfy the users information need. After that, literature on multimedia retrieval is explored. First we focus on image retrieval, because it is closely related to our main research area. Second and last we explore existing literature on our main research area, namely video retrieval.

2.1 Text-based Web Search

When people use an information retrieval system, they are often driven by some sort of information need. Either they need to solve a task or acquire new knowledge on a specific topic or the like. Andrei Broder [12] investigated possible intentions behind web queries. Broder identified a taxonomy of web searches which is grouped in the following three classes:

- **Navigational queries:** Users don't always remember URLs off the top of their head but they know that the page they want to visit exists on the internet. So the intent behind such queries is to find out the URL for a particular page via a search engine. In classic IR such queries are sometimes referred to as "known item" search. Normally, such queries do have just one correct solution.
- **Informational queries:** This is the most common form of retrieval in classic IR and it is a persons desire to gain new information by using exact or related keywords as search terms. The result or solution is either one page with all the necessary information or more pages relating to the topic of interest. Informational queries vary in their broadness because some informational queries are very wide (e.g. houses) and some are very narrow (e.g. Hiybbprqagate[1]).
- **Transactional queries:** A transactional query is given when a user gets directed from a result list of a query to a site where one is able to perform further interaction to reach his goal in mind. Examples for such queries are: *(i)* find an online-shop for a certain product, *(ii)* find an international phone book, *(iii)* find a site to compare flights, or *(iv)* download videos or images.

Two methods led to this taxonomy with three classes. First, Broder performed a user survey, where a *pop-up* window was shown to random users during a search session. Users were asked to answer certain questions on what they were trying to do, why they performed this search, and what they were looking for. Informational queries were defined as queries that are neither transactional, nor navigational. This resulted in 24.5% of navigational queries, more than 22% of transactional queries (estimated 36%) and an estimate of 39% as informational queries. The second method they used was an analysis of a daily AltaVista[2] log. They selected a random set of 1,000 queries consisting of English and non-sexually oriented queries only. The results of this analysis where 20% of navigational queries, 48% of informational queries and 40% of transactional queries.

[1] Explanation of "Hiybbprgagate": `http://blog.searchenginewatch.com/110204-092959` (accessed May 17, 2011)

[2] AltaVista: `http://www.altavista.com/` (accessed May 16, 2011)

Both results indicate that less than 50% of the queries typed into search engines are informational in nature.

Similar research was done by Rose and Levinson [13] where they proposed a framework for understanding the underlying goals of user searches. They came up with a stable and comprehensive framework of main goal categories and sub-categories. This was accomplished by looking at a sample of queries from the AltaVista search engine, own experiences and previous query analysis at AltaVista. Three iterations of classification tests, each time with new query samples, were made. Goal categories were refined during every iteration. The result was a hierarchical structure where the top level of the hierarchy resembles Broders taxonomy [12] but they exchanged *transactional* queries with a more general *resource* class. So the three main search goals are *Navigational*, *Informational* and *Resource*.

If a user already knows that there is an URL for a specific website of an institution, a brand or an organization, he seeks a *navigational* goal. In this category, no child-classifications are available. The *informational* goal is subdivided in *'Directed'* goals where a user wants to learn something about a particular topic and *'Undirected'* goals where users want to learn anything about a topic. *'Directed'* goals could be *'Closed'* (only one unambiguous answer) and *'Open'* (answer to an open-ended question or one with unconstrained depth). Furthermore, users are probably searching for *'Advice'* (ideas, suggestions, or instruction), or they want to *'Locate'* some service or product in the real world. The last subcategory of *'Informational'* is *'List'*, where the user gets a list of plausible websites, which can help to solve a rather unspecified goal. In the *'Resource'* classification the goal is to obtain a resource which is available on web pages. The subcategory *'Download'* represents the goal where a user downloads a resource like an image, video or program for further usage on a computer, while the *"Obtain"* goal is reached when the resource does not require a computer to use. The *'Entertainment'* goal is achieved when the desired resource is simply for enjoyment. *'Interact'* can be compared to Broders *'Transaction'* class where users search for a service to perform further interactions.

Three sets of 500 randomly selected queries from different days at the same time were used to manually classify the queries with a user interface where a participant could select one of the goal categories as explained before. In the final analysis, nearly 40% of queries were non-informal. In most cases of informal queries, users tried to locate a product or service instead of wanting to learn about it. Queries with a research background (e.g. advice seeking, undirected questions) constituted for over 35%. All three sets of queries showed almost the same distribution. Compared to Broder, they found a greater proportion of informational queries and a smaller proportion of navigational/transactional queries. The difference in the informational and transactional class can be explained by the different definitions of those categories whereas the difference in the navigational queries are due to the fact that Broder sampled all queries instead of just session-initial queries.

Another approach to classify user intent was done by Jansen et al. [11] where they present a methodology to automatically classify the user intent. They were the first who tried to automatically classify queries in a large scale environment with 1.5 million queries from a Web search engine log. Each query presents certain characteristics, which were identified in order to properly classify them according to a taxonomy almost identical to Broders taxonomy [12] or Rose and Levinson's hierarchical taxonomy [13]. Characteristics for each class were based on special keywords in the search term, query length, session length, number of query reformulations, and result pages viewed. To be more precise, some of the following characteristics for query identification were taken into account:

- **Navigational:** Names of companies, business, organization, and people; domain suffixes; queries containing the word "Web"; number of query terms less than 3; and user viewing the first results page.
- **Transactional:** Queries containing terms related to movies, songs, lyrics, images etc.; queries with terms related to 'obtaining', 'download', 'audio', 'entertainment', 'interact'.
- **Informational:** Use of natural language phrases, use of question words like 'what is' or 'how to'; users viewing multiple result pages; number of query terms greater than 2; queries which are not navigational or transactional.

Automatic classification of all these queries resulted in 80% of informational queries, and 10% for each navigational and transactional queries. A manual classification of 400 sample queries showed that their automatic classification has an accuracy of 74%. Lee et al. [14] automatically identified query goals using *past user-click behavior* feature and *anchor-link distribution* feature and classified it as either *informational* or *navigational*, based on the taxonomies identified in [11, 12]. Due to the lack of consensus in [11, 12] on the third class, namely *resource* or *transaction*, it is not covered in [14]. *Past user-click behavior* learns from actions of past users. If a goal is *navigational* then users mostly clicked on a single website they have in mind, if it is *informational* users may have clicked on many result items, this is referred to as *average number of clicks per query*. The other measure for this feature is referred to as *click distribution* and measures how often a user has clicked on an answer. If the click frequency for a certain result item is high and for all others low, then the goal is probably of navigational nature, if it is low or flat and identical for most of the result items the goal is probably of informational nature. The second feature is called *anchor-link distribution* which observes destinations of links with the query keyword as anchors. To accomplish this, all HTML links with the same anchor text as the query are extracted, analysed and counted. If a dominating portion of these links point to a single authoritative website the query is expected to be *navigational*, if it points to a number of different destinations it is expected to be *informational*. Link spams and mirror sites could distort the results of that feature and introduce undesirable noise. Each of this features resulted in an accuracy of about 80% and the combination of these two features resulted in 90% of correct identified goals.

Users express their goals while searching the web differently, often they start with an initial query formulation, if the result is not satisfying enough they refine and alter it, generalize it or even use a different goal formulation. Strohmaier et al. [15] analysed selected search sessions from a search engine log in order to gain insight into how users express their goals and how these goals can be represented in a semi-formal way. Based on one selected search session of a user, they represented a search goal as a semi-formal goal graph based on the agent and goal-oriented modelling framework i* [16]. Using such goal graphs in search would be a big benefit for users, because suggestions for refinement or pointing out related goals on a topic could be made and further support the user to reach a goal. An open research question is still the creation of large-scale goal graphs. Suggestions which could help to overcome this issue would be usage data analysis, user involvement (explicit/semi-automatic) or collaborative model construction efforts.

2.2 Image Retrieval

Image sharing platforms are widely used and very popular throughout the online community. Therefore analysing and understanding such platforms can significantly improve image retrieval platforms for the benefit of the user. To better understand image sharing platforms, the authors in [17] were interested in the growth and popularity of the *Flickr*[3] online social network. The authors examined, analysed and evaluated detailed growth data and link formation processes. Data was collected over a three months period and encompassed 950,143 new users and over 9.7 million new links, covering 58% of growth in *Flickr* user population and over 63% growth in the number of links. Growth characteristics in *Flickr* showed that link additions exceeded link removals in their dataset at a rate of 2.43:1. Additionally, links tend to be created by users who already have many links, also referred to as the *'rich get richer phenomenon'*. By using *least-squared-error linear fit*[4] analysis for link creation, they found that on average users create one new

[3]Flickr: `http://www.flickr.com` (accessed January 27, 2012)

[4]Linear least squares is an approach to fitting a mathematical or statistical model to data in cases where the idealized value provided by the model for any data point is expressed linearly in terms of the

link per day for every 227 links they already possess, and that users receive one new link per day for every 370 links they already possess. Another interesting fact showed that users tend to quickly respond (83% within 48 hours) to incoming links by creating links back to the initial source. Examining the shortest path distance between newly created links shows that over 80% of new links connect users that were only two hops apart. That means that the destination user was a friend-of-a-friend of the source user for the newly created link and that it is far more likely to link to other users who are already close in the network.

More insight into the friendship graphs and user interactions in *Flickr* are given by Valafar et al. [18] by examining user-specific attributes (posted photos, associated fans and their arrival times, favourite photos and associated owners) of 122.000 random *Flickr* users. In their data-sample 3.482.000 photos were posted by those users. The authors characterized indirect interactions (i.e., relationship) between fans and owners of photos. Their results showed that only a very small fraction of users in the main component of the friendship graph is responsible for the vast majority of fan-owner interactions and that these interactions involve only a small fraction of photos in *Flickr*. More specifically, a very small fraction of users are fans of a very small fraction of photos which in turn are owned by a very small fraction of users. The vast majority of fan-owner interactions ($> 95\%$) are between a small fraction of users in the largest component of the friendship graph. Additionally, active users appear to form a core in the interaction graph and there is a clear correlation between the level of activity of a user as a fan and as an owner. While older photos can reach higher popularity, there is no strong correlation between age and popularity for a majority of photos. Newer photos appear to reach their target popularity much faster than older photos. A majority of fans for a photo will be received during the first week, whereas older photos experience a lower average fan arrival rate simply due to a longer inactive period. Looking at temporal properties, the authors show that there is no strong correlation between the popularity of a photo and its age, meaning that knowing the popularity of a photo does not provide any strong indication about its age and vice versa. Most photos gain a majority of their fans during the first week of their posting, but older photos have more opportunities to attract fans and are more likely to have a higher popularity.

Several papers have already been published ranging from image production [2] to a classification scheme for user intentions in image search [19]. Research for intentions in image retrieval can be distinguished in three categories, namely production, annotation

---

unknown parameters of the model. The resulting fitted model can be used to summarize the data, to predict unobserved values from the same system, and to understand the mechanisms that may underlie the system. (`http://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)`) (accessed January 27, 2012) and consumption. An initial exploratory study on user intentions in image retrieval was done by Kofler et al [1]. Ten different retrieval tasks were prepared, each one was specified to match one of the classifications according to Jansens taxonomy [11]. Users participating in this study were asked to use *Flickr* to solve each goal. User requests and responses to *Flickr* were logged via reverse HTTP proxy and Apache TCPMon in order to reconstruct click path and to perform further evaluation. Based on their evaluation, Kofler et al. assumed that in the context of multimedia retrieval the query terms do not contain the intention of the user, moreover search terms express and describe the content of the image. Users indirectly show their intention through their search behavior like number of clicked images and average duration time to solve the task. They concluded that tasks of directed informational character and advises are rather short in duration (duration $\leq$ 2:00) and only a small amount of images was clicked (≤ 2), on the other hand undirected informational tasks are rather long in duration (average 4:10) and a high number of viewed/clicked images (average 8.5). Informational "list" tasks identified themselves as rather short search times (duration $\leq$ 2:30) and a large number of images clicked (between 3 and 5 images). Navigational tasks lasted longer (between 3:00 and 4:00) and had a rather large number of clicked images (3.0 and 3.2). They also state that existing taxonomies in information retrieval [11, 12] are not sufficient to describe user intent in image retrieval and that a new model is needed. Lux et al. [**?**] proposed a classification scheme for user intentions in image search, based on the work of Kofler et al. [1]. In a user experiment, 20 users (students and co-workers) were asked to solve ten tasks regarding image search, afterwards quantitative interviews were performed. Intentions are rarely expressed on query level, but intentions are expressed implicitly through search behavior. Broders taxonomy [12] was adapted to fit the needs in image retrieval. Another set of studies was performed with nine people who use *Flickr* intensively to discuss open problems and evaluate the work completed so far. The taxonomy in image retrieval is different because classes can overlap. The proposed taxonomy with the overlapping classes can be seen in the Venn diagram in figure 2.1. After gathering the results of the users studies, a renaming of classes was performed: *informational* class was renamed to *knowledge orientation*, *navigational* to *navigation*, *transactional* to *transaction* and a new class *mental image* was introduced. The classes itself were defined in more detail because they include the users need prior to search, during search and the characteristics of the search result. Within the *knowledge orientation* class a user wants to gain knowledge from the images and he needs to extract information to answer a question. In the *mental image* class they assume that a user has a certain picture in mind and he needs to match this image to the retrieved images. If a user knows that images of his goal exist but he does not know about the visual contents they are falling into the *navigation* class. The last class is *transaction* where a user wants to

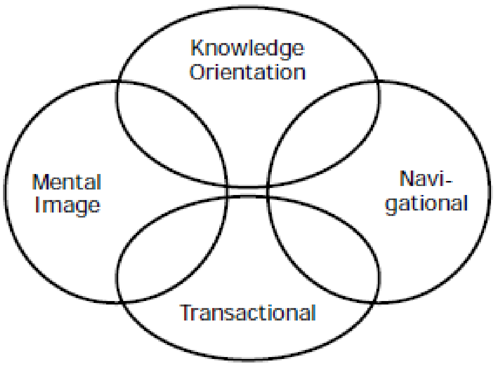

Figure 2.1: Taxonomy for intentions in image retrieval taken from [1]

find a picture to download or buy it for further use.

Lux et. al [2] investigated user intentions on why arbitrary people take pictures with arbitrary devices. The authors wanted to find out the actual and explicit motivations during individual face-to-face qualitative interviews without focusing on a specific classification scheme. Ten participants (7 male, 3 female) reported on 40 different specific photo taking situations, at which every participant had a different experience in photography. Afterwards, a classification scheme for image production scenarios was created. People always take pictures for a reason, either (i) for private use to remind oneself of a special moment, (ii) to share pictures with friends or family to tell a story, or (iii) for public use so that others can conceive an opinion on the photos. Regarding the 40 scenarios of taking photos, all participants reported affective use (28 affective scenarios), whereas functional use has been reported by 6 out of 10 users resulting in a total of 9 situations. Three situations could not be classified due to ambiguous information. Participants who reported functional intention scenarios mostly (3 out of 5) owned a mobile phone camera. In only 7 reported scenarios an individual purpose was the trigger to take a photo, whereas in 29 situations the photos were presented to other people, the remaining situations where ambiguous and could not be classified. In figure 2.2 the authors show a classification scheme for image production scenarios with generic devices. The dashed line indicates a possible overlap between social and individual intentions as people want to first share and afterwards review the pictures on themselves. Also the authors hypothesize that there is a possible overlap between functional and affective intentions because a users stated that he wanted to take a good photo, but at the same time also wants to test his camera. An additional fact is that people tend to use the best camera available in terms of image quality regardless of the intention. It is assumed that behind most photo taking situations an implicit individual intention is present. One big

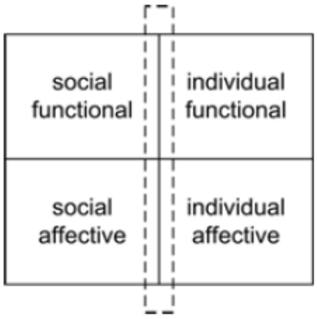

Figure 2.2: Classification scheme for image production scenarios [2]

difference to the intentions in image retrieval is that emotions are a key factor for the intention to take a photo. No one reported to search for pictures out of certain emotion, but when capturing special moments to remember emotions play a big role.

2.3 Video Retrieval

Multimedia retrieval, especially video retrieval is one of the fastest growing and most important research areas in multimedia technology nowadays. Lots of challenges are still ahead, one of the biggest is still bridging the semantic gap, thus connecting the low level features of images and videos to high level features perceived and interpreted by humans in a given situation. Current efforts to bridge the semantic gap are still not sufficient enough to increase retrieval performances significantly. Sebe et al. [20] discuss remaining challenges and problems within a literature study. The main areas of interest are mostly related to semantic e.g. the semantic gap, understanding the semantic meaning and interpretation of the emotional subtone behind a query, as well as the inclusion of personal preferences of users. Discussions on research regarding video indexing, summarization, analysis, annotation, and browsing in addition to the human computer interaction issues in visual information retrieval are conducted. We will focus on various research approaches and cover literature on surveys regarding video retrieval and user behaviour as well as analysing log files and network traffic of known video platforms. Technical in-depth analyses of log files and video metadata are a different method to explore user behaviour and cover other relevant aspects of video retrieval that cannot be covered by surveys or interviews. User studies on mobile video watching behaviour are already existing [21, 22, 23] but none of them covers the broader area of general user behaviour on video platforms while consuming online video content. Our exploratory studies cover a different set of users, we have a broader user community in mind where we want to find out in more general why people watch videos online and what their motivation is. Furthermore, we investigate and analyse usage patterns, sharing behaviour and specific video watching scenarios of users. For more details see the exploratory study in Chapter 3. Nevertheless, previous research in this field, even it is not explicitly about user intent, are informative for us.

Query logfile analysis was done by Jansen et al. [24] to identify characteristics of user queries for multimedia content and to point out differences and similarities between general search queries. A data set from Excite[5], a text-based web search engine, containing 1.025.908 queries from 211,058 different users was analysed with regards to image, audio and video queries. Based on terms related to multimedia content (e.g. 'audio', 'concerts' 'photo', 'movie'), significant multimedia queries were identified: Audio 0.37%; Video 0.75%; Image 2.65%. The authors findings were, that the terms used per query is generally higher for multimedia searches (median 3 terms for video and image, 4 terms for audio) than for general web searches (median 1 term). Also the amount of queries per user differs to general web searches with a median of 2.85 queries, audio searches have a median 2.44 queries, video 2.91 queries and image 3.27 queries. Therefore, multimedia searches are generally longer than general web sessions and queries.

Hopfgartner [25] examined user interactions using different video retrieval interfaces. They evaluate different implicit indicators which may be used to improve retrieval performance by inferring the users intention. Implicit feedback from users should be used to reduce the semantic gap and fulfill the users information need. Based on this work, Hopfgartner and Jose [26] evaluated implicit feedback models for adaptive video retrieval. Using a simulated evaluation methodology, five different interfaces with a combination of different implicit feedback mechanisms were examined and evaluated regarding their performance. For the simulated test runs the 2006 TRECVID dataset was used. The authors identified six implicit feedback categories:

- **Highlighting:** Moving the mouse over a keyframe to show additional information (text/neighboring keyframes). This indicates further interest in a keyframe.
- **Click on a keyframe:** Results in playback of a video shot or perform other actions. Indicates interest in the video which is represented by the keyframe.
- **Using the sliding bar:** Users tend to use the sliding bar when the initial shot is not the one they wanted but they think that the rest of the video might contain other relevant shots.
- **Looking at video metadata:** Indicates that users show a higher interest in the current shot.

[5] Excite Web Search: `http://www.excite.com` (accessed March 19, 2011)

- **Browsing through a video:** Is similar to using the sliding bar. This feedback expresses higher interest in the shot by the user.

- **Playing duration:** Expresses interest in the content of the video.

The authors designed five user interface scenarios consisting of different combinations of the six implicit features, one also supported explicit feedback from the user. Based on a simulated user study with a set of implicit actions generated randomly, the authors conclude that implicit features do have an impact on video retrieval results. Those models who only supported a few implicit features showed the worst results in the simulation. Two models supported the *browsing through a video* and both models returned better results than others, therefore this feature for example boosts retrieval results. The model with explicit feedback feature support in combination with implicit features returned the best results, therefore the authors concluded that a combination of explicit and implicit features show a better overall performance of retrieval results.

Analysing user behaviour in large video databases similar to *YouTube* was done by Qiu et al. [27] and Mongy et al. [28]. In [27] the authors provide an analysis of user behaviour in online streaming, using a trace database from MSN video from the year 2007 and 2008 covering 5,000,00 video streaming sessions. The authors identified behaviour patterns and uncover relationships between those patterns. Seven behaviour patterns (Start, Stop, Jump forward, Jump backward, Replay, Pause, Return) and their estimated transition probabilities were reported. They also found that video popularity and the behaviour patterns in previous sessions affect user behaviour in general. Users typically spend more time on popular videos than on less popular ones. Also some of the behaviour can be predicted for future sessions by analysing patterns in previous sessions using transition probabilities. These measures can be used to improve caching of video material.

Mongy et al. [28] analysed the behaviour of users in the context of movie production of large video warehouses and used video usage mining to suggest a two-level model based approach to model user behaviour on intra- and inter-video level. A main task in movie production in the professional sector is to find existing video sequences in order to reuse them in the creation of new films. This strongly relates to the online streaming behaviour of users. The authors want to understand which, how and why each video has been viewed by users and define two behaviour units, namely *video sequence viewing* and *session*. A *video sequence viewing* unit defines the behaviour of a user while watching a video (intravideo level) while a *session* defines the behaviour while browsing for videos of interest (intervideo level). The outcome will be used to improve the quality of retrieved data by analysing the behaviour of the users and creating user profiles. The clustering method used is an adaptation of the *k-means* algorithm which uses models instead of means. First-order nonhidden Markovian models[6] are used because they are useful in the prediction of requests and in the prediction of the next action to be requested. It was found for example, that when a user is playing a video there is a probability of 8% that he pauses the video in the next second. Additionally they showed that their algorithm is able to differentiate sessions dealing with the same video and results but in different manners and contexts and that there are clusters of videos that are not detected by basic subsequence approaches but with their approach.

Miyauchi et al. [29] explored people's different behavior on live mobile TV and mobile video usage in Japan by performing qualitative interviews with 11 participants. They compared usage on commuter vehicles, usage at home and usage in experience-sharing and found differences in attitudes concerning mobile TV usage and mobile video. They observed that Mobile TV users on commuter transport preferred light entertainment after work to relax, whereas mobile video users watched content which they prepared beforehand (e.g. the night before) because they want to use the time while commuting for a useful purpose (e.g. watch how-to/education videos) but also they watch entertainment programs. When being at home, both user groups do not like to watch content on small screens of mobile phones or similar devices, but sometimes mobile TV users sacrificed larger displays (TV, monitor) for the convenience of a portable viewing experience. Watching programs together (while commuting) was more convenient for mobile TV users because they did not prepare the content for specific purpose and therefore did not have to consider any privacy problems. Mobile video users instead prepared the content to their needs and this content may be private to some extent.

Several other user studies concerning mobile TV or mobile video usage in various countries exist and where conducted in Finland [30] [31], US and UK [32], Korea [21], and Australia [33]. There are significant differences in users behavior among those studies, due to cultural and lifestyle reasons. The results of the study conducted in Finland by Södergard [30] showed that people prefer to watch short television programs or pieces of longer programs while waiting for something or just to pass time. This was not restricted to a specific task or location, people did this for example while waiting for the bus in public or at home. The main advantage was considered to be availability - anywhere, anytime. Repo et. al. [31] found that the private experience by using a mobile phone to watch videos is also an important factor. Also in the US and UK habits are different. People prefer to watch light entertainment (e.g. cartoons, comedy) instead of videos covering self-development or work topics during commutes. One reason could

[6]A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. (`http://en.wikipedia.org/wiki/Hidden_Markov_model`) (accessed January 27, 2012)

be that public commuter transport is not used as much in the UK and US as in other countries. In Korea mobile TV devices are often used to avoid boredom, to stay up to date with current events, or to play games. People also use their mobile TV devices much more at home, even if a big TV screen is available.

Another survey [34] gives a comprehensive overview on the state of the art and open issues for mobile TV, focusing on the users' needs and experiences. The authors combined two research methods to analyse the results obtained by manifold studies on Mobile TV from a users' perspective: Human Centered Design with the Quality of Experience (QoE). Subject of research was the user itself, including his motivation, future trends and needs, as well as the mobile device and the context where, when and how long Mobile TV is consumed. Furthermore they investigated which content is consumed, and the technical performance (regarding to audio, video and transmission) of the service. They summarized some open issues in the field. First, users are of no typical gender, it highly depends on the content and service. Users show an urge to share and modify content on their mobile devices. Open problems are often fee types of services as well as the technical performance. Sometimes the quality is not good enough (especially for sport broadcasts) and the channel switching response time is too high.

O'Hara et al. [22] performed ethnographic interviews and participating users also had to keep a diary about their usage. They highlight a range of motivations and values when using mobile video in different settings and circumstances. They point out how mobile video affects peoples every day life and how they use it besides just watching TV content anytime and anywhere just to pass time. Even though consuming video in mobile devices is a privatizing technology, it also might facilitate togetherness in the home as people can watch their own content while being close to friends and family. It also allowed participants to bring content via mobile phone to social situations and places where conventional TV was not available. Finally, the authors report that users found it frustrating to copy video content on mobile devices and conclude that a better integration among devices is needed.

Another research direction related to this topic is the analysis of query log files or datasets providing metadata of videos (rating, number of views, external links etc.) of various video platforms or content providers. In [35] the authors analyse online video search and sharing behaviour on *YouTube*. They used a web crawler and the *YouTube* API to create a database with all the metadata and statistics provided by *YouTube*. Using this database, they analysed people's search behaviour and determined what makes some videos more popular than others. Videos with longer titles, better description and more meaningful tags receive more views than videos with less metadata, therefore this information is an important factor to improve video popularity. A video receives the highest number of views and clicks in the first few days after the upload and generally after that the popularity decreases. The number of views of a video does not follow a Zipf distribution as observed in other types of web surfing paradigms like desktop and mobile browsing contexts where very few pages have a large number of hits and many pages have very few hits. Distribution of *YouTube* video views suggest that users in this scenario actually visit more pages within the site or within a session. However, when the distribution of uploads and number of favourites was investigated a Zipf-like distribution was indeed found. The same authors also explored the social dynamics and the social interactions between users of the online media sharing platform *YouTube* [36]. Key figures suggest that (1) more users tend to visit the site to view videos rather than uploading new ones; (2) users tend to save videos rather than uploading new ones; (3) many users do not exploit social features like subscriptions, adding friends, commenting etc. A relatively low level of individual participation in *YouTube* is shown.

Cha et al. [37] performed an extensive data-driven analysis on *YouTube* from the perspective of user-generated content (UGC). Subject of the study was to analyse the popularity distribution and evolution for various categories and videos and their relationship with video age. Additionally they provide insights into current UGC distribution systems and propose more efficient designs, namely caching and peer-to-peer designs. In the case of popularity evolution, they showed that if a video does not receive enough requests during the first days it is unlikely that it will get many requests in future. Predicting near-future popularity would be useful for service providers to cache videos before they get popular to overcome bottlenecks. By comparing the first few days of video views with views after a certain time (5, 7, 90 days) the results showed that the second day record gives a good estimation about future popularity with a relatively high accuracy (correlation coefficient above 0.8) even for predicting distant future popularity up to three weeks. Another study [38] presents a systematic and in-depth measurement study on the statistics of *YouTube* videos. By examining 3 million *YouTube* videos the authors found differences to traditional video systems regarding length of videos, access patterns, popularity trends and active life span. They argue that the social network of *YouTube* videos substantially contributed to the success of this service.

A deeper investigation of user roles (namely *Standard*, *Director*, *Comedian*, *Guru*, *Musician*, or *Reporter*) and user aggregated behavior in *YouTube* was done by Biel and Gatica-Perez [39]. Special user roles (all but *Standard*) tend to interact more with the *YouTube* community (through views, subscriptions and comments) and share videos of a wider interest, whereas *Standard* users at most share videos with some relatives or friends and therefore also gain much fewer views. Special user roles are also getting more attention from the *YouTube* community based on views and subscriptions. Gender specific aspects show that more men (73%) than women (27%) are using *YouTube* and it is also more likely that men are enrolling in special user categories (13% vs. 9%).

Usage of social-oriented features show very different numbers. Both genders participate in a similar way regarding uploads and videos watched per user, but on average women with special *YouTube* user roles have a more social-driven behaviour than man. They have 22% more subscribers, 75% more subscriptions, and double the median number of friends that man have. The authors have shown that the *Standard* user role and special user roles are different in their participation and interaction patterns (that is how social they are). A prototype showed that an automated classification of users roles solely based on the user profile can be made with a 68.9% $\pm$ 1.2 classification accuracy rate (CAR).

Benevenuto et al. [40] analysed *YouTube*'s social network by video interactions to uncover typical user behavioural patterns. Additionally they point out evidence of anti-social behaviour such as self-promotion and other types of content pollution. Effective video content classification can be used to identify videos with malicious content such as spam, pornography or copyright protected material. *YouTube* allows video responses to existing videos which establishes some kind of conversation and discussion through video. They showed network characteristics in a directed network with in-degree (k_{in}) and out-degree (k_{out}) distributions. In *YouTube* networks the in-degree exponent is smaller than the out-degree exponent, which suggests a link asymmetry in the directed interaction network which is similar to web graphs. In a Web graph, pages with high in-degree tend to be authorities and pages with high out-degree act as hubs directing users to recommended pages [41]. On the other hand, other social networks exhibit a significant degree of symmetry [42]. The *YouTube* interaction graph also does not show strong evidence for the formation of a social community. Normally social networks posses a topological structure where nodes are organized into communities, measured through the values *clustering coefficient* (CC)[7]. A CC equals 0.047 shows the presence of small communities in the whole video-response network. Also the higher values of the clustering coefficient occur among low-degree nodes, which suggest the lack of large communities around high-degree nodes. Low-degree nodes might explain the formation of very small communities, composed of a few people like a family or a group of friends.

Content pollution or spam in social video networks occur when a video is posted as a response to an opening video but its content is unrelated to it. This is just done to boost its own ranking to make it highly visible in the social network. Video spam takes up more system resources (e.g. bandwidth, storage, ...) than textual spam or spam in comments and it also compromises user patience and satisfaction with the system as

[7] Clustering coefficient of a node i, $cc(i)$ is the ration of the numbers of existing edges over the number of all possible edges between i's neighbours. The clustering coefficient of a network, CC, is the mean clustering coefficient of all nodes

a whole. With a measure called Inter-reference distance (IRD)[8], the authors examine temporal patterns of the users participation in a sequence of video responses. Users that upload many responses in a short time span (like a mechanical process) might be a candidate for further investigation. A large number of video responses per video and a small IRD points to unsocial behaviour and can be used in heuristics to combat anti-social behaviour in video interactions. Using an adapted PageRank [43] algorithm named UserRank, which uses the video response user graph, the authors determine the importance of users in terms of their participation in video interaction. The authors found a strong correlation between the UserRank and the total number of views. Most of the high rank users are *Directors* or other special user groups and therefore rank among the most responded and viewed while on the other hand low rank users have a small number of views and post video responses to some videos but receive no or few responses from the video community. Therefore, UserRank can be used to detect users boosting a video ranking, namely for self-promotion and other forms of content pollution.

Similar research was done in [44], where the authors used *YouTube* to understand the characteristics of internet short video sharing. They used traces crawled in a 3 month period in early 2007 and collected data about 2,676,388 videos and more than 1,000,000 users to present an in-depth and systematic measurement study on the characteristics of *YouTube* videos and its social network. Focus of the analysis were length and access patterns and the active lifespan of videos, ratings, and comments. The authors noticed that there are noticeable differences compared to traditional streaming videos. In their dataset, the most popular categories are *Music* with 22.9% of all videos, second is *Entertainment* (17.8%), followed by *Comedy* (12.1%) and *Sports* (9.7%). The biggest difference to traditional media content servers are the video length. In their dataset, 97.8% of the videos are within 600 seconds, and 99.1% are within 700 seconds whereas in traditional media content servers the video length is typically 30 minutes to 2 hours. It should be mentioned that the numbers from this study highly depend on the *YouTube* restrictions at the point of time where the analysis was performed. Restrictions were being changed by *YouTube* afterwards and for example since 2010 the 10 minute restriction was extended to 15 minutes per video[45] and since September 2011 *YouTube* even rolled out a feature which lets verified users upload videos longer than 15 minutes[46]. The distribution of video lengths on *YouTube* exhibits three peaks: (1) is within one minute and contains more than 20% of the videos (2) between 3 and 4 minutes with 16% (3) near the maximum of 10 minutes. The first peak indicates that *YouTube* is mainly for short videos, the second peak can be explained because *Music* is one of the most popular category within *YouTube* and music videos are within this range, and the third

[8] Inter-reference distance: Sequence of users that upload video responses to video i as the total number of responses that appear between two video responses from the same user

peak is caused by the length limitation for *YouTube* videos which was 10 minutes at that time. Some users split longer content into several parts to circumvent this 10 minute restriction. Deeper analysis of video length and their corresponding categories showed that *Music* videos have a very large peak between three and four minutes, *Entertainment* has a similar but smaller peak and *Comedy* as well as *Sports* videos have a peak within two minutes, which probably corresponds to 'highlight' type clips.

In the dataset, the file size of 190,000 videos was retrieved and analysed and it showed that 98.8% of videos are less than 30MB. Not surprisingly the distribution of video sizes is similar to the distribution of video length. Also the bit-rate of videos showed three apparent peaks, (1) most videos have a bit-rate around 330 Kbps, (2) around 285 Kbps, and (3) around 200 Kbps. This is an indication that *YouTube* balances quality and bandwith by having a moderate bit-rate for most of their videos. An analysis of the views and ratings of videos showed that, as many other researches have already shown (e.g. in [47]), the number of views in correlations with its video rank does not follow a Zipf distribution which indicates that there are not so many less popular videos as Zipf's low normally predicts. Investigation of the social network in *YouTube* showed that 58% of users have no friends. Additionally, the authors found small-world networks in *YouTube* videos (main video and their related videos) which indicates that the *YouTube* network of videos is a much closer group. Small-world networks refer to the principle that objects (people/videos) are linked to all others by short chains of acquaintances (six degrees of separation). It describes networks that are neither completely random, nor completely regular, but possess characteristics of both. The clustering coefficient is still large as in regular graphs, but the measure of the average distance between nodes is small as in random graphs [48]. *YouTube*'s network of related video lists definitely show small world characteristics.

Possible enhancements of social sharing on video platforms, especially on asynchronous video manipulation, was investigated by Cesar et al. [49]. The authors present a more social approach to share and view media. An architecture and implementation for content selection, (re)organization and sharing within a heterogeneous user community and devices is explained. Some users want to share their video experience in a more personal and differentiated manner and want to edit or enrich existing video content to share their version of the video with different social groups like family or friends. Social sharing features proposed by the authors are: (1) *Fragment.* One or more excerpts of an existing clip; (2) *Annotate.* Add user-generated notes or comments in form of audio, text, image or the like to a fragment; (3) *Enrich.* Add new temporal links, captions, subtitles to the new video object. Another challenge addressed in their paper is the dynamic distribution of media content and control to the most suitable device at home. To model user enriched content, the prototype makes use of *Synchronized Multimedia*

Integration Language (SMIL)[9]. A qualitative analysis in two countries, namely the UK and Belgium, were performed in order to evaluate the functionality presented by their architecture. The results conclude that, when extended sharing features are available, it will be widely used. Users from the survey liked the feature to create personalized clips from videos, either for better navigation or to be able to send specific parts of a program to someone. Users liked the idea of a secondary screen (e.g. mobile device) for other usages than controlling the media on the main screen. It allows them to edit and send content without interrupting co-viewers who are watching television at the same time. Therefore the authors propose that existing popular video sharing systems should extend their functionality with video manipulation and sharing features such as video fragmentation, annotation, and enrichment.

Geerts et al. [50] looked at television genres and how they play a role in the use of social interactive television systems. They developed a prototype and performed a user study on their system with 36 participants in a simulated environment with two rooms (one living room, one separate room) and a mix of devices for sending and receiving enriched video fragments to and from various devices. Users were asked to fill out a number of different questionnaires afterwards, followed by personal interviews which lasted about twenty minutes. They discuss which genres are preferred for talking while watching, talking about after watching and for sending the video to users with different devices. Their prototype provides the following functions: (i) *Distributed Rendering/-Control* of the media in order to render the video to a number of devices ranging from high-definition television sets to mobile phones and at the same time the user can control the videos (e.g. Play, Pause, Fragment, Enrich) on a number of interactive devices like a mobile phone, synchronized with watching the media content; (ii) *Media Enrichment* so that users can fragment videos to share certain parts of it and additionally enrich them with other media overlays. The actual video will not be modified; (iii) *Social Sharing* to share the enriched media with other peers of a social network and their devices (mobile phones, TV sets, uploads to personal blogs or social web sites). Through the user study, participants showed mixed results on a device preference to receive video clips. They reported that several factors influence this decision, namely clip length, content quality and immediacy. Mobile phones were preferred for short clips like weather forecasts or breaking news, for higher quality content the mobile phone screen was considered too small by most of the participants, a TV screen would be a more comfortable setting.

[9]SMIL is a XML markup language to describe multimedia presentations. It defines markup for timing, layout, animations, visual transitions, and media embedding, among other things. It allows the presentation of media items such as text, images, video, and audio, as well as links to other SMIL presentations, and files from multiple web servers. (`http://en.wikipedia.org/wiki/Synchronized_Multimedia_Integration_Language`) (accessed January 27, 2012)

Participants raised a fourth factor, namely *privacy*, during the quantitative interviews. Some participants mentioned that a feature, which indicates the recipients status (busy, DnD etc.), would be helpful to decide if it is acceptable to send personal video content. If the recipient is with another person or busy otherwise or the sender does not know who else is watching a video may not be shared at this time. Some participants considered E-mail to be more useful and private than immediate sharing through TV and mobile phones in such cases. A summary of device preferences according to video properties can be seen in Table 2.1. Genres, based on the "EBU system of classification

	Length	Content quality	Immediacy	Privacy
Mobile Phone	Short	Low	High	High
E-mail (PC)	Medium	Medium	Medium	High
TV	Long	High	Low	Low

Table 2.1: Summary of Device Preferences According to Video Properties

of RTV programmes" [51] were analysed regarding (i) genres during which people talk least, (ii) genres people talk most about (at work, at school, ...), (iii) genres people like to share videos of and (iv) genres during which people talk most. Results of (i) are Film, News, Drama Series, Documentary, Standup Comedy, and News Magazine; (ii) Film, News, Soap, Sports, News Magazine, and Comedy Series; (iii) News, Film, Documentary, Music Programme, News Magazine, Standup Comedy, Soap, Sports, Animation Film, Comedy Series; (iv) News, Soap, Quiz, Sports, Reality show, Talk show. The authors found genre similarities between those four groups (i-iv), which means that people would like to share videos with each other of program genres that are discussed at work or at school, but those are genres during which people usually do not talk while watching (namely Film, News Magazines and Documentaries). Strong positive correlations ($p < 0.01$) where found between genres people like to talk most about and people would like to share. Genres people like to talk about or which are their favourite genres are also the genres people tend to share most, except weather forecasts and talk shows. Genres which people talk about during the show and also afterwards are: Soap operas, Sport, News Magazines or Docusoaps. Sport and Soaps also have a strong correlation between talking during watching and sharing. Suitable genres for social iTV systems (genres during which people talk most) are News, Soap, Quiz, Sports, Reality show, Talk show. Genres with more plot structure (e.g. Film, Drama, or Documentary) are preferred to be watched on television, whereas genres with less plot structure (e.g. Quiz, Sport, or Reality show)) can be watched on a mobile phone.

Based on a preliminary study of a quantitative usage study on the Zync Messenger[10], Shamma et. al. [52] studied which genres of web videos were shared most or watched together in a synchronized manner over the web. 2,814 unique users obtained Zync and consented to data collection. The top *YouTube* genres of their most shared videos are *Music*, *Comedy*, *Entertainment*, *Film & Animation*, *Games*, *People & Blogs* and *Sports.* Zync sessions with heavy chat activity were, on average longer in length (average of 304.9s) than the top 20 most watched videos (average of 197.7s). Also videos with heavy activity (pause, rewind etc.) were typically longer (average 345.9s) than the average. Participants reported to share such an experience of synced video watching and additional conversation (either text, voice, etc.) with close friends or family members.

In a three months period, researchers monitored the *YouTube* traffic from a local campus network (University of Calgary[11]) and observed 25 million transactions between users and the *YouTube* community including 625,593 (GET and POST) video requests [53]. Of interest were the network resources consumed by *YouTube* as well as the viewing habits of the campus *YouTube* users. The authors examined usage patterns, file properties, popularity and referencing characteristics and transfer behaviours of *YouTube* and compared them to traditional Web and media streaming workload characteristics. Locally, 3% of the monitored HTTP requests were for *YouTube*, however 99% of the total bytes transferred corresponded to these HTTP responses. In this study, more than 50% of the video requests and their corresponding bytes transferred were for previously requested videos which indicates that in-network caching has the potential to reduce bandwidth demands for *YouTube* or video content in general. As already known through previous research, most of the users just consume video content, users rarely upload video content (133 upload attempts were monitored). The majority of POST requests (overall 28,655) appeared to be the result of users posting comments or rating videos. In their study, the authors observe more requests during day time than during night time. From 8 am to 1 pm requests rise steady and the peak traffic is between 2 pm and 6 pm, followed by a decline from 7 pm to 7 am. Therefore access patterns strongly correlate with human behaviours, as traffic volumes vary significantly by time of day or day of the week. Long term popular videos (weekly, monthly, and all time categories) showed durations well below ten minutes, their mean and median values range between 2.5 and 3.5 minutes. 52.3% videos in the all time popular category are between 3 and 5 minutes long. On campus, the observed mean video duration is 4.15 minutes with a median on 3.33 minutes (COV is approximately 1). Observing the encoded bit rates[12] of

[10]System that allows users to share a synchronized video and chat at the same time, based on *YouTube*. Both participants have equal control over the video's playback features

[11]Campus consists of approximately 28,000 students and 5,300 faculty and staff members

[12]Estimation of the encoded bit rate as the ratio of a video's file size and its duration

the videos requested, only a small number are encoded at extremely low bit rates (e.g. below 100 Kbps), and similarly very few videos are encoded at high bit rates (above 1 Mbps). Videos on the campus network showed a mean bit rate of 394 Kbps and a median bit rate of 328 Kbps. 97% of the accessed videos have bit rates below 1 Mbps, 62.6% have bit rates between 300 Kbps and 400 Kbps. This indicates that videos are encoded so that typical broadband users can begin playback with a minimal start-up delay. Video category trends on campus showed that the top 4 categories are *Entertainment*, *Music*, *Comedy* and *Sports*, least popular categories are *Autos & Vehicles*, *Howto & DIY*, *Pets & Animals*, and *Travel & Places*. Videos in the all time favourite category have similar trends and shows that most of the videos are either in the *Comedy*, *Entertainment* and *Music* category. The authors therefore conclude that users viewing videos on *YouTube* are looking for entertainment rather than information on specific topics.

Boll [54] investigated where current multimedia platforms and Web 2.0 technologies could grow closer together. She points out that not much of the scientific research results found its way into the multimedia area of Web 2.0, although everybody nowadays can create, watch, and share videos on the web and that it is an everyday medium. One positive example, which found its way from research to a practical feature is the integration of *auto-tagging* and *tag-recommendation* on the video upload page on *YouTube*. Famous platforms still lack support of media content analysis, semantic content classification and annotation, or structured multimedia authoring. Platforms could employ simple shot segmentation, which would provide e.g. *YouTube* users with small bookmarks for videos to allow easy navigation between different video scenes instead of just using the slider which has no connection to different video scenes. A feature to find similar or even the same videos and group them on content level instead of just finding them based on keywords could be done via speaker or sound recognition and video shot comparison. Users who upload videos could get a notification before they upload a duplicate video. Different edited versions of the same video could also be interlinked and refer to each other. Finally, Boll suggests a *"Multimedia 2.0 that enables communication, applies the techniques researched in multimedia to commercially relevant application and boosts new start-ups that help people with the results from the research community. There should be a shift from the medium serving as the center of information conveyance to the medium being just a means for conveying the story, the event, or the message."*

CHAPTER 3 Exploratory studies

Our goal for this exploratory studies is to find useful correlations and artefacts within a user's video search and retrieval behaviour and it aims at understanding how and why online videos are watched and what the users actual motivation is regarding video retrieval, search and sharing. Normally a search query is interpreted as it is typed, but the actual intent of a user and the goal in mind is not taken into account yet. If multimedia systems could infer possible trigger events which led to video content consumation, it would be possible to provide a better overall experience for the user. These findings could be used to support and enhance current video retrieval systems, search interfaces and algorithms in order to improve the overall user satisfaction and experience. An exploratory study of why people retrieve, search and share videos online has not been done before, but such groundwork is crucial to optimize search, retrieval, navigation, annotation and recommendation systems.

The study itself is separated in two parts. The first part covers our semi-structured interviews, described in chapter 3.1, where semi-structured interviews were performed to talk about participants experiences regarding video retrieval. This first iteration gave us insight and information about common user behaviour. The second part was the quantitative survey, described in chapter 3.2, where we used an online survey tool (SurveyMonkey[1]) to perform the survey with a broader and therefore bigger and more diverse audience. The main reason for the quantitative survey was to support our results from the first part and to identify outliers. In chapter 3.3 our results of both studies will be presented. Additionally, correlations and differences from the evaluated literature to our present work are described within chapter 3.3.

Participants in both studies were not given any money, vouchers or gifts as a compensation for the time invested in the interviews. Participants of both surveys took part in the interviews on a voluntary basis.

[1] SurveyMonkey: `http://www.surveymonkey.com` (accessed January 27, 2012)

3.1 Semi-structured interviews

To find out the users intention in the field of video retrieval, we first needed to understand their basic motivation and the environment in which one is acting while performing a video retrieval task. We use semi-structured qualitative interviews as a starting point for our research. The in-depth interviews we performed for this survey are more flexible than other methods because they allow participants to talk freely about their video watching and sharing experiences, without being restricted to predefined answers. Moreover, the interviewer decides, based on the participants response, which question to ask next or even how to ask the next question. If necessary, the interviewer has the possibility to dig deeper into a certain topic if it of interest. The analytical objectives for qualitative interviews are: to describe and explain relationships, variation, individual experiences and, if possible, to describe group norms.

This first part of the exploratory study will help us understand how and why video platforms are used and what the user's behaviour and his motivation actually is. We assume that the proposed method and its reliance on personal, one-on-one interviews allows for more expressive answers and deeper insights into user aspects of video watching than the ones that would be captured by automated solutions, e.g., monitoring proxies or network traffic analysis. This is a very unique aspect of the qualitative interview method. The interview structure and questions are based on the insights and knowledge given in [55] for performing and evaluating qualitative surveys.

3.1.1 Methodology

Semi-structured interviews were performed with a selection of random people and we talked about their experiences regarding video retrieval. Each participant was questioned about their experiences related to searching, retrieving, watching and sharing videos online. Especially important was the aspect of *why* they watch videos online, what is their motivation in the first place; and which event caused the action to watch a video in the past? Interviews were based on a semi-structured questionnaire (a mix of quantitative questions with qualitative elements) which allowed a dynamic conversation while still covering our set of questions. We approached students, relatives and friends in a spontaneous manner and asked them if they have time for an interview without mentioning what the survey was about.

Each interview was conducted on a face-to-face approach and lasted about 20 to 30 minutes and the participant was questioned about their experiences related to searching, retrieving, and watching videos online. The questionnaire included questions about demographics; general use of video search engines; communication actions based on videos (e.g., recommendations to other friends to watch the videos); and past situations where the users watched a video online, called *instances*. For these instances of the video retrieval use case interviewees were asked: which event has triggered the action of watching the video in the first place, where and when the videos were watched, which device was used, the length of the video, and whether it was shared with someone else (and how) or not. Interviews were recorded on paper and analysed afterwards. The complete questionnaire with all the questions in detail can be seen in Appendix A.1. In the following paragraphs the four major blocks of questions are described:

- **Demographics:** Consists of questions about job, gender, age, general interests, and hobbies.

- **General usage:** We wanted to find out which online video platforms a user knows and on which he is active and how often he is using such a platform. Also a rating about platform preference (from 1 to 3) was performed. It was also of interest on which devices a user is consuming such digital content. Part of this block was an open question about the user's motivation to search for videos and watch them. Additionally, questions about the usage of diverse features of video platforms were asked (e.g. subscriptions, related feature, and suggestion feature), or if the quality of a video is a criteria for watching it.

- **Communication:** In this block, we wanted to find out how a user is sharing his video experience with other people in his social environment. It was of interest with whom and why one is sharing online videos. Also of interest to us was the content of the video being shared and the intention behind sharing the video link. Questions about the frequency of sharing and the mechanisms used were part of this block.

- **Instances:** Instances are a description of typical specific scenarios a user has experienced in the past. It was of interest, which event has triggered the action on why a user is watching a video online. For example one either searched for it himself, or got a recommendation from another person, or used a platform feature. Information about the video content and the intention behind watching it were crucial questions in that block. Also questions about the setting (place/time), device used and the length of the video were asked, as well as if the video was shared with someone else. Each interviewee could report up to three instances. We pointed out that we are preferably interested in the last recent online video encounters.

3.1.2 Participants

Within our study, 22 participants – all of which were German-speaking Austrians with different background, occupation and hobbies – were interviewed. The group of participants consisted of 19 males and 3 females. Their occupations included students (9) and professionals (13) and their hobbies varied on a broad scale. Participants reported from memory on 1-3 video watching situations, particularly on the following aspects: why they watched a video online, what the video was about, and where they watched it. In the end, the 22 participants reported on 49 different situations, an average of 2.23 scenarios per participant. Table 3.1 gives the demographics of participants along with gender, age range, frequency of online video watching and devices used. The age distribution of the participants can be seen in Figure 3.1. The vast majority of participants ($\sim$ 78%) is younger than 30 years of age.

	Gender	Age	Frequency	Device
P01	m	25-30	1	PC/TV
P02	m	25-30	3	PC
P03	f	18-24	3	PC
P04	m	25-30	1	PC/MD
P05	m	25-30	1	PC/MD/TV
P06	m	25-30	1	PC/TV
P07	m	25-30	1	PC/MD
P08	m	> 40	3	PC/TV
P09	m	25-30	2	PC
P10	m	31-40	2	PC/MD/TV
P11	m	25-30	1	PC/MD
P12	m	< 18	2	PC
P13	m	25-30	2	PC
P14	m	18-24	1	PC/MD
P15	m	25-30	2	PC
P16	m	25-30	1	PC/MD
P17	f	18-24	2	PC/MD
P18	f	18-24	2	PC
P19	m	18-24	2	PC/MD/O
P20	m	> 40	2	PC
P21	m	31-40	4	PC
P22	m	> 40	2	PC

Table 3.1: Summary of 22 participants: gender, age, frequency of watching videos online classified as daily (1), more than once a week (2), more than once a month (3), and once a month or less (4). Used devices are classified as PC/laptop/netbook (PC), mobile devices (MD), TV with streaming capabilities (TV) and others (O).

3.2 Quantitative questionnaire

Based on the preliminary findings of the qualitative survey, we decided to design and perform an online quantitative survey to test and support our findings with a larger and

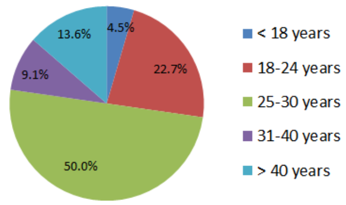

Figure 3.1: Age distribution of participants in the semi-structured interviews

more demographically diverse audience and to gain more insight in user behaviour and user preferences. After the qualitative interviews were finished, we started to design an online survey with SurveyMonkey. We decided to reuse the questions from the semi-structured interviews and provide predefined answers (inspired by the answers from the semi-structured interviews) to them. Additionally, we introduced new questions that were of interest. The complete questionnaire with all questions can be seen in Appendix A.2 in detail. The interview structure and questions are based on the insights and knowledge given in [55] for performing and evaluating surveys. The online survey was distributed throughout different communication channels like email, instant messaging, Reddit[2], Twitter[3], Facebook[4], Slashdot[5], and two personal blogs; a SurveyMonkey mechanism that distinguishes among them was configured. By doing so, we reached 270 people from various countries, from which 217 people (80.4%) completed the whole survey.

3.2.1 Methodology

We used SurveyMonkey to design an online structured questionnaire, which included questions with multiple-choice answers following an ordinal scale, where items were given a value from 1 to 5 which let us calculate an average rating, as seen in many of the upcoming tables and figures in section 3.3. We completely redesigned the *'instances'* part to fit the quantitative interview because we did not want to rely on the memory of the participants. We introduced new questions, for example questions about the user's favourite genres (based on *YouTube* categories), most common events which caused consuming an online video, places where and when videos were watched, and other preferences like picture quality, content quality, sound quality and so forth. Additionally, some of the questions offered the participant an open text box so that they can provide us with further information or a more detailed answer.

[2]Reddit: `http://www.reddit.com` (accessed January 27, 2012)
[3]Twitter: `http://www.twitter.com` (accessed January 27, 2012)
[4]Facebook: `http://www.facebook.com` (accessed January 27, 2012)
[5]Slashdot: `http://www.slashdot.com` (accessed January 27, 2012)

3.2.2 Participants

270 participants took part in the survey, out of which 217 (i.e., 80.4%) completed and answered all questions. For our evaluation, only data of completed surveys was be used, hence the data of the completed questionnaires from 217 participants. The response rate per communication channel from all 270 participants can be seen in Figure 3.2. Most of the respondents attended the survey through email (173), followed by the Reddit community (48). Among the overall 217 persons who finished the survey 69.4% were

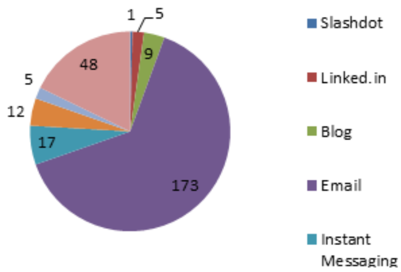

Figure 3.2: Response rate by distribution-channel in number of overall participants

male (150) and 30.6% were female (66), one person skipped the gender question. The age distribution in Figure 3.4 shows once again, that the vast majority of respondents (~ 81%) are less than 30 years of age. Their country of origin (determined via `http://www.ip2country.cc`) can be seen in Figure 3.3: the majority of participants came from Austria (136) followed by the US (50).

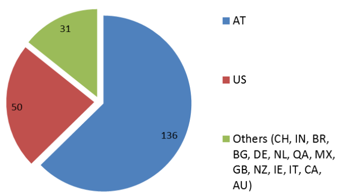

Figure 3.3: Participants per country of the quantitative study

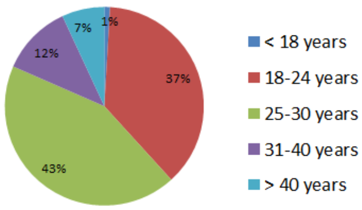

Figure 3.4: Age distribution of participants of quantitative study

A comprehensive overview and comparison from both studies about demographics, education and age can be seen in Table 3.2.

		Qualitative	in %	Quantitative	in %
Demographics	Overall participants	22	100.0%	217[1]	100.0%
	Female	3	13.6%	66[2]	30.4%
	Male	19	86.4%	150	69.1%
	Country of origin	AT: 22	-	AT: 136 US: 50 Others[3] 31	-
Education	Graduate degree	3	13.6%	47	21.8%
	Bachelor, college or university degree	7	31.8%	85	39.4%
	High school	5	22.7%	68	31.5%
	Elementary/secondary school	6	27.3%	5	2.3%
	Others	1	4.5%	11	5.1%
Age	< 18	1	4.5%	2	0.9%
	18-24	5	22.7%	81	37.7%
	25-30	11	50.0%	94	43.3%
	31-40	2	9.1%	25	11.5%
	> 40	3	13.6%	15	6.9%

Table 3.2: Demographics overview

[1] Completed the survey
[2] One person skipped this question
[3] Countries of origin of other participants: CH (1), IN (1), BR (2), BG (2), DE (5), NL (3), QA (1), MX (1), GB (3), NZ (1), IE (1), IT (1), CA (2), AU (1)

3.3 Study results

In this section we summarize and compare the results from both studies. We also highlight interesting facts and figures from both surveys, as well as personal comments from the participants. This section is structured in three parts: in Section 3.3.1 we explore how participants of the studies used video platforms in general; in Section 3.3.2 we provide insights into their video sharing behaviour, and finally in Section 3.3.3 we analyse the specific scenarios reported in the semi-structured interviews and attempt to validate those findings with the data from the quantitative survey by using statistical analysis. Each subsection has been structured using short paragraphs preceded by questions that cover selected topics of interest from our studies.

3.3.1 General usage of video platforms

Which platforms and how often do you use them?

Both questionnaires covered questions about the general usage of video platforms, including questions about most favourite platforms, usage patterns of platforms regarding frequency of visits, reasons for using the platform, etc. Figure 3.5 shows a summary of the preferred video platforms used by participants on a regular basis. *YouTube* appears as the most-used and best-known platform, followed by *Facebook*. The high rating of *Facebook* is somewhat surprising, due to the fact that it is not primarily a video platform. Moreover, *Facebook* posts often contain links to other platforms such as *YouTube* and that could be the cause for its popularity. This finding provides insight on the fact that users do not seem to care much about which platform hosts or streams their videos, but instead which platform (or, in this case, social network) acts as a portal through which they learn about such videos. This is also true for most blogs or news sites, which don't necessarily host their own content or use their own media player. Overall, the percentages reported in our qualitative study are higher than the quantitative counterpart, which is most likely due to the differences in demographics. A notable exception to that pattern is the fact that no one mentioned *Netflix*[6] in the qualitative study, which is clearly due to the fact that in Austria the service is not available yet. Some users have

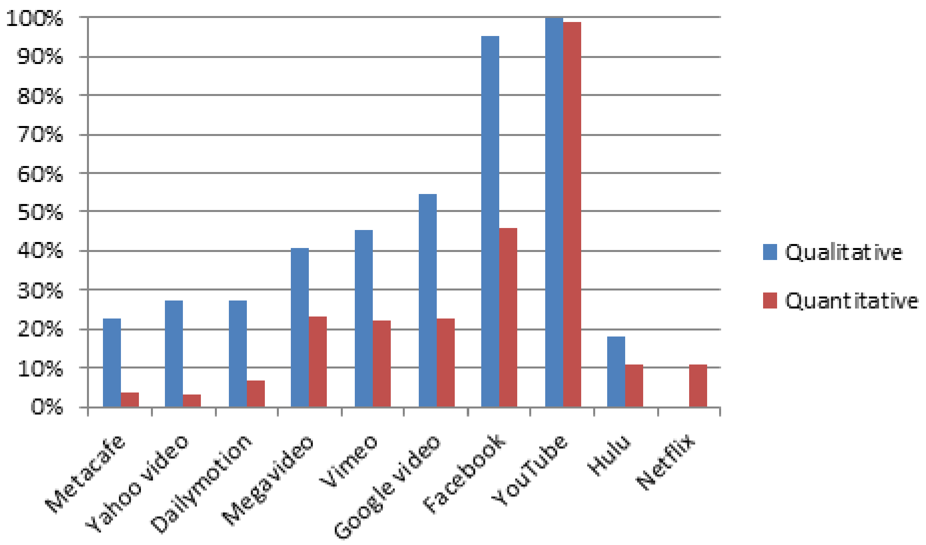

Figure 3.5: Most frequently used video platforms

(in some cases, completely) moved away from cable TV because they consumed all of

[6]Netflix: `http://www.netflix.com` (accessed January 27, 2012)

their video content online.

> *"I do not have cable TV, so all of my video watching is in lieu of television"* (male, US, 31-40)

> *"No cable TV, no bunny ears. All video is strictly streaming from Hulu, Netflix, or other leading services"* (male, US, 25-30)

Regarding frequency of usage, in both studies the majority of participants (80% or more) reported watching videos online at least several times a week or more than once a day (Figure 3.6), reinforcing the notion that Internet-based video platforms are becoming a part of everyday life.

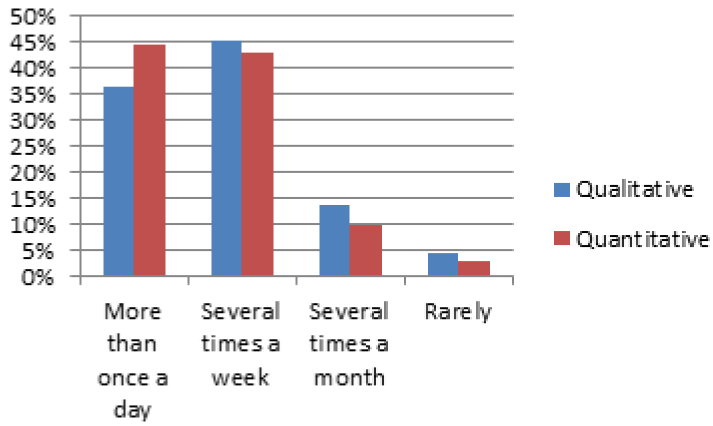

Figure 3.6: Frequency of platform usage

Which devices do you use to watch video contents?

We are also interested in learning about the devices used or preferred among the users to consume online video content. In the qualitative interviews we found that all 22 participants used their PC/laptop/netbook to watch videos online, nine reported to watch videos on a mobile device, and five also watched videos on a TV. One interviewee even watched online videos on his car PC. In the quantitative study we asked about the most preferred devices of participants, giving them the option to quantify the usage for each device by selecting an item from an ordinal scale between *Never (1)* and *Always (5)*. This resulted in an average rating (Figure 3.7) that shows that most of the participants used laptop/notebook/netbook devices to watch online video content followed by the classical PC and smartphones. As new TV sets and gaming consoles have the ability to stream online content and access online video services, the high rating of TV is not unusual. Tablets are still not used very often, but this may be due to the fact that the market for such devices is still very young and growing. Some people also mentioned

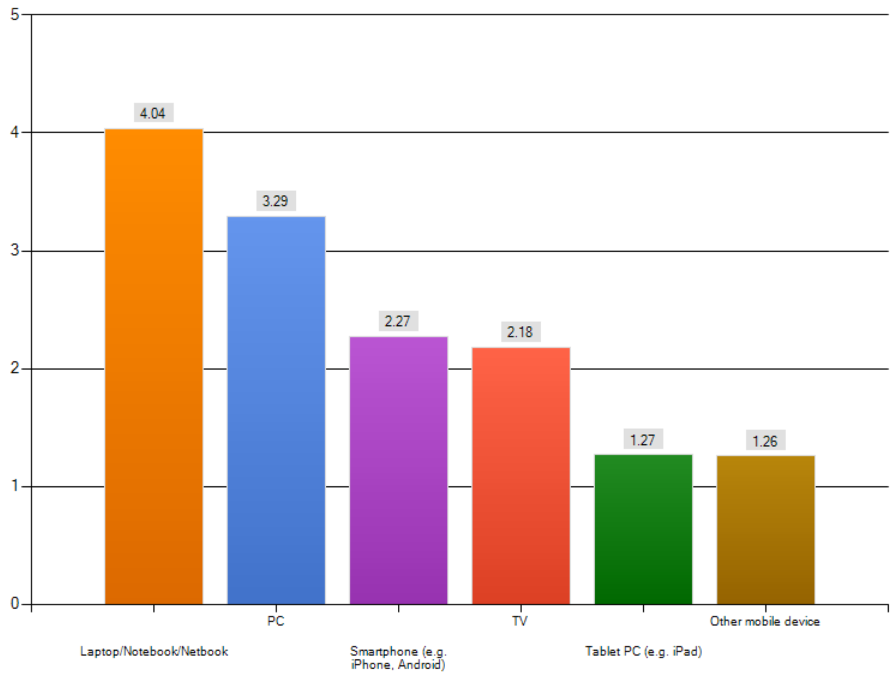

Figure 3.7: Average rating of device usage (based on an ordinal scale ranging from *Never (1)* to *Always (5)*)

other devices and ways to consume online video content:

> *"I watch videos on the web less frequently after I could stream Netflix on my gaming consoles"* (female, US, 25-30)

> *"I have a WDLive Drive. Basically, you set it up with a TV and can access Netflix, YouTube, your stored videos, etc. Makes YouTube sharing much easier"* (male, US, 25-30)

Professional or private usage?

The growing availability of video contents for educational and professional use may also lead to a change in attitude in the mindset of employers and employees. Educational video content is cheap or even free and easy to access and the information presented in a video is often more easy to understand than reading about the same topic. In the semi-structured interviews four participants reported to use video content for professional reasons. Three out of the four were students who answered yes because they watched video tutorials and lectures on specific topics they need to have knowledge about. We followed the same question in the quantitative survey in a more precise manner. Participants reported to use online video content for professional use once a day or more in 3.8%, once a week or more in 14.2%, once a month or more in 21.2% of all cases. The majority reported to use videos for professional reasons rarely (37.7%) or never (23.1%). These numbers seem low, however they indicate an upcoming trend as more and more people recognize the opportunity to use video as a convenient tool for learning.

What about quality?

The quality of a video encompasses various dimensions like picture, sound and content quality but not every person has the same quality requirements, it depends on many factors like the device someone is currently using or which video one is actually watching and the reason why. This is also covered in literature by Geerts et al. [50] where they summarize device preferences of their participants, which can be seen in Table 2.1. In our qualitative interviews 15 people reported that quality is a very important factor for them to watch a video. They reported that if the video or sound quality does not match their expectations they either close the video immediately or they search for another one with the same content but in a better quality. The requirements for picture and sound quality also heavily depend on the content of the video. Users reported that picture quality is very important for video tutorials and how-to videos (because often you need to read or follow a text or the like), and sports/action videos but not that important for most entertainment or music videos. Similar problems were also reported in other surveys like [34] where participants stated that the technical performance in general or the quality for e.g. sport broadcasts is sometimes not satisfying enough. For music videos the sound quality is often more relevant than the video quality because music videos are often used just as background entertainment like radio. In general, users make exceptions if the task to watch a specific video is very important and no alternative video that covers the same content is available. Other quality factors or factors to watch a video are loading and buffering time and sync audio.

> *"When a video is repeatedly interrupted or paused due to slow download, I always press exit or delete because I lose patience and interest quickly. As this happens quite often, I rarely watch videos at all"* (female, AT, > 40)

Since there are also an emerging number of people using high-end equipment at home, they also stated that they often explicitly search for HD videos to watch it either on their TV or in full screen mode on their computing devices which support HD.

> *"Because of 30Mbps download speeds and all of my monitors and internet connected TV's are 1920x1080 resolutions, I prefer highest quality videos. However, while traveling and using iPhone or other slower connected products, I prefer medium to low quality so buffering is not an issue"* (male, US, 31-40)

We also asked similar but more specific questions in the quantitative questionnaire, whose results are summarized in Table 3.3. Here, once again, the majority of users prefer high visual and sound quality, but consider the content of the video the most important factor to impact their decision to watch it.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)	Rating average	Rank
I prefer videos with high visual quality	1.9%	1.9%	18.1%	**42.1%**	36.1%	4.09	2
I prefer videos with good sound quality	0.9%	0.9%	20.1%	**47.2%**	30.8%	4.06	3
My decision to watch a video is based on the content	1.9%	1.4%	13.9%	34.3%	**48.6%**	4.26	1

Table 3.3: Quality preferences among users in quantitative questionnaire

Quality delivered from the video content provider (e.g. *YouTube*) is analysed by other researchers, i.e. in [44] the distribution of the bit rate of 190,000 *YouTube* videos was collected and analysed afterwards. Three peaks were identified whereas most videos have a bit-rate around 300 Kbps, followed by a peak around 285 Kbps, and lastly around 200 Kbps. These findings imply that *YouTube* implements a good balance between quality and bandwidth using a moderate bit-rate. The same results where found in [53] by analysing the *YouTube* traffic in a university campus network. Only a small number of videos are encoded below 10 Kbps, and similarly very few videos are encoded at high bit rates above 1 Mbps. The analysis showed a mean bit rate of 394 Kbps, a median bit rate of 328 Kbps, and 62.2% of the accessed videos had bit-rates between 300 to 400 Kbps.

Why and how are you watching videos?

We also asked the participants of the semi-structured interviews why in general they searched for videos. Certainly and as expected, participants often told us that they use such platforms when they have time, are bored and need to kill time or have nothing better to do and want some entertainment. In this case they watch all kinds of videos ranging from entertainment and fun videos to videos related to their hobbies and interests (such as sports videos). Another interesting fact is that some people use *YouTube* solely as a radio station. They load up a queue of their favourite music videos and just play it in the background, while giving no attention to the video at all. They state that it is one of the easiest ways to listen to some new or favourite music. It should be noted that this cannot be done in some countries due to copyright laws or disputes between the video portals and special interest groups (e.g. Music labels) in that country. For example in Germany, a lot of music videos or videos containing (background) music which is licensed under the GEMA[7] (which is a body represented by 60.000 artists) is blocked in the country. GEMA and *YouTube* could not agree during negotiations on how much licensing fee should be payed from *YouTube* to GEMA per video streamed[56]. Besides entertainment, videos are used to learn different things and get information about certain topics. Using video tutorials and how-to videos to learn new or rehearse known content is very common nowadays. For example, two students used videos that explained certain math paradigms to prepare for an upcoming exam. Others wanted to improve at their hobby/sport so they watched videos about special technique training for rock climbing. Others watched videos to gain knowledge on how something works (a specific microcontroller, a drilling machine, our solar system etc.). They all agree that videos are an easy way to gain new knowledge, learn about unknown things or to refresh knowledge about a certain topic when needed. Some users reported having used online video platforms to see content that they missed on normal broadcast TV programs like summaries of sport events (e.g. missed goals of a soccer game) or the latest local news. It is also an easy way to see how some things look like (e.g. a place). One participant watched videos about a specific cruise ship because he planned to do a cruise vacation in future and he wanted to see what the accommodation looks like and which amenities what the ship has to offer. He also does this with hotels where he is planning to stay while on vacation just to get an impression beforehand. Others used online videos to watch reviews about a product that they wanted to buy because it is much less stressful than reading a product review online.

Participants from both surveys also reported to use video platforms to watch videos not accessible in their country or to watch old TV shows from their youth, because it is the only way to see that content again and most of the old TV shows etc. will not be aired on regular TV again.

> *"I belong to the 2nd generation of TV viewing, and I use YouTube to reminisce my childhood viewing, and to watch original, undubbed, non-censored, or content not available in my country"* (male, US, > 40)

> *"I especially enjoy watching videos on the net that couldn't be accessed otherwise for whatever reasons"* (female, AT, 25-30)

Certain users also reported to watch movie trailers or trailers/advertisement for upcoming events (like concerts, movies, musical premieres etc.) online. A few people also just watch all sorts of videos to have a background noise (i.e., using it as a radio station as mentioned before); they don't pay attention to the video itself that much because they are busy with more important work.

[7] GEMA (Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte): `http://de.wikipedia.org/wiki/GEMA` (accessed January 27, 2012)

"I have dual-monitors on my main computer. I'm more likely to watch videos if I can continue doing my current activity on one screen and have the video play on the other" (male, US, 18-24)

"I'm working from home, alone, but after a while I need something in the background, sometimes its music, sometimes it is a Movie. When I need to work and I choose a movie, it is always a movie I've already seen 3 or 4 times before, so I dont get distracted. So I guess in average I watch a lot of videos, by not watching them" (male, AT, 31-40)

To be more exact and summarize the answers from our participants, the most frequent responses from the semi-structured interviews can be categorized as follows:

- Information search (16 out of 22), e.g., watch missed broadcast news, watch product reviews, see how a place looks like, research and follow-up on a certain topic.
- Educational purposes (8 out of 22), e.g., how-to videos, learn math, tactics for video games, and video tutorials on diverse topics.
- Entertainment (19 out of 22), e.g., sports, comedy, listening to music or watch music videos, movie trailers, discontinued/old TV material.

To be able to compare these findings with our quantitative study, we introduced a similar question and asked participants why they search for videos and watch them by showing them nine predefined events. We let them express how often on average they watch a video based on such an event by offering ordinal scale items from *Never (1)* to *Always (5).* Detailed results are shown in Table 3.4.

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)	Rating average	Rank
For leisure/entertainment reasons	2.3%	4.2%	14.9%	**59.5%**	19.1%	3.89	1
Out of boredom	13.9%	21.3%	26.2%	**33.7%**	5.0%	2.95	6
To listen to music or to watch a music video	3.8%	10.8%	24.4%	**44.6%**	16.4%	3.59	2
To get information about certain topics	4.3%	14.2%	**40.3%**	8.1%	8.1%	2.27	4
To get product reviews	22.2%	**31.4%**	30.0%	14.0%	2.4%	2.43	8
To see missed news-broadcasts	22.2%	**32.9%**	29.0%	14.0%	1.9%	2.41	9
To learn something new	4.3%	14.3%	36.7%	**38.1%**	6.7%	3.29	3
To see how something looks like	11.2%	**28.8%**	27.8%	27.8%	4.4%	2.85	7
To see how something works	5.4%	15.8%	**41.9%**	32.0%	4.9%	3.15	5

Table 3.4: Why do people watch videos?

A close inspection of the *'Rank'* column in Table 3.4 shows that most people use video platforms for entertainment in their leisure time. Entertainment in this case could be whatever a person enjoys, e.g. watching a video about their favourite sport, TV show, comedy or the like. Second, people often use video platforms to listen to music because it is easy accessible and mostly in good quality because music companies or the artists themselves promote their new songs over such platforms. Third, people like to use video platforms to learn something new, to gain new knowledge and to educate themselves on some level. We also heard statements like this in our qualitative interviews quite often.

Do people like platform features?

When asked about which on-site (platform-dependent) features participants use, and how often, participants of the semi-structured interviews indicated a frequent use of features such as *'video suggestions'* or *'related video'*. They often use it when they have enough time and want to get further information about the topic they currently looking at. In some cases they also used it if the first video for their search was not the expected one, instead of refining the query terms they looked at the suggested or related video section first to see if they find the video there. In the quantitative interview we also evaluated the usage of that feature and confirmed that most participants used it, only 3.2% reported having never used it, see Figure 3.8. Table 3.5 summarizes answers to

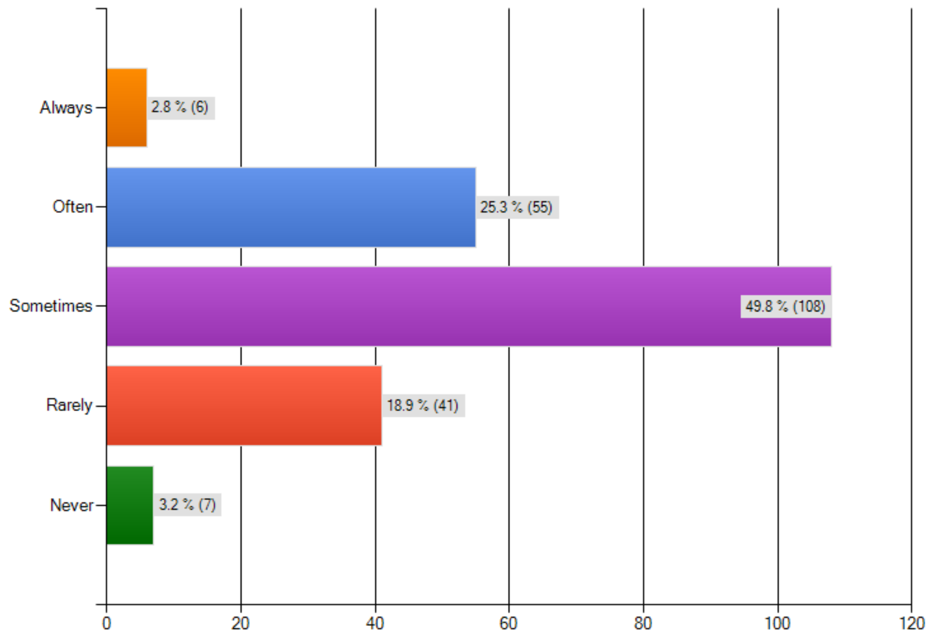

Figure 3.8: How often do people follow recommendations?

questions related to on-site features, such as *'channel subscriptions'* and *'navigation through categories'* from the quantitative study. During the semi-structured interviews, by contrast, only two participants reported having used *'channel subscriptions'*, whereas *'navigation through categories'* was not used at all because it is more convenient and more efficient to search by terms and select a video from the result list. Even if most

	Never (1)	Rarely (2)	Once a month or more (3)	Once a week or more (4)	Once a day or more (5)	Rating average	Rank
I watch video recommendations coming from on-site features (channel subscription, related videos, video suggestions)	9.3%	**30.8%**	29.4%	25.7%	4.7%	2.86	1
I browse through categories and select videos from there	17.8%	**32.9%**	24.4%	18.8%	6.1%	2.62	2

Table 3.5: General feature usage

of the participants used these two features rarely, the overall percentage of respondents who use them is quite significant (82.2% or higher), which suggests that these features are used more often than expected before and that they are somewhat important for the users video experience.

In the quantitative questionnaire, we also found interesting statements that people are missing features on video platforms. For example downloading a video in case it gets removed or just to have an offline copy for further usage. We know that Vimeo offers such a feature for registered users, but on *YouTube* it is only possible with specific browser plugins.

> *"I prefer to check out sites that allow easy download of said videos. I hate streaming sites (buffering)"* (female, MX, age 25-30)

> *"For YouTube I use youtube-dl. Vimeo offers downloads for registered users for some videos, but for some time I had trouble logging in at their webpage, so no more Vimeo. Luckily youtube-dl works most of the time and some people host their videos on their own sites so you can simply download them."* (male, AT, 25-30)

> *"On another note, I tend to download online videos if they are important enough to me and if they are likely to be removed from the respective hosting site in the future. ... "* (male, AT, 18-24)

> *"It will improve the experience if an index would be available of the video. Most of the times you just want to navigate to the part you are interested in, especially when the video is too long. For sports, sometime the statistics of the game is enough"* (male, US, 25-30)

Indications for the requirement of new features were also mentioned by Boll [54] in order to satisfy the users of video communities. For instance, an index or bookmarks and a comprehensive overview would be beneficial for easy navigation within the video instead of just using the slider which has no connection to the content at all.

What do people like to watch?

In our semi-structured interviews we analysed our 49 scenarios reported by the participants and matched each scenario to a *YouTube* genre. The highest fraction of watched videos belonged to the *Music* genre with 34.7%, the *Sports* genre constituted for 28.6% of videos watched. Other genres watched where *Film & Animation* and *Science & Technology* with 8.2% of all cases, *Autos & Vehicles*, *Comedy*, *Education*, *Gaming* with 4.1%, and *News & Politics* and *Travel & Events* with 2.0%.

If we have a closer look at our results from the quantitative survey where we asked participants to express their interest in various genres (by genres we adopted the *YouTube* categories) by selecting one of the following items: *Not interested (1); Rarely interested (2); Neutral (3); More interested (4); Very interested (5).* The average rating based on the values can be seen in Figure 3.9. Once again, the *Music* genre is the most preferred

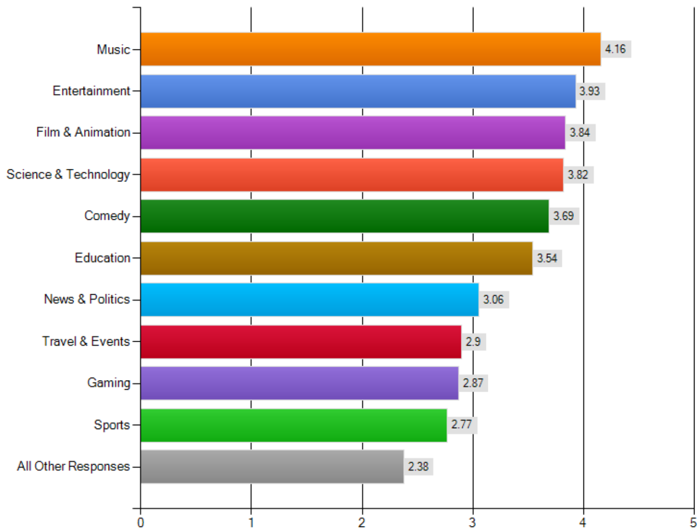

Figure 3.9: Average rating from the favorite genres of the overall population

one, but followed by other genres not *Sports* as our result from the semi-structured interviews suggested. Based on the comparison of both studies, we can easily argue that the rating of the preferred genres is heavily based on the survey population. In literature we found not identical but strongly similar results of user genre preference. Again, differences can be explained by survey population, size of the dataset and the environment and year (both in 2007) in which the analysis was performed. Also in both studies a more technical method, the monitoring and analysis of network traffic data, was used. In [53] the authors analysed the *YouTube* traffic of a campus network and identified the top 4 categories, namely *Entertainment*, *Music*, *Comedy*, and *Sports* whereas the worst 4 categories were *Autos & Vehicles*, *howto & DIY*, *Pets & Animals* and *Travel & Places*. Xu Cheng [44] reported that the most popular and interesting categories are *Music* with 22.9%, *Entertainment* with 17.8%, *Comedy* with 12.1%, and *Sports* with 9.7%. Cheng's study is based on trace data of about 2,676,388 videos which was collected in a 3 months period.

But how do preferred genres change when restricting to certain demographic items or various usage patterns? For this matter we further analysed age groups and frequency of usage regarding genre preference. It shows that older people getting more interested in the genre *Science & Technology* instead of *Entertainment* for example. Table 3.6 shows the first five most preferred genres based on the age of the participants. Investigating

Rank	**< 18 years**	**18 - 24 years**	**25 - 30 years**	**31 - 40 years**	**> 40 years**
1	Entertainment	Music	Music	Science & Tech.	Science & Tech.
2	Comedy	Entertainment	Science & Tech.	Music	Comedy
3	Music	Film & Anim.	Film & Anim.	Entertainment	Entertainment
4	Film & Anim.	Comedy	Entertainment	Film & Anim.	Film & Anim.
5	News & Politics	Science & Tech.	Education	Comedy	Music

Table 3.6: Genre interest per age group

genre preferences based on platform usage frequency did not vary as much and did not show any specific preferences. Table 3.7 shows the first five most preferred genres based on the frequency of platform usage. No matter how often a user visits video platforms the *Music*, *Entertainment* and *Film & Animation* genre seems to be popular among all users except for people who rarely visit such platforms. They seem to be more interested in *Education*.

Rank	**Rarely**	**Several times a month**	**Several times a week**	**More than once a day**
1	Education	Music	Music	Music
2	Music	Entertainment	Entertainment	Entertainment
3	Comedy	Film & Anim.	Science & Tech.	Film & Anim.
4	Science & Tech.	Science & Tech.	Film & Anim.	Comedy
5	Entertainment	Comedy	Comedy	Science & Tech.

Table 3.7: Genre preference based on frequency of usage

What are the annoyances of online video?

Online video experience is not always satisfying for every user all the time and on every level; there are always drawbacks. On some video platforms advertisements are shown before the actual video or even during the video. Our quantitative study did not cover this question per se, but participants frequently reported in the open text box that they are getting annoyed by the upcoming trend of advertisements before, during, or within videos. Often the placement of advertisement covers important parts of the video.

> *"I don't think this is interesting per se but I hate when a site puts ads that obstruct the video I am trying to watch"* (female, US, > 40)

> *"Advertisement links, banner or movies which are shown before the real movie or clip starts is very annoying"* (female, AT, age 25-30)

Users (especially those outside the US) found it particularly annoying that some video content providers (e.g., *Hulu*[8] and *Netflix*) limit access to their resources to specific countries or regions.

> *"I hate it with passion when I'm not able to watch video because I'm in the "wrong" country and there is a copyright issue (YouTube, etc.). How something like this is still possible in this time and place is a mystery to me."* (male, DE, age 25-30)

> *"I watched YouTube since the very early start and enjoyed it all the time. The only thing which starts to annoy me are advertisements in front of spots which you cannot skip or that I am unable to use some YouTube main sites like Vevo out of my country. Same is true for sites like Hulu or FOX video which scan your IP to determine your location and block you"* (male, AT, 25-30)

3.3.2 Sharing behavior

Within our survey we also talked about the sharing behavior of participants. What is their main motivation to share video links with other people, who are the people they share the links with and why? 16 participants (73%) of our semi-structured interviews reported to share video links in general. Their motivations why they shared a link varied, but it can be summarized as follows:

- So that others can laugh and be amused about some entertainment video as well..
- People did find a new or interesting video, where they know that a selected group of persons (friends, family or colleagues) are interested as well because they share the same hobbies/interests or movie/music taste and the like.
- To inform people about breaking news, upcoming events or products (product reviews) .

[8] Hulu: `http://www.hulu.com` (accessed January 27, 2012)

- To show others what one has accomplished, or what one has experienced. Such videos are mostly filmed by themselves and serve as self-representation.

If we compare this to the results of the quantitative study where we asked the same question (based on an ordinal scale), we can clearly see some similarities. People are often sharing entertainment and fun videos (rating average of 4.08) as expected, but still an interesting fact is that most people don't forward links without any special reasons. They think more consciously before sending, considering which other persons could be interested in the video, and then only forward it to a certain group. People do share

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)	Rating average	Rank
I want others to learn about things, transport knowledge	3.5%	20.6%	**36.2%**	31.2%	8.5%	3.21	3
I want others to be entertained, have fun watching the video	1.4%	1.4%	8.5%	**65.5%**	23.2%	4.08	1
I assume that friends I forward the link are interested in this video	1.4%	2.1%	19.7%	**58.5%**	18.3%	3.90	2
To share information about products or events	10.0%	33.6%	**35.0%**	17.1%	4.3%	2.72	5
To share an experience, to share something I have experienced	12.9%	28.1%	**31.7%**	23.7%	3.6%	2.77	4
To share emotions (like anger, disgust, beauty, love etc.)	21.9%	27.0%	**30.7%**	15.3%	5.1%	2.55	6
Someone asked me to share this video	**36.5%**	30.7%	22.6%	0.2%	0.0%	2.07	7
No special reason	**36.8%**	28.8%	27.2%	5.6%	1.6%	2.06	8

Table 3.8: Why are people sharing links?

video links with various groups, mostly friends, family, colleagues and the public. The participants of the semi-structured interviews shared video links with the groups seen in Figure 3.10. The most common group that people share links with is friends (55.56%), followed by an identical share of family members and colleagues (18.52%). This also

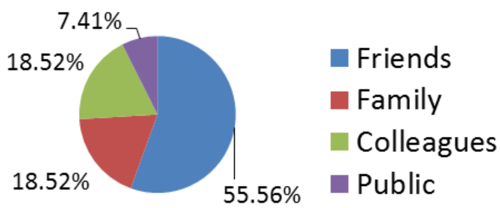

Figure 3.10: Distribution of sharing groups from the qualitative interviews

strongly relates to our findings in the quantitative study, where we asked the same question. Of the 217 participants, 143 (65.9%) shared video links in general with the groups seen in Table 3.9. Also in the quantitative study, friends are the most preferred

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)	Rating average	Rank
Friends	0.7%	0.0%	7.8%	**59.6%**	31.9%	4.22	1
Family members	9.6%	22.1%	**31.6%**	30.9%	5.9%	3.01	3
Colleagues	6.6%	27.7%	25.5%	**35.0%**	5.1%	3.04	2

Table 3.9: With whom are people sharing links?

group to share links with, resulting in an average rating of 4.22, followed by an almost identical average rating of family members with 3.01 and colleagues with 3.04. This correlates very well with our findings in the quantitative study. Two participants of this study also wrote in the open text-box that they post video-links on *anonymous* social sites (like *Reddit*, *Digg*, etc.) and that they share it with *Facebook* friends[9].

In our exploratory studies we also wanted to find out how often and by which means video links get shared with other people. Results from our semi-structured and quantitative survey showed that most people share video links several times per month (Figure 3.11). Only in rare cases people reported to share video links more than once a day. Results from our semi-structured interviews showed that the most common mechanism

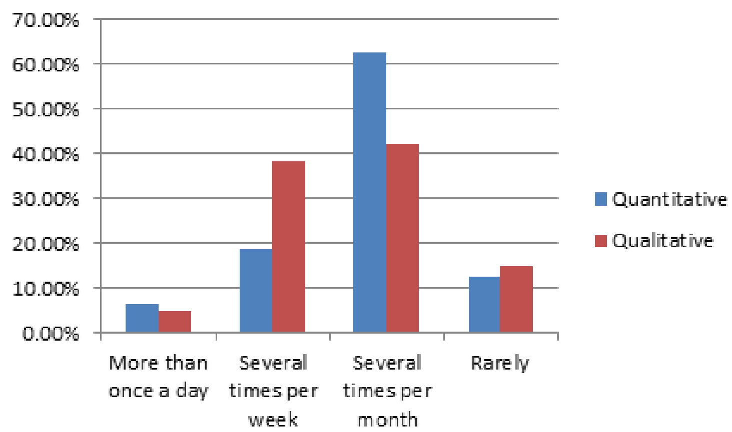

Figure 3.11: Link sharing frequency

[9]This seems to beg the question: is a difference in sharing behavior between normal friends and Facebook friends?

to share video links is still email, followed by instant messaging. Social media or on-site mechanisms (like the *'Share button'* on *YouTube* etc.) are not very popular. Some people also just share a link or give a video recommendation while talking to another person. They simply explain where to find it or just give hints for search terms to find the video, which is also some kind of sharing. A summarization of the results from the semi-structured interviews is given in Figure 3.12. The quantitative study shows some

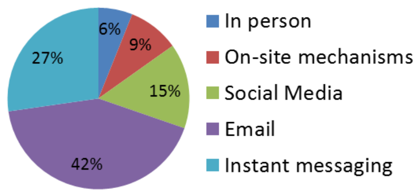

Figure 3.12: Distribution of link sharing mechanisms

differences in the users preferences. Participants preferred sharing links over social media the most, followed by email and instant messaging (Table 3.10). We also validated if this result may be biased through the great response from the *Reddit* community, but this is not the case. When selecting all responses except those coming from Reddit, we nearly get the same results.

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)	Rating average	Rank
In person by showing them on a computer screen or mobile device	9.5%	**36.5%**	35.8%	16.8%	1.5%	2.64	5
On-site mechanisms (share button)	25.9%	17.8%	22.2%	**27.4%**	6.7%	2.71	4
Social Media (Twitter, Facebook, Reddit, Digg etc.)	18.0%	5.8%	19.4%	**43.2%**	13.7%	3.29	1
Email	14.3%	25.7%	22.1%	**31.4%**	6.4%	2.90	2
Instant Messaging	20.6%	17.6%	**28.7%**	23.5%	9.6%	2.84	3

Table 3.10: Which mechanisms do you use to share video links?

What do people do when they receive a video link?

Our study results showed that people watch videos from recommendations (e.g. personal, via email, etc.) quite often. The majority of people (93) watch videos from recommendations once a week or more, followed by 67 people who watch them once a month or more (25 once a day or more, 26 rarely, 3 never).

In our quantitative survey, we asked how people react when receiving a video link: the average rating is shown in Figure 3.13. In most cases people try to watch the video right away and avoid delaying it or they watch it as soon as possible. Very few people ignore it and watch it later or make a note or bookmark for that video to watch it when they have time for it.

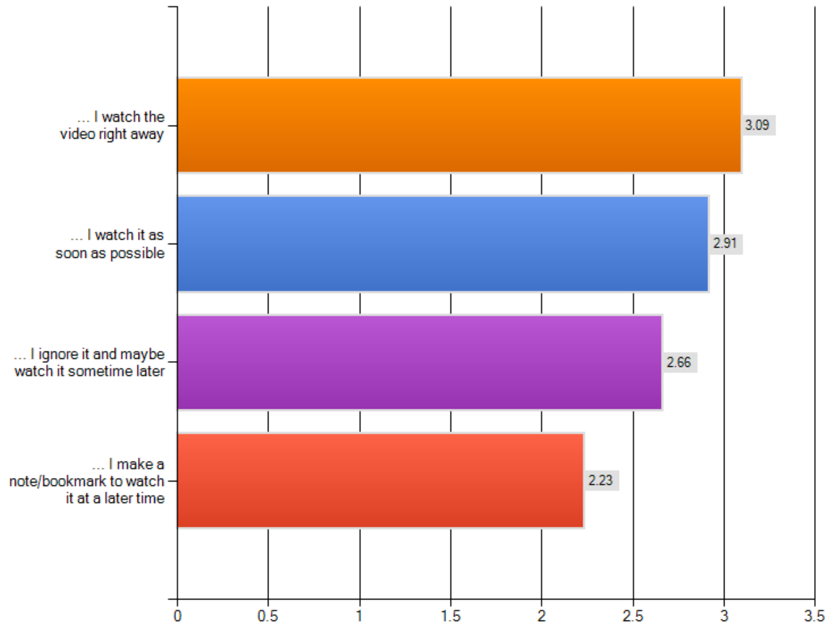

Figure 3.13: Average rating (ordinal scale from *Never (1)* to *Always (5)*) for different behavior patterns when receiving a video link

3.3.3 Online video consumption behavior

In this section we look at scenarios or situations, which are associated with video watching. A scenario or situation is defined as the set of conditions under which video consumption takes place. It encompasses the event that triggered the intention to watch the video content as well as the time, place and device in which it was watched.

Based on our semi-structured interviews, each user reported on up to three scenarios when they consumed video content. Overall, we collected 49 specific scenarios that is an average of 2.2 scenarios per user. In this section we summarize our findings and report on the most interesting situations experienced by our participants. We then also compare our findings with the quantitative questionnaire in order to confirm that some of the behavior is common throughout video retrieval. The most important difference in methodology between the two studies is that in the semi-structured interviews users were asked open questions and relied on memory of past events, whereas in the quantitative survey, they were asked more specific questions regarding their viewing preferences, interests, events which caused them to watch a video, as well as preferred place and time to watch videos.

In summary, active search by a user was performed in 67% of the reported scenarios and that in the remaining 33% of the scenarios a direct link to a video was received (via *Facebook*, email, instant messaging, or face-to-face conversations). In the semi-structured interviews, more than half of the videos associated with the 49 reported scenarios were watched in the evening or at night, followed by the afternoon and morning (Figure 3.14). We also found similar results in our quantitative study, but we introduced separate choices for evening and night (Figure 3.15). These findings are not very unusual due to the fact that most participants are at work or at school during the morning and most of the afternoon.

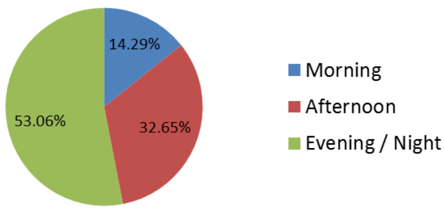

Figure 3.14: Distribution of the time of the day when people prefer to watch online videos (from semi-structured interviews)

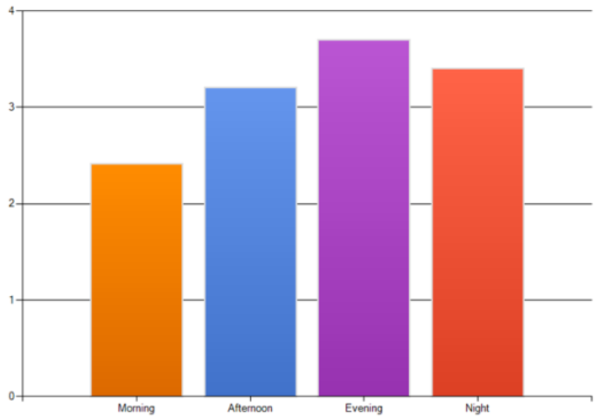

Figure 3.15: Average rating of preferred time to watch online videos (quantitative)

Additionally we investigated where videos are being watched. Using the same 49 situations in the semi-structured interviews, we learned that in 40 cases the videos were watched at home, four videos were watched at the University and only three videos where watched at work. In two situations the video was watched at a friends house while hanging out. One can argue that the number of videos watched at work might in reality be higher than reported due to the fact that people may not want to share that experience if it is not work related. Within our quantitative questionnaire, we asked where people prefer to watch videos by giving them ordinal scale items ranging from *Never (1)* to *Once a day or more (5)*. The rating average can be seen in Figure 3.16. Clearly watching videos at home is the most preferred one, while all other choices scored significantly lower. An interesting fact is that watching videos on the go is not yet very popular, despite the growing usage of smartphones or tablets with online video capability.

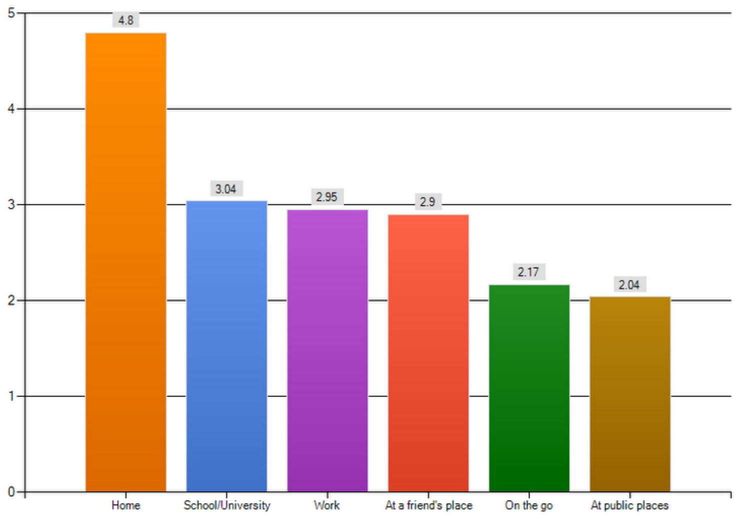

Figure 3.16: Where do people prefer to watch videos?

In the semi-structured interviews, for all the 49 reported scenarios the device used in 43 of the cases was either a PC/laptop or notebook, five times participants watched the video on a TV and just in one case a mobile device was used.

In the semi-structured interviews, most of the videos watched in the reported situations were of short length: 45% are only 1 to 4 minutes long and 35% are 5-10 minutes long (Figure 3.17). During the quantitative surveys, in order to test the hypothesis that most of the videos watched are of short length, we specifically asked the question if participants prefer short videos (up to 5 minutes) by letting them express their level of agreement. Out of the 217 participants, 1.4% strongly disagreed, 2.8% disagreed, 30.0%

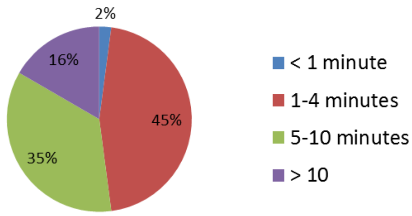

Figure 3.17: Distribution of average video lengths

were neutral where 37.1% agreed and 28.6% strongly agreed to this statement. If we compare these findings to results in literature, we can see certain similarities. Based on the analysis of the traffic of a campus network [53], mean and median values of long term popular videos showed durations between 2.5 and 3.5 minutes. In [44] the authors analysed the video length distribution within *YouTube*. Three peaks where found whereas the first one is within one minute and contains more than 20% of the videos, the second is between 3 and 4 minutes with 16% and third peak is near maximum of 10 minutes which is caused by the *YouTube* 10 minute video limitation for normal users. We can see that the first and second peak indicates that *YouTube* is mainly for shorter videos, which is also preferred by the majority of participants in our survey.

Is there a correlation between events and genres?

While analysing and evaluating the interview transcripts from the semi-structured interviews, we chose to assign each scenario to one of the 15 predefined categories by *YouTube*. To accomplish this, we tried to comprehend which video the user would have watched by analysing the interview transcripts. It was possible for all scenarios to assign a clear and non-ambiguous category. Furthermore, we characterized seven events that caused the video consumption in the first place. These seven events are described as follows:

- **Leisure:** Videos in this category were watched in leisure time, mostly out of boredom, for entertainment, or to kill time. Active search by the user is performed and a video is selected from the results.
- **Pull services:** Video links pulled by the users, e.g., instant message, email. No search was needed because it was a direct link to a specific video.
- **External events:** Follow-up active search by users triggered by an event on TV, radio or newspaper, etc. They then decided to research and follow-up on this topic by actively searching.

- **Face-to-face communication:** Users talked to another person about a certain topic or about a specific video already. At a later time, they then researched or followed up on that topic by actively searching. So these videos were watched as a result of talking to another person.
- **Task:** Users need an answer or they had to solve a specific task. A video is retrieved to accomplish the goal.
- **Facebook/RSS/Feature usage:** *Facebook*, *RSS* and any feature usage can be seen as a push service. Users get the information and click on a video link if it sounds interesting for them. No active search is performed by the user in this case.

The results of this matching between events and genres can be seen in Table 3.11. The table shows only 10 columns because none of the situations matched any of the following *YouTube* categories: *Entertainment*, *Howto & Style*, *Nonprofits & Activism*, *People & Blogs*, *Pets & Animals*. We can see in our result that the most watched videos belong

	Autos & Vehicles	Comedy	Education	Film & Animation	Gaming	Music	News & Politics	Science & Technology	Sports	Travel & Events		
Leisure				2	1	10			3	1	17	34.7%
Push services									4		4	8.2%
External events	1	2		2		3	1	1	3		13	26.5%
Face-to-face						1		1			2	4.1%
Task	1		2		1			2	1		7	14.3%
Pull services						3			3		6	12.2%
	2	2	2	4	2	17	1	4	14	1	49	
	4.1%	4.1%	4.1%	8.2%	4.1%	34.7%	2.0%	8.2%	28.6%	2.0%		

Table 3.11: Cross tabulation showing the co-assignments of genres (columns) to trigger events (rows)

to the genre *Music* with 34.7% and to the genre *Sports* with 28.6%. The most frequent trigger event was *'Leisure'* with 17 nominations (coloured grey), closely followed by *'External events'* with 13 nominations. For example, the *Entertainment* genre was not assigned to any of the videos reported by the interviewees and yet one can claim that a big goal of viewers spending their leisure time is to entertain themselves. The trigger event having the biggest spread in categories (seven categories out of 10) is *'External events'*. This can be easily explained with the broad semantics of the external event. More specifically, we do not know about the semantics, topic or content of the external event that triggered the retrieval process. The same goes for task solving. As the goal of the task is independent from the fact that there is a task to solve the category spreads across six genres and shows no peaks in absolute numbers. Even more interesting is the inverse investigation: how do genres relate to trigger events. Again *Music* gives a good example due to its dominance in the reported instances, coloured cyan. There is an indication that retrieving music videos is very much related to the *'Leisure'* cluster. *Sports* genre on the other hand features nearly all of the six trigger events, coloured yellowish.

In our quantitative questionnaire we also wanted to find out more about events that triggered video consumption. Therefore we asked how often users follow any of the events which are shown in Table 3.12.

	Never (1)	Rarely (2)	Once a month or more (3)	Once a week or more (4)	Once a day or more (5)	Rating average	Rank
I watch videos when I am bored	6.9%	21.7%	18.9%	**30.4%**	22.1%	3.39	1
I watch videos (or follow video links) I get in emails	12.6%	**30.7%**	21.9%	30.2%	4.7%	2.84	6
I watch videos (or follow video links) I get in instant messages	20.8%	**27.4%**	19.8%	25.5%	6.6%	2.70	7
I search for videos to follow up something on TV/radio/newspaper/magazine	9.8%	24.7%	**37.2%**	24.7%	3.7%	2.88	5
I search for videos to follow up something after I talked to another person about it	3.3%	22.2%	**42.9%**	27.4%	4.2%	3.07	3
I search for videos to get help on what Im going to do (e.g. how to)	6.6%	23.5%	**38.0%**	27.2%	4.7%	3.00	4
I watch videos published on RSS streams/feed-s/blogs	**28.9%**	27.0%	24.2%	16.6%	3.3%	2.38	8
I watch videos posted on Facebook (MySpace, Twitter, etc.)	21.6%	17.4%	14.6%	**32.4%**	14.1%	3.00	4
I search for videos to learn how something works / looks like	4.2%	21.5%	**41.1%**	27.1%	6.1%	3.09	2

Table 3.12: Events that lead to watch videos online

Additionally we performed Spearmans rank correlation analysis and our quantitative survey data to further support that peoples different motivation sources to watch videos are correlating with genres they are watching. Table 3.13 shows the nine most significant events triggering video consumption (one per row) and the 15 *YouTube* categories used for video genres in our surveys (one per column). The Spearmans correlation coefficient

rs for each combination of event and genre/category and its two-sided significance p are also shown. High values of rs, together with a high statistical significance ($p < 0.05$) indicate a relation between two variables. The three strongest relations of events and genres for each event are marked green, and the three weakest relations are marked red in Table 3.13. Based on those correlations we postulate that if users use the video platform with a specific intention, some categories will likely contain a larger number of relevant results than other categories. Besides the ranking of categories, we want to point out other insights derived from the numbers in Table 3.13, which include:

- The strongest correlation ($rs = 0.404, p = 0.000$) was found between the event *'I search for videos to get help on what I'm going to do'* and the *Howto & Style* category.
- The *Music* category has very poor statistical significance with $p > 0.094$ across all events in active search. One explanation can be found in statements given by our survey participants. Often they choose to just listen to music on *YouTube* without paying attention to the video at all while working on other, more important tasks.
- The *Education*, *Film & Animation*, and *Non-profits & Activism* categories did not make it to the top 3 list for any event either.
- All negative correlations appeared with poor values for statistical significance ($p > 0.139$).
- The event *"I search for videos to follow up something on TV/radio/newspaper/-magazine"* has the biggest number of significant positive correlations, namely 11 (out of 15).
- The event *"I watch videos (or follow video links) I get in emails"* has the lowest number of significant correlations, namely 6 (out of 15).

		Autos & Vehicles	Comedy	Education	Entertainment	Film & Animation	Gaming	Howto & Style	Music	News & Politics	Nonprofits & Activism	People & Blogs	Pets & Animals	Science & Technology	Sports	Travel & Events
Active search																
I watch videos when I am bored	rs	-0.078	0.362	0.159	0.232	0.217	0.227	0.114	0.111	0.197	0.143	0.262	0.221	0.039	0.187	0.092
	p	0.257	0.000	0.020	0.001	0.001	0.001	0.098	0.108	0.004	0.038	0.000	0.001	0.569	0.006	0.180
I search for videos to follow up something on TV/radio/newspaper/magazine	rs	0.139	0.040	0.180	0.219	0.229	0.096	0.212	0.115	0.282	0.209	0.326	0.202	0.086	0.206	0.253
	p	0.044	0.561	0.009	0.001	0.001	0.165	0.002	0.094	0.000	0.002	0.000	0.003	0.212	0.003	0.000
I search for videos to follow up something after I talked to another person about it	rs	0.141	0.177	0.172	0.169	0.222	0.122	0.159	0.089	0.055	0.134	0.157	0.232	-0.004	0.257	0.251
	p	0.041	0.010	0.012	0.013	0.001	0.078	0.021	0.179	0.431	0.053	0.024	0.001	0.950	0.000	0.000
I search for videos to get help on what I'm going to do (e.g. How-to)	rs	0.237	0.037	0.154	0.053	0.022	0.203	0.404	0.025	0.033	0.084	0.222	0.174	0.189	0.267	0.132
	p	0.001	0.593	0.026	0.447	0.754	0.003	0.000	0.713	0.635	0.228	0.001	0.110	0.006	0.000	0.057
I search for videos to learn how something works / looks like	rs	0.276	0.042	0.279	0.071	-0.023	0.200	0.400	-0.067	0.056	0.208	0.290	0.211	0.331	0.146	0.174
	p	0.000	0.544	0.000	0.302	0.734	0.004	0.000	0.336	0.421	0.002	0.000	0.002	0.000	0.034	0.120
Direct Link Access																
I watch videos (or follow video links) I get in emails	rs	0.124	0.116	0.064	0.094	-0.010	-0.010	0.216	0.147	-0.048	0.048	0.211	0.142	0.036	0.258	0.185
	p	0.072	0.092	0.351	0.170	0.886	0.882	0.002	0.032	0.488	0.493	0.002	0.040	0.602	0.000	0.007
I watch videos (or follow video links) I get in instant messages	rs	0.110	0.223	0.103	0.171	0.157	0.224	0.136	0.189	-0.102	0.024	0.140	0.105	0.074	0.334	0.175
	p	0.095	0.001	0.136	0.013	0.023	0.001	0.049	0.006	0.139	0.725	0.044	0.128	0.286	0.000	0.011
I watch videos published on RSS streams/feeds/blogs	rs	0.172	0.130	0.154	0.065	0.018	0.155	0.215	-0.007	0.199	0.235	0.372	0.222	0.100	0.238	0.274
	p	0.013	0.061	0.026	0.354	0.793	0.027	0.002	0.919	0.004	0.001	0.000	0.001	0.148	0.001	0.000
I watch videos posted on Facebook (MySpace, Twitter, etc.)	rs	0.041	0.187	0.216	0.184	0.125	0.108	0.140	0.289	0.024	0.077	0.303	0.118	-0.029	0.309	0.251
	p	0.556	0.006	0.002	0.007	0.069	0.120	0.043	0.000	0.725	0.268	0.000	0.088	0.677	0.000	0.000

Table 3.13: Event-genre correlation coefficients and statistical significance.

CHAPTER 4 Prototype

With our prototype we want to show a first case study on how our most promising findings can be used for enhancing user experience in video retrieval. We propose a software prototype that implements an adaptive video retrieval system, that utilizes the users' intentions to provide better search results in a user interface adapted to the intentions and needs of users. Regarding the topic of view adaptation in user intention-based video retrieval, to the best of our knowledge, there are just a few published papers. One example would be the work from Hopfgartner et. al [26] where the authors evaluated implicit feedback models for a better adaptive video retrieval experience. Therefore we suggest that this topic deserves deeper and further investigation.

Based on our survey results, we have started developing a prototype whose main goal is to show the impact and benefit of the integration of user intentions in search and retrieval processes in multimedia information systems. It is based upon statistically relevant usage patterns and insights derived from our survey data. We anticipate that the following benefits will be achieved through such a prototype:

- Better search results with higher precision with respect to a user's context based on intentions. By e.g. a dynamic adaptation of relevance functions in search
- Better understanding of queries through intentions driving information need, including query reformulation, query extension, and relevance feedback.
- Dynamic hiding of options and view adaptation based on the users' intention.

Our first step in the development of the prototype, which is the main part of this chapter, has been focused on improving the precision of search results by exploiting the correlation between intentions and genres. The additional benefits listed above will be addressed at later stages. The basic block diagram for the prototype can be seen in Figure 4.1 and the main portions of it are explained in the remainder of this section. As it can

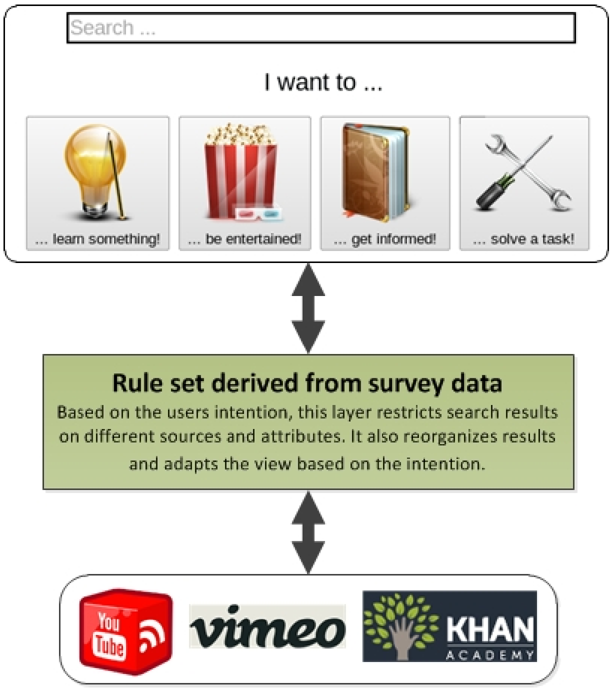

Figure 4.1: Block diagram of the prototype.

be seen in Table 3.13, certain genres show a higher correlation coefficient for certain situations. Based on these results we can implement ad-hoc hard-coded rules so that videos of certain categories with higher correlation coefficients are ranked higher in the search results.

At a later development stage of this prototype, the resulting rule set needs to be influenced by additional aspects of the video retrieval process which were not mentioned earlier, such as:

- the video quality of the resulting clips: e.g., participants of our survey reported to prefer HD video quality when they want to watch a video to learn something;
- the inclusion of other sources (e.g., Vimeo and Khan Academy), in the search results;
- view adaptation based on the users' intention: e.g., survey participants reported that not every option and metadata needs to be displayed at all times for every intention.

For better understanding, we outline a simple example scenario: Let us assume that a user wants to learn about a specific topic, therefore he would use our prototype and type in query terms about the topic of interest. Since the user has a *learning intention* in mind, he clicks on the *'I want to learn something'* button. Our pattern-based rule set then optimizes the search to certain sources and categories. In this example case, *YouTube* and the *Khan Academy* will be used as video sources and the videos will be ranked by categories from the strongest to the weakest correlation. Occasionally, based on our survey results, users are more interested in HD videos when they want to learn about specific topics. Additionally the result view will be adapted to fit the users' intention. In the case of a *'learning-intention'* the final result view will probably show a detailed description of the video along with the user rating and tags.

First, section 4.1 will cover details about the technical background of our prototype. The following sections will briefly explain the main screen in section 4.2, the results view in section 4.3, and section 4.4 will elaborate on the video view screen. Afterwards, in section 4.5 we will discuss the survey setup, talk about the task descriptions, and the technical monitoring features we implemented within the prototype. In the last section 4.6 we present a summary of the prototype survey results.

4.1 Technical background

We decided to develop a web application in PHP5 to implement our prototype because it is the most natural environment for a user to work with a video retrieval platform. In the first stage we focus on *YouTube* as a video source since they offer an extensive API to exploit their resources, namely their videos and the videos metadata. The whole application underlies a small database schema which stores necessary data like offered intentions, correlation data (as seen in Table 3.13), possible adaptation based on intention and the logging of user actions. The database-management-system used is MySQL 5 and an overview of the schema can be seen in Figure 4.2, the SQL script (without data) is shown in Appendix C.

We also ran into some limitations of the YouTube API by Google during the development our prototype. When performing many data requests in a short period of time the API will return a **too_many_recent_calls** error because it uses a quota system to protect against excessive traffic or flooding. It ensures that one person's actions do not negatively impact all other users[1]. To display the results per category, we used a jQuery plugin called jCarousel[2] which let us display a list of items (in our case thumbnails of

[1] More on the YouTube API limitations can be read here:
`http://code.google.com/apis/youtube/faq.html#operation_limits`
`http://code.google.com/apis/youtube/2.0/developers_guide_protocol.html#Quota_errors` (both accessed on May 4, 2011)

[2] jCarousel by Jan Sorgalla: `http://sorgalla.com/jcarousel/` (accessed October 17, 2011)

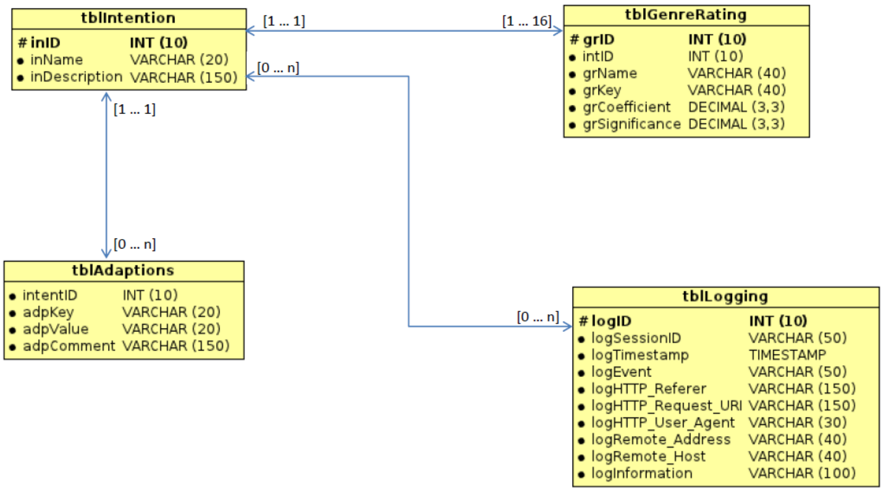

Figure 4.2: IntentDB database schema

videos) in horizontal order which can be scrolled back and forth.

View adaption

At the current stage of the prototype this feature will not be included for the evaluation, but we already prepared the feature for future tests. Each page will look the same for each intention during this survey since we do need further investigation which features and metadata should be shown for different intentions. Nevertheless we explain this feature for completeness. Each adaptation option for an intention is expressed as one row in the table *tblAdaptions*. Let us consider that for a certain intention, the rating of a video in the *Video view* should be shown, therefore a row as follows could be inserted in this table:

```
IntentID = "1"
adpKey = "VV_SHOW_RATING"
adpValue = "1"
adpComment = "Show video rating in the video view"
```

The column *IntentID* is a reference to the column *inID* of the table *tblIntention*, the column *adpKey* is a key which is used to identify the adaption option for each intention and the column *adpValue* holds the value of the certain adaptation which can be either true or false or some other value like *pixels*, *number of items* or the like. These values can then easily be queried by a PHP function to perform the adaption on the different pages. In the final developement stage, these values should be adapted dynamically through machine learning algorithms. For now, the following adaptions are possible which are shown in the format *"Description (adpKey, adpValue)"*:

- **Result view**
  - Size of the video thumbnail (RL_THUMBNAIL_SIZE, 200px)
  - Show user rating of the video (RL_SHOW_RATING, 1)
  - Show short video description in result view (RL_SHOW_VIDEO_DESC, 1)
- **Video view**
  - Show related videos in video view (VV_SHOW_RELATED, 1)
  - Show user comments in video view (VV_SHOW_COMMENTS, 1)
  - Show video rating in video view (VV_SHOW_RATING, 1)
  - Video size of videos in video view (VV_VIDEO_SIZE, 1)

4.2 Main screen

Our prototype gives the user a minimalistic intention-based user interface, consisting of a query text-box and one button for each intention (Figure 4.3). After providing the search terms, the user has to explicitly state his intention by clicking one of the provided search buttons. By doing so, our rule-set layer populates the most relevant videos for the user's search and redirects him to the result view, which is explained in more detail in the next paragraph.

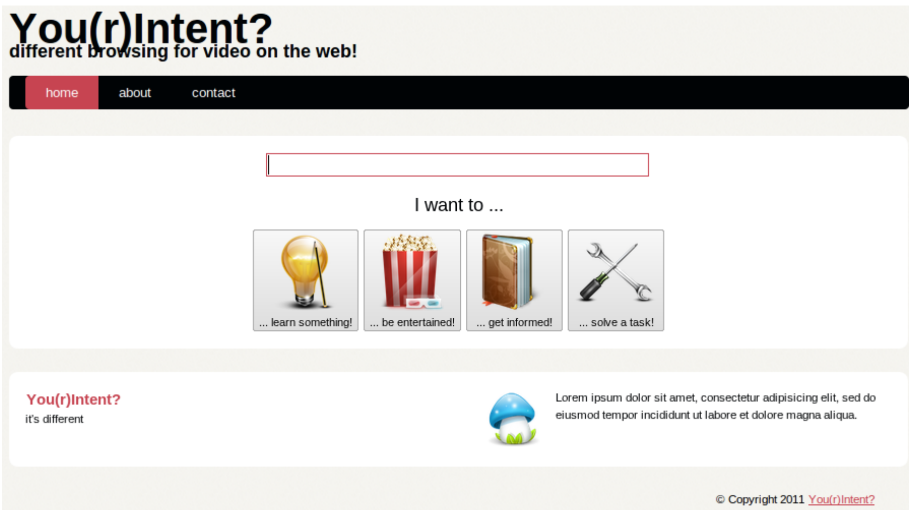

Figure 4.3: Main screen of the prototype

4.3 Result view

Displaying adapted results for different intentions can be beneficial for the end user. We decided to design an adapted view for each intention which is not yet covered in this prototype. For now we only included YouTube as a video source and we show the intention selected, the query term and the prioritized categories based on the selected intention. In the final development stage it will show which video sources are used, which restrictions to video quality and other features are applied and the like. These restrictions and views will also be changeable by the user as needed. After the user clicks on a video from the shown result, the *Video view* (see Section 4.4) screen will be shown.

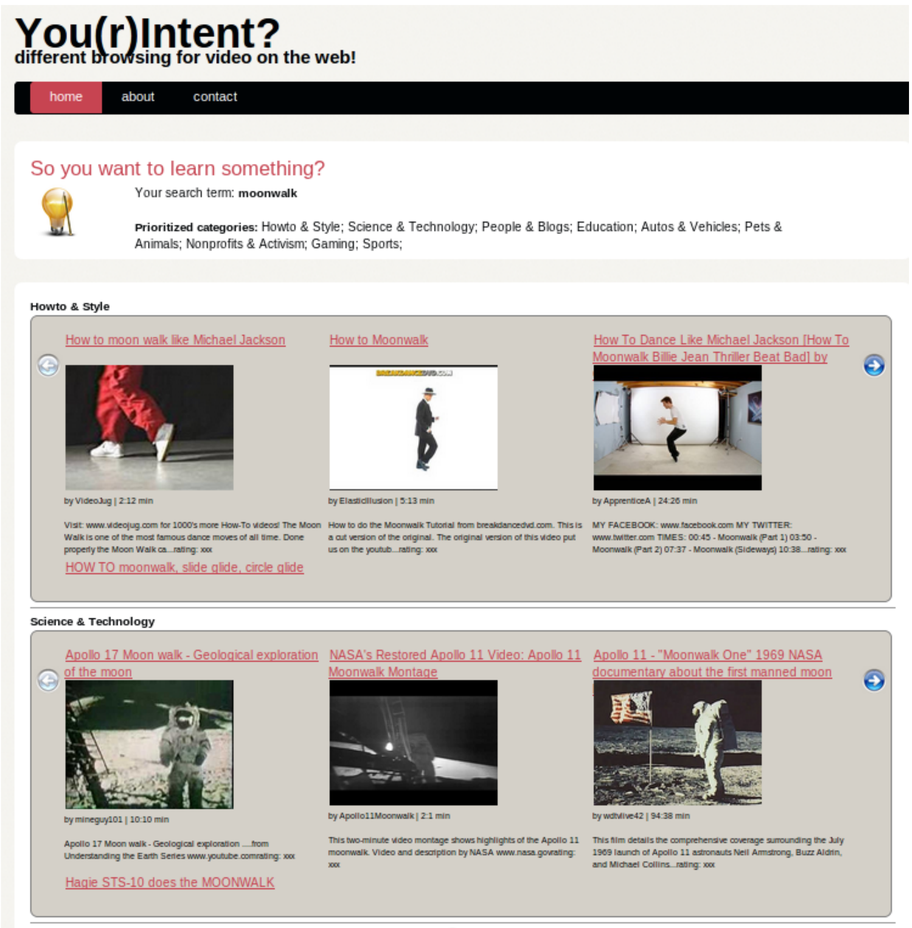

Figure 4.4: Result screen of the prototype

4.4 Video view

The selected video will be shown with the default embedded *YouTube* player. At this stage of the prototype we also show the user who uploaded the video, the length of the video and the related videos which are determined via the *YouTube* API. In the final version of the prototype, also the video view adapts to the users' intention as the screens before and is based on the same principles. Based on the intentions of the users, the visibility and detail of video description, user comments, ratings and related videos will be determined and adapted. This option is not implemented at this stage of the development of the prototype.

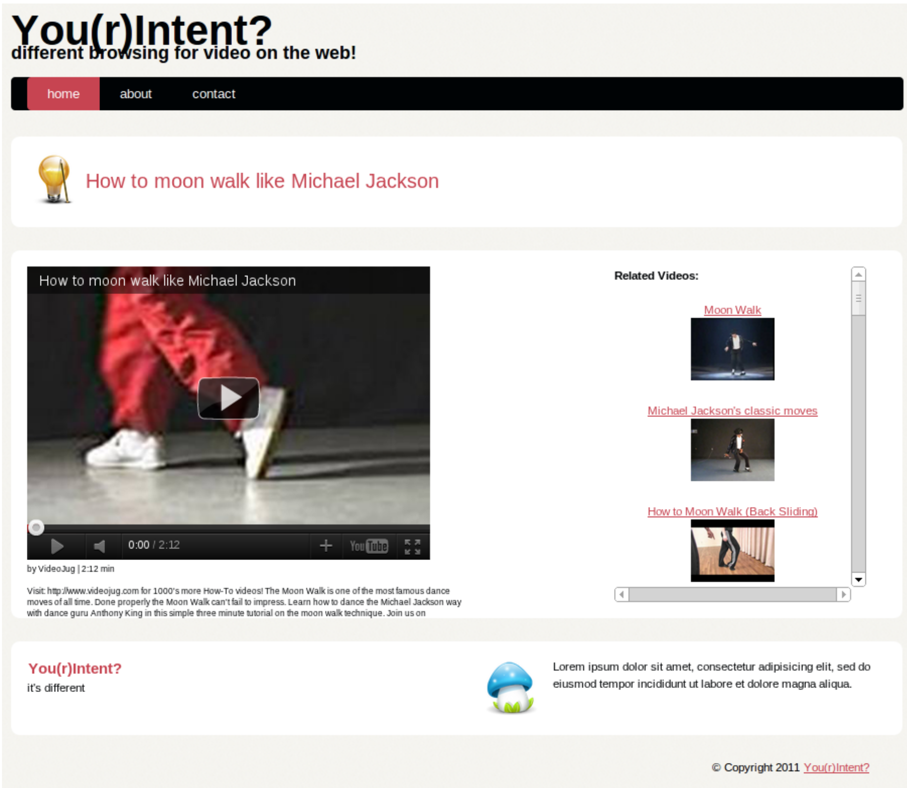

Figure 4.5: Video view screen of the prototype

4.5 Evaluation of the prototype

After finishing the development of the initial stage of the prototype, we performed a user survey in order to determine how satisfied participants are with the overall performance and the search results. Evaluation methods used for this survey are observation of the interviewee, the analysis of mouse tracking heat maps, activity logging, and semi-structured interviews with the study participants.

Five participants worked with our "You(r)Intent" Prototype one after another and each one had to solve five predefined tasks. We explained to the participant on how to use our prototype beforehand if this was necessary. Each one performed the tasks under our supervision and we took notes on their behaviour. We also acquired additional data about their behaviour through our build-in logging in the prototype. Every action will be logged in a table (see *tblLogging* in the IntentDB database schema, see Appendix C) and we also used heatmaps to display the areas of our pages which are most frequently clicked. After each participant was finished, personal interviews were performed in order to get feedback on the qualitative performance of our prototype eg. the participants satisfaction with the ordering, the accuracy and relevancy of our presented search results. The questionnaire can be seen in Appendix A.3 and consists of relevant demographic questions, five questions about the prototype itself and three short questions per task. The questionnaire served mainly as a guideline for the interviewer. Questions were asked immediately after the participant was done with all tasks. The whole survey took up to 30 minutes per participant. Again, no gifts or vouchers were given to the participants to perform the survey.

Task description

With five predefined tasks we want to cover the relevancy of our four intentions: *'... learn something!'*, *... be entertained!*, *'... get informed!'*, and *'... solve a task!'*. Our tasks are defined as follows:

1. **Moonwalk**
   You are in a discussion with a friend about the first moonwalk of the Apollo 11 mission. You want to show him a video about it.

2. **Juggling**
   You see a street artist who performs some great looking juggling tricks. The audience around him is fascinated, so are you. When getting back home, you want to try it our yourself. Search for some videos which show you how to do it.

3. **Ski accident**
   A famous Austrian skier had a crash at the Nagano Olympics in 1998. What was his name and how often did he flip over?

4. **Bonsai**
   You are the new owner of a Bonsai tree. Those trees demand more work than

normal plants. You need to know things on how to take care of those trees, how to trim them etc. Search for videos which help you take care of your Bonsai.

5. **Lego Mindstorms**
   What is the most creative *'Lego Mindstorms'* video you can find?

We especially wanted to find out which intentions were chosen by the participants for the different tasks and at which position they found a proper solution for each task.

Technical monitoring

Heatmaps:

Heatmaps are created for the start page and for each intention on the result page and view page. This is a total of nine different heatmaps for our prototype analysis but we will only take into account the start page because the *Result view* and *Video view* page are not of interest for now. We use an existing solution to create heatmaps which is called *ClickHeat*[3]. This is an easy to implement java script which just needs an individual java call on each page (and if needed also one call for each intention). For the analysis, each heatmap can be customized and restricted in order to show only the data of certain browsers, screen sizes, or certain days, weeks or months. An example heat map of our start page can be seen in Figure 4.6. For additional heatmaps, the following

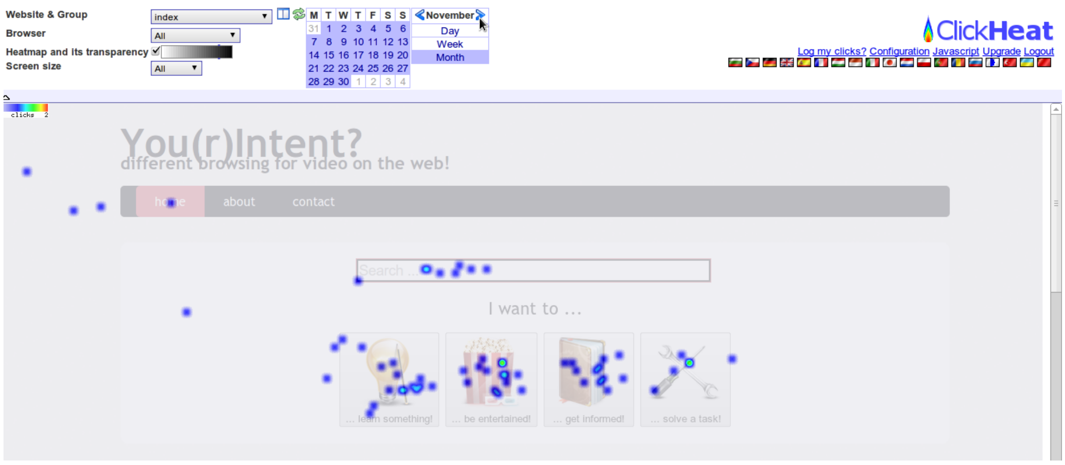

Figure 4.6: Example heatmap of the start page

JavaScript code must be inserted or changed:

```
<script type="text/javascript">
```

[3]ClickHeat: `http://www.labsmedia.com/clickheat/index.html` (accessed November 8, 2011)

```
<!--
  clickHeatSite = 'yourintent';
  clickHeatGroup = 'index';
  clickHeatServer = 'http://localhost/clickheat/click.php';
  initClickHeat();
//-->
</script>
```

The variable *clickHeatGroup* can also be assigned dynamically so that a heatmap for each intention will be created.

Logging:

Each page will log the actions of a user session in the table *tblLogging*. A user session begins when a users clicks on one of the four buttons of intentions at the start page. A new user session begins when one of the buttons is clicked again and therefore a new search is performed. During such a session, each page change (ie. from the *Result view* to the *Video view*) inserts one log entry into the table. A possible row in this table could look like this:

```
logID = "1337"
logSessionID = "ggh3spc74lspsqhr2db6fssf54"
logTimestamp = "2011-11-26 18:05:20"
logEvent = "SEARCH_VIDEO"
logHTTP_Referer = "http://localhost"
logHTTP_Request_URI = "/search.php?txtQuery=test&btnIntent=Entertain"
logHTTP_User_Agent = "Mozilla/5.0 (X11; Linux x86_64)"
logInformation = "intent=ENTERTAINMENT;query=test"
```

Until now, implemented *logEvents* are *USER_SESSION_START*, *SEARCH_VIDEO*, *WATCH_VIDEO*. Further logging of events and actions can be implemented easily if necessary, see the short PHP example beneath. Again, the second and third parameter for this function call can be assigned dynamically.

```
$db->log(session_id(),
         'SEARCH_VIDEO',
         'intent=' . $_SESSION['intentID'] .
                  ';query=' . $_REQUEST['txtQuery']
);
```

4.6 Evaluation Results

After finishing the development of the prototype we randomly asked five people to work with our prototype so that we can gather some informational data. A survey with a bigger audience will be performed after all proposed features are implemented and tested. Demographics of our participants can be seen in table 4.1.

ID	Occupation	Gender	Age	Days of usage/week	Language answered
1	Student, bank employee	female	25-30	3	German
2	Student	male	25-30	7	German, English
3	Employee	male	31-40	1	German
4	Student	male	25-30	7	German
5	Student (PHD) *	male	25-30	7	German, English

Table 4.1: Participants demographics of the prototype survey

* Participant 5 has prior knowledge to research in the field of intentions on image retrieval.

Tasks

This section reports on the evaluation of the five task. A more detailed description of the tasks can be seen in section 4.5. In the coming paragraphs we present a table for every task which lists the query words each participant used to search and the amount of query refinements, if needed. Also we show in which category the video was found (counted from the top) and the position of the video in that category which was selected as a solution. For each task we asked two further questions. The first one (Q1) was about the relevancy of the task *in everyday life* where the participant could rate from *'Irrelevant (0)'* to an *'Every day task (5)'*. With the second question (Q2) we wanted to know if the participant thinks that the task could be solved easier with Google than with a video platform search (*'1=Strongly disagree'* to *'5=Strongly agree'*).

First off, all five participants solved each task and found a proper video with our prototype which they considered as a proper solution. The interviews with our participants afterwards showed some interesting insights. A list of statements for each question given by each test candidate can be seen next[4]:

[4]Statements of the participants are translated from German to English.

What is your first impression of the prototype? What was different, what did you like, what did you dislike?

- **Participant 1:** I liked the graphics of the buttons for the intention and the handling of the prototype was easy.
- **Participant 2:** It would be nice to have a more detailed description of the intentions, how are those intentions defined, what is their difference? Before you perform a search you have to think what you want to do.
- **Participant 3:** I recognized the intentions and I also noticed that you really have to think beforehand what you want to do and what your intention is. This was new to rather me.
- **Participant 4:** I always just pressed the *"Enter"* button at first after I typed my search query because I wasn't familiar that I have to explicitly click on one intention.
- **Participant 5:** I liked the shown intentions and that only a subset of categories is shown and ranked according to my query and selection.

Did you miss any intentions? What else are you doing on video search engines except the things you have seen on our prototype

- **Participant 1:** For me everything is information, I could use this intention for every task. You have to think beforehand what you need. With some tasks it is clear from the beginning with other tasks not so much.
- **Participant 2:** What is the difference between each intention? In my opinion the *Learning* and *Information* intention could be merged. On YouTube I also use the features: *Watch Later*, *Favourites*, and *Playlists* etc.
- **Participant 4:** I have never needed or used the *"I want to get informed!"* or *"I want to solve a task!"* intention to solve the five tasks.
- **Participant 5:** I can't think of an intention which could be missing but it could be that their is a difference (for me) when I watch a video in private or with other persons and that this is not considered. [...] What is the difference between the *"I want to learn something"* and *"I want to get informed"*? There maybe also is a difference if I search for theoretical or practical knowledge in the case of the *"I want to learn something!"* intention.

How did you like the ordering of the results? Would it be an improvement for some searches?

- **Participant 1:** I did not really pay attention to this detail, but thinking about it afterwards it is certainly helpful for some tasks.
- **Participant 2:** I liked the separation into single categories because I could easily skip categories when I know that a video is most likely not in a certain category. For example a video is not related to sports so I can skip the *Sports* category.
- **Participant 3:** Nothing special caught my eye.
- **Participant 4:** The breakdown into categories is beneficial and also useful for some searches.
- **Participant 5:** Yes it would be a benefit for some searches/tasks but one problem may be that you cannot influence and control which category a user assigns to his video upload. This can cause *invalid* videos (e.g. a snippet of serious news broadcast in the *Comedy* category) and therefore loss of precision in the retrieval results.

For question 4 it was just mentioned by participant one that there should be an informational message if no video was found an. Participant 5 noted that it would be beneficial if there is a better description for the four intentions and probably an explanation how the ranking is calculated. This will be implemented in the next version of the prototype. Question 5 was not answered by any of the participants.

Moonwalk

Looking at results in table 4.2, we can see that participants used two different intentions to solve this task, namely *Information* (3 times) and *Learning* (2 times). In every test a video was found in the first displayed category and four out of five times the first video of that category was selected. Concerning the relevancy of the task, participants rated very diverse, resulting in an average of 3.0, which can be interpreted as slightly important in every day life. Looking at Question 2, the average 2.4 indicates a tendency that such tasks are solved easier with a video platform than with a Google (Web-) query.

Juggling

Table 4.3 shows the result details for this tasks. This task was solved with three different intentions, one even used the *Entertainment* intention and found a proper video. Again, in all cases the video was found in the first category shown and each participant found

Participant	Query terms	Query Refinements	Intention clicked	Found in Category	Video No.	Solved	Rating Q1	Rating Q2
1	*moon walk*	0	Information	1	2	Yes	3	2
2	*moonwalk apollo 11*	0	Information	1	1	Yes	5	3
3	*moonwalk apollo 11*	0	Learning	1	1	Yes	1	2
4	*first moonwalk*	0	Learning	1	1	Yes	4	1
5	*moonwalk apollo 11*	0	Information	1	1	Yes	2	1
	Average:			1.0	1.2		3.0	2.4

Table 4.2: Results Task 1: "Moonwalk"

a proper video within the first three items. Looking at the relevancy for this task with an average of 1.8, it shows that this can be considered less relevant on a day to day basis. Question 2 with an average of 2.0 again shows that such tasks are more likely to be solved with a video platform than with a Google (Web-) query.

Participant	Query terms	Query Refinements	Intention clicked	Found in Category	Video No.	Solved	Rating Q1	Rating Q2
1	*jonglieren*	0	Solve a task	1	1	Yes	2	2
2	*jonglieren 3 baelle*	0	Learning	1	1	Yes	4	2
3	*street artist jonglieren*	0	Entertainment	1	2	Yes	1	2
4	*jonglieren lernen*	0	Learning	1	3	Yes	3	3
5	*juggle*	0	Learning	1	1	Yes	4	1
	Average:			1.2	1.8		1.8	2.0

Table 4.3: Results Task 2: "Juggling"

Ski accident

Results for this task can be seen in table 4.4. Again, the two intentions used were *Information* (4 times) and *Learning* (1 time). In one case a participant needed four query refinements but this was due to typing errors. In one case a test candidate found a proper video in the sixth category which is the worst result of all tests conducted over all tasks, in all other cases a proper solution was found at least in the second category. Looking at the position of the videos within each category, again a solution was at least found at the second position. Participants rated the relevancy of this task with an average of 3.6 which can be considered to be a more relevant task in every day life. For question 2 the average rating was 2.4, which shows that people neither agree or disagree if the task could be solved easier with a Google (Web-) query.

Participant	Query terms	Query Refinements	Intention clicked	Found in Category	Video No.	Solved	Rating Q1	Rating Q2
1	*oesterreichischer skifahrer nagano unfall 1998*	0	Information	1	1	Yes	5	1
2	*nagano olympics ski accident*	0	Information	6	1	Yes	4	4
3	*nagano 1998 skiunfall*	0	Information	2	2	Yes	2	1
4	*oesterreichischer skifahrer nagerno unfall;* *oestereichischer skifahrer nagerno unfall;* *oestereichischer skifahrer nagano unfall;* *skifahrer nagano unfall;* *ski fahrer nagano 1998 unfall;*	4	Learning	1	1	Yes	3	5
5	*hermann maier nagano*	0	Information	7	1	Yes	4	1
	Average:			3.4	1.2		3.6	2.4

Table 4.4: Results Task 3: "Ski accident"

Bonsai

Table 4.5 shows the results of the fourth task participants had to solve. Also for this task the two intentions selected were *Information* (3 times) and *Learning* (2 times). For this task, one participant used query refinements because no videos were found with the initial query terms. Only in one case a proper solution was found in the sixth category. Within the categories a video was found on the second position in the *worst* case. The average rating of the task relevance is 3.4, which indicates a slight tendency to an every day task. The second questions average is 2.6 and therefore it can be considered that such tasks are probably easier solvable with a video platform.

Lego Mindstorms

For this last task, participants had to find what is for them the most creative Lego Mindstorms video. A summarization of the results can be seen in table 4.6. Four out of five participants solved this task by using the *"... want to be entertained!"* intention and just one participant used the *".. want to learn something!"* intention which also showed the worst category number and position of the video, namely in category three and video position two. For this task we can see a strong relevancy with an average of 4.2 as an everyday task. This fact also relates to the numbers in figure 3.9 which

Participant	Query terms	Query Refinements	Intention clicked	Found in Category	Video No.	Solved	Rating Q1	Rating Q2
1	*Haltung von bonsaibaum; Umgang mit bonsaibaum; Umgang mit bonsai;*	2	Learning	1	1	Yes	4	1
2	*bonsai pflege*	0	Information	1	1	Yes	4	3
3	*bonsai baum pflege*	0	Information	2	2	Yes	2	1
4	*bonsai pflege*	0	Learning	1	2	Yes	3	5
5	*bonsai pflege*	0	Information	6	2	Yes	4	3
	Average:			2.2	1.4		3.4	2.6

Table 4.5: Results Task 5: "Bonsai"

also shows that videos in Entertainment categories are the most preferred and watched ones. Participants also indicate with an average of 1.8 that it is easier to solve such (Entertainment) tasks with a video platform than with a Google (Web-) query.

Participant	Query terms	Query Refinements	Intention clicked	Found in Category	Video No.	Solved	Rating Q1	Rating Q2
1	*lego mind storms*	0	Entertainment	1	1	Yes	5	1
2	*lego mindstorms rubic cube*	0	Entertainment	1	1	Yes	5	3
3	*lego mind storms*	0	Entertainment	2	3	Yes	4	1
4	*lego mindstorms*	0	Learning	3	2	Yes	3	5
5	*lego mindstorm*	0	Entertainment	2	4	Yes	4	1
	Average:			1.8	2.2		4.2	1.8

Table 4.6: Results Task 5: "Lego Mindstorms"

Heatmaps

Since we also logged each click during the survey, we also can show the resulting heatmaps. Heatmaps for the *Result page* and the *Video view page* are not shown because they are not representable. In the case of the result pages this is caused by the dynamic nature of the ordering of the categories due to the different ranking based on the select intention and the individual videos shown by each search query. In the case of the *Video view page* we did not include the graphic because it showed no interesting data which could be relevant for this thesis. In figure 4.7 the heatmap of the start page is shown.

Figure 4.7: Heatmap of the start page after the survey

Given a bigger survey population, such heatmaps would certainly be of interest. They will play a bigger role and will be more informative during the next prototype survey in the future. In the current situation we have the results already in *written* form on how the intentions were used but nevertheless we can see that the two intentions *"... learn something!"* and *"... get informed!"* almost share the same amount of clicks, whereas the *"... solve a task!"* just got one click. Since we just had one task which was solved by almost everybody with the intention *... be entertained!* we can also explain the smaller amount of clicks for this area.

Prototype Conclusions

Overall the participants were very satisfied while working with the prototype and everyone solved each task more or less easily. Also the position at which a solution video was found was satisfactory throughout all tests. A summary of the average category position in which the video was found and the position of the video within that category can be seen in table 4.7. The numbers indicate that on average a video was found at least in the second category (average of 1.92) and a video was found at least at the second position within that category (average of 1.56).

Only in two cases query refinements were necessary to find a proper video and in one of those two cases typing errors were the cause. If we have a look at how often each intention over all tasks was used, the *Information*-intention was the most preferred one and was selected 10 times. The *Learning*-intention came close afterwards and was selected nine times. During the interviews some participants mentioned that they don't understand the difference between those two intentions or that these two intentions could be merged. Maybe this close result is another indicator for this conclusion. Since we just

Task	Average category position	Average video position within category
Moonwalk	1.00	1.20
Juggling	1.20	1.80
Ski accident	3.40	1.20
Bonsai	2.20	1.40
Lego Mindstorms	1.80	2.20
Average:	1.92	1.56

Table 4.7: Average category and video position per task

had one task which really needed the *Entertainment*-intention the amount this intention was selected was five times. Only one time the *Solve-a-task*-intention was used, therefore this intention can be considered the least important one.

As already mentioned before, based on the personal interviews with the participants, their is no clear distinction between certain intentions. Their was an overall confusion between the differences of certain intentions like *Information* and *Learning*. One participant even commented that he thinks that he could solve each task with *Information*-intention because everything is information. We can therefore conclude that there is an overlap between certain intentions and therefore we need to refine our tests to identify those. Also a more detailed description for each intention must be given to help users understand the difference between the intentions itself. Participants also noticed that it was new and hard for them to first think beforehand on what they need and then which intention they should choose. Concerning the *Result view* most test candidates noticed and liked the ordering and the breakdown into categories and consider it useful for some searches.

CHAPTER 5 Summary, Conclusion & Future Work

In this last chapter we present a comprehensive summary, point out open questions and conclude on the work described in this thesis. Finally we give a short insight to planned future work regarding the topic of intentions in video retrieval.

How can we improve video retrieval and the precision of search results at a time where digital video content on the internet is growing constantly? We have to start understanding how users behave on video platforms and get to know their intentions. Based on that knowledge we can develop better video retrieval models and systems. With this Master thesis we want to introduce intentions in the field of video retrieval on the web. The qualitative and quantitative studies reported in this thesis aimed at collecting insightful data on online video watching behaviour and how they relate to users' intentions. They addressed a number of aspects, from the triggering event that caused the video to be watched in the first place, to the time/place/device in which it was watched, to the way the video was shared with others.

First we wanted to find out how and why people are using video platforms. *YouTube* was mentioned as the most popular platform and most people are using video platforms almost on a daily basis. The majority of participants still preferred to consume online video on laptop or PC devices, but in the future more and more content will be watched on a TV or on various mobile devices like smartphones and tablets. In both of our studies most of the participants watched video content at home during evening and night times. We also noticed a shift towards using online videos for educational/professional purposes.

We further addressed the actual user's behaviour and motivation to retrieve videos. Various aspects influence a user's motivation to watch videos online. The most obvious reasons for video retrieval have some kind of entertainment background or are of informational or educational nature. Depending on the content of the video, different quality factors are important, namely picture, sound and content quality. Investigating overall favorite genres of participants, we found that mostly entertainment categories like *"Music"*, *"Entertainment"*, and *"Film & Animation"* are preferred.

We also addressed the question on how users share their video experience and by which means. Videos are either shared with a closed group of known people through various media (e.g., instant message, *Facebook*, etc.) or with an open community (e.g., *Digg*, *Reddit*). The most frequent group to share videos with is friends, followed by family, colleagues and the public at large. Reasons to share videos are to amuse other people, send relevant videos they could be interested in, inform people about news, events or the like, or to share a personal experience. Based on our quantitative survey, users on average send video links at least several times per month mostly using social media functionality, email or instant messaging.

Last, we investigated relationships between video genres and events. The classification of situations into video genre and trigger event allowed for exploratory analysis of relations between genres and event classes that actually lead to video retrieval. Our qualitative results showed that there is a relation between genre and trigger event. The biggest peak in absolute numbers can be observed in instances when people watched music videos for leisure, with 10 such instances (see Table 3.11). Moreover, only five out of ten genre categories were reported for the *"leisure"* trigger event. This indicates that for users who want to entertain themselves, a subset of categories (which could be grouped under the label *"fun"*) might be more relevant than others (which could be grouped under the label *"serious"*), e.g., politics, science or education. In order to investigate this matter, we performed statistical analysis (Spearman's Rho) on our quantitative survey data to show by numbers that genres correlate well with our predefined events, see Table 3.13 for details.

One of the merits of finding relations between trigger events and genres is that such relations can also be used to circumvent negative effects caused by the fact that *YouTube* categories overlap and do not constitute a neatly organized taxonomy. Based on our results, we designed and implemented a prototype which resembled a basic video retrieval platform. The big difference of our approach was, that we offered

i. the user four predefined intentions which could be selected when performing a video search, and

ii. a different ordering of the results based on the selected intention.

We then planned a small user survey with five participants to test our prototype. Participants had to solve five predefined tasks which resembled real-live search problems. Every task was solved by each participant and the position of the video they found as a solution was satisfiable throughout the whole survey. In the personal semi-structured interviews afterwards, the participants expressed an overall satisfaction with the results and the prototype in general. Other open topics which we observed during the survey and which also needs to be addressed in future are the following:

i. It may be possible that users cannot match their intention to the ones provided by our prototype.

ii. They may also want to express more complex intentions or even multiple intentions at the same time, which we cannot handle yet.

iii. It can also be possible that users do not want, or cannot give their explicit intentions after all.

iv. The abstraction of the intention needs more effort than just the query definition itself.

Additional lessons were also learned from the prototype survey including our observations and the interviews afterwards. As some participants noted, a more detailed description of the four intentions shown in the prototype is needed. It was not clear to them how the intentions were different and how each one influences the search results. One participant noted that he just needed the *Information*-intention to solve all the tasks. Therefore and based on these statements and observations we conclude that the abstraction of the various intentions might be a problem and that we should and need to help the user comprehend the difference. We need to support the users understanding by pointing out the cause-effect relationship for each intention. Also the meta-data quality was noted by one participant, but unfortunately we cannot influence how users categorize, tag or describe their videos while uploading them to a video platform.

In summary, the work reported in this thesis represents a first step in the direction of understanding why people retrieve and watch online videos and how we can capitalize this knowledge to improve video retrieval systems. If a multimedia system can infer the possible trigger events based on the viewing behaviour of a user, this knowledge can be employed to provide more relevant and precise search results to users or filter irrelevant results and recommendations and also meaningful view adaptations for the user.

Future Work Although the results in this Master thesis look promising, still a lot of work needs to be invested into the concept of user intentions in online video retrieval. Due to the exploratory nature of our work we can identify promising research directions and questions for the next steps in research. The presented results of the exploratory studies in this thesis still offer many opportunities for future work in several directions. For example we will pursue further statistical analysis of the numerical results to uncover stronger and more interesting correlations in video search and retrieval behaviour, which should help us build more sophisticated and promising multimedia information retrieval systems towards intelligent systems capable of taking the user context like intentions, tasks, and goals into account. On the other hand deriving rules from descriptive statements given by the survey participants is a somewhat vague approach, which should be complemented with more sophisticated methods, e.g., deriving rules from user behaviour through log file analysis.

Also a larger study needs to be conducted to statistically support correlations between genres and trigger events more solid. We also need to fine tune our selection of intentions so that confusions are avoided already at the beginning. The granularity and selection of genres needs to be discussed in this context. Finally the classification of trigger events needs to be revisited also in terms of granularity, coverage, and a possible overlap. Also the development of additional studies, focused on specific needs and demographics which are not covered by the present work could be beneficial. Furthermore, we might have missed crucial intention categories due to our sample of survey participants.

A main element in our future work plans is the further development of the current prototype. We aim to better understand video queries through intentions which are driving the users information need. This includes query reformulation, query extension, and relevance feedback. Another important feature will be the dynamic hiding of options and view adaptation based on the users intention and behaviour. We are particularly interested in a dynamic rule set and a dynamic view adaptation method that use machine learning algorithms because survey participants reported that not every option and metadata needs to be displayed at all times for every intention. At a later development stage of the prototype, the resulting rule set needs to be influenced by additional aspects of the video retrieval process which were not mentioned earlier, such as

i. the video quality of the resulting clips (e.g., users of our survey reported to prefer HD video quality when they want to watch a video to learn something), and

ii. the inclusion and selection of other sources (e.g., Vimeo and Khan Academy), in the search results.

Also of interest is the inclusion of the results from the direct link access scenarios as well, but this certainly remains an open topic for now. In the final development stage, the prototype will display to the user which video sources are used, which restrictions to video quality and other features are applied and the like. These restrictions and views will also be changeable by the user as needed. We can again monitor the users behaviour and use this information to learn more about their preferences for certain intentions and reuse this data for e.g machine learning algorithms.

In the end, after all proposed features are implemented and tested, another survey with a bigger and more diverse audience needs to be done. The survey should include similar but more detailed user acceptance and satisfaction tests as explained in this thesis. A direct comparison of the results and performance of our prototype with popular video portals, notably to *YouTube* or *Vimeo*, should be part of the tests. Evaluation methods again should also include a deeper analysis of heatmaps and semi-structured interviews with the survey participants. This procedure allows for further evaluation of our prototype, extended user experience, and a better understanding of study results.

Bibliography

[1] C. Kofler and M. Lux. An exploratory study on the explicitness of user intentions in digital photo retrieval. In *I-KNOW '09: Proceedings of the 9th International Conference on Knowledge Management and Knowledge Technologies*, 2009.

[2] Mathias Lux, Marian Kogler, and Manfred del Fabro. Why did you take this photo: a study on user intentions in digital photo productions. In *Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access*, SAPMIA '10, pages 41–44, New York, NY, USA, 2010. ACM.

[3] The Official YouTube Blog. Holy Nyans! 60 hours per minute and 4 billion views a day on YouTube. `http://youtube-global.blogspot.com/2012/01/holy-nyans-60-hours-per-minute-and-4.html`, January 21, 2012 (accessed January 27, 2012).

[4] YouTube.com. YouTube Press Room - Statistics. `http://www.youtube.com/t/press_statistics/`, accessed July 27, 2011.

[5] Cisco Systems Inc. Cisco Visual Networking Index: Forecast and Methodology, 2009–2014. `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf`, June 2010 (accessed January 11, 2011).

[6] Cisco Systems Inc. Hyperconnectivity and the Approaching Zettabyte Era. `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf`, June 2010 (accessed January 11, 2011).

[7] Mark Robertson. YouTube Video SEO Tips From SES Chicago. `http://www.reelseo.com/youtube-seo-tips/`, 2009 (accessed April 14, 2011).

[8] Vannevar Bush. As We May Think. *The Atlantic Monthly*, 176(1):101–108, 1945.

[9] Mooers Calvin. The theory of digital handling of non-numerical information and its implications to machine economics. In *Meeting of the Association for Computing Machinery at Rutgers University*, Cambridge, MA, USA, March 1950.

[10] Michael Lesk. The seven ages of information retrieval. In *Proceedings of the Conference for the 50th anniversary of As We May Think*, pages 12–14, Cambridge, MA, USA, 1995. MIT Press.

[11] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. Determining the informational, navigational, and transactional intent of web queries. *Inf. Process. Manage.*, 44:1251–1266, May 2008.

[12] Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

[13] Daniel E. Rose and Danny Levinson. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, WWW '04, pages 13–19, New York, NY, USA, 2004. ACM.

[14] Uichin Lee, Zhenyu Liu, and Junghoo Cho. Automatic identification of user goals in web search. In *Proceedings of the 14th international conference on World Wide Web*, WWW '05, pages 391–400, New York, NY, USA, 2005. ACM.

[15] M. Strohmaier, M. Lux, M. Granitzer, P. Scheir, S. Liaskos, and E. Yu. How do users express goals on the web? - an exploration of intentional structures in web search. In *Proceedings of the 2007 international conference on Web information systems engineering*, WISE'07, pages 67–78, Berlin, Heidelberg, 2007. Springer-Verlag.

[16] Eric Siu-Kwong Yu. *Modelling strategic relationships for process reengineering*. PhD thesis, University of Toronto, Toronto, Ont., Canada, Canada, 1995. AAINN02887.

[17] Alan Mislove, Hema Swetha Koppula, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Growth of the flickr social network. In *Proceedings of the first workshop on Online social networks*, WOSN '08, pages 25–30, New York, NY, USA, 2008. ACM.

[18] Masoud Valafar, Reza Rejaie, and Walter Willinger. Beyond friendship graphs: a study of user interactions in flickr. In *Proceedings of the 2nd ACM workshop on Online social networks*, WOSN '09, pages 25–30, New York, NY, USA, 2009. ACM.

[19] Mathias Lux, Christoph Kofler, and Oge Marques. A classification scheme for user intentions in image search. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 3913–3918, New York, NY, USA, 2010. ACM.

[20] Nicu Sebe, Michael S. Lew, Xiang Zhou, Thomas S. Huang, and Erwin M. Bakker. The state of the art in image and video retrieval. In *Proceedings of the 2nd international conference on Image and video retrieval*, CIVR'03, pages 1–8, Berlin, Heidelberg, 2003. Springer-Verlag.

[21] Yanqing Cui, Jan Chipchase, and Younghee Jung. Personal tv: a qualitative study of mobile tv users. In *Proceedings of the 5th European conference on Interactive TV: a shared experience*, EuroITV'07, pages 195–204, Berlin, Heidelberg, 2007. Springer-Verlag.

[22] Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. Consuming video on mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 857–866, New York, NY, USA, 2007. ACM.

[23] Petteri Repo, Kaarina Hyvönen, Mika Pantzar, and Päivi Timonen. Users inventing ways to enjoy new mobile services – the case of watching mobile videos. In *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 4 - Volume 4*, HICSS '04, pages 40096.3–, Washington, DC, USA, 2004. IEEE Computer Society.

[24] Bernard J. Jansen, Abby Goodrum, and Amanda Spink. Searching for multimedia: analysis of audio, video and image web queries. *World Wide Web*, 3:249–254, 2000. 10.1023/A:1018743024123.

[25] Frank Hopfgartner. Studying interaction methodologies in video retrieval. *Proc. VLDB Endow.*, 1:1604–1608, August 2008.

[26] Frank Hopfgartner and Joemon Jose. Evaluating the implicit feedback models for adaptive video retrieval. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, MIR '07, pages 323–331, New York, NY, USA, 2007. ACM.

[27] Fan Qiu and Yi Cui. An analysis of user behavior in online video streaming. In *Proceedings of the international workshop on Very-large-scale multimedia corpus, mining and retrieval*, VLS-MCMR '10, pages 49–54, New York, NY, USA, 2010. ACM.

[28] Sylvain Mongy, Fatma Bouali, and Chabane Djeraba. Analyzing user's behavior on a video database. In *Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data*, MDM '05, pages 95–100, New York, NY, USA, 2005. ACM.

[29] Koji Miyauchi, Taro Sugahara, and Hiromi Oda. Relax or study? a qualitative user study on the usage of live mobile tv and mobile video. *Comput. Entertain.*, 7:43:1–43:20, September 2009.

[30] C. Sodergard. Mobile television - technology and user experiences. *VTT Information Technology*, 2003.

[31] Petteri Repo, Kaarina Hyvönen, Mika Pantzar, and Päivi Timonen. Users inventing ways to enjoy new mobile services – the case of watching mobile videos. In *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 4 - Volume 4*, HICSS '04, Washington, DC, USA, 2004. IEEE Computer Society.

[32] W. A. Vorbau, A. S. Mitchell, and K. O'Hara. "My iPod is my Pacifier": An Investigation on the Everyday Practices of Mobile Video Consumption. In *Mobile Computing Systems and Applications, 2007. HotMobile 2007. Eighth IEEE Workshop on*, pages 29–33, 2007.

[33] Wei Song and Dian Tjondronegoro. A survey on usage of mobile video in australia. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, pages 5–8, New York, NY, USA, 2010. ACM.

[34] Shelley Buchinger, Simone Kriglstein, and Helmut Hlavacs. A comprehensive view on user studies: survey and open issues for mobile tv. In *Proceedings of the seventh european conference on European interactive television conference*, EuroITV '09, pages 179–188, New York, NY, USA, 2009. ACM.

[35] Martin J. Halvey and Mark T. Keane. Analysis of online video search and sharing. In *Proceedings of the eighteenth conference on Hypertext and hypermedia*, HT '07, pages 217–226, New York, NY, USA, 2007. ACM.

[36] Martin J. Halvey and Mark T. Keane. Exploring social dynamics in online media sharing. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 1273–1274, New York, NY, USA, 2007. ACM.

[37] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In *In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07*, 2007.

[38] Xu Cheng, Cameron Dale, and Jiangchuan Liu. Statistics and social network of youtube videos. In *IWQoS'08*, pages 229–238, 2008.

[39] Joan-Isaac Biel and Daniel Gatica-Perez. Wearing a youtube hat: directors, comedians, gurus, and user aggregated behavior. In *Proceedings of the 17th ACM*

international conference on Multimedia, MM '09, pages 833–836, New York, NY, USA, 2009. ACM.

[40] Fabricio Benevenuto, Fernando Duarte, Tiago Rodrigues, Virgilio A.F. Almeida, Jussara M. Almeida, and Keith W. Ross. Understanding video interactions in youtube. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, pages 761–764, New York, NY, USA, 2008. ACM.

[41] Jon M. Kleinberg. Hubs, authorities, and communities. *ACM Comput. Surv.*, 31, December 1999.

[42] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 29–42, New York, NY, USA, 2007. ACM.

[43] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web 7*, WWW7, pages 107–117, Amsterdam, The Netherlands, The Netherlands, 1998. Elsevier Science Publishers B. V.

[44] Xu Cheng. Understanding the characteristics of internet short video sharing: Youtube as a case study. In *Procs of the 7th ACM SIGCOMM Conference on Internet Measurement, San Diego (CA, USA), 15*, page 28, 2007.

[45] The Official YouTube Blog. Upload limit increases to 15 minutes for all users. `http://youtube-global.blogspot.com/2010/07/upload-limit-increases-to-15-minutes.html`, July 29, 2010 (accessed January 27, 2012).

[46] The Official YouTube Blog. Additional Creator Tools from YouTube. `http://youtube-global.blogspot.com/2011/09/additional-creator-tools-from-youtube.html`, September 21, 2011 (accessed January 27, 2012).

[47] Hongliang Yu, Dongdong Zheng, Ben Y. Zhao, and Weimin Zheng. Understanding user behavior in large-scale video-on-demand systems. In *Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems 2006*, EuroSys '06, pages 333–344, New York, NY, USA, 2006. ACM.

[48] D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.

[49] Pablo Cesar, Dick C.A. Bulterman, David Geerts, Jack Jansen, Hendrik Knoche, and William Seager. Enhancing social sharing of videos: fragment, annotate, enrich, and share. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, pages 11–20, New York, NY, USA, 2008. ACM.

[50] David Geerts, Pablo Cesar, and Dick Bulterman. The implications of program genres for the design of social television systems. In *Proceeding of the 1st international conference on Designing interactive user experiences for TV and video*, UXTV '08, pages 71–80, New York, NY, USA, 2008. ACM.

[51] European Broadcasting Union. Escort: EBU System of Classification of RTV Programmes. Technical report, European Broadcasting Union, 2007.

[52] David A. Shamma, Marcello Bastea-Forte, Niels Joubert, and Yiming Liu. Enhancing online personal connections through the synchronized sharing of online video. In *CHI '08 extended abstracts on Human factors in computing systems*, CHI EA '08, pages 2931–2936, New York, NY, USA, 2008. ACM.

[53] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 15–28, New York, NY, USA, 2007. ACM.

[54] Susanne Boll. Multitube–where web 2.0 and multimedia could meet. *IEEE MultiMedia*, 14:9–13, January 2007.

[55] Jürgen Bortz and Nicola Döring. *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. Springer, Heidelberg, 4., überarb. edition, 2006.

[56] TGDaily.com (DD&M Inc.). YouTube blocks GEMA music videos in Germany. `http://www.tgdaily.com/business-and-law-features/41913-youtube-blocks-gema-music-videos-in-germany`, April 1, 2009 (accessed January 27, 2012).

APPENDIX A

Questionnaires

A.1 Qualitative questionnaire

On the next pages, the pre-made questionnaire used for semi-structured interviews (quantitative survey) can be seen.

This survey is part of a research project investigating intentions of users retrieving online video content. Further questions? Dr. Mathias Lux (mlux@itec.uni-klu.ac.at) or Christoph Lagger (clagger@edu.uni-klu.ac.at)

Questionnaire: Why do people search and watch videos online?

Interviewer / Place	**Date & Time**

1. Demographic

Name *(will be anonymized, just to make sure not to have 2 interviews of the same person)*	
Job	
Gender	o Male o Female
Age	o < 18 o 18-24 o 25-30 o 31-40 o > 40
What are your general interests / hobbies?	

2. General use

Which online video platforms do you know?	☐ YouTube.com ☐ vimeo.com ☐ video.google.com ☐ video.yahoo.com ☐ hulu.com ☐ megavideo.com ☐ miscellaneous blogs	☐ dailymotion.com ☐ blip.com ☐ metacafe.com ☐ break.com ☐ facebook.com ☐ Others…
How often do you use these platforms?	o more than once a day o several times a week o several times a month o rarely	
Which online video platforms do you use? *Please perform a ranking based on your liking (1 to 3)*	☐ YouTube.com (____) ☐ vimeo.com (____) ☐ video.google.com (____) ☐ video.yahoo.com (____) ☐ hulu.com (____) ☐ megavideo.com (____) ☐ miscellaneous blogs (____)	☐ dailymotion.com (____) ☐ blip.com (____) ☐ metacafe.com (____) ☐ break.com (____) ☐ facebook.com (____) ☐ Others…
On which device do you watch videos online?	☐ PC / Laptop / Netbook ☐ Mobile device	☐ TV ☐ Others…

Type of use	☐ Professional usage ☐ Private usage
Why do you search for videos and watch them? *(open question)*	
Do you use the "video suggestion" or "related video" feature?	☐ Yes. How often do you use it? _______ ☐ No
Do you browse through categories (and watch videos from it)?	☐ Yes ☐ No
Is the quality of a video a criteria for watching it?	☐ Yes. Which factor of quality is important to you? Give examples! ☐ No
Do you have an account on one of these platforms?	☐ Yes (on which platforms?) Do you subscribe to channels (e.g. YouTube)? ◯ Yes ◯ No ☐ No

3. Communication

Did you forward video links to others (in person) lately? **<u>If yes</u>: why do you send videos (or video links) to other people?** **<u>If yes</u>: To whom (friend, family, colleague, public etc.?** **What were the videos about?**	
How often do you send video links (e.g. Email, MMS, via YouTube etc?)	o More than once a day o Several times per week o Several times per month o Never
If you send video links, how do you send/share them?	☐ In Person ☐ On-site mechanisms (in a Channel etc.) ☐ Social Media (Twitter, Facebook, Digg etc.) ☐ Email ☐ Instant Messaging ☐ Others

4. Instances

<table>
<tr><td colspan="2">Think of the last 2-3 videos you watched.</td></tr>
<tr><td>Video 1</td><td>How did you get to this video?

☐ Recommendation from a friend via ___________________
☐ Searched for it on platform ___________________ with the following terms:

____

☐ Browsed for it on platform _____________________ and used
☐ Categories
☐ Channel subscription
☐ Related feature
☐ Suggestion feature
☐ Others: ___________________________________

What did the video show, what was it about?

Why did you watch the video (out of a necessity or out of situation)?

Where & when did you watch the video?

Which device did you use?

What was the length of the video?

Did you only watch selected parts of it? <u>If YES</u>, how did you decide which parts to watch?

Did you show / send the video link to someone or publish it or plan to do so?
<u>If YES</u>, why?

If you just got recommendations for videos in the past, maybe you can think of the last video you intentionally searched for? <u>If YES</u>, what was it about and how did you search for it?</td></tr>
</table>

A.2 Quantitative questionnaire

On the next pages, the questionnaire which was used in the online survey can be seen.

Survey on video retrieval

1. Introduction

Thank you for taking the time to participate in this survey. The goal of this survey is to find out more about a user's video search and retrieval behavior as well as video sharing habits. The collected data will be used to enhance current video retrieval systems, search interfaces, and algorithms, in order to improve the overall user satisfaction and experience.

This survey should only take about 10 minutes of your time. Your answers will be completely anonymous.

In order to progress through this survey, please use the following navigation links:
- Click the Next >> button to continue to the next page.
- Click the Previous >> button to return to the previous page.
- Click the Exit the Survey Early >> button if you need to exit the survey.
- Click the Submit/Done >> button to submit your survey.

Who is behind this survey?
I'am a student at the University of Klagenfurt and currently working on my masters thesis and this survey is part of my work. If you have any questions or want more information regarding this survey, please contact me via email at clagger(AT)edu.uni-klu.ac.at

Survey on video retrieval

2. Demographics

1. What is your gender?

- male
- female

2. How old are you?

- < 18
- 18-24
- 25-30
- 31-40
- > 40

3. What are you hobbies and interests?

- Sports
- Music
- Science
- Literature
- Travel
- Movies & Cinema
- Technology
- Home & Garden

Others (please specify)

4. What is your highest education level?

- Graduate Degree (PhD, Masters, MD, JD)
- Bachelor/College/University Degree
- High School
- Elementary/Secondary School
- Others

Survey on video retrieval

3. General usage

* 1. Which online video platforms do you use on a regular basis?

- YouTube
- Dailymotion
- Vimeo
- Blip
- Google video
- Metacafe
- Yahoo video
- Break
- Hulu
- Netflix
- Megavideo
- Facebook

Others (please specify)

* 2. How often do you visit and use any of these platforms?

- more than once a day
- several times a week
- several times a month
- rarely

3. How often do you use such platforms for either professional or personal reasons?

	Never	Rarely	Once a month or more	Once a week or more	Once a day or more
Professional usage					
Personal usage					

* 4. Why do you search for videos and watch them?

	Never	Rarely	Sometimes	Often	Always
for leisure/entertainment reasons					
out of boredom					
to listen to music or to watch a music video					
to get information about certain topics					
to get product reviews					
to see missed news-broadcasts					
to learn something new/to gain new knowledge					
to see how something looks like (e.g. place or product)					
too see how something works					

Others (please specify)

Survey on video retrieval

4.

*** 1. Did you forward or recommend video links to others lately?**

- Yes
- No

Survey on video retrieval

5. Sharing habits

* 1. Why do you forward video links to others?

	Never	Rarely	Sometimes	Often	Always
I want others to learn about things, transport knowledge					
I want others to be entertained, have fun watching the video					
I assume that friends I forward the link are interested in this video					
To share information about products or events					
To share an experience, to share something I have experienced					
To share emotions (like anger, disgust, beauty, love etc.)					
Someone asked me to share this video					
No special reason					

Other (please specify)

2. With whom do you share video links?

	Never	Rarely	Sometimes	Often	Always
Friends					
Family members					
Colleagues					

Others (please specify)

* 3. How often do you send or share video links with others?

- More than once a day
- Several times per week
- Several times per month
- Rarely

* 4. Which mechanisms do you use to share video links?

	Never	Rarely	Sometimes	Often	Always
In person by showing them on a computer screen or mobile device					
On-site mechanisms ("share" button)					
Social Media (Twitter, Facebook, Reddit, Digg etc.)					
Email					
Instant Messaging					

Others (please specify)

Survey on video retrieval

6. Usage

* 1. Which videos are you interested in?

	Not interested	Rarely interested	Neutral	More interested	Very interested
Autos & Vehicles					
Comedy					
Education					
Entertainment					
Film & Animation					
Gaming					
Howto & Style					
Music					
News & Politics					
Nonprofits & Activism					
People & Blogs					
Pets & Animals					
Science & Technology					
Sports					
Travel & Events					

* 2. Look at the following statements and tell us how often you do any of the following (watching/searching video refers to online videos):

	Never	Rarely	Once a month or more	Once a week or more	Once a day or more
I watch videos when I am bored					
I watch videos (or follow video links) I get in emails					
I watch videos (or follow video links) I get in instant messages					
I search for videos to follow up something on TV/radio/newspaper/magazine					
I search for videos to follow up something after I talked to another person about it					
I search for videos to get help on what I'm going to do (e.g. how to)					
I watch videos published on RSS streams/feeds/blogs					
I watch videos posted on facebook (MySpace, Twitter, etc.)					
I search for videos to learn how something works / looks like					

* 3. How often do you do any of the following?

	Never	Rarely	Once a month or more	Once a week or more	Once a day or more
I actively search for videos					
I watch videos from recommendations from other persons					
I watch video recommendations coming from on-site features (channel subscription, related videos, video suggestions)					
I browse through categories and select videos from there					

* 4. How often do you follow recommendations of a video page after you have seen a video? (e.g. "related videos" or "similar videos")

- Always
- Often
- Sometimes
- Rarely
- Never

5. Where and how often do you watch online videos?

	Never	Rarely	Once a month or more	Once a week or more	Once a day or more	N/A
Home						
Work						
School/University						
At a friend's place						
At public places						
On the go						

Others (please specify)

6. When do you watch videos?

	Never	Rarely	Sometimes	Often	Always
Morning					
Afternoon					
Evening					
Night					

* 7. Which devices do you use to watch videos?

	Never	Rarely	Sometimes	Often	Always
PC					
Laptop/Notebook/Netbook					
Smartphone (e.g. iPhone, Android)					
Tablet PC (e.g. iPad)					
TV					
Other mobile device					

Others (please specify)

* 8. If I receive a video link (via Email, IM, feed, etc.), ...

	Never	Rarely	Sometimes	Often	Always
... I watch the video right away					
... I make a note/bookmark to watch it at a later time					
... I watch it as soon as possible					
... I ignore it and maybe watch it sometime later					

* 9. How do you relate to the following statements?

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I prefer short videos (up to 5 minutes)					
I prefer videos with high visual quality					
I prefer videos with good sound quality					
My decision to watch a video is based on the content					

Survey on video retrieval

7. End

1. One last question! Can you tell us something special about your video watching experience that would be of interest?

Survey on video retrieval

8. Thank you!

When you click on the "Done" button your survey response will be submitted.

Thank you for completing this survey and taking the time!

A.3 Prototype questionnaire

On the next pages, the questionnaire for the prototype survey and the personal interviews afterwards can be seen.

This survey is part of a research project investigating intentions of users retrieving online video content. Further questions? Dr. Mathias Lux (mlux@itec.uni-klu.ac.at) or Christoph Lagger (clagger@edu.uni-klu.ac.at)

Questionnaire: *You(r)Intent* Prototype

Interviewer / Place	**Date & Time**

1. Demographics

Name *(will be anonymized, just to make sure not to have 2 interviews of the same person)*	
Job	
Gender	o Male o Female
Age	o < 18 o 18-24 o 25-30 o 31-40 o > 40
How many days per week do you use a video platform?	

2. Prototype questions

What is your first impression of the prototype? What was different, what did you like, what did you dislike?
Did you miss any intentions? What else are you doing on video search engines except the things you have seen on our prototype?
How did you like the ordering of the results? Would it be an improvement for some searches?
Do you have concrete suggestions for improvements?-
Anything else you want to add?

3. Tasks

Task 1	Did you solve the task?	( ) Yes ( ) No ( ) I don't know
	Relevance of the task in every day live:	irrelevant (0) (1) (2) (3) (4) (5) every day task
	Do you think that you could solve this task with Google easier?	(1) Strongly disagree (2) Disagree (3) Neither agree or disagree (4) Agree (5) Strongly agree
Task 2	Did you solve the task?	( ) Yes ( ) No ( ) I don't know
	Relevance of the task in every day live:	irrelevant (0) (1) (2) (3) (4) (5) every day task
	Do you think that you could solve this task with Google easier?	(1) Strongly disagree (2) Disagree (3) Neither agree or disagree (4) Agree (5) Strongly agree
Task 3	Did you solve the task?	( ) Yes ( ) No ( ) I don't know
	Relevance of the task in every day live:	irrelevant (0) (1) (2) (3) (4) (5) every day task
	Do you think that you could solve this task with Google easier?	(1) Strongly disagree (2) Disagree (3) Neither agree or disagree (4) Agree (5) Strongly agree
Task 4	Did you solve the task?	( ) Yes ( ) No ( ) I don't know
	Relevance of the task in every day live:	irrelevant (0) (1) (2) (3) (4) (5) every day task
	Do you think that you could solve this task with Google easier?	(1) Strongly disagree (2) Disagree (3) Neither agree or disagree (4) Agree (5) Strongly agree
Task 5	Did you solve the task?	( ) Yes ( ) No ( ) I don't know
	Relevance of the task in every day live:	irrelevant (0) (1) (2) (3) (4) (5) every day task
	Do you think that you could solve this task with Google easier?	(1) Strongly disagree (2) Disagree (3) Neither agree or disagree (4) Agree (5) Strongly agree

APPENDIX

B Source listing of IP2Country.sh

```
# Resolve Country ISO code for IP addresses
# (one IP per line) given in a data.csv
# and output it to ip2country.txt (one ISO code per line).

#! /bin/bash
for line in `cat data.csv`;
do
  echo $line;
  wget --quiet -O /dev/stdout http://www.ip2country.cc/?q=$line |
       grep 'Country Name:' |
       awk '{print $6}' >> ip2country.txt
done
```

APPENDIX C
IntentDB database schema

```
-- phpMyAdmin SQL Dump
-- version 3.4.5deb1
-- http://www.phpmyadmin.net
--
-- Host: localhost
-- Generation Time: Nov 24, 2011 at 06:36 PM
-- Server version: 5.1.58
-- PHP Version: 5.3.6-13ubuntu3.2

SET SQL_MODE="NO_AUTO_VALUE_ON_ZERO";
SET time_zone = "+00:00";


/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8 */;

--
-- Database: `IntentDB`
--

-- --------------------------------------------------------

--
-- Table structure for table `tblAdaptions`
--

CREATE TABLE IF NOT EXISTS `tblAdaptions` (
```

```
  `intentID` int(11) NOT NULL,
  `adpKey` varchar(20) NOT NULL,
  `adpValue` varchar(20) NOT NULL,
  `adpComment` text NOT NULL
) ENGINE=MyISAM DEFAULT CHARSET=latin1;

-- --------------------------------------------------------

--
-- Table structure for table `tblGenreRating`
--

CREATE TABLE IF NOT EXISTS `tblGenreRating` (
  `grID` int(11) NOT NULL AUTO_INCREMENT,
  `intID` int(11) NOT NULL,
  `grName` varchar(40) CHARACTER SET utf8 NOT NULL,
  `grKey` varchar(40) CHARACTER SET utf8 NOT NULL,
  `grCoefficient` decimal(3,3) NOT NULL,
  `grSignificance` decimal(3,3) NOT NULL,
  PRIMARY KEY (`grID`)
) ENGINE=MyISAM  DEFAULT CHARSET=latin1 AUTO_INCREMENT=124 ;

-- --------------------------------------------------------

--
-- Table structure for table `tblIntention`
--

CREATE TABLE IF NOT EXISTS `tblIntention` (
  `inID` int(11) NOT NULL AUTO_INCREMENT,
  `inName` text CHARACTER SET utf8 NOT NULL,
  `inDescription` text CHARACTER SET utf8 NOT NULL,
  PRIMARY KEY (`inID`)
) ENGINE=MyISAM  DEFAULT CHARSET=latin1 AUTO_INCREMENT=5 ;

-- --------------------------------------------------------

--
-- Table structure for table `tblLogging`
--

CREATE TABLE IF NOT EXISTS `tblLogging` (
  `logID` int(11) NOT NULL AUTO_INCREMENT,
```

```
  `logSessionID` varchar(50) NOT NULL,
  `logTimestamp` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON
    UPDATE CURRENT_TIMESTAMP,
  `logEvent` varchar(50) NOT NULL,
  `logHTTP_Referer` varchar(150) NOT NULL,
  `logHTTP_Request_URI` varchar(150) NOT NULL,
  `logHTTP_User_Agent` varchar(30) NOT NULL,
  `logRemote_Address` varchar(40) NOT NULL,
  `logRemote_Host` varchar(40) NOT NULL,
  `logInformation` varchar(100) NOT NULL,
 PRIMARY KEY (`logID`)
) ENGINE=MyISAM DEFAULT CHARSET=utf8 AUTO_INCREMENT=1121 ;
```