How People Find Videos

Sally Jo Cunningham

Dept of Computer Science
Waikato University
Hamilton, New Zealand
+64 7 838 4402

sallyjo@cs.waikato.ac.nz

David M. Nichols

Dept of Computer Science
Waikato University
Hamilton, New Zealand
+64 7 838 5130

dmn@cs.waikato.ac.nz

ABSTRACT

At present very little is known about how people locate and view videos 'in the wild'. This study draws a rich picture of everyday video seeking strategies and video information needs, based on an ethnographic study of New Zealand university students. These insights into the participants' activities and motivations suggest potentially useful facilities for a video digital library.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: *User issues*; H.1.2 [User/Machine Systems] Human information processing

General Terms

Design, Human Factors

Keywords

Video retrieval, qualitative research, user study.

1. INTRODUCTION

The rapid growth of video on the Web is exemplified by sites such as YouTube, which have helped to widen access to multimedia resources. The experience of browsing and searching amongst rich video collections has moved out of the laboratory and into the homes and workplaces of typical Web users. However, at present little is known about how people locate and view videos in natural situations. In this paper we report on a qualitative ethnographic study to shed light on how people find and access video resources on the Web.

After outlining related work in Section 2 we describe our research methodology and then give an overview of the results in Section 4. We report on the participants' interaction strategies in Section 6, on uncovered usability issues in Section 7 and conclude with a discussion on the design of systems to support video access.

2. RELATED WORK

YouTube and related sites have already proven to be a fertile source of research for analysis based on web crawling (e.g. [9], [3], [6]). These statistical studies can give a view of video

retrieval based on the public face of the website but provide only limited insights into the behavior of users. The users who publish videos and comments at a site like YouTube leave traces that can be analyzed by crawler-based studies; however, such methods tell us little about the majority of users who simply search, browse and view video content.

Although the video-oriented digital library literature contains many useful user studies on system design in constrained laboratory environments ([2], [20]), there appears to be little work that attempts to understand users in their everyday use of video on sites like YouTube, GoogleVideo, YahooVideo, and so forth (one related study is [14], a diary study of use of a video news system). In order to design effective video retrieval interfaces we need to balance system development and laboratory studies with an understanding of users' natural behavior 'in the wild'. In this paper we attempt gain understanding into users' video seeking behavior through qualitative approaches that can act as a complement to insights gained through site-specific analysis and web log studies.

3. DATA COLLECTION

Data for this study was gathered through a project assigned to an upper-level undergraduate course on Human-Computer Interaction, at Waikato University in New Zealand. The semester-long project required each student to design and prototype a system to support locating, viewing, and sharing videos. The initial phase of the project involved gathering evidence of the video-related behaviors of potential users for the system. To this end, the students conducted interviews and observed Web searches for videos. The interview summaries and records of the search observations are analyzed in this present study.

The students first created a 'personal ethnography' [5], in which they investigated their own video searching and viewing behaviors—their motivations for looking for videos, the types of video content that interests them, the circumstances under which they look for and view videos, the video-seeking strategies that they use, and so forth. 'Video' was broadly defined for this assignment as any digital video as viewed or located on the Web—for example, a full-length Hollywood movie, a five-minute music video, or a few seconds captured on a home video camera. The students then conducted a similar ethnographic investigation of a friend's behavior in locating and viewing videos. A total of 27 autoethnographies and 27 ethnographies were collected, and are summarized in approximately 180 printed pages.

The students also observed up to four video seeking sessions, manually recording the actions taken by the searcher and categorizing these actions according to set of codes specified in

the course assignment (described in Section 5). As with the ethnographies, some sessions involved self-observations and others were observations of another individual as they searched.

It is probable that some degree of participant self-censorship occurred in the autoethnographic/ethnographic interviews and in the selection of topics for the video searches. This hypothesis is supported by the fact this that none contained any mention of pornography, despite the participants being, for the most part, members of the main demographic for consumption of online pornography (young and male) [17]. We also believe that the data under-reports illegal video downloading and viewing. However, some participants were remarkably open about activities involving copyright infringement—only one participant asked that the name of the pirate video site he visited not be recorded.

The experiences must also be interpreted in the context of the participants' technological environment. While two-thirds of New Zealand households have Internet access, connection speeds are slow by international standards. In 2007 (the period in which this study was conducted), approximately 30% of households relied on dial-up connections to the Internet [19]. June 2007 statistics places New Zealand 20th out of 30 OECD countries for home broadband access [15]. Moreover, home broadband services are slow and restrictive; in 2006, New Zealand Telecom's Basic Broadband Service to households offered 256kbps downstream and 128kbps upstream connection, with a 200MB monthly cap on traffic [18]. In Sections 6 and 7 we highlight ways that slow connection speeds influenced the searching and browsing strategies adopted by participants.

Grounded theory methods [7] were used to analyze the student summaries of their interviews and observations. To preserve the anonymity of participants, each participant is referred to with a randomly assigned letter of the alphabet. Note that quotes from a student participant can be in the first person (referring to the student's autoethnography) or in third person (referring to the ethnographic investigation of another person).

We recognize that this data gathering approach has methodological limitations. In an autoethnography or personal ethnography, applying ethnographic techniques to one's own experiences challenges the researcher to view personal behaviors objectively and then interpret those experiences in the light of applicable theory ([4], [5]). While absolute objectivity is not possible, the ethnographic data can be validated to a degree by comparison with information from other informants (eg, the ethnographies)—and to that end the student observer/participants each explicitly compared their own behaviors with those of the person that they interviewed and observed.

Observers took manual notes during the video searches, rather than video-recording or using screen capture. This limited the ability of the observer to distinguish subtle or brief actions and to identify the participants' focus of attention. Observers also found it difficult to create comprehensive records for extended sessions, and so for several of the longer sessions the observers captured a representative section rather than a complete session. While one goal of this research is to capture authentic video-related behavior, the act of observation interferes with the behavior being observed. This is particularly true of the self-observations, as the participant/observer must interleave recording with searching.

We argue that the strengths of this approach counterbalance its weaknesses. Involving students in data gathering supports the rapid gathering of relatively large amounts of ethnographic and observational data. While this data may be of uneven depth of description, it was of sufficient quality to allow us to identify patterns of behavior, and sufficient quantity to enable cross-validating these behaviors across multiple participants. Manual note-taking in search observations allowed the student participant/observers to go outside the research lab and come closer to our goal of capturing authentic video-seeking behavior 'in the wild'. It also may have allowed us to capture a broader range of behaviors, particularly questionably legal activities—it is unlikely that participants would have been willing to be video-recorded as they visited sites with illegal content.

4. OBSERVATIONS: OVERVIEW

A total of 98 video search session summaries were recorded; these include participant self-observations as well as observations of searches performed by others. 60 of the session participants were male, 19 female, and for 16 of the sessions the participants' gender was not recorded. For each session, a summary was created including the length of time that the session lasted, a list of the websites visited, a brief statement of what the participant was looking for and why, the number of videos viewed during the session, and whether the participant felt satisfied with the videos.

Session lengths (Table 1) range from a quick 5 minute look for a U2 music video, to a marathon 2½ hour viewing of 10 episodes of a TV series. This average session length is considerably longer than that reported for search sessions of text-document digital libraries (for example, an average of 10 minutes for a collection of computing technical reports [10]), but is more closely aligned to behavior in Web search engines ([16] reports session averages of an hour or more, depending on time of day). The range of session times observed in this study are likely to be biased to include more long sessions than shorter, since the observers were less likely to be present for brief, impromptu searches. Interviews confirmed that the participants occasionally take brief breaks from other computing activities to hunt for videos. The study is also likely to miss exceptionally lengthy sessions, as they would be impractical for the observers to record.

Table 1. Summary of session lengths (minutes)

23.5	Median session length	5	Shortest session
29.3	Average session length	152	Longest session

The participants consulted a surprising variety of sources—a total of 44 websites. Table 2 presents the site name or type of site, and the number of sessions that included a visit to that site or category. **YouTube** is clearly the primary site consulted by the participants—it was included in two-thirds of the sessions. General search engines (Google, AltaVista) were used to search for sites likely to contain videos of interest, or to directly search for videos. 'Legal' websites focusing on **commercial television shows and movies** were primarily consulted for reviews and trailers. Given the natural reticence to be observed breaking the law, the number of sessions including a visit to an 'illegal' **Peer-to-Peer or pirate** video site is surprising—and likely under-represents the scale of illegal downloading that occurs among participants. **Sport** sites included those dedicated to a single sport (eg, skateboarding) or to presenting updates on a sporting event (eg, the Tour de France). Sport sites include a variety of information—photos, discussion forums, news updates, and so

on—in addition to video. Similarly, **music** sites consulted included lyrics servers, artist sites, music news, forums, and reviews as well as music videos. Sites dedicated to **computer games** provide commercially produced game trailers and homemade gameplay recordings. Personal pages on **social networking** sites such as MySpace and Bebo include embedded video and links to videos on other sites (for example, YouTube).

Table 2. Categorization of websites visited in sessions

Site / site type	No. of sessions
YouTube	65
Other legal video filesharing sites	9
Search Engines: Google, AltaVista	14
TV / Commercial Movie (legally distributed): trailers, movie reviews, TV channel sites, video purchase, etc.	10
Peer-to-Peer / pirate / illegal	7
Sports related: sport interest, sporting event news	6
Music related: music videos, forums, artist, lyrics	6
Computer game related: trailers, forums	4
Social networking: MySpace, Bebo	3
Other: Wikipedia, digg.com	2

The overwhelming majority (79 of 98) of sessions consulted only one website (Table 3), with 50 of these one-site sessions at YouTube. Approximately 20% (20 / 98) of the sessions involved two, three, or four sites. Section 6.4 explores the searching/browsing strategies employed in multi-site sessions.

Table 3. Number of sites consulted per session

1 site	2 sites	3 sites	4 sites
79	11	8	1

Average number of sites consulted per session: 1.33

The descriptions of video information needs summarized in Table 4 derive from an informal categorization similar to the usability method of open card sorting. Data could be placed in more than one category and related data (the "why" data summarized in Table 5 and the initial query actions) was used to resolve vague information needs. For example, the data "a particular ad from tv that they had seen before" was initially placed solely under the TV category. However, when considered holistically alongside the "why" answer ("to show a friend a funny ad") and the initial query ("german car ad") the response was counted in three categories: "TV", "Humor" and "Cars". Similarly, background knowledge ("Spider Pig" is from the Simpsons Movie) and Internet searches ("assassins creed" is a computer game) were used to resolve other ambiguities in the data.

Table 5 show a summary of the categorization of the "why" data. The categorization was performed in a similar manner (an open card sort) to the information need data although there was less of a need for holistic ambiguity resolution. As with the information need data, "why" data could be placed in several categories: e.g. the response "I wanted to watch the video of the song because it is a song I currently enjoy listening to" was counted under "Visual", "Audio" and "Mental State".

Table 4. Categorization of subject of video information need

Category	Description / Examples	Count
Music	"music video of Iron Maiden live at a festival", "Dave Mathews Band music video"	23
Humor	"humorous video clips", "ipod in a blender"	18
Movies	"Spider Pig", "movie trailers of the new movie transformers"	13
TV	"myth busters clips", "new episodes of House or Boston Legal"	13
Computers	Including computer games: "warcraft PVP vids", "videos demonstrating the new compiz fusion 3d desktop for linux"	9
Cars	"a video that my friend put up of his car", "car clips", "boy racer video shown on news which police used to catch them"	6
Sport	"video about soccer comedy & best goals ("soccer")", "NBA basketball game videos"	4
Other	"any martial arts video or clips", "elearning methods"	22

The searchers' motivation for searching for videos could be straightforward ("bored and wants entertainment", Observation P2) or astonishingly complex. Consider the following two discussions of why the searcher wished to locate a music video:

> If there are any concerts that he is going to attend he likes to see if he can find any video for the particular band that he is going to see, so he can get a much better idea of what they are like performing live on stage, as an example he was searching for Justin Timberlake videos as it is a concert that he is going to soon. He will also often search for videos of concerts that he has been to, so that he can relive the experience what he felt at the concert. [E]

> …music is my passion (I like to compose music) so I seek to analyze music-related videos for learning purposes rather than just viewing it. … I really need to analyze fine-details of music videos like finger positions and motions on guitar performances. [Y]

This present research focuses primarily on behavior in locating videos and to a lesser extent on viewing videos, with a goal of suggesting functionality for video digital libraries. Discussions such as these indicate that a deeper examination of the intended use for the videos located could also inform digital library design; for example, a music video digital library might support participant E in creating a video scrapbook of concerts he has attended, and provide enhanced scene selection, playback, and zooming so that participant Y can more easily learn guitar techniques.

Table 5. Categorization of subject responses about their motivations for locating the video material

Category	Description / Examples	Count
Mental State	Explicit references to subject's emotional state/mood. "bored – to pass time", "entertainment", "casual enjoyment"	47
Visual	Explicit references to visual aspects of the target material: "because she heard that this music video is very creative, so she wants to have a look", "he wants to compare the performance and some of the features of the new desktop because he is interested in installing it on his machine"	12
Audio	Explicit references to audio aspects of the target material: "hear the songs", "to hear songs by an artist before deciding whether or not to buy the CD"	11
Learning	"to learn new techniques in hair styling", "so that they can learn to play the guitar parts of the song"	10
Social	"a friend told me about the video and I wanted to view it so I could literally see what he was talking about." "to share something funny on their bebo site"	7
MSM	References to the mainstream media (MSM): "follows the sport, getting daily coverage not shown on local tv", "watch F1 free. Is no longer on free tv in NZ"	4
Temporal	References to planned events in the future: "because he is going to the concert later on in the year so wanted to see what he was like on stage", "so the user has something interesting to watch later"	4
Other	"to see if they should buy it", 'interested to see what is still available"	24

For each session, the searcher was asked whether their video information need had been satisfied. Table 6 summarizes the responses. An information need was recorded as having been satisfied if the searcher felt that many or all of the videos selected for viewing were relevant to the information need, and unsatisfied if few or no viewed videos were relevant. Two searches identified large video files to download but did not involve viewing any video; for these searches, the participants indicated that they would not be able to make a relevance assessment at that point. In retrospect, it is not surprising that the vast majority—86 of 98—of the sessions satisfied the searcher's original goals, since a major motivation for many searches was simply to pass time in an enjoyable manner.

Table 6. Summary of satisfaction responses

Yes, wholly or partially satisfied	No, not satisfied	Unable to judge	No response recorded	Total
85	9	2	2	98

5. OBSERVATIONS: ACTIONS

For each of the 98 video searching sessions, the observer manually recorded actions taken. The observer briefly described each action, and then categorized the action as one of: Search, Browse, View, Read, or Other. Table 7 summarizes the number of actions in each category over all 98 sessions, and gives example actions drawn from the session records. A total of 1132 actions were recorded, for an average of 11.6 actions per session.

Table 7. Categories of actions in search observations

Action	Description	Count
Search	Manually constructing and entering a query: *ipod in a blender* *martial art video mb*	146
Browse	Looking through search results, site categories, etc: *browsed search results looking for a clip that might interest me.* *user then browsed through list of related videos given to the right of the video clip just watched*	300
View	Watching a video, in whole or part *watched part of a clip realized I've seen it* *participant watched the entire clip*	350
Read	Reading text associated with a video (eg, comments from other users, reviews) or other site text (eg, news stories on bbc.com) *Read comments from other users as well as description by uploader* *user reads the myspace profile at the top of the page*	93
Other	Other actions *participant posted their own comment about the clip* *Downloaded torrent file and loaded it into mtorrent* *visited other sites during the loading process*	243

Observers were asked to record the query for each **Search**, if possible. **Browsing** is notoriously difficult to define [13], and so was construed broadly so as to catch as many related user behaviors as possible. **Viewing** includes watching all or part of a video, by streaming or (more rarely) downloading first. **Reading** involves a longer engagement with text than skimming or glancing. The **Other** category captures actions that do not fit into the previous categories.

There are obvious methodological limitations encountered in recording and categorizing the actions within an observation. The five categories are not 'crisp', and the student observers struggled at times to distinguish between them. For example, is looking through the video descriptions in a search results list a Browse of the list, or a Read of each result in the list? Is clicking on a YouTube clip link that then autoplays one action (View), or two (Other—select clip and View)? An 'action' may be difficult to delimit: is deciding whether a video on the search results list is relevant an action, or does an action occur only if the relevance decision is positive (and the user selects that video for viewing)?

It is difficult to achieve high levels of observational and coding consistency across two or three researchers, let alone the 27 student observers in this study.

5.1 Searching
Search strings were recorded for 115 of the 146 search actions. Queries were short, ranging from one to six terms, with an

average of 2.39 terms per query and a mode of 2. This result echoes earlier studies of searching for text documents in OPACs, digital libraries, and web search engines, which typically find queries averaging between 2 and 3 terms (eg, [10], [16]).

Sessions contain an average of 1.49 searches. Successive query refinement is obviously minimal: typically adding a term or two to narrow a broad topic (eg, 'baby', 'baby fart') or to locate a specific video associated with a prolific performer (eg, 'Justin Timberlake', 'Justin Timberlake what goes around'). As will be discussed in Section 6.1, searches are often used to identify an initial relevant video, and then additional relevant videos are located through browsing.

5.2 Browsing

Browsing emerges in the observations as a significant activity. Video sites recognize this by providing a variety of options for browsing; YouTube, for example, provides browse-able displays of 'related videos', recent uploads, videos from same contributor ('more from the same user'), most popular, most discussed, and so forth. In this present study, looking through search result lists is also counted as browsing.

Participants report that browsing is particularly useful for undirected, one might say aimless, attempts to locate videos:

> Sometimes if I can not think of anything in particular that I want to watch I will just see what videos people are currently watching and what promoted videos there are … and choose a video that looks interesting from one of those lists. Once I have finished watching the video that I have selected I very often will then pick a video from the related video lists… This process can go on for a long time as the related links lead from one topic to another and so on. [P]

Browsing can also be an efficient technique for locating relevant videos. To this end, participants noted that YouTube's 'related videos' feature was particularly useful. This feature prominently displays a set of site-identified 'similar' videos in a panel near a video selected by the user for viewing. This feature seems most useful topics for which the metadata can support strong clustering of related relevant videos—for example, queries naming a particular piece of software or a specific musical artist:

> This is quite useful to me if I am looking at how something works. You get many similar videos so can build up a good idea of what a piece of software is like. [E]

> [Related videos] is handy when you're looking for music videos for a band because once you find one song it has the list of all songs done by that artist. [F]

5.3 Viewing

The number of clips viewed (Table 8) ranged from none (for searches that did not locate a relevant video, or searches that identified videos for downloading and later viewing) to a mammoth 38 (primarily car racing and crash clips). Viewing facilities in YouTube and other sites consulted in this study are relatively standard (play, pause, scrollable scene select). Section 7 examines usability issues encountered in viewing videos.

Table 8. No. of videos viewed per session (in whole or part)

3	Mode viewed	0	Fewest viewed
4.17	Average viewed	38	Most viewed

5.4 Reading

Text examined by participants includes video metadata (its title, tags, description), comments and reviews from other viewers, and text not necessarily related to the video content of the website (eg, news stories on the BBC website). Few observation summaries made specific reference to reading video metadata; it was assumed that participants skimmed the metadata in search result lists as they selected videos to view.

Participants were divided over whether comments are useful in deciding whether a video is worth watching (Section 6.6). While some comments are serious attempts by viewers to critique video content or engage with a serious dialog with other viewers, many are off-topic, ad hominem attacks on the video creators, or attempts to create weak social links within YouTube ([11], [12])—and so dilute the effectiveness of comments as a way to rate the quality or relevance of a video. Comments on peer-to-peer or pirate sites appear more reliable for evaluating relevance than those on YouTube, since on these sites the social networking aspects of user comments are downplayed. Even if they don't address the quality or content of the video, however, reading the comments can help pass time while videos buffer or download ("The comments on www.youtube.com also appeared to be a relatively good "time waster" while waiting for a video clip to stream" [R])—and they may even be read while the video is playing, if it isn't particularly gripping ("read the comments on the video while it was playing"; Observation E4).

A few participants consulted subject-specific or news sites with embedded videos (the official National Basketball Association site, a skateboarding enthusiast site, the BBC News site). For these sites, participants read site text both to identify relevant videos to view, and for the information content of that text.

5.5 Other Actions

The 'Other' category allows observers to capture behaviors other than searching, browsing, viewing videos, and reading. Commonly occurring 'other' actions include downloading videos (discussed in Section 6.8), managing tabs and multi-window displays (Section 6.5, 6.7), and fussing with video play functions (Section 7).

Interestingly, only two of the observations included adding a user comment or rating to a video, although two sessions included references to Instant Messaging (IM) recommendations or discussions with a friend, one session involved face-to-face discussion in a computing lab, and one included using email to send and receive URLs for promising videos. Apparently the participants in this present study are not part of the minority of hardcore 'YouTubers' [11] who frequently comment and rate videos on the site.

6. SEARCHING / BROWSING / VIEWING STRATEGIES

This section describes patterns of behavior for searching, browsing, and viewing videos.

6.1 First Search, Then Browse

Approximately two-thirds of the sessions began by searching (65 of the 99 sessions). If a relevant video was located on the first page or so of search results, then it was selected for viewing; otherwise, another search was entered. None of the participants looked through more than a couple of search result pages.

Once a potentially relevant video was identified, further relevant videos were located primarily by browsing related videos, rather than by returning to the initial list of search results. This pattern is, to a large extent, imposed by YouTube's lack of a graceful way to return to the list of search results—a user must rely on the browser's back button. A reliance on browsing is also symptomatic of the intrinsic mismatch between the representation of the information need (a textual query) and that of the desired result (video); it can be difficult for the user to identify terms used to describe videos [8]. This problem is exacerbated by the sparse textual metadata associated with YouTube videos—typically, brief titles and a handful of tags (a median of 6 per video) [6].

Given the effectiveness of browsing as a strategy for locating videos, why begin a session with a search? An initial search appears to be most useful in large and heterogeneous sites such as YouTube; the query serves to narrow down the overwhelming mass of available video to a more manageable set (the initial page or two of search results), and then the most relevant video(s) identified can be used as a starting point for a traversal of a small 'neighborhood' of the collection, by following 'related video' links. With smaller, more tightly focused video collections (such as a site dedicated to skateboarding), it can be easier to begin by browsing; the site can provide metadata and browsing structures more tightly suited to the interests of the user.

6.2 First Browse

Browsing can be a successful initial strategy if the user's needs are closely matched by either a 'starter' relevant video, or by the browsing categories of the collection.

Participants sometimes sought out interesting initial links themselves, by looking through non-video websites and online forums dedicated to a hobby or interest (for example, a video game review site or video game discussion forum), or by consulting a rating site such as www.digg.com. More frequently, good 'seed' videos are obtained through recommendations from friends; links to interesting videos are passed along by Instant Messaging, email, and even on slips of paper:

> This [browsing] often begins through conversations on instant messenger services with friends while working, and then someone mentions a video or sends a link. This then leads to browsing through the related links, and discussing the videos that are found. [H]

One framework for interpreting video link sharing is to view it as the enabling mechanism of a 'media circuit' [12]. In a media circuit, a technology or media (telephone, email, etc.) is used by members of a social group to maintain a sense of connection with each other. The connection may be restricted to members of the social group—for example, an Instant Message or email—or it may also be accessible by people outside the network, for example through entries on a MySpace page:

> Bob [a pseudonym] also uses MySpace to find videos online, Bob mainly uses MySpace video to find music videos of his favorite bands, he usually does this while browsing his friends MySpace profiles, often his friends will have links to bands MySpace profiles which contain music videos. [K]

Sharing an interesting video link can strengthen the closeness of a relationship—the sender demonstrates an understanding of the tastes of the receiver by identifying videos that the receiver will enjoy. The sender also reveals some aspect of his/her own personality through the content of the video being shared. Musical tastes are particularly significant markers for identification with or definition of social groups, and in the example above Bob's friends use links to musical artists and music videos as part of their public display of their own personalities.

A media circuit can involve bi-directional contact. In the quote from [H] above, an exchange of additional related videos and further discussion strengthens the social ties between friends—figuratively 'completing' the media circuit.

Participants experienced varying degrees of success in relying on built-in site browsing categories. Collections with a tight focus could cater their users' common information needs with appropriate categories: for example, a collection of (illegally distributed) commercial movies included a "most recent movie" category for new releases. This anticipation of user needs is obviously more difficult for heterogeneous collections, though the YouTube 'Comedy' category may be satisfactory for fulfilling a desire for 'something funny' [Observation Y2]. As noted in Section 4, 'humor' queries are often motivated by a desire to kill time, and so the bar is low for satisfying that need. Other, more specific, video needs may be more difficult to cater to with a priori categorizations. The YouTube 'Music' category, for example, was less useful for our participants; its categories ("Rock", "Pop") are far too broad to match to the users' interests.

6.3 Multi-focus Sessions

Several sessions involved searching/browsing for videos on two or more topics. Sometimes this behavior is deliberate: in session G2, the participant first browsed through World of Warcraft spoof movies "for enjoyment & laughs", and then got down to business and located DreamWeaver tutorials "to refresh [his] memory in software usage".

But YouTube's Related Videos feature is a double-edged sword. While it is invaluable in locating additional relevant videos, it also can bring to the user's attention interesting but irrelevant clips. Serendipitous exploration can be enjoyable:

> Usually ending up looking at videos on a very different topic to the one I originally searched for. [B]

Serendipity can also be distracting, as the user drifts away to far-removed 'neighborhoods' in the collection:

> This type of browsing is good if the user is looking to fill in time or entertain themselves but if the user was looking for something specific and got sidetracked it could be quite a nuisance. [W]

We draw a distinction between 'berrypicking' search behaviour [1] and this latter undirected exploration. In berrypicking, the user's exploration of a collection is motivated by an evolving understanding of the underlying information need, and the documents ("berries") selected support the satisfaction of that need. The serendipitous behaviour noted in the observations and ethnographic investigations could more aptly be termed magpie-like, as the participants lose sight of their original goal because a shiny new treasure distracts them.

6.4 Multi-site sessions

A significant minority (21 of 99 sessions) involved consulting more than one website for videos. In the simplest case, the participant could not locate the desired video on the first site, and simply repeated the same search on different sites until the video was identified (or the participant gave up).

Other multi-site sessions were more complex, as the participant attempted to compensate for inadequate metadata or search facilities on one site by identifying supplementary information on a second site. For example, one participant skillfully located music videos on YouTube by first finding artist, title, and lyric details on Google and www.lyrics.com:

> The participant also feels that her way of finding videos such as searching Google, searching lyrics.com then finally linking to her desired video [on YouTube] is quite time consuming, but uses this method as lyrics.com offers an alphabetical searching approach rather than simply searching what you want, this appeals to the participant as she is sometimes unsure what artists names are. Also the inclusion of lyrics from artists' songs is appealing as she can look up songs based on a remembered chorus rather than by name or artist. [F]

S	"Why why Love" episode 9
B	briefly scans the results half way through from the list, and finds episode 9 part 1
V	views episode 9 part 1, when it is half way through
O	presses Pause, and back to the list
B	briefly scans list and finds episode 9 part 2
O	open episode 9 part 2 in the new window, and press "Pause" (Let it load)
V	go back to the window with episode 9 part 1, and finishes the entire part
B	goes back to the list and finds episode 9 part 3
O	open episode 9 part 3 in the new window, and press "Pause" (Let it load)
V	go back to the window with episode 9 part 2, and finishes entire part
B	goes back to the list and finds episode 9 part 4

Figure 1. Fragment of an observation; viewing videos in order

6.5 Locating & Viewing Sequences of Videos

Sometimes the order in which several videos are viewed is important—the user locates multiple videos, but the information need is not satisfied unless the user watches the videos in a logical sequence. In 2006 YouTube established a limit of 10 for most user uploads, which forces uploaders to break longer videos into several clips [3]—and it can be a challenge to locate all of the clips and reassemble them into the correct order. Time sequencing, primarily on the video's date of production or release, may also be important: for example, so that a drama unfolds coherently through the episodes of a soap opera, or so that a sports enthusiast can quickly find the latest update from the Tour de France and then backtrack through the race coverage.

The tasks of locating all the videos in a series and sequencing them appropriately for viewing are not supported well in YouTube. Consider a fragment from the observation notes (D2) for a YouTube search for a single episode of the Chinese drama, "Why Why Love" (Figure 1). The episode has been broken into 10 video clips, and the clips do not appear in logical order in the search results list. The participant has to identify the next clip in the sequence on the search page, manually create a tab and load the clip into it, then stop each clip as it Autoplays so that it can buffer in the background—for each of 10 clips! This activity is interleaved with viewing the clips, and surely must have distracted from the participant's enjoyment of the show.

It is easier to locate multiple videos that have a logical time-based viewing sequence on sites with a subject focus (for example, a sports site) or that provide access to commercially produced videos. These sites tend to provide more specific and reliable date metadata (e.g., date of movie release), and to clearly label video fragments and present the fragments in sorted order. However, these sites still push onto the users the problem of managing the serial display (eg, through tabs).

6.6 Making Relevance Judgments

With a fast Internet connection, relevance judgments can be made based on the video clips themselves, rather than relying on document surrogates. One participant with a particularly good broadband connection reported that he simply opens "many videos in different tabs", then flips through the tabs and "evaluates the video based on the beginning of the clip instead of reading the comments and ratings" [N].

When Internet connections are slow, being able to make fast and correct relevance decisions in video selection becomes crucial to a satisfactory user experience. The penalties for an incorrect decision—wasted minutes as a page loads and video buffers—can quickly eat through the user's free time.

Participants reported relying on a variety of types of information to decide whether a particular video would be worth the download time. Thumbnails offer immediate visual feedback on the content and resolution quality of a video, and for some searches appear to be used exclusively to identify a relevant item ("…at times they would not even read the titles or summaries of the clips" [J]).

For more open-ended searches, such as when the participant simply wanted a funny or 'good' video, thumbnails are a less reliable indicator of relevance/interestingness/quality. At this point the participants reported relying on a variety of textual resources to make relevance judgments: video titles, ratings, view counts, and comments from other viewers.

Video titles can be particularly useful in making relevance judgments on searches involving what one participant referred to as "specific content" [Y]—that is, videos described by a commonly accepted title (eg, "Family Guy episode"), an associated person's name (eg, "Justin Timberlake), a narrowly defined technical term (eg, "squirrel suit"), or other highly distinctive descriptor. Titles are less reliable for general searches based on, for example, emotional impact: how much trust can be placed in a title including the words "hilarious" or "amazing"?

For these latter searches, viewer ratings or view counts can be a more trustworthy measure of whether a video is indeed funny, cool, or bizarre than the titles and descriptions created by the video's creators/uploaders. Viewer ratings may be thought of as slightly more reliable because the raters have to make a conscious decision about their feelings about the video. View counts are useful too, as a video with a high view count is "almost guaranteed to be either really good or really controversial" [A].

Viewer comments are less likely to be useful in making relevance judgments in general video sites such as YouTube; YouTube comments often include relatively meaningless "cool" comments or "you suck!" flames of the video uploaders [11]. Viewer commentary on topic-focused sites (for example, a skateboarding

enthusiast site) is given more weight, as it is perceived as being more likely to concern the content of the video rather than the video's creator. Comments are crucial to relevance decisions when downloading rather than streaming; given that a full length commercial movie may take several days to download, the participants want to be positive that the file isn't corrupted and that it is of viewable quality. Having made a decision to view this particular video, the participants are less concerned with whether the commenters enjoy the video or think that it sucks ("Comments related to full length clips will generally always be checked only to try and determine video quality." [G]).

When considering whether the content of a lengthy video might be personally interesting to the participant, small snippets of the original video can be invaluable. For example, several observations involve searches for trailers to commercial movies, so that the participants could decide which new releases to download or view in a theater.

In YouTube, duplicate or near-duplicate videos are common, particularly for popular clips ("Lots of different titles which all contained the key words had the same content in videos" [X]). These redundant copies are not marked as such or filtered from search results lists. Shortened or home-edited version of the same video may also clutter results lists. Users found it "frustrating" [I] and "annoying" [C] to browse the artificially longer hit lists, and particularly resented the potential for wasted time if they downloaded a video similar to one that they had previously viewed. On the other hand, at times the presence of duplicates could be useful; Participant I noted that videos with duplicate content can have radically different playback quality, and so "These duplicate search results allow users to quickly find a better quality video if the quality of their first selection is poor."

Determining relevance can be more difficult if the user is searching for a particular type of media in a site containing more than one media type. Participant R, for example, explains that he preferred video-only sites when searching for music videos:

> I browsed for music videos on the site [YouTube], unaware that users could also upload plain audio and wasted time and effort in waiting for the clips to stream… In contrast to the other two sites that were used in the sessions … where it is clear that all you will find is video clips and not just audio files "posing" as a video clip. [R]

On the other hand, for some music-related searches the presence or absence of video was irrelevant, as the primary goal was to locate the audio (for example, the participant in one search observation wanted "to hear songs by an artist before deciding whether or not to buy the CD").

The language of a video's audio content might, or might not, affect a searcher's relevance judgment. One participant who did not understand Japanese found the language to be 'an issue' when a video search for "funny videos" (Observation J3) unearthed a clip with Japanese audio. Another participant [V] searches for "Japanese pranks" for amusement; presumably the physical comedy transcends his inability to comprehend the audio. Of course, the language of a video's audio might be altered from the original, and whether this change has occurred can affect a video's acceptability to a given user. One participant searched for "Japanese anime" (Observation J3) but was only interested in videos with associated English audio, while Participant D was interested only in Chinese content with the original audio intact.

6.7 Interleaving Multiple Activities

At times, video searching can be a deliberately chosen distraction, a pleasant way to briefly "procrastinate work" [H] or "to take a break from the other drier things he was doing on the computer… a change of scenery" [W]. These participants reported that they work/study on their computer, and interleave these serious activities with brief sessions of video browsing. Alternatively, other computing activities can fill in the time as videos slowly download or buffer—two observations noted participants checking email or visiting other websites while they waited for the video to begin playing.

6.8 Downloading or Streaming?

The preference for downloaded or streamed video depends on several factors: the technical expertise of the individual, the length and quality of the video, how frequently the individual intends to view it, and the circumstances in which it will be viewed.

Downloading typically requires a greater degree of technical expertise than viewing streaming video—and frequently, as noted above, a willingness to ignore copyright laws (Section 3). Locating the requisite sites, software, plugins, and so forth is not as straightforward as clicking on a link in YouTube, and may constitute a barrier for the less technically capable ("They [a friend] only view streaming video for their convenience where as I don't mind downloading/saving videos, maybe because I am more computer literate than they are." [Y]).

Commercially produced videos (for example, movies) tend to be downloaded rather than streamed; these videos are longer (40+ minutes), and may download in the background or overnight. Whether a shorter commercial video (for example, a TV episode) might be streamed or downloaded further depends on the desired viewing quality of the video ("I prefer to download videos first and watch them later to get a higher standard of videos" [W]).

If a video is to be watched more than once, or in a social setting with friends, quality is important and downloading is preferable. For one-off viewing, a streamed version is more convenient:

> …he didn't really mind the quality and that it was far more important to him that the videos were quickly available than high quality. He also said that he would only watch a video once so it wasn't worth him downloading them. [W]

Finally, some events can only be enjoyably viewed as they actually occur, and must be streamed rather than downloaded for later viewing regardless of any degradation of the streamed video:

> I'm a big fan of Formula One racing… The only way I like to watch it is live… I have tried watching it on replays, … [but] even though the download is in better quality than the streaming version it just doesn't have the same appeal to me because I already have herd [sic] on the news what has happened and who has won. [L]

7. USABILITY ISSUES
7.1 Disruptive Buffering

With slow or erratic Internet connections—common in New Zealand [19]—achieving continuous video play can be a bit of an art. Smooth play is particularly important for music related videos (a major target category for the observations; Section 4). The user needs to allow the video to buffer before clicking start, but has to guess how much buffering is sufficient: 10%? 40%? The user is

impatient to see the video, and so does not want to over-estimate the buffering time—but the penalty for under-estimation is being forced to "watch the video in chunks" [Y]. And, of course, the act of judging when the correct amount of buffering has been achieved itself interferes with the enjoyment of viewing the video:

> This is an annoyance; the user shouldn't have to guess when to begin playback, the website playing back the content should be able to determine at what percent download a movie will give continuous feedback. [AA]

7.2 Missing Warnings and Alerts

Several of the video websites (including YouTube) give inadequate feedback to the user when problems occur:

> …some of the videos did not load and one of the video clips was taking a very long time to load. There were no comments or warnings about the video clip being disabled, no longer available or any indication that it would not load. [R]

YouTube's buffering progress bar was also a source of annoyance to users, who found it difficult to use the bar to estimate the speed of buffering or the time required to fully buffer the video. Participants who used a site (tv-links.co.uk) that reports buffering status in terms of percentage of the video buffered found this display to be more easy to understand, and were reported to be "a lot more relaxed" about buffering waits than YouTube users [R].

7.3 Autoplay

On YouTube, and other video-specific sites, selecting a video causes it to load and autoplay. This is obviously useful if the user's focus is on a single window and a single activity (locating and viewing a video)—in this case, autoplay is a one-click solution for an otherwise two-step process (select and play).

Many of the participants were working with multiple windows/tabs of videos, as well as moving between looking for videos and other computing activities; autoplay detracted from the user's current task as the sound from multiple videos on background tabs unexpectedly blared. As the browser does not indicate which tab is producing audio, the user had to locate the tab and video, stop the video, and then later restart it for viewing:

> It's common that videos take a while to load … the user will want to keep browsing in a new window, whilst this movie loads, unless perhaps it is a very short clip. … We found that videos would start by themselves, even when the webpage with the video was hidden. This really doesn't make sense, and this is very frustrating to the user, especially if they have to rewind the movie, and wait for it to rebuffer again. [AA]

7.4 Coarse Grained Video Skip Function

When viewing videos for educational purposes—for example, to learn new hairstyling techniques [Observation S4] or to find out exactly which guitar fingering techniques Paolo Nunlini uses [Observation Y3]—it is not uncommon to repeatedly re-view an especially interesting portion of a video. The simple skip and seek functions provided in the YouTube video player do not provide the fine-grained control required by some users:

> One of the users noted that YouTube's player was really annoying because they could not skip to an exact time where they wanted to replay from; it only let them seek through the video in small amounts of time, sometimes

that [sic] had to watch about 20 seconds of video that they weren't interested in watching again. [Y]

7.5 YouTube's Missing 'Back' Button

As noted in Section 6.1, at the time these observations were conducted YouTube lacked a 'return to search results' button. In a session that included searches, 'related video' and visits to other browsing spaces, participants quickly became confused about how many times they needed to click 'Back' to return to the search—or in a multi-tab session, which tab contained their search.

Additionally, returning to a previous page might cause a video to autoplay, on some video sites; this could be startling as well as time-consuming, as the streaming video placed its burden on the connection bandwidth ("This resulted in unexpected loud outbursts of sound and caused the process of getting back to the original search results to be delayed." [I]).

8. DISCUSSION AND CONCLUSIONS

The primary contribution of this paper is that we believe it is the first naturalistic study of video seeking behavior on the Web. This qualitative research complements laboratory-based usability studies of specific video digital libraries [2], qualitative study of a specific type of video content [14] and large-scale statistical analyses of usage and system behavior [3, 9]. We captured authentic information behavior by moving outside lab and asking broad questions about 'how people locate videos'. By doing so, we develop a fine-grained understanding of video seeking strategies and the difficulties that the participants encountered in locating and viewing their desired videos. This picture of user motivations and frustrations points up opportunities for further development of video digital libraries. We recognize that these insights must be viewed through the lens of these participants, and may not be fully generalizable. Future work should include further fine-grained investigations over diverse user groups.

Participants experienced a number of minor, but irritating, usability issues when locating videos (Section 7). These types of problems are best addressed in early stages of system development, and underscore the importance of incorporating user testing with the development lifecycle [2].

Lightweight facilities supporting users in creating video 'playlists' on the fly (a sequence of videos to be viewed in a particular order) would be welcome. It can be difficult to locate all the videos for a sequence on sites with relatively sparse metadata and only simple facilities for sorting search results. Once the videos for a 'playlist' have been located, the user should not have to manually create tabs or other viewing structures, personally monitor buffering, or be distracted by videos on hidden tabs auto-starting.

Internet connection speed is an issue for many New Zealand users and in many other locations connection speed is still likely to lag behind the users' demands. When downloading, users need to be able to test a video's technical quality and content—for example, by streaming arbitrary portions of the video file. Users need better support when judging relevance of videos before beginning to stream or download, to minimize potentially time-expensive interaction with irrelevant videos. Thumbnails are useful, but a single static image can only go so far in conveying the gist of a video—suggesting the usefulness of sophisticated video summaries and video surrogates (c.f. [20]).

For our participants, locating and viewing videos are deeply embedded in their lives: they mix in a bit of video surfing when

their interest in study flags, explore YouTube while dinner is cooking, and IM chat about videos they've just found. It's difficult to keep track of these different activities; juggling multiple tabs and windows, remembering which tasks are associated with which video searches and resources—indicating a need for structured support for multiple sessions and multiple goals. Without this support, users can become distracted and fall prey to the 'magpie syndrome' (Section 6.3), as serendipitous exploration overwhelms more directed information seeking.

One specific group of users were interested in accessing the audio aspects of their results (Table 5), the visual portion of the results appeared to be secondary. For these users, a zero-hit video search could be supplemented by audio results which could at least partially satisfy their information need. Similarly some visual aspects of queries might be satisfied by static images. Regarding video as a composite media (sequences of static images plus audio) suggests that system designers consider 'fall-back' strategies to simpler media types for unsuccessful video-only retrieval attempts. A 'fall-back' strategy might also involve passing a failed query over to an external audio-specific collection. Conversely, failed audio queries might be satisfied by routing the request to a video-specific digital library.

Some users (Table 5) wished to view resources that they could not access through their local mainstream media, either because of costs or geographical limits. These users were time-rich (in that they were expending effort to locate the relevant video) and possibly cash-poor (in that they weren't using pay-TV options). For content providers, the existence of these users suggests that there is some market for advertising-supported multimedia access.

Many of our participants' video queries were driven by their mood or emotional state. There are clearly many difficulties for computers in accurately discerning users' mental states. The field of affective computing is relatively new and some of the physiological measures it uses are not likely to be widely deployed in the near future. However, if the user's goal is really only to be pleasantly distracted for five minutes before starting to work then video resources are clearly not the only results that might satisfy their information need.

9. ACKNOWLEDGMENTS

REFERENCES

[1] Bates, M.J. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13(5): 407-424.

[2] Christel, M.G. and Conescu, R.M. 2005. Addressing the challenge of visual information access from digital image and video libraries. *Proc. of JCDL '05*, 69-78.

[3] Cheng, X., Dale, C., and Liu, J. 2007. Understanding the characteristics of internet short video sharing: YouTube as a case study. *Procs of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego (CA, USA), 15 – 28.

[4] Crawford, L. 1996. Personal ethnography. *Communication Monographs* 63/2, 158-170.

[5] Cunningham, S.J., Jones, M. 2005. Autoethnography: a tool for practice and education. *Proc. of the 6th New Zealand Int. Conf. on Computer-Human Interaction*, 1-8.

[6] Geisler, G., and Burns, S. 2007. Tagging video: Conventions and strategies of the YouTube community. *Proc. of JCDL '07*, Vancouver (BC, Canada), 480.

[7] Glaser, B., and Strauss, A. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago University Press.

[8] Goodrum, A. 2003. If it sounds as good as it looks: Lessons learned from video retrieval evaluation. *SIGIR 2003 Workshop on the Evaluation of Music Information Retrieval (MIR) Systems*. http://www.musicir.org/evaluation/wp3/wp3_goodrum_if.pdf

[9] Halvey, M. J. and Keane, M. T. 2007. Analysis of online video search and sharing. *Procs. of the 18th Conf. on Hypertext and Hypermedia)*. ACM, New York, NY, 217-226

[10] Jones, S., Cunningham, S.J., McNab, R.J. and Boddie, S. 2000 A Transaction Log Analysis of a Digital Library. *International Journal on Digital Libraries*, 3(2), 152-169.

[11] Lange, P.G. 2007. Commenting on comments: Investigating responses to antagonism on YouTube. Presented at: *Society for Applied Anthropology Conference*, Tampa (FL, USA). URL: http://sfaapodcasts.files.wordpress.com/2007/04/update-apr-17-lange-sfaa-paper-2007.pdf

[12] Lange, P.G. 2008. Publicly private and privately public: Social networking on YouTube. *Journal of Computer-Mediated Communication*, 13, 361-380.

[13] Lee, H., Smeaton, A.F., O'Connor, N., and Smyth, B. 2006. User evaluation of Fischlar-News: An automatic broadcast news delivery system. *ACM TOIS*, 24(2), 145-189.

[14] McKay, D., Shukla, P., Hunt, R., and Cunningham, S.J. 2004. Enhanced browsing in digital libraries: three new approaches for browsing in Greenstone. *International Journal on Digital Libraries*, 4/4, 283-297.

[15] OECD Broadband Statistics to June 2006. URL: www.oecd.org/sti/ict/broadband .

[16] Ozmutlu, S., Spink, A., and Ozmutlu, H. 2004. A day in the life of Web searching: an exploratory study. *Information Processing & Management*, 40/2, 319-345.

[17] Quinn, J.F., and Forsyth, C.J. 2005. Describing sexual behavior in the era of the internet: a typology for empirical research. *Deviant Behavior*, 26, 191-207.

[18] Saarinen, J. 2006. Update: Business NZ's bogus broadband report. *Computerworld NZ*, 24 February 2006. URL: http://computerworld.co.nz/news.nsf/news/29BFF57CD51A236CCC25711E00732501 .

[19] Statistics New Zealand 2006. Household Use of Information and Communication Technology Survey: 2006. Published April 2006. URL: http://www.stats.govt.nz/NR/rdonlyres/4698FAB1-8395-4FE1-B844-75BAA95E0A4B/0/householduseofict2006mr.pdf

[20] Yang, M. & Marchionini, G. 2005. Deciphering visual gist and its implications for video retrieval and interface design. *Proc. of CHI '05,* 1877-1880.