Time Shifting Patterns in Browsing and Search Behavior for Catch-up TV on the Web

Mika Rautiainen[1], Arto Heikkinen[1], Jouni Sarvanko[1],
Konstantinos Chorianopoulos[2], Vassilis Kostakos[1], Mika Ylianttila[3]

[1] Mediateam, University of Oulu
P.O.BOX 4500
FIN-90014 University of Oulu, Finland
+358 294 48 0000
firstname.lastname@ee.oulu.fi

[2] Ionian University
Plateia Tsirigoti 7, Corfu
49100, Greece
+30 26610 87707
choko@ionio.gr

[3]Center for Internet Excellence
P.O.BOX 1001
FI-90014 University of Oulu, Finland
+358 8 553 7651
mika.ylianttila@cie.fi

ABSTRACT
Catch-up TV services on the Web have facilitated time-shifted TV viewing. However, there is limited information about user search behavior with regard to recently time-shifted versus archival TV content. We deployed two distinct content-based web services to explore information retrieval of time-shifted TV content. The first web service is based on a browsing metaphor, while the second is based on free text content search metaphor. We analyzed more than 5000 user sessions from 12 months of logs and found that the programs accessed via browsing categorized program content summaries were typically less than one week old. In contrast, the programs accessed via free text search on subtitle content were typically more than a week old. Our findings provide a first assessment of user behavior in accessing time-shifted and archival TV content. Further research should develop the user experience for content-based TV access and explore the sharing patterns of archival TV content on social networks.

Categories and Subject Descriptors
H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems, H.3.5 [INFORMATION STORAGE AND RETRIEVAL]: Online Information Services H.2.4 [DATABASE MANAGEMENT]: Multimedia databases

General Terms
Design, Experimentation, Human Factors

Keywords
User behavior, content-based retrieval, catch-up TV

1. INTRODUCTION
Digitalization and the Internet have created many opportunities for time-shifted TV viewing [4]. It is necessary to study what types of information seeking behavior span over long periods of time in the archive and for how long. The above issues are important to new digital service development for interactive TV. Looms [14] identified four distinct modes of TV viewing: 1) *Broadcast viewing* occurs in real-time to the broadcast 2) *Off-set* or *buffer viewing* occurs in almost real-time to the broadcast, 3) *Time-shifted viewing* occurs at least one hour after the broadcast, and 4) *Archival viewing* occurs over one week after the broadcast. Carey [3] noted that archival viewing has not been utilized due to user interface issues with solutions available at the time. Whereas traditional TV schedule is influenced by the expected popularity of the content, time-shifting TV services provide an opportunity to reach a wide audience for niche archival TV content [1]. Van den Bulck [16] suggests that it may be more productive to look at time shifting as another TV channel. Also Ferguson [7] and Kang [11] point out that new technologies such as video recorders, cable television and the Web have increased the channel repertoire of TV viewers.

In addition to Looms' modes of TV viewing, several information seeking and retrieval models can describe users' TV viewing behavior. The classical information retrieval model consists of a query with an explicit information need and process of matching the query to document representations to identify maximally relevant document set. Bates [2] proposed an alternative model called "berrypicking" where user needs and search strategies are constantly evolving during the seeking process. Cove and Walsh [5] described a serendipitous model as an unstructured random activity potentially leading to accidental discovery of interesting information. Marchionini [15] introduced exploratory search as a three-stage information behavior process consisting of lookup search (fact retrieval), learning search (exploration and comprehension of neighbourhood) and investigative search (analysis, synthesis and evaluation). According to the definition by Hjorland [8] browsing implies relevance assessment and decision about acquisition or selection of some of the information objects. We adopt Hjorlad's definition of relevance in our experiments: if a user clicks any program related information on the user interface, we interpret this action as an indication of relevance. In this paper we introduce new user services to facilitate flexible content-based searching and browsing strategies for finding TV programs and catch-up TV streams for time-shifted and archival viewing.

There are several studies that examine user behavior in content-based TV and video archives. Huurnink et al. [10] studied the search behavior of media professionals with audiovisual archives and suggested that increased support for fine-grained access to audiovisual material through e.g. content-based analysis might be beneficial for the archive use. In another study, Huurnink et al. [11] recommend audiovisual archives to invest in embedding content retrieval into their work-flow and that audiovisual archives prioritize video retrieval using transcripts. Lee et al. [12]

EuroITV'13, June 24–26, 2013, Como, Italy.

introduced Fischlár-News as one of the first automatic, content-based broadcast news analysis and archival systems that allowed users to search, browse, and play it in an easy-to-use manner with a conventional web browser. Informedia Digital Library is the landmark research program in video retrieval that was started in 1994. Christel [6] calls for longitudinal studies with broader user populations, broader selection of unstructured content genres and experimenting with browsing tasks in addition to retrieval.

2. NEW CONTENT-BASED SERVICES FOR BROADCAST TELEVISION

In order to collect a representative data-set, we developed an online TV platform that indexes and summarizes TV programs in near real-time using detected subtitle text from seven free-to-air TV channels. Our service architecture consists of three parts and an integrated browser component for viewing program content excerpts.

The *Service backend* processes content dynamically in DVB-C video data streams for seven national channels. Subtitle transcripts are detected and recognized from the video stream, and program descriptions are de-multiplexed from the MPEG transport stream.

The *TV Program Search* service provides free text queries on program subtitle content, program title, and broadcasting channel. Program subtitles and descriptions have been extracted from DVB broadcast. Title search finds programs by title, while channel search is used to filter out programs from other channels.

Figure 1 shows the results of conducting a content search using the query *"Fiat 500" auto.* The service allows free text searching of recent and relevant TV programs. Users can use the service to find specific information from the metadata of 180 000 broadcasted programs and use it to find web streams of relevant programs from the broadcaster sites.

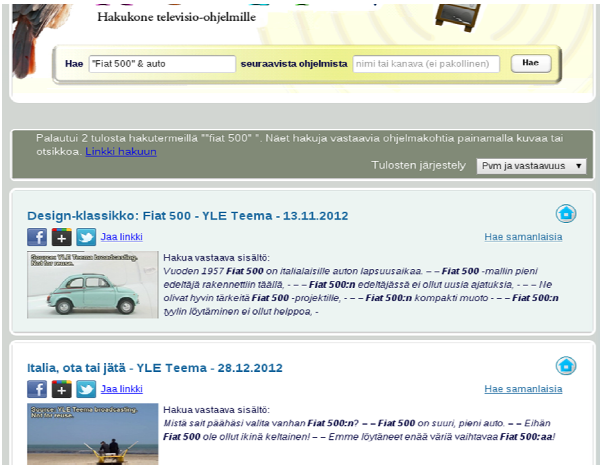

Figure 1. Program Search interface depicting program query with terms "Fiat 500" & auto

Catch-up TV Guide is a service that collects, categorizes and analyzes TV programs from seven linear Finnish TV channels. The service supports browsing and following programs on the Web, and programs can be watched online if the catch-up web stream is available on the broadcaster's site. The service relies on machine learning and data mining techniques to detect novelty keywords in program subtitle content, and allows users to view dynamic quotes from the program. The service allows users to browse Finnish TV broadcasting according to predefined program categories (*news*, *cooking*, *nature, living, documentaries, science* etc.) and time scales (*most recent*, *past week* and *past month*). Figure 2 shows the user interface of Catch-up TV Guide with a program and highlighted novelty words that allow users to view excerpts of the program content using Program Content Browser.

The *Program Content Browser* allows users to inspect program content information dynamically. It assists users in determining relevance of content without requiring playback of the TV program. The Program Content Browser is an integrated popup window that opens when users access program-related links from either Catch-up TV Guide or TV Program Search service. It displays basic information about the program and provides search and browsing mechanisms to inspect parts of program through dynamically extracted picture quotes. Figure 2 depicts the Program Content Browser with a quote from a program. The quote shows "grain fields", a concept that has been highlighted automatically as a novelty word in a TV program about national landscapes of Finland. Automatically highlighted novelty words help users to examine parts of content information before viewing the actual program stream at the broadcaster site.

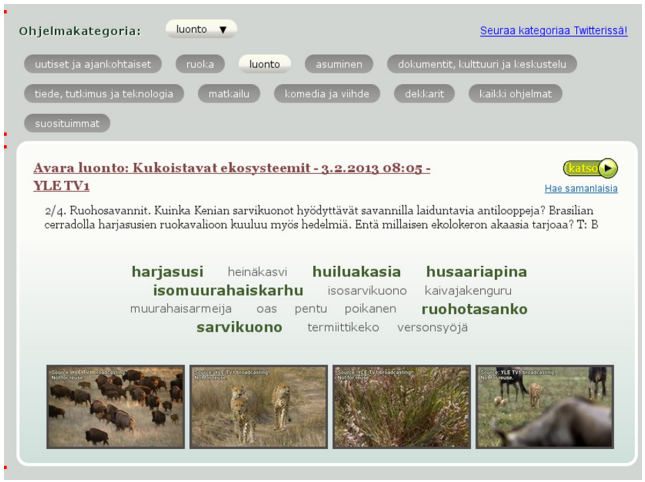

Figure 2. The Catch-up TV Guide user interface showing most recent program from the category Nature.

Figure 3. Program Content Browser shows an excerpt of "grain fields". It is part of the automatic program summary.

3. LONGITUDINAL EVALUATION OF A LIVE DEPLOYMENT

People searching for TV programs have traditionally used TV guides that show only upcoming programs. Today catch-up TV services and video-on-demand are liberating users from the present, i.e. linear broadcasting schedules. So will users change their behavior? Is recency still important in catch-up and archival viewing?

We deployed our platform on the Internet and conducted a 12-month evaluation of the system with real users. The users were members of the public and were not directly recruited by the researchers. Users were free to interact with the service using their own equipment at any time suitable to them. Our service did not stream any copyrighted media but instead offered users pictorial quotes from the program metadata.

We wanted to investigate what is the distribution of age of the programs that people find through the services. Our user logs span 12 months (02/2012-02/2013). We collected 5032 user sessions for analysis: 2500 sessions from the Catch-up TV Guide service and 2738 sessions from the Program Search service. Over 46000 activity events were logged, of which over 18000 were activities in the Program Content Browser. Next we present an analysis of user behavior in the Catch-up TV Guide and Program Search services respectively. For each session we recorded the behavior of users by logging activities at the service, i.e. pressing buttons and clicking URL access links that result in further examination of program content information with Program Content Browser.

3.1 Differences between browsing and searching strategies

The functionality for searching program information is different in the Catch-up TV Guide and Program Search services. Whereas the former provides recent, weekly and monthly summaries of programs in different categories for browsing, the latter allows free search of programs from 180 000 program database containing broadcasted program metadata of almost three years' time. Here we describe how these differences affect usage logs.

The premise of the strategies is that users are browsing most recent programs via Catch-up TV Guide service and using Program Search service for archival search. It is possible to use the Search service to retrieve most recent programs as well, but Table 2 shows that this is not the case.

Table 2 gives basic statistics of mean, median and percentiles of the respective age distributions (ADs).

Table 2. Age statistics of the accessed programs. Numbers are in days.

	Catch-up TV Guide	Program Search
mean	1.7546	196.371
stdev	2.05322	275.558
25th percentile	1	2
median	1	19
75th percentile	2	358
95th percentile	4	777

A significant sign of heavy tail effect can be seen in the Program Search AD where the differences between median, mean and standard deviation are large. The 95th percentile shows the heaviness of the tail: the oldest programs that are needed to cumulatively reach 95% of the accessed programs are 777 days i.e. 2.1 years old. In contrast, 95% of the accessed programs in Catch-up TV Guide service were up to four days old.

Figure 4 displays a cumulative age histogram of accessed programs in Catch-up TV Guide and Program Search service with programs up to 30 days of age. The age of a program is defined as the difference between access time in the service and original broadcast time. As can be seen from the figure, the majority of the programs accessed via the Catch-up TV Guide fall within the age of seven days, whereas only 42.8% of the programs accessed in Program Search service are less than seven days old. The age distribution of both histograms follows a power law, with the distribution of the Program Search having a heavier tail than the Catch-up TV Guide distribution.

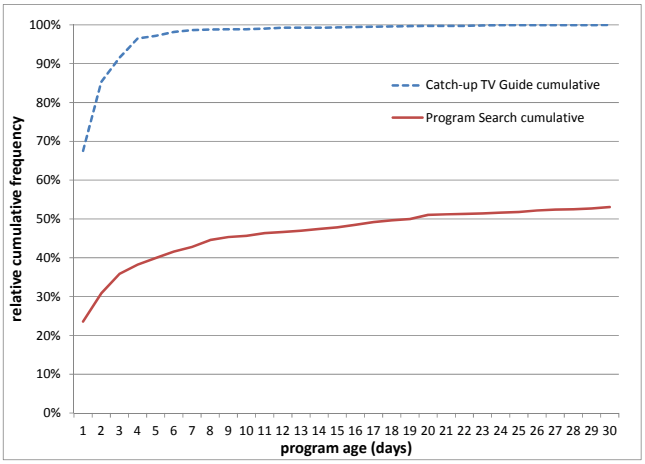

Figure 4. Cumulative age distribution of accessed programs in Catch-up TV Guide and Program Search (first 30 days)

Figure 5 illustrates the full cumulative frequency distribution of program ages in the Program Search service. The oldest accessed programs were 986 days (2.7 years) old.

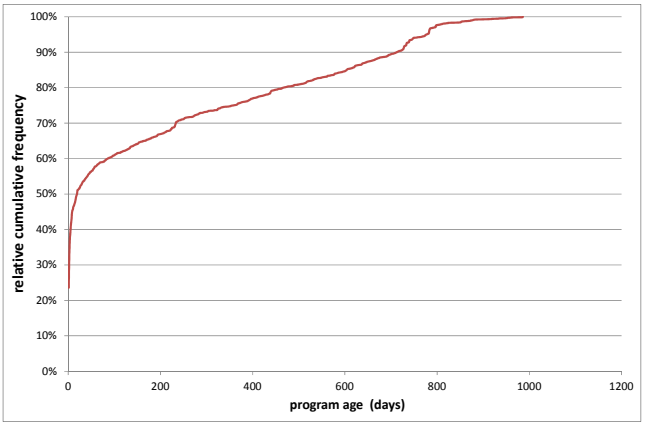

Figure 5. Cumulative age distribution of accessed programs in Program Search

4. DISCUSSION

The differences between the distributions of age of the accessed programs, namely the much heavier tail in Program Search than the Catch-up TV Guide, demonstrate that our users retrieved and

accessed programs of any age with the Program Search service while used the Catch-up TV Guide to find the most recent programs within days. This is very likely due to the design of the services, but also indicates that people treat browsing TV programs with strong temporality in mind: 93.6% of the page views in the Catch-up TV Guide were using *most recent* time scale, the rest used *past week* and *past month* time scales. The 95th percentile of the age of accessed programs is only 4 days, suggesting that users were considering: "let's see what was shown recently on cooking programs." In contrast, a more goal oriented searching reveals a heavy tail effect of relevant information suggesting that user behavior is more age neutral ("Show me cooking programs that use lamb in cooking") even when the likelihood of finding active catch-up TV sources is low.

Previous work has explored time-shifted TV viewing in the short time-frame provided by local storage devices, but they have not regarded engagement with older TV content. In particular, Looms [14] found that the users of digital video recorders use them for pausing broadcast TV content or time-shifting a favorite program for later viewing in the same or the following day. There has been limited research in the scale of weeks and months old archives. Our results indicate that users access older TV content information if it is retrieved and relevant, so broadcasters and content providers should consider this when designing new services for end users. Our findings about the use of archival TV content are in agreement with Levy [13] who examined the use of home video recorders for time-shifting and found that many home recordings are not replayed within a diary week.

5. CONCLUSIONS

In this paper, we presented a fully operational online service for content analysis of Finnish TV broadcasting. We evaluated this framework by introducing new user services with content-based access for broadcast TV content on the web. We studied the significance of temporality in accessing program content information with thousands of user sessions.

We have introduced content-based service platform with new end user tools for accessing digital TV broadcast content on the web. With continuous indexing and 180 000 indexed programs, we believe our service framework provides a significant experimental platform for future end-user TV services development. We described two end user web services for searching and browsing TV broadcasts. The Catch-up TV Guide service supports time-shifted viewing and the Program Search is a content-based search tool for archived TV programs. We analyzed one year of usage data that revealed quantitative differences in user behavior between the two end user services. We showed that user queries retrieve relevant information from TV programs with heavy tailed age distribution indicating that users are not just limiting their search to recent programs, but are more broadly interested of any program content that fits to their search interests. Future work involves improving the content-based user experience and combining social media with content-based TV services.

Better utilization of personal preferences and context pose the need for technological frameworks capable of semantic understanding of media via content-based technologies. This is an important step in maximizing content relevance for the end users, improving user satisfaction with catch-up services, and more efficiently adopting new, non-mainstream content for broader audiences.

6. ACKNOWLEDGMENTS

We would like to thank Academy of Finland for supporting this work.

7. REFERENCES

[1] Anderson, C. (2004). The long tail. Business Books.

[2] Bates, M.J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online review. (5), 407-424.

[3] Carey, J. (2002, May). The evolution of TV viewing. In Proc. 4th Annu. TV Meets the Web Seminar.

[4] Cesar, P., & Chorianopoulos, K. (2009). The Evolution of TV Systems, Content, and Users Toward Interactivity. Foundations and Trends® in Human-Computer Interaction (Vol. 2, p. 95). doi:10.1561/1100000008

[5] Cove, J.F. and B.C. Walsh. (1988) Online text retrieval via browsing, Information Processing and Management, Vol. 24, No. 1, p. 31-37.

[6] Christel, M.G. (2006) Evaluation and User Studies with Respect to Video Summarization and Browsing http://repository.cmu.edu/cgi/viewcontent.cgi?article=1380&context=compsci

[7] Ferguson, D. A. (1992) "Channel repertoire in the presence of remote control devices, VCRs, and cable television," Journal of Broadcasting and Electronic Media, vol. 36, no. 1, pp. 83–91.

[8] Hjørland, B. (2011) Theoretical clarity is not "Manicheanism": A reply to Marcia Bates. In: Journal of Information Science. 37(5), p. 546-552

[9] Huurnink, B., Hollink, L., van den Heuvel, W. and de Rijke, M. (2010) Search Behavior of Media Professionals at an Audiovisual Archive: A Transaction Log Analysis. Journal of the American Society for Information Science and Technology, 61(6):1180–1197.

[10] Huurnink, B., Snoek C.G.M., de Rijke, M., and Smeulders, A. W. M. (2010) Today's and tomorrow's retrieval practice in the audiovisual archive. In Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR '10, pages 18–25, New York, NY, USA.

[11] Kang, M.H. "Interactivity in television: Use and impact of an interactive program guide," Journal of Broadcasting and Electronic Media, vol. 46, no. 3, pp. 330–345, 2002.

[12] Lee, H., Smeaton, A.F., O'Connor, N.E. and Smyth, B. (2006) User evaluation of Físchlár-News: an automatic broadcast news delivery system. ACM Transactions on Information Systems (TOIS), 24 (2). pp. 145-189.

[13] Levy, M. R. (1983). The time-shifting use of home video recorders. Journal of Broadcasting & Electronic Media, 27(3), 263-268.

[14] Looms, P. (2005). Recent Developments with PVRs and the Free-To-Air Television Market in Europe. DR/EBU.

[15] Marchionini G. (2006). Exploratory Search, from Finding to Understanding. Communication of the ACM, 49(4): 41-46.

[16] Van den Bulck, J. (1999). VCR use and patterns of time shifting and selectivity. Journal of Broadcasting & Electronic Media, 43(3), 316-326.