

Studying Extended Session Histories

Chaoyu Ye
Mixed Reality Lab
University of Nottingham, UK
psxycy1@nottingham.ac.uk

Martin Porcheron
Mixed Reality Lab
University of Nottingham, UK
me@mporcheron.com

Max L. Wilson
Mixed Reality Lab
University of Nottingham, UK
max.wilson@nottingham.ac.uk

ABSTRACT

While there is an increasing amount of interest in evaluating and supporting longer “search sessions”, the majority of research has focused on analysing large volumes of logs and dividing sessions according to obvious gaps between entries. Although such approaches have produced interesting insights into some different types of longer sessions, this paper describes the early results of an investigation into sessions as experienced by the searcher. During interviews, participants reviewed their own search histories, presented their views of “sessions”, and discussed their actual sessions. We present preliminary findings around a) how users understand sessions, b) how these sessions are characterised and c) how sessions relate to each other temporally.

Categories and Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

Keywords

HCIR, Interactive, Information Retrieval, Sessions

1. INTRODUCTION

Information Retrieval (IR) specialists are becoming increasingly concerned with users who continue to search beyond a few queries or a few minutes¹. Although Information Retrieval, and even Interactive IR, evaluations are well known, research is recognising situations where people continue to search after finding seemingly useful results [13]. Some might be in a larger session involving several related subtopics, while others may continue to search for entertaining videos until they struggle to find ‘good’ results [3, 1]. Consequently, researchers are interested in how to evaluate, measure, and ultimately better support searchers who continue to search for extended sessions.

Most research into extended search sessions, described in detail below, has focused on analysing search engine logs [1, 4, 8] by dividing the logs using obvious periods of inactivity and either qualitatively [1] or quantitatively [4, 8] characterising them. Some research has investigated human web behaviour and user goals qualitatively through interviews,

¹The recent NII Shonan event and the forthcoming Dagstuhl are both, for example, focused on this topic.

however our research has focused on using such methods to better understand real extended search sessions. This paper begins by first summarising literature on sessions and then describes our research methods and preliminary findings about extended search sessions.

2. UNDERSTANDING “SESSIONS”

Although investigations into web sessions can be dated back to around 20 years ago (e.g. [2]), the concept of a session still lacks clear definition. A number of researchers have generated diverse definitions of a session using different delimiters such as cutoff time, query context, or even the status of the browser windows (e.g. [7]). In 1995, Catledge and Pitkow used a “timeout”, the time between two adjacent activities, to divide user’s web activities into sessions and found that a 25.5 minute timeout was best [2]. Their research, however, was focused on general web activity rather than search sessions, but their 25.5 minutes timeout has been used by many others. He and Goker later aimed to find the optimal interval that would divide large sessions, whilst not affecting smaller sessions [4]. Their analysis found that optimal timeout values vary between 10 and 15 minutes.

In 2006, Spink et al [11] defined a session as the entire series of queries submitted by a user during one interaction with a search engine, and one session may consist of single or multiple topics. Their approach focused on topic changes rather than temporal breaks, yet it is perhaps unclear how they determined “one interaction” with a search engine.

A clear definition has also been cited as an important challenge in other research. While focusing on “revisitation” behaviour, Jhaveri and Rähkä [6] and Tausher and Greenberg [12] found it challenging to differentiate between in-session revisitation and post-session revisitation, for which a clear detection of session boundaries would be useful.

When focusing on searching, rather than web sessions, some use the concept of a “query session”. Nettleton et al defined a query session as at least one query made to a search engine, together with the results which were clicked on and other user behaviours as well [8]. They also evaluated the “session quality” based on the number of clicks, hold time and ranking of selected documents, and they used these measures to help determine the difference between sessions.

To summarise the different approaches used to define sessions, Jansen et al. provided a summary of the three most representative strategies [5], as shown in Table 1. As IP and cookies were utilised to identify a user, the most frequent strategies involve temporal cutoffs and topic change.

The methods summarised in Table 1 are primarily focused on temporal and topical boundaries, but other research has shown clear challenges to these strategies. Mackay et al, in

Table 1: Session Diving Strategies; Jansen et al [5]

Approach	Session Constraints
1	IP, cookie
2	IP, cookie, and temporal cutoff
3	IP, cookie, and content change

2008, examined tasks that frequently occur as multi-session tasks, where something thematically consistent occurs over multiple sessions [7]. Moreover, research into web, browser, and browser-tabs, has found that some users often keep web pages spread out over time, especially in the information gathering tasks, e.g. [10]. These situations indicate that the logged web behaviour may differ significantly from the actual behaviours and intentions of the searchers. This research focuses on the searcher’s experience of web sessions, such that others may continue to develop strategies for more accurately dividing large scale logs into sessions.

3. EXPERIMENT DESIGN

To understand and characterise real extended search sessions, we employed similar interview methods to Sellen et al. [10]. Participants were engaged in a 90-120 minute interview about their own search behaviour. To ground the interviews in real data, participants focused on printouts of their own web history, and we used the card sorting technique [9] to probe their mental models of sessions. The procedure was approved by the school ethics board and pilot tested.

Participants began by providing their web history and they were advised to edit their history in advance should they wish to keep some logged activities private². These logs were gathered by importing their search histories to Firefox (if not already there), and creating an XML export using “History Export 0.4”³. This log was then structured and preliminarily processed using a) automatic methods to find search URLs, and b) manual investigation to find possible sessions to discuss in the interview. After providing demographic information, participants spent around 20 minutes examining the structured printout of their history, using a pen to mark sessions. These sessions, unless duplicates of prior sessions, were written onto separate cards for later sorting until around 20 cards were produced. Each card had a number, a title, activity purpose, included history items from the history list and also whether it has been completed successfully or not; an example is shown in Figure 1.

The remainder of the interview involved first open, and then closed card sorting. Open card sorting allowed the participants to classify and group the sessions according to their own ideas, whilst closed card sorting allowed us to make sure the following dimensions were considered: purpose, for whom, with whom, location, duration, difficulty, importance, frequency, and priority. This exercise was to help explore the session feature in a more detailed way. For example, studying frequency helps to find out the most frequent sessions and elicit the pattern of user’s web activity.

²Although this means we have likely missed common search sessions, like the lengthy adult sessions observed by Bailey et al [1], it was considered an important ethical provision.

³addons.mozilla.org/en-us/firefox/addon/history-export/

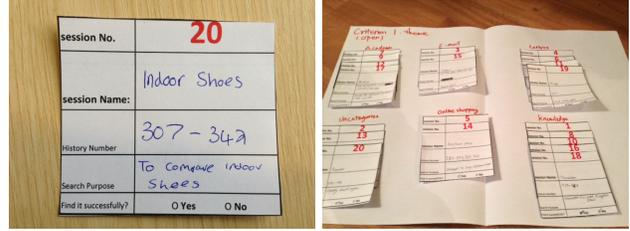


Figure 1: Session Card Information

In addition, the reasons for leading to non-success and difficulty can be investigated via the card sorting of difficulty, and the difference of user’s web behaviour in different environments can also be examined by the sorting of location. The entire interview was audio recorded, and physical copies of the card sorts were kept for analysis.

This paper describes our preliminary analysis of the first phase of the study, which involved 11 interviews. Phase two, which is still under way, involves a slightly refined methodology to capture more information about topics that emerged from the initial analysis described below. A more comprehensive analysis of both phases will be published later.

4. PRELIMINARY FINDINGS

Based on our preliminary investigation, some potentially interesting results relating to perceived duration, time of day, and use of queries were found. We considered each of these below according to two aspects: activity goal and activity context. For activity goal, we used Sellen et al’s [10] 6 categories: ‘finding’, ‘information gathering’, ‘browsing’, ‘transaction’, ‘communication’, and ‘housekeeping’. This approach did not include any email, so this was added as a 7th category. For activity context, we applied Elseweiler et al’s [3] comparison between work and non-work (leisure) activities, involving: ‘work’, ‘serious-leisure’, ‘project-leisure’, and ‘casual-leisure’. At this early stage in the project, the primary author performed the classification individually based on corresponding examples given in the referenced work.

4.1 Defining Sessions

There were 216 sessions in total and 19.6 sessions per person have been studied thus far, as shown as Table 2. Amongst these, 94 were longer than 5 minutes, 99 featured search and only 9 sessions were unsuccessful.

Table 2: All Session Information

Participant	Session No.	Long Session No.	Unsuccessful Session No.	Search Session No.	Query No.
1	18	9	1	13	45
2	30	14	0	11	34
3	20	12	1	12	101
4	20	8	1	9	22
5	16	10	0	6	17
6	26	6	0	16	27
7	30	5	1	0	0
8	17	7	1	12	74
9	10	6	0	6	18
10	10	8	4	4	57
11	19	9	0	10	23
Total	216	94	9	99	418
Avg.	19.6	8.5	0.8	9	38

All participants mentioned that activities with the same purpose and subject should be grouped into one session, as shown in Table 3. In addition, 8 of the 11 suggested that similar tasks happened in different time periods should be classified as a single session, rather than them being tem-

Table 3: Session Delimiters Summary

Participant	Topic	Type of Source	Differ time-> Differ Session	Emotion
1	+	+	-	-
2	+	-	+	-
3	+	-	-	-
4	+	-	-	-
5	+	-	-	-
6	+	-	+	-
7	+	-	-	+
8	+	-	+	-
9	+	-	-	-
10	+	-	-	-
11	+	-	-	-

porally connected. Some participants said that they always kept the browser windows open when doing long-term tasks. Finally, 1 participant advised that they care about the emotion involved within these web activities, even when they were doing the same task, such as “buying a pair shoes”. In particular, this participant indicated that one topically consistent session should be divided between two disappointingly unproductive and excitingly productive phases.

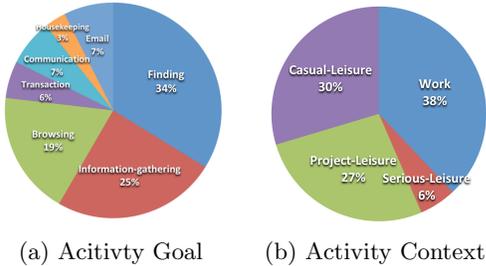


Figure 2: Session Categories

Finally, besides the pre-defined dimensions, participants also came up with some unique sorting dimensions as shown in Table 4, and these may benefit in exploring the session’s delimiters and features in new perspectives.

Table 4: Unique Dimensions

Unique Dimensions	
Google it or Go to Website directly	Content contributor
National	Certain topic or not
University related or not	Based on old knowledge or brand new
Amusement	Preference
Result Satisfaction	Eyes Ears Needed
Security	

4.2 Duration

As duration is one of the targeted dimensions, all participants were asked for their own definition of what constitutes a “long session”. 45% of participants defined the session where the duration is more than 5 minutes, whereas 27% went with over 30 minutes, 18% more than 1 hour, and 1 participant chose over 2 hours.

Because participants first defined what they considered to be a long session, and then later sorted their sessions into length categories, we investigated the difference between sessions that met their definition of long, and ones they remembered as being long during the card sorts. Participants frequently grouped ‘defined short’ sessions as long and vice-versa. Consequently, we investigated both ‘over-estimated’ and ‘under-estimated’ sessions in addition to ‘defined long’, ‘long’, ‘actual long’, ‘defined short’, ‘short’, and ‘actual short’ as given in Table 5.

Table 5: Duration Categories

Group	Detail
Defined Long	Sessions defined as Long by Participant
Long	Session whose actual duration is ≥ 5 mins
Actual Long	Session defined as Long and its actual duration is ≥ 5 mins
Over-estimated	Session defined as Long but its actual duration is less than 5 mins
Defined Short	Session defined as Short by participant
Short	Session whose actual duration is less than 5 mins
Actual Short	Session defined as short and its actual duration is less than 5 mins
Under-estimated	Session defined as Short but its actual duration is ≥ 5 mins

Table 6: Duration, by Activity Goal

	Defined Long	Over-esti	Defined Short	Under-esti
Finding	24	17 (70.8%)	36	3 (8.3%)
Info-gathering	35	15 (42.9%)	7	4 (57.1%)
Browsing	28	17 (60.7%)	5	0
Transaction	4	2 (50.0%)	5	2 (40.0%)
Communication	9	3 (33.3%)	5	0
Housekeeping	0	0	1	1 (100.0%)
Email	7	6 (85.7%)	7	0

Firstly, considering activity goals given in Table 6, the number of ‘information-gathering’ sessions defined as long was 5 times as that of those ‘defined short’, as was the same with ‘browsing’. On the contrary, the number of ‘finding’ sessions defined as short was 1.5 times the number defined as long. Overall, nearly 70% of ‘finding’, 42% of ‘information-gathering’, 60.7% of ‘browsing’, 50% of ‘transaction’, and 85.5% of ‘email’ sessions defined as long were overestimated by users. Moreover, under-estimation occurred with ‘finding’, ‘information-gathering’, and ‘housekeeping’ although over-estimation was more frequent with ‘finding’, ‘browsing’, ‘communication’, and ‘email’ sessions.

Table 7: Duration, by Activity Context

	Defined Long	Over-est.	Defined Short	Under-est.
Work	38	22 (57.9%)	31	2 (6.5%)
Serious-Leisure	8	2 (25%)	1	0
Project-Leisure	22	15 (68.2%)	23	5 (21.7%)
Casual-Leisure	39	21 (53.8%)	11	3 (27.2%)

Table 7 above shows that the number of ‘casual-leisure’ sessions defined as long was as 3 times as that of those ‘defined short’ and that 57.9% of ‘work’, 68.2% of ‘project-leisure’, and 53.8% of ‘casual-leisure’ sessions defined as long were over-estimated by users with lower levels of under-estimation occurring. This encouraged a further study on the feature of each kind of web activity to determine the main cause for an incorrectly perceived length.

4.3 Time of Day

Figure 3 shows that most the ‘information-gathering’, ‘finding’ and ‘housekeeping’ sessions seem to occur between 10:00 and 16:00 whilst more ‘browsing’, ‘email’, and ‘communication’ activities were done between 22:00 and 0:00, which was labelled “before bed time”. Additionally, there is a peak around 14:00, in which more ‘finding’ and ‘information-gathering’ happened rather than other kinds of sessions. Finally, at 23:00, general ‘browsing’ is most prevalent.

Figure 4 shows that most of the ‘serious-leisure’ sessions occurred between 18:00 and 22:00. Most of the ‘work’ activities happened between 11:00 and 18:00, which seems to fit in within a typical working day. In the time ‘before bed’,

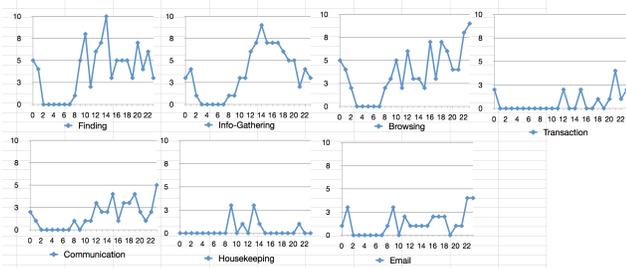


Figure 3: Time of Day, by Activity Goal

the most frequent activity is ‘casual-leisure’.

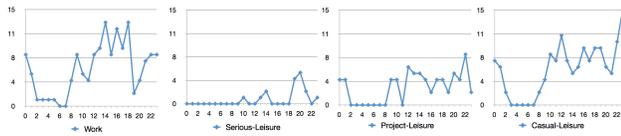


Figure 4: Time of Day, by Activity Context

Combined with the two comparisons above, there seems to be some overlap between ‘information-gathering’, ‘finding’, ‘housekeeping’ and ‘work’. There was also some overlap between ‘browsing’ and ‘casual-leisure’. Furthermore, these tend to suggest that there may be some patterns for user’s web activity in their daily life.

4.4 Search Queries

In Figure 5 below, sessions with more search queries tend to be classified as ‘defined long’, ‘long’, and ‘actual long’ than those with fewer queries. An interesting observation is that what the user defined as a long session features a relatively low average number of search queries compared with ‘long’ and ‘actual long’ sessions. Equally, sessions defined as ‘short’ by the user actually feature relatively more queries compared to ‘short’ and ‘actual short’. This may indicate that the user did not consider the number of queries performed when defining the duration of sessions and failed to realise the effect of this behaviour.

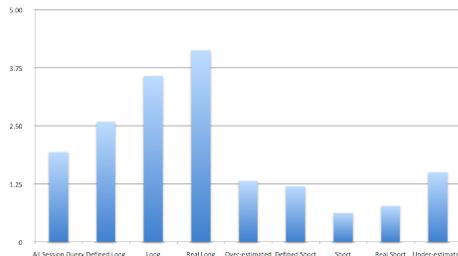


Figure 5: Average Number of Search Queries

5. CONCLUSIONS

Although this paper only describes a preliminary analysis of over 200 sessions from 11 participants, we have begun to see some potentially interesting early findings. Initially, participants varied greatly in their opinions about their own sessions, with some matching topical divisions, some temporal divisions, and some a combination of the two. The majority of participants judged “long sessions” as being longer than 5

minutes, but many had inaccurate recollections of the length of sessions. Long sessions were typically a mix of casual and serious leisure that often involved information gathering and browsing behaviour, while the majority of work related sessions were typically short. We also noticed that some of these activities may also be related to certain times of the day. All of the findings will be further explored after phase two of the study, but early insights suggest that real extended search sessions could be more accurately modelled based on additional factors such as: time of day, activity goal, activity context, and number of queries.

6. REFERENCES

- [1] P. Bailey, L. Chen, S. Grosenick, L. Jiang, Y. Li, P. Reinholdtsen, C. Salada, H. Wang, and S. Wong. User task understanding: a web search engine perspective. In *NII Shonan Meeting on Whole-Session Evaluation of Interactive Information Retrieval Systems*, Kanagawa, Japan, October 2012.
- [2] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the World-Wide web. *Computer Networks and ISDN Systems*, 27(6):1065–1073, 1995.
- [3] D. Elsweiler, M. L. Wilson, and B. K. Lunn. Understanding casual-leisure information behaviour. In A. Spink and J. Heinström, editors, *Library and Information Science*, pages 211–241. Emerald Group Publishing Limited, 2011.
- [4] D. He and A. Göker. Detecting session boundaries from Web user logs. *Methodology*, pages 57–66, 2000.
- [5] B. J. Jansen, A. Spink, C. Blakely, and S. Koshman. Defining a session on Web search engines. *JASIST*, 58(6):862–871, 2007.
- [6] N. Jhaveri and K.-J. Rähä. The advantages of a cross-session web workspace. In *CHI2005 Ext. Abstracts*, page 1949. ACM Press, 2005.
- [7] B. Mackay and C. Watters. Exploring Multi-session Web Tasks. *Time*, pages 1187–1196, 2008.
- [8] D. Nettleton, L. Calderon-benavides, and R. Baeza-yates. Baezayates, analysis of web search engine query sessions. In *Proc. WebKDD 2006*, 2006.
- [9] G. Rugg and P. McGeorge. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80–93, 1997.
- [10] A. J. Sellen, R. Murphy, and K. L. Shaw. How knowledge workers use the web. In *Proc. CHI2002*, pages 227–234. ACM Press.
- [11] A. Spink, M. Park, B. J. Jansen, and J. Pedersen. Multitasking during Web search sessions. *IP&M*, 42(1):264–275, 2006.
- [12] L. Tauscher and S. Greenberg. How people revisit web pages: empirical findings and implications for the design of history systems. *IJHCS*, 47(1):97–137, 1997.
- [13] E. G. Toms, R. Villa, and L. McCay-Peet. How is a search system used in work task completion? *Journal of Information Science*, 39(1):15–25, 2013.