A User Defined Taxonomy of Factors that Divide Online Information Retrieval Sessions

Chaoyu Ye
Mixed Reality Lab
University of Nottingham, UK
psxcy1@nottingham.ac.uk

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

ABSTRACT

Although research is increasingly interested in session-based retrieval, comparably little work has focused on how best to divide web histories into sessions. Most automated attempts to divide web histories into sessions have focused on dividing web logs using simplistic rules, including user identifiers and specific time gaps. This research, however, is focused on understanding the full range of factors that affect the division of sessions, so that we can begin to go beyond current naive techniques like fixed time periods of inactivity. To investigate these factors, 10,000 log items were manually analysed by their owners into 847 naturally occurring web sessions. During interviews, participants reviewed their own web histories to identify these sessions, and described the causes of divisions between sessions. This paper contributes a taxonomy of six factors that can be used to better model the divisions between sessions, along with initial insights into how the divided sessions manifested in web logs. The factors in our taxonomy provide focus for future work, including our own, for finding practical ways to more intelligently divide and identify sessions for improved session-based retrieval.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - *search process.*

Keywords

Sessions, Web History, Log Analysis, Qualitative

1. INTRODUCTION

Recent research has moved beyond trying to provide optimal results for a current or evolving set of queries, towards trying to model and support a "search session" [25]. Bailey et al, for example, identified a number of sessions that typically last longer than 5 minutes, including: adult, how-to, and entertainment sessions [2]. Most current approaches to detecting the start and end of sessions, however, have

used simplistic techniques, such as identifying users in search engine logs, and separating their activities by 25 minutes of inactivity [12]. While other papers have investigated alternative methods of identifying sessions in logs, such as modeling clear changes in the focus of queries [10], these papers have typically used an artificial corpus of uniform search sessions from TREC. Real sessions, however, are rarely uniform and human web behaviour is highly dynamic, and so this work focuses on understanding the full range of factors that relate to the boundaries of sessions.

Much research has shown that people interleave many different activities within single web episodes [28, 29, 15, 32], such as email, social networking, and information gathering. Further, research has shown that users also spread more notable tasks, such as vacation planning, research, and complex purchasing, across multiple sessions [13, 19, 16]. Despite aims to support these multi-session tasks, systems have struggled with multitasking [19] or have been retired [7]. Consequently, this research has sought to build a richer understanding of the factors that cause sessions to start and stop, by analysing 847 real web sessions, self identified by their owners in their own terms. In particular, our research questions were:

RQ1) What factors affect the end of a session?
RQ2) What factors relate to the start of new sessions?
RQ3) What factors divide apparently single sessions?
RQ4) What factors join two seemingly separate sessions?

To better understand the boundaries of search sessions against the sessions they occured between, the study investigated all web sessions, including non-search sessions, from personal browser histories. We define "web sessions" as sessions of general web history from participants, "search sessions" as those web sessions that involve web search queries, and "browse sessions" as web sessions without web search.

The following sections first present an overview of how sessions have previously been determined, analysed, and supported. Our interview study is then described in Section 3, and the results are presented as a taxonomy of six factors of session boundaries in Section 4. We conclude with a discussion about better modeling web session boundaries, regardless of their temporal relation to each other.

2. RELATED WORK

The notion of sessions started in the form of query sequences represented in search systems. Early work on DIALOG [31], for example, kept track of a searcher's queries and allowed them to reuse them by reference. Such systems were about supporting longer tasks within specific collections of

documents, rather than web search, which is an aim still held by recent research (e.g. [24]) to support extended episodes of Exploratory Search [34] and sensemaking [27]. Investigations into *web* sessions, however, can be dated back to the mid 90s (e.g. [5]). Despite this history, there is increasing focus on web sessions, where search engines are keen to better support searchers who continue to search for more than a few queries or minutes [35, 33, 2]. Queries can be disambiguated, for example, given a user's query history, but more specifically against current queries if the bounds of the current session are known. Ozmutlu (2006) found about 28% of queries were reformulations of previous queries [23], while Jansen et al (2007) reported that about 37% of search queries were reformulations when repeated queries were not considered [12]. Similarly, query analysis in user experiments has also found that users are more likely to submit reformulations in more complex search tasks [18]. Despite these ideas, we still know very little about what constitutes a session, nor how to determine the start and end amongst the highly dynamic behaviours we exhibit online [29, 28].

2.1 Determining sessions

A number of researchers have generated definitions of a session using different delimiters such as cutoff time, query context, or even the status of the browser windows (e.g. [19]). In 1995, Catledge and Pitkow suggested a 25.5 minute "time-out", the time between two adjacent activities, was best to divide logs into sessions [5]. Although their research was focused on identifying contiguous periods of general web activity, rather than homogenious search sessions, 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 [11]. Their analysis found that optimal timeout values vary between 10 and 15 minutes. Spink et al [29] 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 "one interaction" was determined as a contiguous period. Going beyond simplistic time divisions, google defines a session boundary based on three issues: 1) 30 minutes interval, 2) end of a day, and 3) traffic source value change [6].

To summarise the different approaches used to define sessions, Jansen et al. provided a summary of the three most representative strategies [12]. As IP and cookies were utilised to identify a user, the most frequent strategies involve temporal cutoffs and topic change. Other surveys of session boundary detection methods have been provided by Wolfram [36]. Gayo-Avello [9] provided a comprehensive summary of previous search session detection methods involving both temporal and lexical clues based on query logs, however it only focused on the "consecutive" search activity without considering interleaving.

2.2 Understanding sessions

Taking a user-focused approach, Sellen et al [28] investigated the different activities that people perform online, including information gathering, browsing, transacting, communicating, and housekeeping. Many others have tried to categorise the types of activities, and thus perhaps sessions, that people engage in online (e.g. [15, 32]). Broder divided web search behaviour into three main categories: navigational, informational, and transactional [3]. Although these types of taxonomies help us to understand the types of things people do online, they do not practically help search engines to identify and support real web sessions, because they are highly interleaved and dynamic in nature. Consequently, researchers resort to the techniques described above to divide search engine logs and investigate them.

Understanding the nature of longer sessions, however, can help provide results relevant to the current session. In analysing Bing logs, Bailey at al identified several key examples of sessions that typically lasted more than a few minutes, or involved more than a short sequence of queries [2]. Their analysis showed, for example, the nature of adult search sessions, and other long sessions types including: researching how to do something, and finding pictures or watching entertaining videos. Elsweiler et al [8] also investigated these latter casual-leisure sessions, highlighting a) their tendency to be long, b) that participants continue to search despite already finding good results, and c) that participants typically stop when they cannot find good results. Further, Kotov et al analysed multi-session search tasks [16]. Such findings highlight the importance of providing relevant results for a whole session [25].

2.3 Supporting sessions

While research contiues trying to identify and determine sessions, researchers have used the available techniques to collate examples of sessions and find ways to better support them. The aim of the TREC session track, for example, is to improve retrieval accuracy over an entire session [14], rather than optimising for one query at a time, by taking into account recent query history and other logged behaviours. To do this, a series of real sessions were extracted from search engine logs, however they were identified using similar timeout techniques described above, and are typically homogenous in topic or style. Such corpuses of sessions, however, have allowed researchers to determine how to use query change to find possible session boundaries [10]. Conversely, Adeyanju et al [1] aimed to determine which pages people typically end up in during sessions. By identifying the likely session motivating the query, they can try to return key results earlier in the search, despite not being relevant to the earlier queries. Similarly, Raman et al [25] identified patterns for "how-to" searches, and aimed to return results that matched the likely phases of the sessions.

Research has also produced systems that try to support searchers during their sessions. For a while, Yahoo! developed SearchPad, which provided searchers with a note-taking facility for use during longer sessions [7]. Work by Mackay and Watters aimed to support people in tasks that span multiple sessions, by allowing them to explicitly specify their current sessions in a tool bar [19]. Alternative approaches have tried to break web history into sessions in order to make them easier to review. SearchBar, for example, let people manipulate their search histories as being related to certain topics or sessions [20]. Many other browser extensions exist for sessions management and alternative views of web history.

The research above reinforces that real web sessions are highly dynamic and that using notions of time gaps in search engine logs are likely to be too simplistic for automatic detection of session boundaries. Our research, therefore, is focused on understanding how real human web sessions, and their boundaries, relate to each other, and what factors must be considered to (semi-)automatically identify them.

3. EXPERIMENT DESIGN

To understand and characterise real web sessions, we employed similar interview methods to Sellen et al. [28]. 20 participants engaged in a 90-120 minute interview about their own web histories. To ground the interviews in real data, participants focused on printouts of their own web history, and we used the card sorting technique [26] to probe their mental models of sessions. Although these methods do not allow us to analyse web sessions at a large scale, they are conducive with building a better, richer understanding of web sessions and their boundaries, and so we can focus on insights rather than scale. The procedure was approved by our school's ethics committee and pilot tested.

3.1 Procedure

Preparation. Participants began by providing their web history and were advised to edit it in advance should they wish to keep some logged activities private[1]. These logs were gathered by importing their web histories into Firefox (if not already there), and creating an XML export using "History Export 0.4"[2]. This log was structured and pre-processed using a) automatic detection of search URLs, and b) manual identification of periods of interest to discuss in the interview.

Examining History Logs. After providing demographic information, participants spent around 20 minutes examining the structured printout of their history, using a pen to mark out "sessions". During the study, the term "session" was left to be ambiguous as possible in order to avoid influencing their mental modal. The only precaution taken was to make sure participants did not simply categorise entries rather than identify sessions, i.e. simply classifying all social networking entries into one large 'social networking' session that spanned their entire log. Participants comprehensively identified sessions from the most recent 500 entries in their web history, which varied between 2 to 5 days of history, depending on the individual. Consquently, 10,000 history items were manually analysed into 847 sessions for later analysis. All participants also put around 10 sessions onto individual cards, unless single queries or similar in nature to previously carded sessions, for later sorting. Each card had a number, a title, activity purpose, included history items, and whether it has been completed successfully or not.

Interview. After participants marked their own web histories, the interview began by discussing participants' session boundaries. Participants were given the chance to review each session boundary, however this discussion typically focused on unclear boundaries or sessions that the participants or researchers found interesting or worth discussing. This phase provided three benefits: 1) allowing the participants to review and revise session demarkations, 2) allowing the researchers to begin to understand the ways that participants understood sessions, and 3) supporting the participants to begin producing criteria for the subsequent card sorting.

Card Sorting. The remainder of the interview involved first open, and then closed repeated single-criteria card sorting [26]. 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: duration, difficulty, importance,

---

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

[2]addons.mozilla.org/en-us/firefox/addon/history-export/

and frequency. The interviews were audio recorded, and physical copies of the card sorts were kept for analysis.

3.2 Participants

The 20 participants were recruited broadly from across a university in the United Kingdom, including students and staff from both non-technical and technical backgrounds. 9 were male and 11 female; all were aged between 18 and 30. 18 out of 20 said they search online everyday, while the remaining two participants indicated they search online every 3 to 5 days. Participants were given £15 remuneration for the time they gave to the study.

3.3 Analysis

Three types of data were collected and analysed during the study: logs, interview data, and card sorts.

Quantitative Analysis. We were able to produce summative data about 847 sessions, such as average size of sessions, temporal gaps between sessions, number of queries, and so on. Some dimensions of card sorts were also summarised, and used to summatively analyse the carded sessions.

Qualitative Analysis. Interview data was transcribed from the audio recordings and was analysed using an open inductive form of Grounded Theory [30]. Initially, one interview was coded using open coding by two researchers, such that process and focus of the coding could be discussed and compared. After discussing and reaching agreement on the focus of the coding process, the remainder of the interviews were analysed using open, axial, and selective coding, which were reflected upon at multiple stages as coding progressed. Codes were collected, given definitions, and associated with sample pieces of text, and then considered collectively. According to the Grounded Theory process, these codes were assessed in order to produce categories and then themes within the data. Disagreements were discussed carefully and codes were merged or divided as their definitions, and the definitions of the categories and themes, developed. A final taxonomy of the factors involved in differentiating between sessions is presented in the results. To assess the stability and reliability of the taxonomy, a copy was provided to an independent researcher, alongside a sample of 58 quotations from the text. The indepenent researcher firstly spent ten minutes reading and discussing the taxonomy, until they felt comfortable with each part. The independent researcher then categorised the 58 samples according to the taxonomy, which was compared to the categorisations chosen by one of the primary researchers. A Cohen's Kappa score of 0.796 was reached between the independent and primary researcher, which is considered as *Substantial Agreement* [17].

4. RESULTS

The 20 participants identified an average 42.35 sessions each, creating a total of 847, which are summarised in Table 1. Sessions involved an average of 13.3 mins of active web behaviour (not including gaps), with a standard deviation of 31.25, but 33.9% of them lasted for less than 1 min and 76.9% of them lasted for less than the average length. 5.3% of them lasted for more than 60 mins, where the longest recorded session of web activity (excluding breaks) lasted for 303 mins. Sessions included an average of 10.1 history log entries, which we take loosely to be pageviews - although some dynamic updates may not have been captured. We divided these sessions into three sets: short - less than 15

Table 1: Analysis of sessions recorded

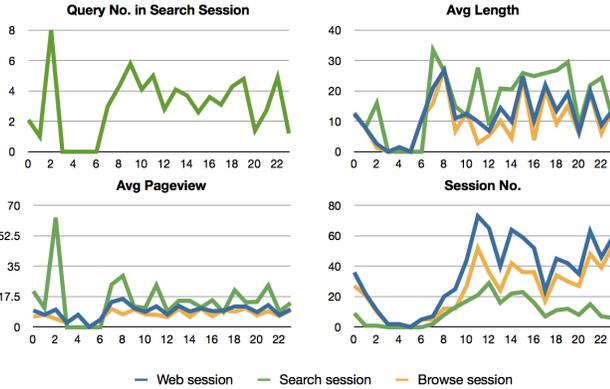|	All	Short	Medium	Long
Number	847	658	144	45
avg. Length	13.3	2.6	29.6	118.1
avg. Pageviews	10.1	5.9	21.6	34.7
Search Sessions				
Number	246	158	65	23
avg. Length	20.0	4.1	28.8	104.1
avg. Queries	3.6	2.3	3.6	12.4
avg. Pageviews	15.6	6.8	26.0	46.4
Browse Sessions				
Number	601	500	79	22
avg. Length	10.6	2.1	30.3	132.7
avg. Pageviews	7.8	5.5	18	22.5

Figure 1: Time of the Day Analysis

mins, medium - between 15 mins and 60 mins (inclusive) and long - more than 60 mins; these are the median numbers in the duration definition of session given by participants.

601 of our sessions did not include a search query, which we call *browse sessions*. They were shorter than the average length of all web sessions and the vast majority (83.2%) were short, indicating a large proportion of short navigational episodes in our dataset. Notably, 3.7% of the browse sessions lasted longer than 60 mins, without a single query.

246 of the sessions involved search URLs and those were longer than the average of all sessions; we call these *search sessions*. Although it seems hard to have a long session without a query, the average number of queries for long sessions was 12.4, at around one for every 9 mins, as the average length was 104.1 mins. The longest search session lasted for 246 mins (4.1 hours), but had only 15 queries, which was one every 18.9 mins. Conversely the session with the largest number of queries, 42, lasted for only 190 mins (3.2 hours), which was one query every 4.5 mins. Only 21 sessions had more than one query per minute, and 19 of these we classed as short; none were classed as long.

The time of the day for each session was also studied. As shown in the Figure 1, between 1-3am, people had more pageviews and queries than other times in search sessions. However, the duration of each single search sessions was lower than the average 12.0 mins. Therefore, the search sessions "before bed" involved more queries and more pageviews but took shorter time. The longer search sessions always happened in the morning between 8-9am, which also involved 6 queries per search session. The number of pages viewed in the morning was much lower than late at night. Participants spent longer viewing pages in the mornings, which Nettleton et al [21] said is indicative of 'good quality' search.

Many of our participants took breaks during sessions. 77 sessions involved inner-breaks longer than 10 mins, with an average length of 288.3 mins. 62 had inner-breaks longer

than 1 hour and 3 of them even had day-break that was longer than 24 hours. Further, between sessions, 456 involved breaks of inactivity of less than 10 mins, leaving 378 that included more notable breaks. 302 of those had breaks lasting for more than 26 mins, indicating that simple divisions of logs, using 25.5 mins as proposed by papers like Catledge and Pitkow [5], would have only divided 35% of our sessions. In addition, more than 30% of our sessions involved discontinuous activities, with interruptions from other web activities or real-life (e.g. cooking), indicating that session identification is not only important for consecutive activities but also interleaving activities.

4.1 Understanding Session Boundaries

Our qualitative analysis identified 6 key factors that are involved in determining different sessions: Topic, Task, Phase, Group of People involved in the activity, Time gap, and Multitasking. Table 2 summarises these key factors, with detail about when they cause a change in session, when there is an exception, and when they override other factors. The Topic, Task, and Phase refer to the lexical clues about the activities grouped into sessions. The differences among them can be presented as: 1) Topic refers to the broad aim of a series of activities, which may consist of one or more specific tasks/phases; 2) Tasks related to one topic are content-relevant, such as the different *(task)* questions search about *(topic)* "Java Programming", but not neccessarily sequential; 3) Phases related to one topic are sequential, e.g. booking a flight ticket *(topic)* may firstly browse cross sites *(phase 1 - info gathering)* to get the info before making the final decision and doing the transaction *(phase 2 - transaction)*.

These 6 factors are interrelated, and they represent common themes discussed by participants, rather than rules that can directly applied. It is not, however, a matter of *how often* each factor applies, but instead *how much* each factor applies at different times.

4.1.1 Topic Change

The main topic was found to be one of the primary session delimiters in this study. Topic typically refers to the main idea of a user's intention, or their higher-level Work Task [4]. It may consist of one or multiple specific Tasks or Phases.

Creates a Session Boundary. Most participants discussed the topic of their work when marking boundaries in their web log. P14, in Table 2, said: "*session 7 is about online shopping, and session 8 is related to my academic study, they are topical difference*".

Exception. However, if the topic is too trivial to be identified as a single session, topic-change may fail in causing session change and participants may just group trivial activities from different topics into one session instead: P8 said "*I grouped all of these [free-browse online shopping, social network] into one session because they are just free browsing, I don't have any particular purpose, just to relax.*" In addition, if the topic is broad and its tasks are easily dividable, the tasks may be put into separate sessions rather than grouped together; described further in the "Task" section below.

Overrides another Factor. Sometimes, topic may become dominant and override other factors. P14 said:"*all of info search in trip plan to europe before 1st August should be put into one session, including accommodation, ticket, and places of interests searching.*" In this case, the session was expanded through days and the large time gap was not

Table 2: Dimensions that affect session boundaries, with common exceptions and overides.

	Factor	Creates a Session Boundary	Exception	Overrides another Factor
Content-Relevance	**Topic change**: Topic refers to the user's main intention, or higher-level work task, and may consist of one or more tasks or phases.	When Topic-change led to session change. Users may start a new session when the topic shifts. • P14: "*session 7 is about online shopping, and session 8 is related to my academic study, they are topical difference.*"	Topic-change may not lead to session change when the topic is too trivial or too broad. • P8: "*I grouped all of these [free-browse online shopping, social network] into one session because they are just free browsing, I don't have any particular purpose, just to relax.*"	Topic may override factors and join seemingly separate sessions. • When it overrides timeout and task-change: P14: "*all of information search in trip plan to europe before 1st of August should put into one session, including the accommodation, ticket , and places of interests searching.*"
	Task change: One topic may span several specific tasks, e.g. corresponding to distinct specific questions related to a big topic.	When task-change led to session change. A big topic-based web acitivity may have mutliple specific tasks, which are relevant but different to each other. • P15: "*all of the specific problems searching related to the topic "Matlab" are put into separate sessions.*"	Task-change may not lead to session change when the task is closely integrated or too small. • P17: "*these [topical-related] tasks are for different questions, but I want to group them together because some of them are just quick search and have only one query.*"	When the task or phase overrides other factors and bridge seemingly separate session together. • When it override time gap: P16: "*when I did some search yesterday and continue doing some more search on the same thing, I will put them into one session. Even if they have longer time gap.*" • When it override topic: the Matlab example from P15.
	Different phases: One topic may be made up of phases, they are more sequentially dependent when compared with the "Task".	When phase-shifts in one topic leads to session change. In a topic-based web activity, it may be identified as mutliple phases. • P4: "*In the flight ticket booking, Looking for information and final purchase are two different phases, because from checking price to purchase, I need a decision making and it takes time.*"	Phase-change may not lead to session change when phases may be too small to have a separate session. • P1: "*Searches on 'Burn a DVD' has two parts: 'How to burn it' and 'a software resource searching', and I put them into one session, because they are relevant and I didn't spread many sites.*"	
Social	**Different People** The group of people involved in the activity. e.g. different collaborators or clients for different projects.	When group of people involved in the activity is changed, it can indicate a session change • P11: "*The gmail and uni emil should be put into different session, because I use the uni one to contact with my classmates and colleagues, and use the gmail for friends and family.*"	Some participants grouped all of adjacent activities across different social networks together. • P3: "*so while waiting, I will either to check my personal emails, or because I use google chat a lot, or facebook, chat with my uni friends, and all of these should be put into one session, because they are just a break for me*"	When people override other factors and bridge seemingly separate sessions. • When it override topics, task, time: P6: "*I put all of the web activities from the same mailbox into one session, because the people I contacted with via the same mailbox are from same group.*"
Behaviour	**Time gap**: The time gap between web activity, as is traditionally the main technique used to divide session.	When the time gap is big enough to lead to session change, depending on other factors, such as task size and type of interruption. • P6: "*For the video, the acceptable time interval is less than 45mins. and for facebook, probably 1-2mins, and in academic search is less than 1h*" • P15: "*I put these two activities on one specific questions into one session, even they have more than 2 hours gap and it exceeds my acceptable time gap, but I knew it is interrupted by lunch.*"	When time gap is not considered as a factor in session division, especially for bigger, more Important activities. • P10: "*I don't mind the time, because they are for the same purpose, it is the same duty. So I put them together.*" • P14: "*because some of my information search may spread over days, for example, the information gathering on schengen visa takes me about three days, I will put all of them into one session even with days break.*"	When the time gap override other factors and bridge seemingly separate session together, such as the comments from P6 in the "Multi-tasking" factor.
	Multitasking: Sometimes, users may do multiple things concurrently.	Enough characteristically-diverse behaviour creates a session of "diverse activity", or a multi-tasking session. a session of un-connected web behaviour. • P6: "*I may feel borded when doing some task, so I probably stop and then go through my facebook, emails, and or stream to have a break. I will put all of these during that period into one session - break session.*"	When the scale of interleaving activities are not trivial and they can be easily dividable. • P19 said "*my initial aim is to do academic info search then I switched to browse property info, and go back to academic again after a while. The property viewing in the middle should be put into a different session.*"	N/A

considered as important as the connectivity of the larger topic of trip planning. Further, P14 bridged different specific tasks: accommodation, ticket booking, and places of interests searching into one session because of the one topic - "trip plan", which overrides the Tasks - "accommodation" and "ticket booking" below.

4.1.2 Task Change

Participants often divided periods of activity into different tasks, where descriptions indicated that this was when these were more easily dividable, or larger in size.

Creates a Session Boundary. Specific tasks can be used to divide web activities into separate groups, even if related to a big topic. The tasks shifting may create a session boundary. In a big topic-based web activity, there may be multiple specific tasks, which are topically relevant but different to each other. For example, the "Matlab" example from P15 above. The tasks like questions searching on "what

does error XXX mean in Matlab" and "how to declare a variable in matlab" are both related to "Matlab" but grouped into different sessions, because he thought they were two different "how-to" tasks.

Exception. When the scale of the tasks for one topic are relevantly small, participants were less likely to divide tasks as sessions, but as complementary or supporting "missions" to the main Topic. A comment from P17 described that several difference but relevant tasks within single search query should be put into one session, because he thought a session with single item was meaningless.

Overrides another Factor. Task is clearly related to main topic in some form, and so projects that try to model common tasks in sessions would help to determine thresholds and task detection. Task may override topic when they are easily dividable such as specific tasks in the "Matlab" example above, and it overwrites the rule of "putting topic-related activities into one session". Task also has association with specific collaborators and time impacts, especially as they grow to the size of smaller topic sessions. Similarly, larger task sessions can also begin to tolerate brief divergent web activity or temporal gaps. P16 decided to group the continuous searching on one technical problem solving accross multiple days into one session, despite spanning overnight breaks, because the task was unchanged. The challenge in delineating between small but similar tasks, means that when trying to model human web sessions, systems may need to retrospectively consider relative thresholds before deciding if they were in the same session.

4.1.3   Different Phases

Some types of activities have clear phases [25], for which progression can be predictable. There may be multiple phases related to one topic, for example. Compared with "Tasks", they are more sequentially dependent with each other. Our participants also reported this behavior, adding weight to the idea of whole-session relevance. These phases can be hierarchical or sequential, and participants noted that one phase may affect the activity in another one.

Creates a Session Boundary. One common example of this type was participants dividing periods of research and option comparison, as a separate session to then finding the best place to buy a product and then purchasing it. P4 said: *"In the flight ticket booking, Looking for information and final purchase are two different phases, because from checking price to purchase, I need a decision making and it takes time."* Another described a two contiguous activities involving banking and bill paying as separate phases and thus separate sessions.

Exception. Not all phase-shifting lead to session changes. Like with Task, some participants said that although there were clear phases in the process, they were too small to be considered as separate sessions, as P1 said:*"Searches on 'Burn a DVD' has two parts: 'How to burn it' and also 'a software resource searching', and I put them into one session, because I think they are relevant and I didn't spread many sites."* The phases in this are easily identifiable, however, P1 thought the size of each phase was not big enough to warrant a separate session. P1 also said that if the downloading of software had involved learning and researching, that they would have become separatable phases, highlighting the importance of size and delineatable aims for phases as well as tasks.

Overrides another Factor. It is feasible that phases are simply a sequential instance of tasks, but this was not easy to determine from the qualitative data collected. The finding does have implications for projects looking at supporting sessions with phases [25], which if grow to contain phases across separate sessions would have to adapt.

4.1.4   People

The group of people involved in the activity was also a common theme in the interviews, although heavily related to others like task. It is mainly applied in the online communication, e.g. the group of people a user communicates with via their email or social network. Related people could help identify a topic/task, but other contiguous periods of web activities were divided simply by the collaborators alone.

Creates a Session Boundary. 70% participants preferred to put activities from different mailboxes and social networks into different sessions as they utilised them for contacting different groups of people. P11 described how contiguous use of email could be divided by people involved:*"The gmail and uni emil should be put into different sessions, because I use the uni one to contact with my classmates and colleagues, and gmail for friends and family."*

Exception. A small number of participants grouped all of the activities across different social networks into one session when they were adjacent to each other. For example, P3 said: *"so while waiting, I will [...] either to check my personal emails, or because I use google chat a lot, or facebook, chat with my uni friends, and all of these should be put into one session, because they are just a 'break' for me"*.

Overrides another Factor. Sometime, the group of people may override other factors, such as talking to specific people about multiple topics or tasks. P6 grouped all of the web acitivities from one mailbox even within a big time gap or activity-differentiation into one session as P6 said: *"I put all of the web activities from the same mailbox into one session, because I think the people I contacted with via the same mailbox are from same group."*

4.1.5   Time Gap

As with most research in this area, Time gap has clearly been associated with methods to divide sessions. Large time gaps were repeatedly mentioned as separating sessions in our interviews, but the findings most notably highlight that they vary dramatically according to context. In addition, the type of web activity can affect the tolerance of temporal gaps.

Creates a Session Boundary. Temporal gaps between activities were a common cause of separating sessions, whereas topics and tasks were frequently cited as anchoring sessions over a time gap. Large gaps, such as overnight breaks, usually divided sessions, P5 said:*"I did academic search about the "Learning enviroment" in different days, and I put them into seperated sessions."*. Further, the acceptable time gap varies from types of web activity and the length of invested time, and some people even suggested non-web-activity gaps from a real life interruption may need to be as long as a few hours to divide a session. P6 said:*"For video, the acceptable time interval is less than 45mins. For facebook, probably 1-2mins, and in academic search is less than 1h"*, and P15 said:*"I put these two activities on one specific questions into one session, even they have more than 2hs gap and exceeds my acceptable time gap, but I knew it is interrupted by lunch."*

Exception. In relation to other factors above, time did not divide temporally distant web activity, when another factor became the overriding one. P14 highlighted how a topic can tie over a large period of activity: *"[...] because*

some of my information search may spread over days, for example, the information gathering on schengen visa takes me about three days, I will put all of them into one session even with days break."

Overrides another Factor. The time gap may override other factors and bridge some separate sessions together, even when they are unrelated to each other, such as the "break period activities" from P6 above and he grouped all of the unrelated casual activities happened in the break period into one session because of the short time gap and trivial tasks. As a result, scaling the acceptable size of break in accordance with the size of the session determined so far might be a better way to model inactivity periods as a factor.

4.1.6 Multitasking

Participants frequently described activity in their logs as being caused by multi-tasking. Multi-tasking often accounted for divergent behaviour amongst larger sessions, however participants also entered *states* of multitasking.

Creates a Session Boundary. Enough characteristically-diverse behaviour creates a session of "diverse activity", or a multi-tasking session; a session of unconnected web behaviour. P6 said he may also check his facebook and email simultaneously to have a break, during the serious working period. In this case, he preferred to put the break activity inside of the working session. The model of this situation is similar with "one mainstream activity with some other trivial activities".

Exception. When the scale of interleaving activities are not trivial and they can be easily dividable, e.g. P19 said "*My initial aim is to do some academic information search, then I switched to browse property information, and go back to academic again after a while. The property viewing in the middle should be put into a different session.*" The model of this situation seems to be "two or more mainstream activities interleave with each other", and the topical difference causes the session division.

Overrides another Factor. To handle multi-tasking, during other sessions (created by main Topic), some approaches have simply ignored them and focused on things that match the current topical focus of the session (e.g. [10]). People may multi-task in natural breaks, like between Parallel tasks. These approaches to avoiding session changes seem relevant, but for sessions that are *identified* for multi-tasking it would be important to learn to model them to avoid unwanted incorrect support. The multi-tasking factor perhaps best highlights the risks for supporting sessions.

These comments and findings highlight that these 6 interrelated factors have effects on determining a user's session boundaries in different situations. The first three mainly focused on the content of the web activities and make the decison based on the lexical relavance. Telling the scale difference between them, however, is still a big challenge for deciding when tasks should become separate phases. The People factor is added mainly for the activities involving other people, such as the online communication via email or social networks, but because People can be closely related to different work activities, it can become a good indicator of web content. Time gap is a temporal technique applied in most existing research, but we find that a "fixed time gap rule" may be insufficient without modeling the size of sessions that precede them. Scale varies dramatically according to the feature of the activities themselves and also individual preferences. The final factor, Multitasking, reflects that

human behaviour is also related to the session division.

4.2 Understanding Sessions with Card Sorts

To understand what people thought about different types of sessions, we first asked participants to sort their cards, or sessions, according to their own criteria. Table 3 shows the range of critieria chosen to use by participants in open card sorting. Nearly every participant began by using purpose to divide their sessions, creating groups like: work, entertainment, and social networking. From the open card sorting, we received some unexpected dimensions that differentiated sessions, such as Willingness to do the activity, and whether sessions involved refinding via bookmarks. Interestingness was an unexpected but commonly used dimension. Although interestingness was defined differently to willingness, the separation of sessions was similar. Although difficult to utilise directly, these different dimensions may help us to investigate other factors of session boundaries in the future.

Table 3: Dimensions in Open-Card Sorting

Purpose (19)	Willingness (1)
Language (2)	Academic or Not (2)
Text VS. Audio+Video (1)	Concentration (1)
Personal Use VS. Sharing (1)	Device (1)
Interestingness (5)	Region (Physical) (1)
Money related or not (1)	People involved (1)
Free VS. Paid (1)	Familiarity (1)
Long VS. Short Term (2)	Usefulness (1)
Google it or Not (1)	Successful or Not (2)
Search Engine Differ (1)	Via Bookmark or Not (1)

During closed card sorting, if not already used in the open process, we asked participants to divide their sessions according to the following 4 critiera as shown in Table 4: Importance, Frequency, Difficulty, and Perceived Length, in order to obtain the relation between people's perception on the session scale and those four dimensions. Further, we noticed that some classified sessions were longer or shorter than objective measurements.

4.2.1 Importance, Frequency, and Difficulty

Importance. Search sessions in the High Importance group were longer and had notably more queries but fewer pageviews than other groups. This indicates that the query number and the length of single pageviewing time may be an indicator of search session importance. Conversely, browse sessions in the High Importance group had many more pageviews than other groups, indicating that the pageviews is related to the importance of browse sessions.

Frequency. Search sessions in the Low Frequency group had more queries, indicating that frequent searches had fewer queries. 35 out of 46 sessions in High Frequency group were browse sessions with longer than average length, perhaps because of some daily casual-leisure sessions like video streaming.

Difficulty. The majority of High and Medium Difficulty groups were search sessions, which implicitly indicated that query input may lead to higher difficulty. The search session in the High Difficulty group lasted much longer, and had more queries and pageviews. However, longer length and more pageview,s in browse sessions, did not often lead to higher Difficulty, compared to the Browse sessions in the Medium Difficult group, which lasted longest and had far more pageviews.

As a result, search session with more queries input may lead to higher importance and difficulty, but happen less

Table 4: Analysis of 146 card sorted sessions, by importance, frequency, difficulty, and perceived length.

	Importance			Frequency			Difficulty			Perceived Length			
	Low	Med	High	Low	Med	High	Low	Med	High	PS	AS	AL	PL
Number	67	31	48	75	25	46	71	35	40	8	52	6	45
Avg. Length	15.7	18.6	17.3	15.8	24.6	14.2	9.7	22.3	24.8	32.3	4.2	73.3	18.9
Avg. Pageviews	19.2	15.5	16.9	17.7	20.7	15.9	12.0	22.5	23.4	15.0	8.6	56.5	23.7
Search Sessions													
Number	37	24	23	53	20	11	23	30	31	6	27	4	26
Avg. Length	19.1	22.0	22.1	19.2	28.9	13.1	9.3	22.8	27.1	31.5	4.9	72.8	21.2
Avg. Queries	2.8	2.3	4.9	3.7	2.5	2.2	1.8	2.6	4.8	1.5	1.8	13.3	3.7
Avg. Pageviews	25.5	16.2	15.5	19.7	22.2	18.3	11.1	19.6	27.2	14.0	7.9	52.3	28.4
Browse Sessions													
Number	30	7	25	22	5	35	48	5	9	2	25	2	19
Avg. Length	11.5	7.1	12.8	7.7	7.6	14.5	9.8	19.2	16.6	34.5	3.4	74.5	15.7
Avg. Pageviews	11.5	13.1	18.2	12.9	14.8	15.2	12.5	39.8	10.2	18.0	9.3	65.0	17.2

frequently. Browse sessions with more pageviews may lead to higher importance, and their length does not seem to have notable differences.

4.2.2 Perceived Length

We built two main categories to analyse perceived length: sessions that were actually the type that they specified (Actual Short (AS) and Actual Long (AL)), and categories that were perceived to be long or short when objectively not in those categories (Perceived Short (PS) and Perceived Long (PL)), as shown in Table 4. Participants were more likely to over-estimate (45 PL) rather than under-estimate (8 PS) the session length.

PS and AS. The sessions in PS were perceived as Short but their actual length were located in either the Medium or Long groups. First, their average length and pageviews were much higher than the sessions in the AS group. However, the query number in PS was lower than in AS. This indicates that query number in Search Sessions may affect the perceived length and lower query numbers may lead to under-estimation.

PL and AL. The sessions in PL were perceived as Long but their actual length were located in either Medium or Short Length group. The sessions in AL laster much longer and had more pageviews, and had many more queries if it was a search session. There were no clear indicators for the reason causing the over-estimation of shorter sessions.

PS and PL. In the search session comparison between these two, the query number and pageviews in PL were more than twice of these in PS, which indicates that query number and pageview may have effects on the length perception, and more queries and pageview may lead to over-estimation.

4.3 Combining our Results

Our main taxonomy presents the factors that were associated with the boundaries of sessions, which indicate that several factors may be relevant depending on the scale of the sessions being divided. Further, using objective analyses of these sessions, the Importance, Difficulty, and Frequency analysis indicated the query number, pageviews, and length may have some effects on the *perception* of activity scale. The "over-estimation" and "under-estimation" in the length of sessions, however, indicates that we should estimate perceived scale of session, combining time and activity as indicators of importance and difficulty, rather than objectively measure them directly. There are several insights that can be drawn from the taxonomy and session features. When thinking about the appropriate factors for different situations, the features of the sessions play an important role. For example, the tolerance of time gap for dividing sessions varies from

types of web activity, and it was much higher for bigger and more important activities. In Search Sessions, we found a relationship between the number of queries and Importance, where more queries may lead to higher Importance and probably over-estimation on length. In Browse Sessions, more pageviews lead to higher Importance. It seems that search sessions with more queries, or browse sessions with more pageviews, should have higher tolerance for time gaps. In addition, from the study on time of the day in Figure 1, the search sessions that happened "before bed" typically had more queries, and may also lead to the higher tolerance of time gap. Similarly, number of queries was also a good indicator of frequent search sessions, where High Frequency sessions had fewer queries and Low Frequency sessions had more queries. This may help to identify routine activities that can be more easily bounded as common sessions.

5. DISCUSSION - APPLYING THE TAXONOMY

The results above have provided three perspectives on how to determine the start and end of web sessions: 1) a taxonomy of factors that relate to the boundaries of sessions, along with notable exceptions and overrides, 2) insights into how those sessions manifested in web logs, and 3) how users perceived and categorised these sessions. Core to our contribution is that these factors have not been determined by researchers, but elicited from the users who created them. In relation to the four RQs set out in the Introduction, we discovered that the triggers cause that start, end, divide, and join session are dependent on 6 inter-related factors. The priority of each factor involved in different session boundaries needs further study, as it is highly related to the scale of web activities themselves. Determining the scale of the activity could be one of the more challenging parts in future work. Although, for example, common triggers like large time gaps are typically considered to divide sessions, we saw sessions spanning overnight periods, if the nature of the work task was large or important enough. Secion 4.3 also presented some insights into, for example, what make a session important; search sessions seem to be considered important if they included more queries according to the data in Table 4, and more page views were indicative of important browse sessions.

As the factors in our taxonomy were drawn from qualitative methods, and are abstract themes that each relate to the boundaries discussed by our participants, the subsequent challenge is to put the factors from our taxonomy into practice. It is not the case that simple rules can be derived from our six factors, as each factor may play some amount of influence on a session ending. Consequently, the challenge for

applying the taxonomy is to learn how to measure and track each of these factors, and then to discover how their thresholds, in combination, create a boundary. This challenge is both the reason we cannot yet quantify the importance of each factor in our taxonomy, and thus the primary motivator of our on going and future work. Below, however, we provide an initial discussion of potentially applying our taxonomy.

5.1 Implications for Systems

As the taxonomy captures factors, rather than specific trigger events, putting our results into practice means not detecting events in specific factors, but monitoring each factor in combination. Time, for example, has been commonly modeled in research [5, 11], but our findings highlight that timeouts are closely related to other factors, such as the size or importance of a task; the analysis of card-sorted sessions provided insights into the nature of important search and browse sessions. Conversely, however, many sessions changed without a notable time gap, based upon topic, phase, or task change. A more intelligent time gap calculation is required, as the initial finding from our study is that acceptable time gaps varied by query numbers in search sessions, and pageviews in browse sessions. Recent work has also studied topic change as a means to detect the end of sessions [10, 23], but these only focused on consecutive search activity and did not consider any activity resumption and multitasking.

From our study, session boundaries usually occured when the activity was either completed or interrupted. Participants reported being interrupted by a number of triggers, including: 1) non-web demands, such as sleeping or cooking; 2) internal demands, such as feeling bored and needing a break (e.g. social network); and 3) interruption from concurrent tasks. Different factors should be considered for each. For example, in 1), longer time gaps could be accepted around normal meal times and over night. This may be especially true if the system has identified the current task as being especially large, and detects the following morning's web activity as being related. In 2), multitasking, time gap, and size of task should all be considered, as users often grouped lots of small diverse trivial tasks activities during a break into one session. The scale measurement of "Task", "Topic" and "Phase" is also challenging as it is highly dynamic and subjective, such as our user who was searching for matlab related content, but these spanned across separate tasks.

Detecting the change of people involved in the activity as a factor may be one of challenging parts of applying the taxonomy, especially if this information is not accessible for search providers. The most obvious examples in our dataset were from users who engaged with notably separate groups via work and then personal email, even though the nature of the web activity appeared to be similar. Both of these two sessions may have covered several topics, and involved some web search as part of the response to emails. There are, however, means to determine a notable change of group of people. First, network analytics [22] may indicate when users notably switch between sections of their network, even within a single service. Further, document editors like Google Docs, may list specific collaborators in the permissions, and so it may be simple, in some cases, to detect notable switches. The largest challenge, in relation to People as a factor, however, is tying these people to other factors like topic, task, or phase. These may reinforce boundaries, where a user moves from one topic and set of collaborators to another. Conversely, they may conflate each other, with users covering several

small tasks while working in their email client.

5.2 Achievability

One concern for the taxonomy is that some factors require access to data that services may not have, especially because it was based upon client data. Conversely, we argue that modern services should have access to each of these factors. Nearly all major search engines provide browsers, email services, social networks, document repositories, toolbars, etc that would allow them to monitor our six factors in combination. In fact, current major search engines are well placed to model sessions according to our factors and provide relevant results to dynamically evolving sessions.

5.3 Individual Differences

Personalisation may be an important aspect of applying the taxonomy. Within our study, we interviewed a range of technical and non-technical participants, where some users had much less overall web activity than others. Less-frequent web users, for example, typically indicated a higher tolerance of longer gaps, than those users with dense web history. Similarly, frequent technically-minded participants more commonly had smaller multitasking activities within or in parallel to larger tasks. The potential implication for systems, or search services, is that session detection needs to be relative to each user's normal web behaviour, however a much larger sample should be investigated to see whether these differences can be consistently and automatically identified. This issue highlights, though, that simplistic time-gap dividers, for example, have notable implications for session studies.

6. CONCLUSIONS

Supporting users with more session-relevant results is a common shared objective for IR (e.g. [7]), but with a lack of effective approaches to automatically determine sessions, most research has either focused on systems that let people explicitly label their sessions (e.g. [19]) or by presuming, for now, that sessions have been well determined (e.g. [10]). This research has focused on trying to develop our understanding of how real human web sessions relate to each other, such that sessions can be better identified, instead of using single naive measures like average timeouts. Our primary contribution is a taxonomy of six key factors that relate to the boundaries of sessions, with insights into how they relate, exceptions, and when they override each other.

Beyond our primary contribution, we have also contributed an objective analysis of the 847 real human web sessions that were analysed when discussing the factors relating to session boundaries. Finally, by analysing our card sorting data, we have identified additional categorations of sessions that are classfied as difficult, important and frequent, and that participants perceived some types of tasks as longer.

There are several avenues of future work that can build on our work, aside from the development of systems that attempt to implement a model based upon our taxonomy. First, our resource of 847 real sessions can be examined more comprehensively according to our taxonomy, and taxonomies from other papers like web activity [28] and casual leisure [8]. This process would help us to quantify and examine both the prevelence and the interrelation of our factors on a larger dataset. Further, it would be extremely valuable to investigate much larger search engine logs based upon our taxonomy, to detect their prevelance across many more users than we could study qualitatively in our interviews.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] I. Adeyanju, F. M. Nardini, M. Albakour, D. Song, U. Kruschwitz, et al. Rgu-isti-essex at trec 2011 session track. In *Proc. TREC 2011*. NIST, 2011.

[2] 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.

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

[4] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Inf. Process. Manage.*, 31(2):191–213, Mar. 1995.

[5] 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.

[6] T. Claiborne. Update to sessions in google analytics, 2011.

[7] D. Donato, F. Bonchi, T. Chi, and Y. Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proc. WWW2010*, pages 321–330, New York, NY, USA, 2010. ACM.

[8] D. Elsweiler, M. L. Wilson, and B. K. Lunn. Understanding casual-leisure information behaviour. In A. Spink and J. Heinstrom, editors, *Library and Information Science*, pages 211–241. Emerald Group Publishing Limited, 2011.

[9] D. Gayo-Avello. A survey on session detection methods in query logs and a proposal for future evaluation. *Information Sciences*, 179(12):1822 – 1843, 2009.

[10] D. Guan, S. Zhang, and H. Yang. Utilizing query change for session search. In *Proc. SIGIR2013*, pages 453–462. ACM, 2013.

[11] D. He and A. Göker. Detecting session boundaries from Web user logs. *Proc. 22nd BCS-IRSG*, pages 57–66, 2000.

[12] B. J. Jansen, A. Spink, C. Blakely, and S. Koshman. Defining a session on Web search engines. *JASIST*, 58(6):862–871, 2007.

[13] N. Jhaveri and K.-J. Räihä. The advantages of a cross-session web workspace. In *CHI2005 Ext. Abstracts*, pages 1949–1952. ACM, 2005.

[14] E. Kanoulas, B. Carterette, M. Hall, P. Clough, and M. Sanderson. Session track 2011 overview. In *Proc. TREC 2011*, 2011.

[15] M. Kellar, C. Watters, and M. Shepherd. A field study characterizing web-based information- seeking tasks. *JASIST*, 58(7):999–1018, 2007.

[16] A. Kotov, P. N. Bennett, R. W. White, S. T. Dumais, and J. Teevan. Modeling and analysis of cross-session search tasks. In *SIGIR 2011*, pages 5–14, New York, NY, USA, 2011. ACM.

[17] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[18] C. Liu, J. Gwizdka, J. Liu, T. Xu, and N. J. Belkin. Analysis and evaluation of query reformulations in different task types. *Proc. ASIST 2010*, 47(1):1–9, 2010.

[19] B. Mackay and C. Watters. Exploring multi-session web tasks. In *Proc. CHI2008*, pages 1187–1196. ACM, 2008.

[20] D. Morris, M. Ringel Morris, and G. Venolia. Searchbar: a search-centric web history for task resumption and information re-finding. In *Proc. CHI2008*, pages 1207–1216. ACM, 2008.

[21] D. Nettleton, L. Calderon-benavides, and R. Baeza-yates. Baezayates, analysis of web search engine query sessions. In *Proc. WebKDD 2006*, 2006.

[22] M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.

[23] S. Ozmutlu. Automatic new topic identification using multiple linear regression. *Inf. Process. Manage.*, 42(4):934–950, 2006.

[24] P. Qvarfordt, G. Golovchinsky, T. Dunnigan, and E. Agapie. Looking ahead: query preview in exploratory search. In *Proc. SIGIR2013*, pages 243–252. ACM, 2013.

[25] K. Raman, P. N. Bennett, and K. Collins-Thompson. Toward whole-session relevance: exploring intrinsic diversity in web search. In *Proc. SIGIR 2013*, pages 463–472, New York, NY, USA, 2013. ACM.

[26] 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.

[27] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proc. CHI1993*, pages 269–276. ACM, 1993.

[28] A. J. Sellen, R. Murphy, and K. L. Shaw. How knowledge workers use the web. In *Proc. CHI2002*, pages 227–234. ACM.

[29] A. Spink, M. Park, B. J. Jansen, and J. Pedersen. Multitasking during Web search sessions. *Inf. Process. Manage.*, 42(1):264–275, 2006.

[30] A. Strauss and J. Corbin. Grounded theory methodology. *Handbook of qualitative research*, pages 273–285, 1994.

[31] R. K. Summit. Dialog: An operational on-line reference retrieval system. In *ACM 1967*, pages 51–56. ACM, 1967.

[32] H. Weinreich, H. Obendorf, E. Herder, and M. Mayer. Not quite the average: An empirical study of web use. *ACM Trans. Web*, 2(1):5, 2008.

[33] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proc. WWW2007*, pages 21–30. ACM, 2007.

[34] R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.

[35] D. Wolfram, A. Spink, B. J. Jansen, T. Saracevic, et al. Vox populi: The public searching of the web. *JASIST*, 52(12):1073–1074, 2001.

[36] D. Wolfram, P. Wang, and J. Zhang. Identifying web search session patterns using cluster analysis: A comparison of three search environments. *JASIST*, 60(5):896–910, 2009.