The Characteristics of Casual Sessions in Search Behaviour Logs

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

The premise of Searching4Fun literature is that people engage in search sessions for the enjoyment of searching and passing time, rather than to find specific information for a larger goal-oriented task. Identifying these cases can be difficult, however, because what one person may search for, for fun, may be the important work task of another. In this paper, we provide initial evidence that behavioural data, combined with time-of-day, maybe be a good indicator. To do this we initially analyse high- and low-importance sessions, as identified by their owners. We argue, based on these preliminary results, that such aspects could be studied in more depth to automatically identify people's sessions when they are searching for fun.

Categories and Subject Descriptors

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

Keywords

Sessions, Web History, Searching for Fun, Casual-Leisure

1. INTRODUCTION & RELATED WORK

The premise of Searching4Fun literature is that people engage in search sessions for the enjoyment of searching and passing time, rather than to find specific information for a larger goal-oriented task. Stebbins [4] argued that there are 3 types of leisure activities: serious leisure such as volunteering and committed hobbies, project leisure such as planning a holiday or buying a car, and casual-leisure such as passing time or playing games. Identifying these cases can be difficult, however, because what one person may search for, for fun, may be the important work task of another. Playing games, for example, may be casual fun for one person, a serious hobby for another, and the work role of a professional. Further, these roles may be overlapping; Tate and Russell-Rose [5] argued that casual tasks were not tied to leisure, but at one end of a spectrum of seriousness for all task domains; i.e. a computer scientist may search for the latest apple laptops, which may be considered casual, for both fun and for remaining on top of current developments in their profession.

> *The Cadillac Example*
> In our previous work, we analysed sessions that were self-identified by their owners [8]. Whilst qualitatively investigating these to look at Searching 4 Fun, we encountered an interesting cadillac example. Participant 19, a Computer Scientist at a university in the UK, began to search for information about Cadillacs. Our initial conclusion was that the user had entered a period of searching for fun. Shortly after, however, P19 began to search for information on spare parts and maintenence information, at which point we concluded that they may be engaged in Project Leisure. In total, however, P19 spent 5 hours over 3 sessions, all in one day, engaged in searches related to Cadillacs. Consequently, based on topic alone, we concluded that we did not know a) whether Cadillac's related to secondary employment or leisure time, b) whether it was serious or casual searching, nor c) whether it was stressful or fun.

Elsweiler et al [1] argued that casual-leisure search was different from task-oriented searching models, because users may begin searching without a specified information need, and a successful session was dependent on finding a result (or concluding that there wasnt a result). Further, they are argued that successful casual-leisure episodes were ones where users did not run out of things to find, rather than run out of time to find a result. Similarly, Wilson and Elsweiler [7] argued that such sessions were more exploratory in nature than complex and subjective work tasks, which are traditionally considered to involve Exploratory Search [6].

A conclusion can be drawn, therefore, that it is not the topical nature of searches that is indicative of session types, even if we personalise topical models to the searcher, but that there is reasonable evidence that we can identify casual sessions, and perhaps 'fun' sessions, by their nature. Russell-Rose et al [3] argued that sessions were more easily classifiable by behaviour than theme. In the sections below, we present initial evidence to support the arguement that we may be able to detect sessions of searching for fun, using behavioural data in logs.

2. DATA SOURCE & ANALYSIS

As a preliminary investigation into the nature of casual-leisure sessions, we investigated a set of sessions that were self-identified by their owners in our previous work [8]. In the original study, participants were asked to review their web history and identify and classify sessions into different dimensions; we did not define sessions for them. Including both identifying, classifying, and discussing their sessions, interviews lasted between 90-120 minutes. Twenty participants identified 847 of their own sessions from 10,000 log entries; these log entries were taken from desktop and laptop machines.

One dimension we asked participants to classify their sessions with was *Importance*; again, we did not define this for participants, but left them to use their judgement. For this workshop paper, we chose to focus on the 146 sessions that were classified as being either Low or High Importance. The first assumption of our analysis is that sessions classified as Low Importance, may be a good approximation to 'Casual' sessions[1].

Although we cannot guarentee these sessions involve searching for fun, the Low Importance leisure examples included online shopping and facebooking, while similar work examples included grammer and definition searches. Conversely, high importance examples were focused on information gathering for work tasks, whilst also including some personal tasks like topping up credit on a mobile phone. In the next section, we present results that analyse low and high importance sessions, as identified by their owners, for different periods of the day, to see if they are represented by notably different behaviour patterns in the logs.

3. RESULTS

In order to study the difference between LI (Low Importance Search) and HI (High Importance Search) sessions, according to the time of day, we divided the day into 4 periods: 8:00-16:59 (Working Hours), 17:00-21:59 (Early Evening), 22:00-2:59 (Late Evening or Before Bed) and 3:00-7:59 (Sleeping Time). For each session, we analysed 1) Average Pageviews, 2) Average Length, 3) Average Number of Queries, 4) Average Dwell Time per Page, and 5) Query Frequency. These analyses, broken down by period of the day, are shown in Figures (1 - 5) below.

3.1 Main Findings

In Figure 1, more pages were viewed in LI sessions than that in HI sessions through all of these periods, although especially between 8-16 and 17-21. This difference was less obvious late at night. These results indicate that casual sessions may typically involve more pageviews than more serious sessions.

In Figure 2, the time spent in LI sessions and HI sessions were similar between 8-16 and 22-2. Whilst between 17-22, the time spent in LI sessions was much more than that in HI sessions, which may indicate that longer sessions in the early evening are indicative of casual sessions.

[1]Unfortunately, for this paper, we did not ask participants to directly classify their sessions into dimensions of work vs leisure, and casual vs serious

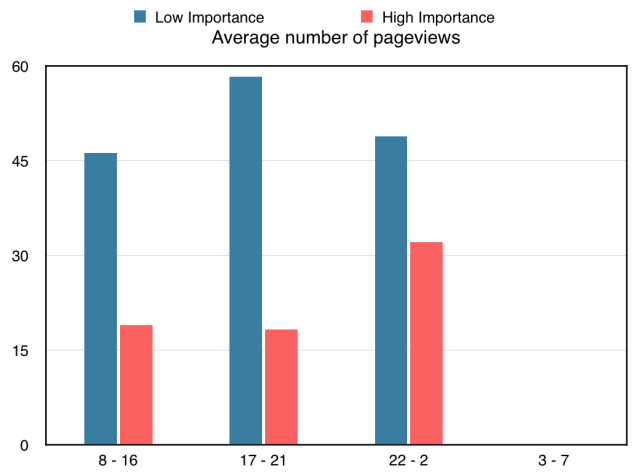

Figure 1: Average Pageview

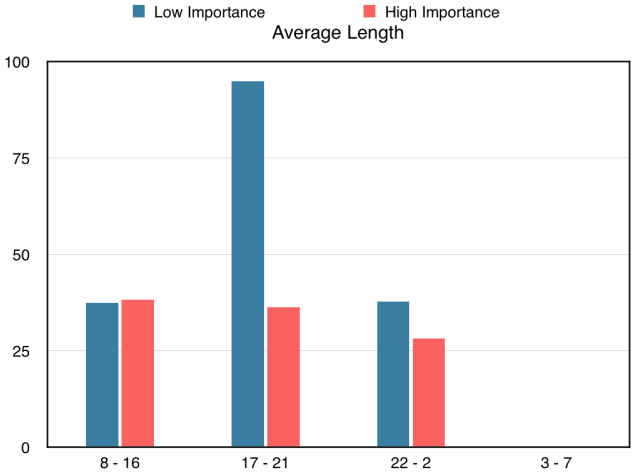

Figure 2: Average Length

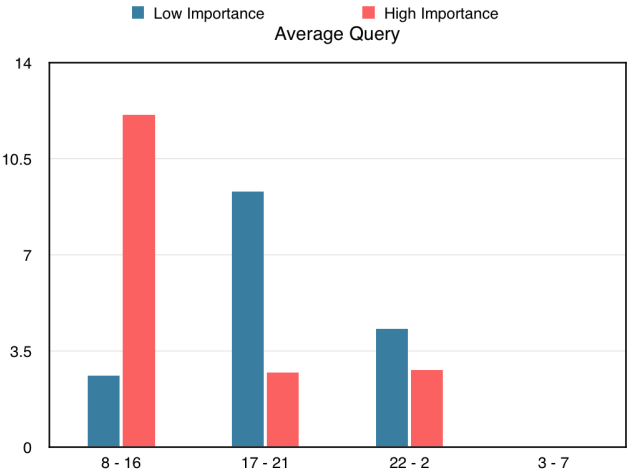

Figure 3: Average Query

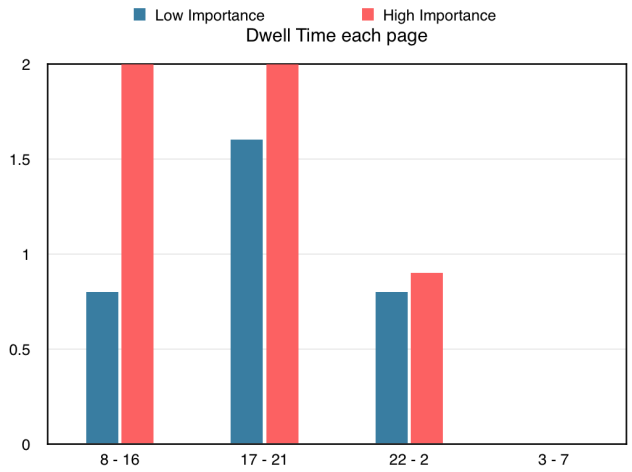

Figure 4: Average Dwell Time per Page

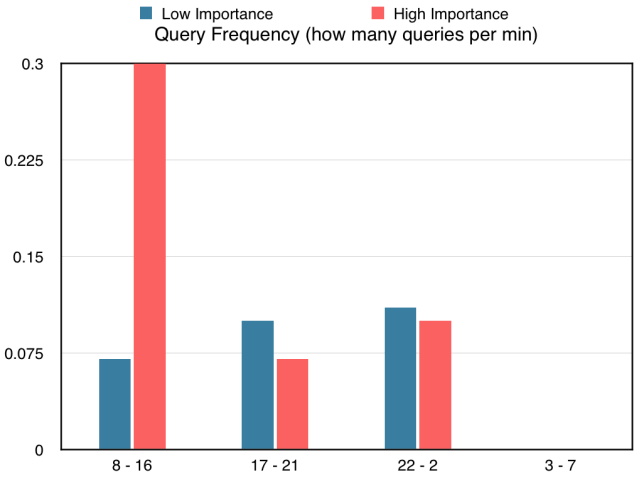

Figure 5: Query Frequency (Queries per minute)

In Figure 3, HI sessions during the 8-16 period involved many more queries than LI sessions, and than in important sessions during the early evening and late at night. Conversely, in the evening between 17-21, LI sessions involved more queries than HI sessions. These time-dependent indicators mean that time of day should be considered if number of queries in a session are used to judge its seriousness; sessions with many queries during the day may indicate a serious nature, while more queries in either early evening and late night may indicate a casual session.

In Figure 4, HI sessions during the day had notably longer per-page dwell times than LI sessions. Although there was a slight difference in per-page dwell times between HI and LI sessions in the evening and at night, the difference was less prominent. These results indicate that high per-page dwell time can be indicative of a serious session, especially during the day.

Similarly, in Figure 5, higher query frequency (queries per minute) may be highly indicative of important or serious sessions during the day, but may not be a useful indicator in the evenings and at night.

3.2 Integration of Findings

Table 1: Comparison of Log Data by Time Period

	8-16	17-21	22-2
Average Pageview	LI > HI	LI > HI	LI > HI
Average Length	LI = HI	LI > HI	LI > HI
Average Query	LI < HI	LI > HI	LI > HI
Dwell Time per Page	LI < HI	LI < HI	LI < HI
Query Frequency	LI < HI	LI > HI	LI > HI

The colored cells in Table 1 represent that there is obvious difference between LI and HI: blue stands for "When data in LI is notably bigger than that in HI", red stands for "When data in LI is notably smaller than that in HI", and white means no obvious difference.

In 8-16, LI has more pageviews, but fewer queries, shorter dwell time per page, and also lower query input frequency than HI, while their lengths are similar.

In 17-21, LI has more pageviews, more queries, and last longer than HI. It also has shorter dwell time per page and higher query input frequency than HI but they are not very obvious.

In 22-2, LI has more pageviews than HI. It also lasts longer, has more queries, shorter dwell time per page, lower query input frequency than HI but not very obvious.

4. DISCUSSION

4.1 Key Findings

From the preliminary results presented above, there are a number of conclusions that may be potentially valuable to study in future work:

1. Viewing more pages may indicate a casual Low-Importance session, especially during the day (8am-16pm) and in the evening (17pm-21pm) (Figure 1).

2. Longer dwell time per page may indicate a serious High-Importance session, especially during the day (8am-16pm) (Figure 4).

3. In the evening (17pm-21pm), very long sessions may indicate a casual Low-Importance session (Figure 2).

4. In the evening (17pm-21pm), a large number of queries may indicate a casual Low-Importance session (Figure 3).

5. During the day (8am-16pm), a large number of queries may indicate a serious High-Importance session (Figure 3).

6. During the day (8am-16pm), a high query frequency may indicate a serious High-Importance session (Figure 5).

4.2 Strengths & Limitations

Although these initial findings could be interesting to study further in future work, there are two key caveats to acknowledge: 1) As we are investigating the notion of casual leisure in a dataset from a study that was not purely focused on searching for fun, we had to work with a refutable assumption, that the notion of "Low Importance" is an approximation towards "Casual". We argue that there may be a big overlap between these concepts. As a result, we believe the initial findings may provide initial insight into the difference between "Casual" and "Non-Casual" web activity. 2) Our data sample is a relatively small, and so we have not focused on statistical evidence, but possible insights. Consequently, we cannot draw hard conclusions from our findings.

One strength of our work, however, is that the session divisions are self-defined by the participants themselves [8], which may be more meaningful than studies of sessions that are artificially separated by default periods of time [2].

5. CONCLUSION

In this paper, we have presented some initial preliminary findings that indicate that casual, perhaps fun, sessions may have different, clear patterns to more serious forms of searching. Although based on a small sample, and relatively simplistic analyses, the results are interesting, and compelling to investigate further in the future. Consequently, we argue that fun sessions are likely to have clear behavioural patterns, which could be easily identified, and perhaps treated differently, by web search systems.

6. REFERENCES

[1] 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.

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

[3] T. Russell-Rose, P. Clough, and E. Toms. Categorizing Search Sessions: Some Insights from Human Judgments.

[4] R. A. Stebbins. Leisure and its relationship to library and information science: Bridging the gap. *Library trends*, 57(4):618–631, 2009.

[5] T. Tate and T. Russell-Rose. The information needs of mobile searchers: A framework. In *Searching4Fun! 2012*, pages 9–11, 2012.

[6] 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.

[7] M. L. Wilson and D. Elsweiler. Casual-leisure searching: the exploratory search scenarios that break our current models. In *Proc. HCIR 2010*, 2010.

[8] C. Ye and M. L. Wilson. A User Defined Taxonomy of Factors that Divide Online Information Retrieval Sessions. In *Proc. IIiX2014*. ACM, 2014.