

What are the Real Differences of Children’s and Adults’ Web Search?

Tatiana Gossen, Thomas Low, Andreas Nürnberger
Faculty of Computer Science,
University of Magdeburg, Germany
{tatiana.gossen,thomas.low,andreas.nuernberger}@ovgu.de

ABSTRACT

We present first results of a logfile analysis on web search engines for children. The aim of this research is to analyse fundamental facts about how children’s web search behaviour differs from that of adults. We show differences to previous results, which are often based on small lab experiments. Our large-scale analysis suggests that children search queries are more information-oriented and shorter on average. Children indeed make a lot of spelling errors and often repeat searches and revisit web pages.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Experimentation, Measurement

Keywords

children information retrieval, web log analysis, web search

1. INTRODUCTION

Internet usage nowadays knows no real age limits. Many children have access to the Internet and explore the web from a young age. Unfortunately, the web adapts slowly to the needs of children. There are many good techniques in IR for adults, but still not much scientific insights on how to design search engines for kids in terms of both user interfaces and underlying algorithms. Thus, the research on children web search is very important. Previous research shows a difference between kids and adults in their search behaviour [1]. However, these findings were based on small user studies with lab experiments and may not show real life behaviour. Only one research group previously attempted to identify children’s queries in a large-scale query log [6]. Here, all queries where the user selected a search result whose domain is listed in the *DMOZ’s kids&teens* directory were regarded as child queries. However, there is a significant chance that such pages were clicked by adults by accident. They also define a children session as one that contains at least one children query entry, which we do not consider as a good threshold. Additionally, AOL is a common search engine. To our knowledge, this is the first large-scale log study of search behaviour on children web search engines.

2. DATASET

For our analysis we used logfiles of the three major german search engines, whose main target group are kids from 6 to 13: *Blinde-Kuh.de*, *FragFinn.de*, *Helles-Koepfchen.de*. The strength of these search engines lies in the retrieval of web documents containing only *child-appropriate* content, which means the kids are able to understand the content and have no access to “adult” web sites. The dataset is composed of 2.5 million requests gathered over one week in January 2011.

The logs contain both *searching interaction* and *click-through* records [8]. We specify a query log as the set:

$$Log = \langle \langle u_i, s_i, t_i, q_i, c_i \rangle \mid 1 \leq i \leq n \rangle \quad (1)$$

where u_i , s_i , t_i , q_i , c_i refer to the user information (user agent, geographic information), session ID, time of interaction, submitted query and its details (correction suggestions, total number of hits, list of result URLs with ranking), click-throughs (URL, referrer). Thus, the logs provide insight into interactions with the search engines’ user interface: queries, navigation, pages visited before and/or after entering the query and viewings of result pages. Although the logs already contained session data, the session definition was not consistent among all three search engines. We refined the sessions by splitting those which had gaps of more than 15 min between session log entries. This is the common approach in logfile analysis, except for some variation of the chosen temporal cutoff parameter (5 to 120 min) [9]. The collected data contains a total of 608,162 sessions.

3. RESULTS

Search engine queries provide an insight into children’s information needs. We analysed a total of 725,846 (226,387 unique) queries. Using the *Google Insights for Search* service¹, we compared the most frequent children queries to the most frequent queries by adults. We retrieved the latter for the same time period and the same location. The most frequent children and adults queries (translated from German) are shown in Table 1.

The results indicate that the information need (in terms of the *Broder taxonomy* [3]) of children differs from that of adults. The children’s queries have a more informational intention. The purpose of *informational queries* is to find information about a topic assumed to be available on the web, in order to read about it. Meanwhile, the adults most frequent queries are *navigational* or *transactional*, with the immediate intent to reach a particular website that the user

¹<http://www.google.com/insights/search/>

Table 1: Most frequent queries from our logs (left) and from Google Insights (right) during the same period

1. games	4. animals in winter	7. dogs	1. facebook	4. ebay	7. amazon
2. sex	5. squirrel	8. animals	2. youtube	5. you	8. web.de
3. electricity	6. whale	9. egypt	3. google	6. weather	9. gmx

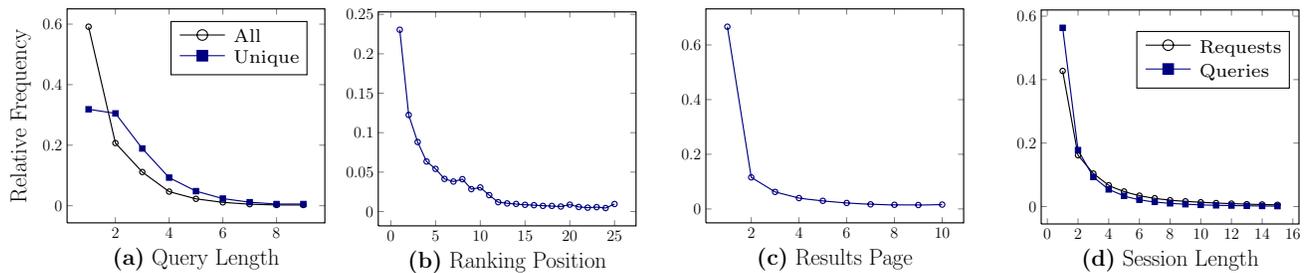


Figure 1: Distributions of relative frequencies of (a) the length of queries in number of words, (b) the ranking position of a clicked result, (c) the number of viewed results' page, and (d) the length of session in both number of all page requests (including queries) and queries only.

has in mind, or even further carry out some transactions, e.g. purchasing a product. Since children may also use the search engines for homework, queries like “egypt” could reflect the curriculum of their schools.

Figure 1(a) presents the distribution of query length. Note that the three search engines do not provide query string auto-completion. Queries contain on average 1.8 terms (2.4 for unique queries), whereas queries by adults have between 2.4 and 2.7 terms [10, 7]. Thus, our results contradict previous findings [1, 6] that children would use longer queries on average.

The three search engines combine textual links to search results with images, which also lead to the result web page. This is mainly done due to the belief that pictures attract children more than text. However, our results indicate that the kids click on a picture only slightly more often (52%) than on text (48%).

Figures 1(b), 1(c) illustrate how often a query result with rank n is clicked and how often the m th results' page is viewed. The search engines provide 10 results per page. Only a few results (2–4) can be seen without scrolling. We found, that most children visited only the first results' page and looked at the first three results. These findings indicate similarity with adult click-through behaviour. More than 70% of the time, adults only view the top ten results [10].

The average session, ignoring identical queries, included about 1.8 queries. Adult studies found that most web search engine users searched one query per session and an average session included about 1.6 queries [10]. We found it similar to children web search behaviour (Figure 1(d)).

Earlier children studies discovered that kids have a *loopy* browsing style, whereas adults' browsing style is *linear* or *systematic* [1]. Children are supposed to click, repeat searches and revisit the same result web page more often than adults. We found evidence for this behaviour in the percentage of repeated URL clicks within a session given the same query (16.6%) and the percentage of repeated queries within a session (20.5%). Adults repeat roughly 12.4% of queries [4].

Another well known fact is that children have difficulty with spelling [2]. About 25% (using *Lingua-DE-Wortschatz*²), 21% (based on small random sampling and manual check-

ing), 40% (with *GNU aspell* dictionary) of all queries contained at least one spelling error, which is significantly more than around 10-15% errors of adults [5].

4. CONCLUSIONS AND FUTURE WORK

This paper presents first results of a large-scale logfile analysis of children's behaviour on targeted web search engines. We compared our findings with previous results about adults' behaviour on common search engines. We found that children tend to formulate informational queries while the adults' most frequent queries are navigational. In contrast to earlier findings, our results show that on average children formulate shorter queries than adults. We did not see any differences in click-through behaviour between kids and adults. We also did several statistics on the “loopy” browsing style of children and spelling errors in queries. We plan to analyse more complex patterns of children search behaviour, such as term co-occurrences and query reformulation.

5. ACKNOWLEDGEMENTS

This research is based on data collected by the German Youth Institute.

6. REFERENCES

- [1] D. Bilal and J. Kirby. Differences and similarities in information seeking: children and adults as Web users. *Information Processing & Management*, 38(5):649–670, 2002.
- [2] C. Borgman et al. Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog project. *J. Am. Soc. Inf. Sci.*, 46(9):663–684, 1995.
- [3] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.
- [4] M. Chau et al. Analysis of the query logs of a web site search engine. *J. Am. Soc. Inf. Sci. Technol.*, 56(13):1363–1376, 2005.
- [5] S. Cucerzan and E. Brill. Spelling correction as an iterative process that exploits the collective knowledge of web users. In *Proceedings of EMNLP*, volume 4, pages 293–300, 2004.
- [6] S. Duarte Torres et al. Query log analysis in the context of information retrieval for children. In *Proc. 33rd Ann. Int. ACM SIGIR Conf.*, pages 847–848. ACM, 2010.
- [7] E. Gabrilovich et al. Classifying search queries using the web as a source of knowledge. *ACM Trans. on the Web*, 3(2):1–28, 2009.
- [8] B. Jansen. Search log analysis: What it is, what's been done, how to do it. *Libr. & Inf. Sci. Research*, 28(3):407–432, 2006.
- [9] B. Jansen et al. Defining a session on Web search engines. *J. Am. Soc. Inf. Sci. Technol.*, 58(6):862–871, 2007.
- [10] A. Spink and B. Jansen. A study of web search trends. *Webology*, 1(2):4, 2004.

²<http://wortschatz.uni-leipzig.de/Webservices/>