

Department of Computer Science and Mathematics

Munich University of Applied Sciences (MUAS)

In collaboration with

Faculty of Informatics and Design Cape Peninsula University of Technology (CPUT)

The research as described in this document, was done by the first author as noted below. He was a student at the Munich University of Applied Sciences (MUAS) in 2014 - the topic was given as a student project. It has not been peer reviewed, but has been edited for basic grammar and accuracy. References have been standardized as far as possible according to the Harvard system. Consider this document as a working paper, to be used for basic referencing but not as seminal source for research work. It could be useful for research in the fields of Website Visibility, Information Retrieval and Search Engines.

Query generation and seek behaviour – how do average users search the net

Author: Roland Widmann (MUAS)

Student Number: 06184012

Mentor: Prof M Weideman (CPUT)

Submitted 30 October 2014 Edited November 2014

A study about how people build queries for Web search and how they behave while searching

Harvard Reference:

Widmann, R. & Weideman, M. 2014. Query generation and seek behaviour – how do average users search the net. *Working Paper*, Munich University of Applied Science, Munich. Available online: <http://web-visibility.co.za/website-visibility-digital-library-seo/>

ABSTRACT

Search engines especially Google cover more and more influence about how information is found all over the Internet. Some studies investigated the searching of Web users by analysing huge search logs of search engines. Others studied smaller educated groups. This working paper investigates a small group of people of average skill at various age which represent the everyday people. It analyses their query generation and behaviour while searching. Participants searched for 10 issues logging all queries, search engines, website visits and auto-completion that have been used. Results show that 100% use Google. 26% of the queries were generated using auto-completion and 1.83 websites were viewed in average per search. Slightly differences between genders were monitored. Google knows how people search and Web users trust in Google is very high. This is both good and bad.

KEYWORDS

search queries, terms, search engines, users search, generate query, query formulation, query generation, search behaviour

1. INTRODUCTION

There's no doubt, Internet plays a central role in our modern time. Life as we live it in our globalized world wouldn't be possible without the Internet. Beneath the communicational aspects, provided by the Internet like email services and social network, the connective infrastructure to interconnect companies' remote locations, the Internet is a giant collection of knowledge and information that wants to be used – and found.

Nachmias (2002) wrote "A recent survey reveals that adult Web users search the Internet more than they engage in any other computer activity" and described the importance of the Internet as a source of information. Brooks (2004) prove alone "Google, with an index of around 25 billion Web pages and 250 million queries a day" and provides an overview about the magnitude of Information seeking in the Internet. For the search for information there are several portals that can be used, like Wikipedia as the global library of knowledge or diverse forums for thematically specified searches. But Greenberg (2000) found that "85 per cent of Web users turn to a search engine when they are trying to find information." This shows that the main search takes place with search engines. These have algorithms implemented, with those search engines search the Web and index the Web in order to provide Web users the right results relevant to their queries. But what are these queries? How do they look like and in what ways do users construct them? For the constructors of search engines it is therefore important to collect information about the structure of search queries. These are the orders of user to become the product of the search engines – the correct information. With knowledge of the structure of users queries and their behaviour while seeking for information search engines can improve their algorithms and possibly react to significant changes or understand possible sections that didn't work satisfying at all.

For Web users it is also important to understand how their often possibly unconscious behaviour in the Internet while searching comes about. This offers the opportunity to change things or optimize behaviour and searching mechanisms.

This working paper wanted to research the queries user type in to search their information. It intended to analyse the behaviour of Web users while searching, the structure, building and development of queries.

2. RELATED RESEARCH

2.1 Gender-specific information behaviour

An interesting paper of Maghferat et al (2010) investigates gender differences in the search behaviour in an empirical study. It analyses general search behaviour and the methods used for deep search. Finding women to search more targeted and being more satisfied with their search results whereas men were found to overrate their ability to search for information.

2.2 Web search behaviour of Internet experts and newbies

Hölscher et al (2000) investigates the types of knowledge which are relevant for information seeking. This study finds differential combined effects on both Web experience and domain knowledge. In particular that Web search novices had severe problems with e.g. formulating reasonable queries. Internet skill trainings should take more focus on web searching.

3. RESEARCH METHODOLOGY

3.1 Questions to be solved

In order to examine how users find their information in the Web, how their search queries look like, how they are structured and developed the author wanted to answer the following questions:

- Q1: Which portals were used for the information seeking
- Q2: How long were the search queries, how many terms were used
- Q3: How were the search queries structured
- Q4: How was the use of Boolean operations
- Q5: What impact did the auto-completion feature have on searching
- Q6: How many result pages and websites have been viewed during the seek

All the questions should in addition be examined, if possible, concerning the gender of the participants.

3.2 Data collection

For this study it was necessary to gather information about the search behaviour of Web users. Therefore a questionnaire has been created in form of an excel sheet. The participants should provide information about their affinity in handling Internet search on a scale from 0 to 5 also the gender was asked.

Further searches for 10 given subjects should be executed. Every single query that was made for finding each subject should be protocolled. Furthermore the following details were provided regarding every single subject:

- Which portal has been used for the search
- Has the auto-completion feature been used
- How many websites have been viewed
- How many result pages have been viewed
- Has a satisfying result been found

At the end of the questionnaire participants provided further personal information about their search behaviour in general and how they groped to their satisfying search result.

3.3 Participants

The questionnaires have been sent to 15 persons, each personally known to the author, via email. To achieve a most sophisticated result as possible, despite the little number of participants, the chosen attendees were at ages between 23 and 55 years. The experience level of the users valued by the author was tried to be fair allocated. All attendees were Germans, hence the searching of this study has been done in German language. Occurring comprehension questions have been answered personally by the author.

4. RESULTS AND DISCUSSION

4.1 Feedback

The questionnaires have been returned by 10 participants, one questionnaire couldn't be used because of unusable filling. So the research based on 9 participants containing 5 men and 4 women.

4.2 Used portals for the search (Q1)

The analysis of the questionnaires resulted a clear picture of which portals are used mainly by Web users to search for information. All queries of this study were invariably entered on the website of Google. Even though some satisfying results have finally been found on other portals, the participants got there via Google. Machill et al (2004) has proven that 91% of Web users use search engines for seeking information and Google has 76% market share. This study reflected these results. Considering the small number of participants it is very likely that the percentage of Google to be used is at 100%. Regarding to the personal comments participants added in the questionnaire many users used Google as an entry portal to access the Internet itself:

„Ich rufe jede Webseite über Google auf“ – („I enter every website via Google“)

Some queries had terms added, indicating search portals like „Wikipedia“ at the end of the query. This shows that participants even assumed to find the satisfying result at a certain portal or place in the Internet, but nevertheless they used Google to get there.

4.3 Length and number of queries (Q2)

By the length of a query in this study is meant the number of terms within it. Therefore every occurrence has been counted, including filler words or Boolean operators.

Schmidt-Mänz et al (2005) describe, referring to an AltaVista study with 150 million queries, that the average number of terms in a query was 2.4. This working paper resulted an average of 3.2 terms per query. Female participants used 16% more terms in average than male (see figure 2). This was related to women having more often searched in questions and so used more filling words. The discrepancy of men using less words in average but more words in sum (figure 1) occurs because there were more men part of the study, but nevertheless, the amount of more used words is little considering that.

Figure 1:

Terms used in the study

Figure 2:

average query length, separated by gender

4.4 Query structure and development (Q3)

While participants searched for the given information 95 queries were submitted. 5 queries were reformed, because the first one didn't lead to a satisfying result. Men reformulated twice by adding in average 0.5 words to the query. Women reformulated three times and added one word in average per reformulation. No reformulation of a query had a decrease of terms. The maximum number of queries per search was two.

Spink, Jansen & Ozmultu (2000) figured out that most Web users only use one query per search, so the average count of queries per session was 1.6. Regarding to Spink & Wolfram et al (2000) 77.6% of sessions had one query, 13.5% had two, and 4.4% had three. The resulting of the searches of actual study's participants provided a value of 1.05 queries per search which is 95% one query, 5% two queries and no search with more than two queries. Having a look on the structure of the queries following that Spink et al (2002) has proven that "(1) 50% of Ask Jeeves queries and less than 1% of Excite were in question format, (2) most users entered only one query in question format with little query reformulation, (3) limited range of formats for queries in question format – mainly "where", "what", or "how" questions, (4) most common question query format was "Where can I find....." for general information on a topic, and (5) non-question queries may be in request format."

Participants in this study only formulated 2 out of 95 queries as a question. 50% by men. The author analysed the amount of filling words used in the queries (figure 3). These were indicators for search queries that did not only consist of keywords. Women use 0.6 filling words per query, men used 0.2. Women tended more than men to form their queries in human spoken language like requests and questions. Nevertheless the number of filling words was low and the participants very often used a summary of keywords to form the query. This may be related to a kind of evolution having used search engines for a long time and mostly the correct use of key words, uncared the filling words, led to satisfying results.

Figure 3: The use of filling words separated by gender

Participants noted "Wenn ich Tipps suche, stelle ich Fragen" ("when searching for tips, I formulate questions"). A conclusion therefor was, that if people search for empirical information, mostly written by private people, they tend to search in questions, requests and sentences.

An effect of the expertise, participants noted they would have, on their query building couldn't be found in this study. Aula (2003) however showed, that experience correlates with both the

number of search terms and the formulation of the users query (broad or precise). This may be related to a more specific fragmentation of the term experience. Especially in considering the users experience in a certain topic that is searched for. An overall specified value for computer and Internet affinity like this study regarded to, was not enough to figure out an influence on query building.

4.5 Use of Boolean operators (AND, OR, NOT, +, -) (Q4)

Spink et al (2004) described that the use of Boolean operators is very seldom. One out of 18 users used them. Every second user of Boolean operators made a mistake when using them. No participant of this study used a single Boolean operator. Some noted on the questionnaire that they generally didn't use Boolean operators. Many Web users, especially the less advanced, may even not know about these operators to exist. These factors prevent Boolean operators from becoming more important in the Web search.

4.6 Use of autocomplete mechanism (Q5)

The analysis of the use of the auto-completing feature to formulate the search queries resulted, that 26.3% of all queries haven been built with use of this technique.

Female participants used auto-completion about 30% more often than male users did (see figure 4). Attendees used auto-completion very little, one out of 20 queries, for searching for a picture. More use of this feature could be observed in technical or political topics.

Participants explained, that auto-completion is often used to ensure correct spelling of words and saving time while not having to type the whole word.

Another thing mentioned by the users of this study was that auto-completion was simply used to collect ideas how to reach a satisfying search result. If users weren't sure how to formulate the query, they typed in one fitting word and looked what suggests were shown. Despite the fact few participants have read it, these results coincide with Googles official intention of implementing this feature (Google Autocomplete Help)

Figure 4: The percentage usage of auto-completion separated by gender

4.7 Visited websites and search result tables (Q6)

Participants of this study watched in average 1.8 websites per query.

Regarding Wolfram et al (2001) back in 2001 people visited 1.7 sites in average. This fact was also approved in this actual study and seems to be a constant factor in the search behaviour of people in the Internet. Spink et al (2004) found that in average 2.4 pages of result were viewed assuming that one page shows ten hits. This corresponds a typical Google result page. Analysing the research data about the search result pages, an average of 1.3 pages was visited. Almost solely result pages indexed larger than one have been visited while searching for pictures. The mode for picture search of search engines provides all results as little displayed pictures. As the participants watched these pictures, it often happened, that the satisfying result was found on the result pages two or later. The switching between the result pages occurs automatically in picture search mode, so users mostly didn't intentionally search more result pages.

5. CONCLUSION

The participants of this study represented the user searching for everyday tasks. Generated queries and the behaviour in searching the web were optimized for reaching results the fastest way. Short query length and few viewed pages per search show that users did not want to compare results or search for better hits. In searches for more complex ~~and experienced~~ topics, this may be different. ~~But~~ reformulation of queries and thus entering multiple queries for the same search are not done very often. This means the search engines provide mostly satisfying results placed in the little area of results a user is willing to view. Users and search engines got adjusted to each other. This is good on one side because users easily find something that fits their query. It also shows that search engines seem to understand how people search. On the other side Web users become very addicted to certain search engines. Considering all attendees of this study used Google as single point of information retrieval, Google has a big influence as almost being the only servant of information for everyday search. Regarding that aspect the auto-completion feature has the opportunity to lead users to a certain kind of results. Kulathuramaiyer (2006) said: "With respect to the quality of the results gained by search engines, users have shown to be overly trusting and often rather naive." In conclusion search engines know very good how to deal with the queries of the users. Web users should in some cases use their opportunity to differ their mechanisms on searching the Web.

6. REFERENCES

- Aula, A. 2003. Query formulation in Web information search. *Proceedings of the IADIS International Conference on WWW/Internet 2003*. 5-8 November. Algrave, Portugal. <http://sites.google.com/site/anneaula/questionnaire.pdf> [November 2003].
- Brooks, T. 2004. The nature of meaning in Age of Google. *Information Research*, 9(3) paper 180 <http://InformationR.net/ir/9-3/paper180.html> [3 April 2004].
- Google 2014. *Search>Help Autocomplete*. <https://support.google.com/websearch/answer/106230?hl=en> [28 October 2014].
- Greenberg, K. 2000. Search Patterns. *Adweek Midwest Edition*: 42, November 9.
- Hölscher, C., Strube, G. 2000. Web search behaviour of Internet experts and newbies. *Computer Networks*, 33(1-6): 337-346.
- Kulathuramaiyer, N., Balke, W.-T. 2006. Restricting the View and Connecting the Dots-Danger of a Web Search Engine Monopoly. *J. UCS*, 12(12):1731-1740.
- Machill, M., Neuberger, C., Schweiger, W., Wirth, W. 2004. Navigating the Internet – A Study of German-Language Search Engines. *European Journal of Communication*, 19(3): 321-347.
- Maghferat, P., Stock, G. 2010. Gender-specific information search behaviour. *Webology*, 7(2). http://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/stock/Gender-specific_information_search_behavior_01.pdf [07 November 2010].
- Nachmias, S, Gilad, A. 2002. Needle in a Hyperstack – Searching for Information on the World Wide Web. *Journal of Research on Technology in Education*, 4(34): 453-473.
- Schmidt-Mänz, N., Koch, M. 2005. Patterns in Search Queries. *Data Analysis and Decision Support*. Heidelberg: Springer-Verlag Berlin.
- Spink, A., Jansen, B. 2004. A study of Web search trends. *Webology*, 1(2): 4.
- Spink, A., Jansen, B., Ozmultu, H. 2000. Use of query reformulation and relevance feedback by Ecite users. *Internet Research*, 10(4): 317-328.
- Spink, A., Ozmultu, H. 2002. Characteristics of question format Web queries: An exploratory study. *Information Processing and Management*, 38(4): 453-471.
- Spink, A., Wolfram, D., Jansen, B. & Saracevic, T. 2000. Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3). <http://onlinelibrary.wiley.com/doi/10.1002/1097-4571%282000%299999:9999%3C::AID-ASI1591%3E3.0.CO;2-R/full#> [12 Dec 2000].
- Spink, A., Wolfram, D., Jansen, B. & Saracevic, T. 2001. Vox populi: The public searching of the Web. *JASIST*, 52(12):1073-1074.