

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/324542818>

The relationship between search query length and user internet exposure [0073]

Research · January 2014

CITATIONS

0

READS

39

2 authors, including:



[Melius Weideman](#)

Cape Peninsula University of Technology

181 PUBLICATIONS 588 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Relationship between Website Usability and Web Metrics [View project](#)



Search Engine Marketing and Business – Present and Future [View project](#)

**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 It has not been peer reviewed, but has been edited for basic grammar and accuracy.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.

**The relationship between search query length and
user internet exposure**

Author: Maximilian Haag (MUAS)

Student Number: 04122312/06

Mentor: Prof M Weideman (CPUT)

Submitted 27 October 2014 Edited November 2014

**Field and Topic: Study about the relationship of user searching behaviour related to
query length and Internet exposure.**

Harvard Reference:

Haag, M. & Weideman, M. 2014. The relationship between search query length and user internet exposure. *Working Paper*, Munich University of Applied Science, Munich.
Available online: <http://www.willbeadded later.com>

ABSTRACT

Today people from all social structures and subjects are using the internet and online search engines. This article examined users searching behaviour. Is there a relationship between user query length and user internet exposure? How do users insert queries into search engines? It was taken a survey about the typically length of search queries and an examination if there is a relationship between query length and user experience. A new research was taken, values like search query numbers and search query length were created by voluntary persons. Different topics were chosen to be searched by users with different levels of experience and compared to some former researches if there are similarities. This article examined today's user internet behaviour and there is a tendency of more experienced users doing shorter queries compared to less exposed internet users.

KEYWORDS

search engines, query length relationship, internet exposure, user experience

1. INTRODUCTION

Today the internet is a great source of information for people in the whole world. People from all different occupations and different ages use search engines to find what they are looking for. "On-line searching by users is now the norm in universities, businesses, and households (Jansen and Pooch, 2001)".

Therefore the users of search engines are more or less experienced in their usage and users have a very different searching behaviour. "It is not only important to know the users' behaviour in order to improve our understanding of search engines, but also to understand what users want from search systems in general (Lewandowski, 2008)".

Search engines analyse the users' queries and try to find out what they are searching for.

"Web search engines can help people find the resources they are looking for by more clearly identifying the searcher's intent behind the query (Jansen and Booth 2007)".

So users create different queries with different intents, proved by former studies.

"From these studies, the aggregate statistics are compared with searching on traditional IR and OPAC systems to highlight possible similarities and differences (Jansen and Pooch, 2001)".

Users are very different in their goals. "Every user has a distinct background and a specific goal when searching for information on the Web (Sieg et al, 2007)".

To find out users intentions and behaviour there is a survey required.

"Questionnaires are a commonly used tool for the user-driven assessment of software quality and usability (Laugwitz et al, 2008)".

By this article a new research in the relationship between query length and user internet exposure was made. Today's user behaviour and typically search query lengths were examined.

A comparison between well experienced and novice search engine users was made depending on their query length and their internet experience level.

2. OTHER RESEARCH

2.1 Users web search behaviour

There were many former studies of web searching behaviour which reported users mostly using short queries for searching. "Web queries are short. About 62% of all queries contained one or two terms. Fewer than 4% of the queries had more than 6 terms (Jansen et al, 2000)". Another study showed that experience of users is dependent on their query length. "An additional factor that tends to constrain query length in real environments is that many computer users have been trained by the text search capabilities in many systems to believe that longer queries will fail to retrieve any document (Croft et al, 1995)". There were studies showing users getting

more experience and learning from their searching results for the future. "Analysis of consecutive queries indicates that more than half of all queries build on the previous query (Jones et al, 1998)".

2.2 Search query length

Former studies showed the past examination of search query length and number.

There was a study which compared users from the USA and users from Europe by their searching formats with two different search engines.

"While only data from Excite and FAST was examined, we provide a baseline for comparing Web searching trends in the United States and Europe (Spink et al, 2002)". The table included the percentages of one, two or more terms in a query and showed that the majority used three or more terms.

Variables	2001 Excite Study (1.2M Queries)	2001 FAST Study (1.2M Queries)
Mean Terms Per Query	2.6	2.3
Terms Per Query		
1 Term	26.9%	25%
2 Terms	30.5%	36%
3+ Terms	42.6%	39%
Mean Queries Per Session	2.3	2.9
Session Size		
1 Query	55.4%	53%
2 Queries	19.3%	18.9%
3+ Queries	25.3%	29%

Table 1: Comparative statistics for FAST and Excite 2001 Web query data sets – one million queries per study (Spink et al, 2002).

2.3 Search query content

This article is about to show if there was a dependency between query length, searching results and the internet searching experience of the user. Former studies showed different search behaviours comparing different user experience levels. "From the expert study one would have expected Web experts to use longer queries. This hypothesis was not confirmed: the queries issued by Web experts were only marginally longer than those of Web novices (2.61 vs. 2.32 words/query). Instead we found a significant effect of domain knowledge: Participants with little domain knowledge made significantly longer queries (average query length: 2.96 vs 1.97 words) (Hoelscher and Strube, 2000)".

2.4 Users searching tasks

"Former studies show that there are different searching tasks between users (Rose and Levinson, 2004)". In the study some user profiles were created for analysing users' behaviour.

"It is interesting to note that nearly 40% of queries were non-informational in every case, and a large fraction of the informational queries appeared to be attempts to locate a product or service rather than to learn about it. In fact, just over 35% of all queries appeared to have the kind of general research goal (questions, undirected requests for information, and advice-seeking) for which traditional information retrieval systems were designed (Rose and Levinson 2004)". So it was to distinguish between informational queries, general queries and their search results.

Another research (White et al, 2009) examined users search queries and their opened document or website after finishing the searching progress. It compared four different subtopics, subdivided into expert and non-expert. The following table shows the results which were obtained.

Domain	In domain		Out of domain		Pearson's correlation coefficient (<u>R</u>)
	Expert	Non-expert	Expert	Non-expert	
Medicine	84.9%	76.8%	75.4%	78.6%	0.55
Finance	84.6%	81.4%	80.1%	82.1%	0.38
Legal	82.3%	79.9%	79.3%	81.1%	0.49
CS	83.6%	69.1%	72.8%	71.0%	0.66
Average	83.9%	76.9%	76.8%	78.2%	0.52

Table 2: Percentage of successful sessions, by expertise. The Pearson's R with domain expertise is reported (White et al, 2009).

3. METHODOLOGY

3.1 Research Methodology

According to analyze today's searching behaviour of the population a new research was made. People were attended to search for ten different topics from everyday life. The main topics in this research were health, technologic, music, news, TV, clothes, books, relationship, education and science. For each main topic there was a subtopic with many details given to the user who had to search for its content.

After searching the user had to record the number of words for his searching process and also the complete sentence length in letters. So it was possible to determine the typical user query length.

Required was also a tag, if the searching process was successful or not. Did the user find, what he was looking for, or not. The users' internet exposures were measured on the period they were already using the internet and search engines. It was depending on their own opinion of technical knowledge to rate the experience.

Topic	Detailed subtopic to be searched for.	Entered user search query.	Number of words of subtopic search query.	Complete length of search query in letters.	Found successful search entry? (yes/no)
Health	Daily fitness tips	staying fit tips	3	17	yes
Technologic	Newest current android version	current android version	3	24	yes
Music	Good internet radio station	radio stream	2	13	yes
News	News of the past week	news of last week	4	18	yes
TV	Current weeks streaming schedule	current tv plan	3	16	yes
Clothes	New trends clothes	clothes trends	2	15	yes
Books	Book ideas and recommendations	current new books	3	18	yes
Relationship	Solving relationship problems	relationship help	2	18	yes
Education	Platform for learning	school learning help	3	21	yes
Science	Technical new inventions	technical inventions	2	21	yes
User internet exposure (1-5):	4				

Table 3: Showing one user questionnaire done for research.

Data of all recorded users was collected and compared. Averages of data were made and analyzed. The following table shows each collected entry which contains topic with number of

words in one query and query length in letters for each topic and the users own intention of his experience. An overall average was made and an average, sorted by users experience. Each user successfully found a search entry matching his expectations.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
User exp.	H. no.	H. len.	T. no.	T. len.	M. no.	M. len.	N. no.	N. len.	TV no.	TV no.	C. no.	C. len.	B. no.	B. len.	R. no.	R. len.	E. no.	E. len.	S. no.	S. len.
3	2	12	3	19	3	23	5	22	3	15	3	16	3	19	3	23	2	15	3	21
4	3	17	3	24	2	13	4	18	3	16	2	15	3	18	2	18	3	21	2	21
5	3	15	2	16	3	22	3	15	2	16	2	11	2	10	2	18	2	18	3	19
2	5	29	3	24	3	20	5	22	3	18	4	21	4	20	3	23	3	20	3	25
3	4	27	3	19	3	16	4	18	4	18	3	18	3	14	2	22	3	19	3	25
4	3	18	3	20	2	15	3	15	3	20	2	12	2	11	2	18	2	14	3	21
2	4	24	3	23	3	20	4	18	4	20	3	17	3	25	3	28	2	18	3	26
3	3	19	3	23	3	23	3	15	3	27	3	19	2	10	2	18	2	14	3	20
4	3	20	3	24	3	23	3	16	3	22	3	18	5	24	2	18	2	14	2	13
3	3	19	2	15	3	20	3	16	4	23	3	18	4	30	3	28	2	17	3	26
Avg. overall																				
3,3	3,3	20	2,8	20,7	2,8	19,5	3,7	17,5	3,2	19,5	2,8	16,5	3,1	18,1	2,4	21,4	2,3	17	2,8	21,7
Avg. by exp.	H. no.	H. len.	T. no.	T. len.	M. no.	M. len.	N. no.	N. len.	TV no.	TV no.	C. no.	C. len.	B. no.	B. len.	R. no.	R. len.	E. no.	E. len.	S. no.	S. len.
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4,5	26,5	3	23,5	3	20	4,5	20	3,5	19	3,5	19	3,5	22,5	3	25,5	2,5	19	3	25,5
3	3	19,25	2,75	19	3	20,5	3,75	17,75	3,5	20,75	3	17,75	3	18,25	2,5	22,75	2,25	16,25	3	23
4	3	18,33	3	22,67	2,333	17	3,333	16,33	3	19,33	2,333	15	3,333	17,67	2	18	2,333	16,33	2,333	18,33
5	3	15	2	16	3	22	3	15	2	16	2	11	2	10	2	18	2	18	3	19

Table 4: Showing complete states of research in numbers.

3.2 Dependency of user experience

For comparing users query length and experience there were created two figures from Table 4. The first compared the average number of words per search for five different levels of user experience. It shows a tendency, that higher experienced users used a smaller number of words than lower experienced users. There were only measures from experience level two up to level five available. The number of words per search was around the count of three, so this number might be the typically query length of internet search engine users.

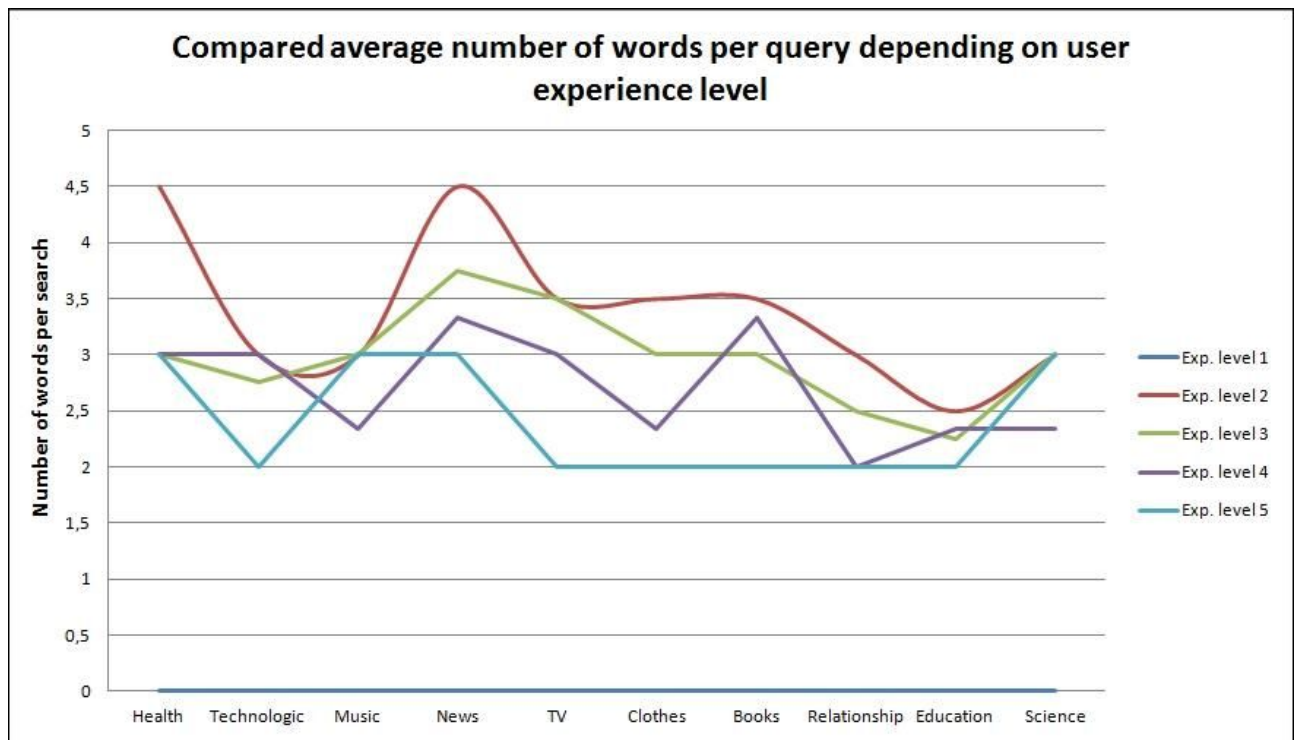


Figure 1: Compared average number of words per query depending on user experience level.

The other figure shows the compared average word length which users took for their query. It was also split into five different levels of user experience. A tendency was observed, that higher experienced users often used shorter words in their queries compared to lower experienced users.

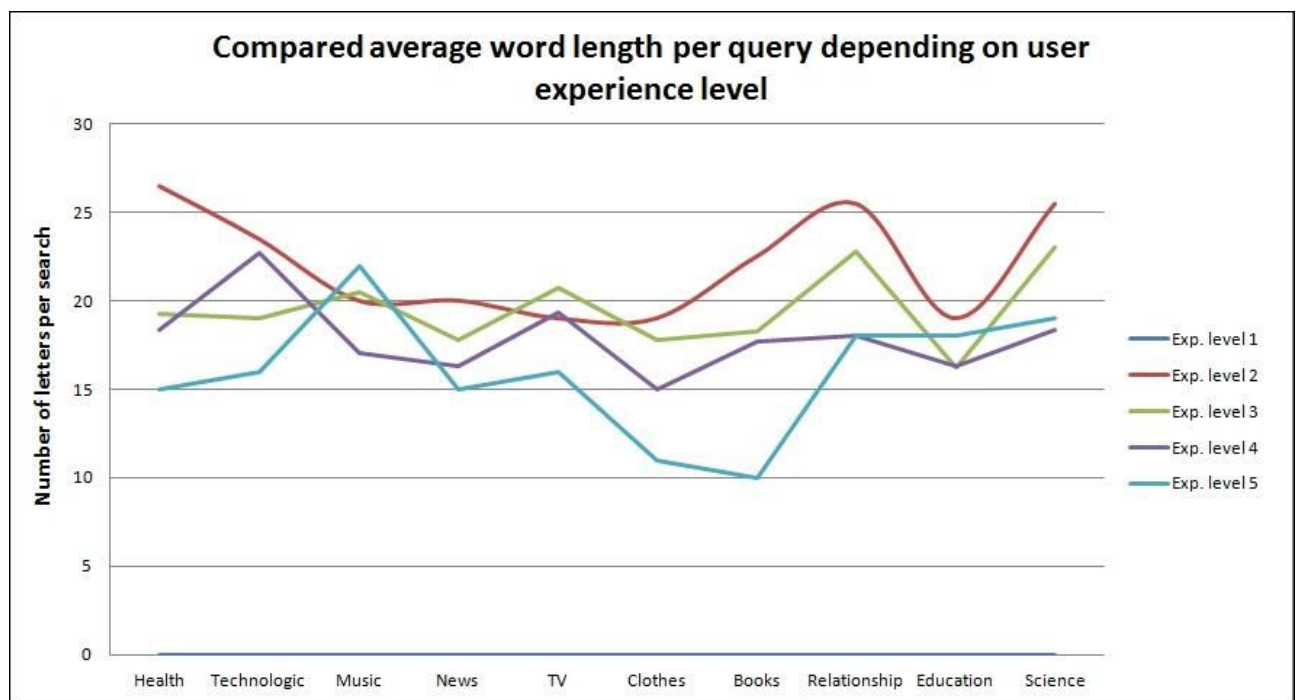


Figure 2: Compared average word length depending on user experience level.

4. RESULTS

4.1 Current research compared to former

Compared to the research of White et al, 2009 which was telling: “We have seen in our study that domain experts employ different search strategies and are more successful than non-experts in four different domains. Given these differences, we believe it is possible to help people better find what they are looking for by taking into account their domain expertise (White et al, 2009)”, this research showed a tendency, that more experienced users used a smaller number of terms in one query and also a smaller number of letters per query for earning good results from the searching engines. This article proved and coincided in some cases with the former research. Each search result was successful, but more experienced users needed less terms and less query length for the same results and the same success in one query than users with fewer experience. By this, a relation between query length and user internet exposure could be determined. This research also showed, that the typically search query contains around three terms and around 15-20 letters. Experienced users took around two with maximum three terms for one query and the number of less experienced users started at about three terms.

This research showed that experience in the usage of searching engines has influence in the searching behaviour, proved by the usage of very different values in the questionnaire and the different results obtained by the examined persons.

5. CONCLUSION

There was proven an existing dependency between user web search queries and the users experience in searching the internet. This article revealed users with less experience writing queries with a greater number of words and a greater number of letters compared to more experienced users which were using less numbers of words in their queries and less letters.

It was proved that there was quite a relationship between query length and user Internet exposure similar to former research studies.

6. REFERENCES

Croft, W. B., Cook, R. & Wilder, D. 1995. Providing Government Information on the Internet: Experiences with THOMAS. *Digital Libraries Conference. 11 – 13 June*. USA, Austin. <http://www.csd.tamu.edu/DL95/papers/croft/croft.html> [25 October 2014].

Hoelscher, C. & Strube, G. 2000. Web Search Behavior of Internet Experts and Newbies. *Computer Networks - COMPUT NETW*, 33(1-6): 337-346.

Jansen, B. J., Booth, D. L. & Spink, A. 2007. Determining the User Intent of Web Search Engine Queries. *World Wide Web Conference Series - WWW*, USA, New York, 8 May 2007: 1149-1150.

Jansen, B. J., Pooch, U. 2001. A Review of Web Searching Studies and a Framework for Future Research Computer Science Program. *Journal of the American Society for Information*

Science and Technology, 52(3): 235-246.

Jansen, B. J., Spink, A. 2004. A Study of Web Search Trends. *Webology*, 1(2).
<http://www.webology.org/2004/v1n2/a4.html> [25 October 2014].

Jansen, B. J., Spink, A. & Saracevic, T. 2000. Real life, real users and real needs: A study and analysis of users' queries on the Web. *Information Processing and Management*, 36(2): 207-227.

Jones, S., Cunningham, S.J. & McNab, R. 1998. Usage Analysis of a Digital Library. *Digital Libraries 98: Third ACM Conference on Digital Libraries*, USA, Pittsburgh, 23-26 June 2008: 293-295.

Laugwitz, B., Held, T. & Schrepp, M. 2008. Construction and Evaluation of a User Experience Questionnaire. *Usability Symposium - USAB*, Austria, Graz, 20-21 November 2008: 63-76.

Lewandowski, D. 2008. Search engine user behaviour: How can users be guided to quality content? *Information Services & Use*, 28(3-4): 261-268.

Ozmutlu, S., Spink, A., & Ozmutlu, H. C. 2003. Multimedia Web searching trends: 1997-2001. *Information Processing and Management*, 39(4): 611-621.

Rose, D. E. & Levinson, D. 2004. Understanding User Goals in Web Search. *World Wide Web Conference Series - WWW*, USA, New York, 17 May 2004: 13-19.

Sieg, A., Mobasher, B. & Burke, R. 2007. Web search personalization with ontological user profiles. *International Conference on Information and Knowledge Management - CIKM*, USA, New York, 6 November 2007: 525-534.

Spink, A., Ozmutlu, S., Ozmutlu, H. C. & Jansen, B. J. 2002. U.S. VERSUS EUROPEAN WEB SEARCHING TRENDS. *Sigir Forum*, 36(2): 32-38.

Spink, A., Wolfram, D. & Jansen, B.J., Saracevic, T. 2001. Searching the web: the public and their queries. *Journal of The American Society for Information Science and Technology - JASIS*, 52(3): 226-234.

White, R. W., Dumais, S. T. & Teevan, J. 2009. Characterizing the Influence of Domain Expertise on Web Search Behavior. *Web Search and Data Mining - WSDM*, USA, New York, 9 February 2009: 132-141.