Google versus Google

Charley Gielkens   Jeroen Hulman

Abstract

This paper hopes to find a difference in user preference and performance using different search techniques. The techniques compared are a regular interface, i.e. a text box into which you type search terms and than perform a query, and a regular interface augmented with a tag cloud. This will give the user frequently occurring terms in the results. After having a group of subjects in, we were unable to find a difference between both conditions.

1 Introduction

In this report we will investigate the usefulness of tag clouds for increasing user performance with search engines. Similar research has been done before, for example by Kuo et al. (2007). By presenting the users with several questions they have to answer, we hope to find a significant difference in performance between a traditional interface and an interface augmented with tag cloud widgets.
We found such an application in `www.deeperweb.com` which augments Google's standard search results with a tag cloud. Thus comparing Google and Deeperweb is a very good way to investigate the effects of tag clouds on user performance, as this eliminates search engine performance from having an effect on user performance.

2 Background

Internet searches

With the rise of the internet, the information age started. Through a virtual web, people got the opportunities to look at billions of webpages with information. To find the right information, people have to use a filtering system. The most successful information retrieval systems on the web have been the search engines. Because of that the most successful search engine, Google, has managed to build a multi-billion dollar imperium.
The most common way to find information across the World Wide Web is by using of search engines. Google is the most popular search engine and performs millions of searches per day. It has gained a market share of over 70% in the US (Garner, 2009)and even over 90% in the Netherlands (Chekit, 2010).
Although Google does not rank the most relevant results at the top, it does have a high degree of usability (van Zwol & van Oostendorp, 2004), which could very well explain the high market share. So although they have a high market share, there is still room for improvement.

Tag clouds

Tag clouds are another way to show a summary of content to users. Tag clouds gather the most common keywords out of a field of content. Therefore people can find information that is hidden deep down a website or web page (Kuo et al., 2007). Tag clouds are still very primitive, but can in the future be developed to help users navigate across the web. An improvement of the tag clouds can be in categorization.

Research from Sinclair & Cardew-Hall (2008) showed when tag clouds can best be used. Results showed that tag clouds are particularly useful when people are searching for a broad subject. Tags should be used for categorization. They gave users an idea of the domain and so could help people with refining their queries. But users commented that the tag cloud was not suitable for finding specific information. Finally, using tags for refining your query requires less cognitive load for a user.
Through visualization methods tag clouds can show more information (Lohmann et al., 2009). Large tags attract more attention than small tags. Tags in the center of the cloud will get more attention than tags at the borders. Further, tags in the upper left quadrant will be found more quickly and are more likely to be recalled. Users scan the cloud more than actually reading it and the layout has a big influence on the users perception.(Bateman et al., 2008; Rivadeneira et al., 2007)

Motivation

Spink et al. (2001) studied trends in web search behavior. They concluded that the interaction of a user with a search engine was short and limited. They pleaded for a generation of more interactive searching tools. Meanwhile, the major search engine Google hasn't changed much.
It is necessary that search engines help people reduce the information complexity (Kao et al., 2008) and provide individual information to users and present the results in ways that are easier to understand.
An experimental search engine was created where search results where combined with a tag cloud of semantically related result (Mirizzi et al., 2010). Although there are a lot of arguments that these tags could help the users to refine their queries, there is still no empirical evidence that confirms these ideas.

Deeperweb

One of the attempts to improve web search is Deeperweb. Deeperweb combines the Google search engine with suggestions for a more advanced query. After an user performed a search, Deeperweb generates a tag cloud from the search results with the most common keywords[1]. Users can choose to add this keywords to the query or to make a search for results without the keyword. Deeperweb generates a tag cloud for the most common keywords, the most common phrases, the most common sites and the most common zones. If a tag shows up more often in the results, it will be bigger in the cloud.
In this research we will look at the implications that Deeperweb has. Does a combination of a tag cloud really help users to find information across the internet and will users accept such a tool? Therefore the following hypotheses are being tested.

Hypotheses

H1 Deeperweb decreases the amount of time needed to find a satisfactory answer

H2 Deeperweb decreases the number of page visits needed to find a satisfactory answer

H3 Participants using Deeperweb will use more queries to find a satisfactory answer

H4 Participants using Deeperweb will give up less

H5 Participants are most satisfied with the best performing search engine

H6 User experience influences performance positively

H7 User experience influences both conditions equally

[1] Deeperweb tutorial: `http://www.deeperweb.com/help/tutorial-howto.php`

3 Method

To be able to support our hypotheses we set up an experiment in which participants were randomly assigned to either Google or Deeperweb. Users were welcomed by the experimenter and given written instructions[2] , which basically said that they should not follow links on the pages found by the search engine. They were also told that we would capture the screen output and that each task would take about five minutes.
Using either of the search engines they were asked to find the answers to six questions. The first was a training question, to get used to the kind of question and the search engine. The following five were administered in random order to cancel out any possible learning effects. Administration took place via a PHP script, which also recorded the time needed to find a satisfactory answer. Users were also asked to submit these answers and the URLs on which they found it. The questions were:

- What was the real name of the singer Pepsi? [3]
- In what dance is a controversial Spanish tradition represented?
- What 1929 car gave rise to an expression still used today?
- How was mister Coke involved in the second world war?
- What was miss Coca famous for?
- Who was president of Bolivia in 1995?

The questions were constructed in such a way that it was unlikely that participants knew the answer and that the answer would not appear on the first page of hits when using keywords from the sentence. In doing so we hoped to stimulate the users to reevaluate their original queries either by coming up with new search terms by themselves or by using the tag cloud in Deeperweb.
After the participants had completed the tasks they were asked to fill out a survey (see appendixB) in order to find out information like experience with search engines and their satisfaction with the used search engine.
The screen captures were later analyzed using InqScribe[4] to retrieve data about the number of pages visited, the number of queries used and the number of times the tagcloud was used.
The experiments took place in the "Sterrenzaal" in the Minnaert building of Utrecht University. Although the environment was not under our control, experience told us that this was a quiet room with adequate facilities and minimum chance of being disturbed.

4 Results

In total 18 people participated in our experiment. Of those we could use 13 for our statistical analysis. The five others were discarded because they either clicked through on a website, used a different search engine or (on one occassion) the script failed and the participant inadvertently started over.

From our survey we learned that all of the valid results were people with an information science background. This possibly contributed to the high number of searches they performed (14,8 searches per day on average) and the amount of time they spend on the internet (on average 4 hours and 50 minutes).

[2] See appendix A
[3] This was the training question
[4] `http://www.inqscribe.com`

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
q5_time	.187	13	.200*	.876	13	.064
q4_time	.274	13	.008	.683	13	.000
q3_time	.154	13	.200*	.941	13	.471
q2_time	.101	13	.200*	.982	13	.987
q1_time	.201	13	.156	.827	13	.015
q1q5_time	.271	13	.010	.851	13	.030

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Table 1: Tests of normality on the time variables

Time

In order to asses if a t-test is allowed to compare means between the groups we tested for normal distribution of the samples and homogeneity of variance. To check for a normal distribution we used the Shapiro-Wilk test and found that the time for questions 1 ($D = 13, p < .05$) and 4 ($D = 13, p < .05$) and the total time ($D = 13, p < .05$) were not normally distributed. The times for questions 2, 3 and 5 ($D = 13, p > .05$) are normally distributed. (see table 1)
Levene's test revealed that for all measured times, the variances are equal. See appendix C.1.
The normally distributed results were checked visually using bar charts with 95% confidence intervals plotted on them (see figure 1). This revealed that confidence intervals overlapped for a large part, meaning that there is no significant difference in performance for questions 2, 3 and 5. For questions 1 and 4 and the total time users needed we perform a Mann-Whitney test in order to investigate if difference between the means exist. (See table 2)

Test Statistics[b]

	q4_time	q1_time	q1q5_time
Mann-Whitney U	19.000	20.000	19.000
Wilcoxon W	40.000	41.000	40.000
Z	-.286	-.143	-.286
Asymp. Sig. (2-tailed)	.775	.886	.775
Exact Sig. [2*(1-tailed Sig.)]	.836[a]	.945[a]	.836[a]
Exact Sig. (2-tailed)	.836	.945	.836
Exact Sig. (1-tailed)	.418	.473	.418
Point Probability	.052	.055	.052

a. Not corrected for ties.

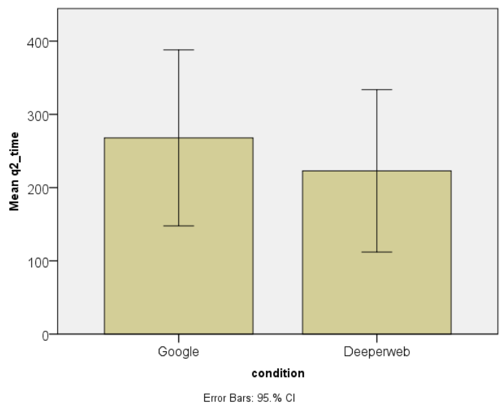
b. Grouping Variable: condition

Table 2: Results of Mann-Whitney test on the times for question 1, 4 and the total time needed

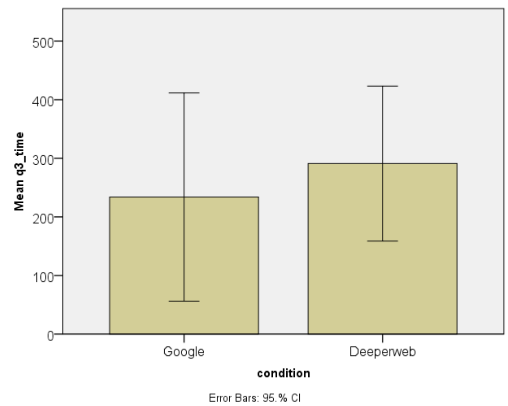
Performance for question 1 ($Mdn_{google} = 184, Mdn_{deeperweb} = 126$) did not differ significantly between conditions $U = 20.000, z = -.143, ns, r = -.04$. Also for question 4 ($Mdn_{google} = 103, Mdn_{deeperweb} = 90$; $U = 19.000, z = -.286, ns, r = -.08$) and the total time needed for questions 1 through 5 (($Mdn_{google} = 728.00, Mdn_{deeperweb} = 730.50$; $U = 19.000, z = -.286, ns, r = -.08$) no significant differences have been found.

Queries and page visits

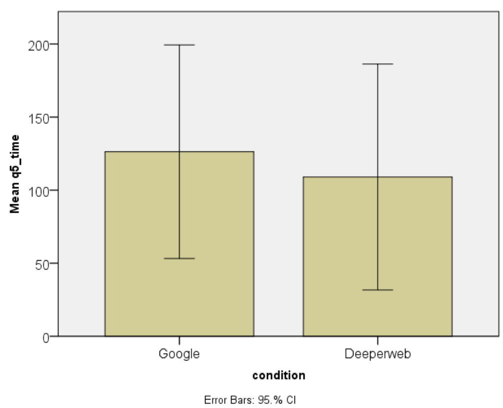
Again, before performing a t-test we checked for normality and homogeneity of variance. We found using a K-S test that the number of queries was significantly not-normal $D(13) = .179, p > .05$. The number of page visits per participant did not differ significantly from the normal distribution



(a) Mean time of participants question 2



(b) Mean time of participants question 3



(c) Mean time of participants question 5

Figure 1: Barcharts with mean times and a 95% confidence interval for the normally distributed times

$D(13) = .308, p < .05$. Levene's test showed that variances were not significantly different between groups for both total number of queries ($F(1, 11) = 0.000, p > .05$) and total number of pages visited ($F(1, 11) = 0.825, p > .05$). This means we can use a t-test to compare the number of page visits, but that we will need to use a Mann-Whitney test to compare the number of queries.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
tot_page	Equal variances assumed	.825	.383	-.011	11	.991	-.048	4.253	-9.408	9.313
	Equal variances not assumed			-.012	8.774	.991	-.048	4.031	-9.203	9.107

Table 3: Results of the Independent Samples T-Test on total number of page visits

Analyzing the number of page visits, we found that participants in both conditions have almost exactly the same average number of page visits. $M_{google} = 12.29$ and $M_{deeperweb} = 12.33$. As could be expected from this small difference, the t-test revealed that it was not significant $t(11) = -.011, p > .05$.

Using a Mann-Whitney test to analyze the number of queries performed in total by participants, we found that there is no significant difference between conditions ($Median_{google} = 18, Median_{deeperweb} = 16,50$). $U = 19, z = -.288, ns$

Test Statistics[b]

	tot_query
Mann-Whitney U	19.000
Wilcoxon W	47.000
Z	-.288
Asymp. Sig. (2-tailed)	.774
Exact Sig. [2*(1-tailed Sig.)]	.836[a]
Exact Sig. (2-tailed)	.808
Exact Sig. (1-tailed)	.404
Point Probability	.024

a. Not corrected for ties.

b. Grouping Variable: condition

Table 4: Results of Mann-Whitney test on total number of queries performed

Satisfaction

We asked users both directly and indirectly about their satisfaction of the search engines they used. This way we could compare verify the measurements by consolidating all indirect questions into an average and compare that to the direct question using a paired-samples t-test. This showed that both questions most likely measured the same information, as the correlation was very high and significant at $p < .01$ (see table 5) and the means ($M_{direct} = 3.31, SE = .237; M_{indirect} = 3.55, SE = .143$) did not differ significantly $t(12) = -1.617, p >, 05, r = .42$. (see appendix C.2)

Correlations

		avg_ sat	generalsat
avg_ sat	Pearson Correlation	1	.801**
	Sig. (2-tailed)		.001
	Sum of Squares and Cross-products	3.177	4.231
	Covariance	.265	.353
	N	13	13
generalsat	Pearson Correlation	.801**	1
	Sig. (2-tailed)	.001	
	Sum of Squares and Cross-products	4.231	8.769
	Covariance	.353	.731
	N	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5: Correlation between the average of indirect questions and the direct question about satisfaction.

Knowing this we feel confident that we can use either measure of satisfaction to test our hypothesis concerning user satisfaction and which engine they used. We choose to use the measure built up of the indirect questions, as investigating peoples minds is generally better done indirectly.
Using a Shapiro-Wilk test we found that the values are normally distributed ($Df(13), p > .05$), which means we can use a t-test to compare the results. Doing so taught us that the minimal difference between user satisfaction for participants that used Google ($M = 3.49, SE = 0.49$) and that used Deeperweb ($M = 3.62, SE = 0.58$) was not significant $t(11) = -.436, p = .671$ and has only a minor effect $r = .13$. (see appendix C.3)

Experience

We hypothesized that experience influences user performance. In order to assess if correlations exist, we first tested if the different variables were normally distributed. As can be seen in table

6, number of hours spent on the internet, number of searches per day and total amount of queries needed by the participant to complete the experiment are normally distributed.
As such, we are able to investigate the correlation between these variables using Pearson's r. This showed that no significant correlation exists between total number of queries needed and number of searches per day or hours spent on the internet per day. (See table 7)

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
internet	,280	13	,006	,885	13	,082
searches	,139	13	,200*	,936	13	,412

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

Table 6: Test of normality

Correlations

		searches	internet	tot_query
searches	Pearson Correlation	1	.664*	-.121
	Sig. (2-tailed)		.013	.693
	N	13	13	13
internet	Pearson Correlation	.664*	1	-.107
	Sig. (2-tailed)	.013		.727
	N	13	13	13
tot_query	Pearson Correlation	-.121	-.107	1
	Sig. (2-tailed)	.693	.727	
	N	13	13	13

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7: Correlation (Pearson's r) between number of searches, hours of internet and total number of queries

Correlations

			searches	internet	tot_page	q1q5_time
Spearman's rho	searches	Correlation Coefficient	1.000	.624*	-.453	-.301
		Sig. (2-tailed)	.	.023	.120	.318
		N	13	13	13	13
	internet	Correlation Coefficient	.624*	1.000	-.456	-.284
		Sig. (2-tailed)	.023	.	.117	.346
		N	13	13	13	13
	tot_page	Correlation Coefficient	-.453	-.456	1.000	.833**
		Sig. (2-tailed)	.120	.117	.	.000
		N	13	13	13	13
	q1q5_time	Correlation Coefficient	-.301	-.284	.833**	1.000
		Sig. (2-tailed)	.318	.346	.000	.
		N	13	13	13	13

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 8: Correlation (Spearman's ρ) between number of searches, hours of internet and total pages visited and time needed for questions 1 through 5

In order to test whether a relationship exists between any of the other variables, numbers of searches per day, number of hours on the internet per day, total number of pages visited and total

time needed to complete the questions, we need to use a non parametric test like Spearman's ρ. Unfortunately this also did not reveal any significant correlations between the variables we wanted to test.
What is revealed in both analyses though, is that there is a correlation between time spent on the internet and number of searches per day. $r = .624, p < .05, R^2 = .389$ Interesting though this is, it is neither surprising nor what we were looking for.

Failures

A Shapiro-Wilk test revealed that our data for this hypothesis are not normally distributed $((D13), P < .05)$. Levene's test showed that the assumption of homogeneity of variance has not been violated $(F(1, 11) = 1.47, ns)$. This mean that we have to use a Mann-Whitney test to investigate the difference in means.
The number of questions given up on by users in both conditions $Mdn_{google} = 1, Mdn_{deeperweb} = 0$ did not differ significantly (see table 9). There is a medium sized effect though. $U = 8.500, z = -1.977, ns, r = -.548$

Test Statistics[b]

	failcount
Mann-Whitney U	8,500
Wilcoxon W	29,500
Z	-1,977
Asymp. Sig. (2-tailed)	,048
Exact Sig. [2*(1-tailed Sig.)]	,073[a]
Exact Sig. (2-tailed)	,078
Exact Sig. (1-tailed)	,053
Point Probability	,049

a. Not corrected for ties.

b. Grouping Variable: condition

Table 9: Mann-Whitney test on how often participants gave up

5 Discussion

The research did not deliver the results as we expected. For this there are a couple of factors that can have influenced the results.
For this research we have tested 18 participants. All participants had an age between 18 and 30, were higher educated and had experience in computer science. The survey showed that the participants perform a lot of searches and spend a lot of time on the internet. A larger and more diverse sample could have made a difference. Because people have experience and are used to search with Google, the need for another search engine can be low.
The search tasks the participants had to perform where made up. The questions where supposed to be found after the use of multiple queries, which turned out pretty well. However, the questions were quite ambiguous and therefore sometimes hard to understand. Participants had to search for very specific information, although tag clouds proved more useful for broad categorization. The influence of Deeperweb on less specific tasks remains unknown.
In this research we haven't checked the answers people gave to the questions. The assumption was made that users would give a satisfactory answer. Through to the enormous amount of information it is hard to say if an answer is right or wrong. Some answers were definitely right, others were definitely wrong. Some of the answers were harder to define. To define these answers as right or wrong, issues like web credibility come to play.
Participants who performed the tasks all spoke Dutch as first language. Everyone knew English as a second language. The tasks where performed in English, because Deeperweb is based on Google in English. Participants who spoke English as a first language could give other results.
In this experiment only short term effects of the tag cloud were registered. People didn't have any experience with Deeperweb. In the long term, Deeperweb might show better results because people learn to work effective with it. Therefore a longer and "deeper" research is needed.
The exact impact of the tag cloud on the searches performed is not totally clear. We have measured the number of clicks on tags, but the inspiration that the tag cloud gave to type in new queries is unknown. In most researches about tag clouds, eye-tracking software is being used, to see the influence. Use of eye-tracking software can also show if Deeperweb can be improved through changing the layout or the positioning of the tag cloud.

6 Conclusion

Tags can be used to improve search queries (Sinclair & Cardew-Hall, 2008) . Deeperweb uses a tag cloud to refine queries for the Google search engine. This research showed that there are no significant differences in the time people have to spend searching for an answer to a question between Deeperweb and Google. There were also no differences found in the user satisfaction of both tools. The differences in results between both test groups were minimal. Therefore it makes no sense to continue this research as it is.
Still there are some leads for future research. The Deeperweb tag cloud gives a quick overview of the keywords out of the results. It gives the user more information to find information. How users respond to this is an interesting phenomenon. Issues like layout, web credibility, type of task and user experience will all play their role in the use of deeperweb. At least we can say one thing for sure. Web search is still not fully optimized and there's still room for improvement. The search for better search tools should continue.

References

Bateman, S., Gutwin, C., & Nacenta, M. (2008). Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, HT '08 (pp. 193–202). New York, NY, USA: ACM.

Chekit (2010). Chekit Nationale Search Engine Monitor.

Garner, R. (2009). Google Ups Share of Search To 72%; Yahoo, MSN and Ask Continue To Tank.

Kao, G. Y.-M., Lei, P.-L., & Sun, C.-T. (2008). Thinking style impacts on Web search strategies. *Computers in Human Behavior*, 24(4), 1330–1341.

Kuo, B. Y.-L., Hentrich, T., Good, B. M., & Wilkinson, M. D. (2007). Tag clouds for summarizing web search results. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07 (pp. 1203–1204). New York, NY, USA: ACM.

Lohmann, S., Ziegler, J., & Tetzlaff, L. (2009). *Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration*, volume 5726 of *Lecture Notes in Computer Science*, (pp. 392–404). Springer Berlin / Heidelberg.

Mirizzi, R., Ragone, A., Noia, T. D., & Sciascio, E. D. (2010). What if Exploratory Search and Web Search meet?. *Proceedings of the Fourth Ph.D.Workshop in CIKM.*

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07 (pp. 995–998). New York, NY, USA: ACM.

Sinclair, J. & Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34(1), 15–29.

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

van Zwol, R. & van Oostendorp, H. (2004). *Googles Im Feeling Lucky, Truly a Gamble?*, volume 3306 of *Lecture Notes in Computer Science*, (pp. 378–389). Springer Berlin / Heidelberg.

A Instructions

Dear participant,

Thank you for taking the time to participate in our experiment, we really appreciate it. The task we have for you is to find answers to certain questions using a search engine. First you will get a training question, so you know what kind of questions you can expect to get and get acquainted with the search engine. After that you will be presented with five real questions. Please take your time to find the correct answers. Per question you will need about five minutes. This means that this experiment will take about 45 minutes in total.

We will capture the screen during the experiment for later analysis. If you have any objections to this, please inform you experimenter.

After you have found the answers to all five questions you will be presented a short survey with some general questions about your background. We will not ask you for any personal information and of course the entire experiment is anonymous. Please feel free to ask any questions you have now or after the experiment.

Kind regards,
Jeroen Hulman and Charley Gielkens

B Survey

See next page.

Search engines

In this survey we will ask you about your experiences with the searches you performed today

* Required

What is your reference number? *

What is your gender? *

Male

What is your field of study? *

How many hours per day do you spend on the internet *

Hours per day

How many internet searches do you perform on a day? *

internet searches per day

Answer the following questions on a scale from 1 to 5, where 1 stands for totally disagree and 5 for totally agree *

	1 Totaly disagree	2	3 Neutral	4	5 Totaly agree
I'm satisfied with the search options offered by the search engine?	○	○	○	○	○
I'm satisfied with the presentation of the results	○	○	○	○	○
I'm satisfied with the relevance of the retrieved documents	○	○	○	○	○
I'm satisfied with this search engine for my search as a whole	○	○	○	○	○
I'm satisfied with the response time of the search engine	○	○	○	○	○
The search engine helped me finding the answer to the problems	○	○	○	○	○
I needed much time to find the right answers	○	○	○	○	○

	1 Totaly disagree	2	3 Neutral	4	5 Totaly agree
The results I found are trustworthy	◯	◯	◯	◯	◯

Do you have any remarks concerning this research?

Submit

Powered by Google Docs

C SPSS output

C.1 Homogeneity of Variance for completion times

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
q5_time	Based on Mean	,038	1	11	,849
	Based on Median	,065	1	11	,804
	Based on Median and with adjusted df	,065	1	8,083	,805
	Based on trimmed mean	,013	1	11	,910
q4_time	Based on Mean	1,891	1	11	,196
	Based on Median	,746	1	11	,406
	Based on Median and with adjusted df	,746	1	7,243	,415
	Based on trimmed mean	1,421	1	11	,258
q3_time	Based on Mean	1,485	1	11	,248
	Based on Median	,426	1	11	,527
	Based on Median and with adjusted df	,426	1	7,541	,533
	Based on trimmed mean	1,369	1	11	,267
q2_time	Based on Mean	,118	1	11	,738
	Based on Median	,083	1	11	,778
	Based on Median and with adjusted df	,083	1	9,437	,779
	Based on trimmed mean	,120	1	11	,736
q1_time	Based on Mean	,549	1	11	,474
	Based on Median	,094	1	11	,765
	Based on Median and with adjusted df	,094	1	6,720	,769
	Based on trimmed mean	,282	1	11	,606
q1q5_time	Based on Mean	,164	1	11	,693
	Based on Median	,100	1	11	,757
	Based on Median and with adjusted df	,100	1	10,624	,758
	Based on trimmed mean	,137	1	11	,718

C.2 Paired-samples t-test for checking consistency of satisfaction within subjects

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	generalsat - avg_sat	-.2417	.5389	.1494	-.5674	.08390	-1.617	12	.132

C.3 Independent samples t-test to check satisfaction between conditions

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
avg_ sat	Equal variances assumed	.213	.653	-.436	11	.671	-0.1293	0.2965	-0.7818	0.5233
	Equal variances not assumed			-.430	9.982	.676	-0.1293	0.3003	-0.7985	0.5400