



DEGREE PROJECT, IN HUMAN-COMPUTER INTERACTION, SECOND CYCLE
DH224X | 30.0 CREDITS , SECOND LEVEL

STOCKHOLM, SWEDEN 2015

Streamlining Search User Interfaces on the Smartphone

AN EXPERIMENTAL STUDY OF COMPARING
DIFFERENT GUI VERSIONS OF KAROLINSKA
INSTITUTET'S SEARCH SOLUTION

JESPER ANNEBÄCK

KTH ROYAL INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE AND COMMUNICATION (CSC)

Effektivisera sökgränssnitt på smarttelefonen

En experimentell jämförelsestudie av användargränssnitt på Karolinska Institutets söklösning

REFERAT

Hur man presenterar sökresultat har blivit allt viktigare i och med dagens av smarttelefoner. Det här arbetet undersökte vad som gör sökresultatssidor effektiva när man använder en smarttelefon. I arbetet utfördes en experimentell studie av Karolinska Institutets söklösning. Tre olika versioner av Karolinska Institutets smarttelefonsgränssnitt utvecklades genom att undersöka potentiella *features* (funktioner, egenskaper) som förbättrar effektiviteten. De olika utvecklade versionerna jämfördes mot det original som Karolinska Institutet använder idag. Detta gjordes genom experiment med studenter från Karolinska Institutet. De kvantitativa resultaten kunde inte visa på någon statistisk signifikant effektivitetsförbättring. De kvalitativa resultaten pekade på att kategorisering och summeringstexter är föredragna *features* i ett sökgränssnitt i smarttelefonen.

Streamlining Search User Interfaces on the Smartphone

An Experimental Study of Comparing Different GUI Versions of Karolinska Institutet's Search Solution

Jesper Annebäck^{*}
KTH Royal Institute of Technology
School of Computer Science and Communication
Stockholm, Sweden
anneback@kth.se

ABSTRACT

The presentation of the search result information has become more important ever since the rapid development and use of smartphones. This thesis investigated what makes search engine result pages efficient and effective on the smartphone screen. For this thesis an experimental study of Karolinska Institutet's search solution was executed. Three different versions of Karolinska Institutet's search engine result pages were developed, by researching potential features that would improve the effectiveness and efficiency. The developed versions were compared to the original search engine result pages through experiments with students from Karolinska Institutet. The quantitative results showed no statistical significance of improvement in effectiveness or efficiency. The qualitative results indicated that categories and summary texts are preferable features in a search GUI in a smartphone. Thus the results were indecisive, the author reflects and provides ideas on how to achieve clearer results.

Keywords

HCI, smartphone, GUI, search engine, responsive web design, experimental design

1. INTRODUCTION

Using search engines has become one of the most efficient ways to navigate through the internet's vast sea of information. A user's behaviour on a search engine has also become more predictable and universal. This is because the search result website design follows the same presentation standard, in a relevance based list [1]. However, the strength of this kind of layout is when the user can find its search goal among the top results [2]. Previous studies have revealed that a user is more likely to try clicking the top result on the screen even if there is a better result further down [3, 4].

The graphical layout of these search result websites shows that it is feasible to influence behaviour through graphical design. User studies with eye tracking have shown that an individual refocuses his or her search pattern on the screen depending on where and how the search results are presented [2]. It is not only behaviour that can be influenced by positioning interface elements in a different way. Search engines can also influence a user's trust by simply presenting the list of the result in a different order [5].

Positioning in the screen and the usage of the space on the screen have become important, since the development of the smartphone. With all the new technology and development opportunities that come with this new smartphone market, usability demands are placed in the top of the users' requirements lists [6]. The users want to be able to access and fit all of their important services and applications on the screen. Also, when using other external services such as a website in a web browser, the usability and the design of the website are of utmost importance.

Responsive web design is a technique which helps with the screen content fitting problem, as it restructures the design of the website depending on the screen resolution [7]. This is also called *liquid layout* and is one of the best practices for search result pages [8], because they tend to contain explaining sentences and populate the available screen space. Smartphone responsive web user interfaces (SRWUI) have different components or features of which some relates to, or are the same as, the desktop web user interface (DWUI). The focus of the SRWUI is often to lower the development costs as well as highlighting the important features of the website [9, 10].

^{*}The user terms of this paper follows the guidelines of master thesis reports at KTH. <https://www.kth.se/en/samverkan/exjobb/studenter/riktlinjer-1.293804>. Written by Jesper Annebäck for Computer Science and Communication at the KTH Royal Institute of Technology in Stockholm, Sweden and Findwise Stockholm AB. March 26, 2015

A responsive search solution with an intuitive information architecture is fairly rare on smartphones. Because of this, there is an interest in what kind of features which makes out an effective graphical user interface (GUI) [11]. This thesis focused on these features and which of them that were important to have in an effective and efficient smartphone web user interface.

Karolinska Institutet's (KI) public web search solution is responsive and was crafted by the search solution consulting company Findwise. In this study the effectiveness and efficiency of a SRWUI will be measured through experiments, involving experimental smartphone GUI designs of the KI search solution.

1.1 Purpose

The purpose of this thesis is to investigate if it is possible to streamline the GUI of a search solution on the smartphone. Furthermore this study will investigate which features that improves the effectiveness and efficiency of a smartphone based search solution. The aim is to get recommendations of which features to use in a search solution for smartphones, by comparing implemented features of the KI search solution.

1.2 Problem statement

If someone uses a search solution today it is not entirely sure that the search result page on the smartphone screen will contain a satisfying answer. Even if it does, the most suitable answer may not be displayed at the top of the screen and the users will have to scroll through the search engine result page (SERP) and search for the target by themselves. In addition, there exists several other features that are shown amongst the search results and the important information might be lost in it. Therefore it is interesting to investigate which features on a SERP that should be shown and which should not.

Thus, the research question of this thesis:

In a search solution which kind of features are important for making an effective and efficient smartphone responsive web user interface?

1.3 Delimitations

To answer the research question of this thesis within the timespan of 20 weeks of full-time studies some delimitations were needed.

As the number of required experiment subjects depended on the number of experiments and were the greatest risk of this work; a number of three (3) self produced experimental designs deemed to be manageable, along with a number of 30 test subjects (10 for each design).

The themes of the FOIs (explained in 2.3 Features of interest, FOI), that were the base for the experimental designs, were limited to the presentation of the search results' cells and striving to save screen real estate, because of the numerous GUI features that exists on a website. The implementations of these experimental designs were limited to only be smartphone GUI versions of the KI search solution.

The variables measured during the experiments were focused on measuring efficiency and effectiveness.

Important to note for this thesis work is that focus will not be on developing the best experimental procedure. The work centers around the experimental designs of the features and which one them that are important in an efficient and effective smartphone GUI.

1.4 Abbreviations

The following list contains abbreviations that are used through this whole paper.

GUI - Graphical user interface

KI - Karolinska Institutet

SRWUI - Smartphone responsive web user interface

DWUI - Desktop web user interface

SERP - Search engine result page

FOI - Feature of interest

PCC - Pearson correlation coefficient

2. RELATED WORK

Previous research within the area of responsive search solution and search engine result page (SERP) evaluation have mostly focused on either just DWUI [5, 12, 13] or SRWUI [14] and not on search. However, there are prior work [15] which focuses on mobile web systems as well as evaluating their designs. The following section presents how prior work have implemented different search GUI layouts, starting with desktop layouts followed by the smartphone layouts. This is followed by the features of interest (FOI) that have affected user behaviour in the experiments of prior work. Finally the definition of variables of effectiveness and efficiency are presented.

2.1 Different types of desktop SERP layouts

In the previous study by Kammerer et al. [5] the authors conducted design experiments to examine the users' evaluation process on a search result page. These experiments based on pre-selected Google SERPs (seen in Figure 1) and eye tracking revealed that a result page presented as a list interface made the users' viewing pattern more linear shaped - comparing to a grid interface layout. The list interface also dragged the attention of the user to the answers at the top. However, a grid interface made the user attend to the search results equally (for gaze time and amount of attention). Similar findings were made from another comparison study by Weisman and Bar-Ilan [13] which also concluded that users seldom jumps to another result page besides the first one. As smartphones have a very limited screen real estate presenting search results in a grid layout is not desirable [8]. A grid layout is a two-dimensional layout and could lead to a *two-dimensional scrolling situation* which is not preferred according to Jones et al. [16].

Furthermore in 2009 a study [17] introduced a new concept to SERPs and an extension of the grid layout. The concept had two parts to it. The first one being that the order of the results were semantic-oriented instead of sorted by the information content and structure. The second part was that the SERP was presented as a grid layout, where the extension was that each cell represented a topic that was based

on the search query. The conclusions that emerged from this prior experimental user study were that the semantic-oriented layout together with visual topic summary could possibly ease the users' search process. Furthermore, the authors also claimed that they could not pin-point certain features that made a successful web search interface.

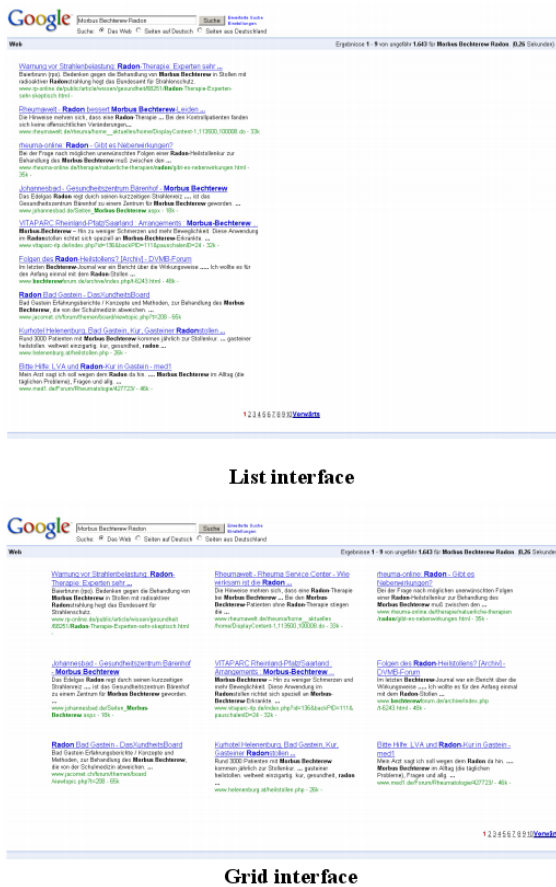


Figure 1: The experimental GUIs of Kammerer et al. [5].

A thorough study by Dumais et al. [12] involving desktop SERP interfaces, was conducted in 2001. In this study the authors compared seven different SERP designs with user experiments. The main difference between these designs were that three of them had a classic search result list layout while the remaining four presented the results in categories. The results from the study showed that the categorisation of result topics was more effective than presenting them in a list. The combination of category names and web page titles proved to be the best among all of the designs. The qualitative results also showed that all the category interfaces were more effective in all cases. In addition the study's post-experiment questionnaire indicated that the participants almost unanimously preferred the same type of interface (category layout) as well. However, since this research was more than a decade old and the experiments were conducted on a desktop web browser, all of its results are not possible to apply on today's technology.

An unfortunate recurrence in the results of the prior studies was that the authors could not reveal any concrete features that were important in a SERP. However, the results revealed examples on features that were appreciated by the users.

2.2 Smartphone GUI research and SERP

A previous study by Schmiel et al. [18] revealed that mobile websites, which were suited for smartphones, got the user to complete tasks 30-40% faster than on the stationary counterpart. By the knowledge of this fact a study by Raneburger et al. [14], which compared two web GUI prototypes on the smartphone, revealed that minimising the number of taps was important for this kind of device and it led to an increased user efficiency. The prototypes of the prior study explored the benefits of vertical scroll and tab-based layouts and revealed also that scrolling was superior to tabs. The authors also concluded that the users completed the task in the experiment better if there existed a well and pre-defined scenario as well as a prototype with only task-focused functionality.

A common thread in the prior research of smartphone GUI is that it is a necessity to both strip down and position features as well as layout. This is not only to save resources, but more importantly to increase the usability and efficiency of the product. Heimonen and Käki [15] concluded this when they investigated two mobile GUI layouts based on categories and lists.

2.3 Features of interest, FOI

Grid layout has shown in the studies of Dumais et al. and Weisman et al. [12, 13] to be an effective way to present search results on desktop resolution, but it was not chosen to be one of the experimental designs tested in this thesis. This is because one of the main guidelines for mobile web GUIs is saving space [8] and also because a desktop web GUI is not just a version of a mobile web GUI with more space. In addition, it is important to note that these types of GUIs (smartphone and desktop) are different and have their own physical UI. The three FOIs which are explained individually below, were chosen based on the guidelines of prior work and on the theme of limiting screen space usage.

2.3.1 Summary text

In previous studies of [12, 19] the omission of a summary text in a web SERP has shown to impact the performance (search time and click pattern) negatively, because of the lack of context. Another study by Granka et al. [20] also concluded that this is where the users focused most of their attention. Furthermore, this could be of interest to implement as a test design since the summary text of one result of KI's smartphone SERP (seen in Figure 2) consumes about one third of the screen (seen in Figure 2a) and saving space is important [8].

2.3.2 Person images

A study by Gossen et al. [21] concluded that images (thumbnails) in the SERP were appreciated by children but not so much by adults. This feature is interesting because it follows the Nudelman's suggestion of saving screen real estate [8]. In the KI search solution the one result type that differs

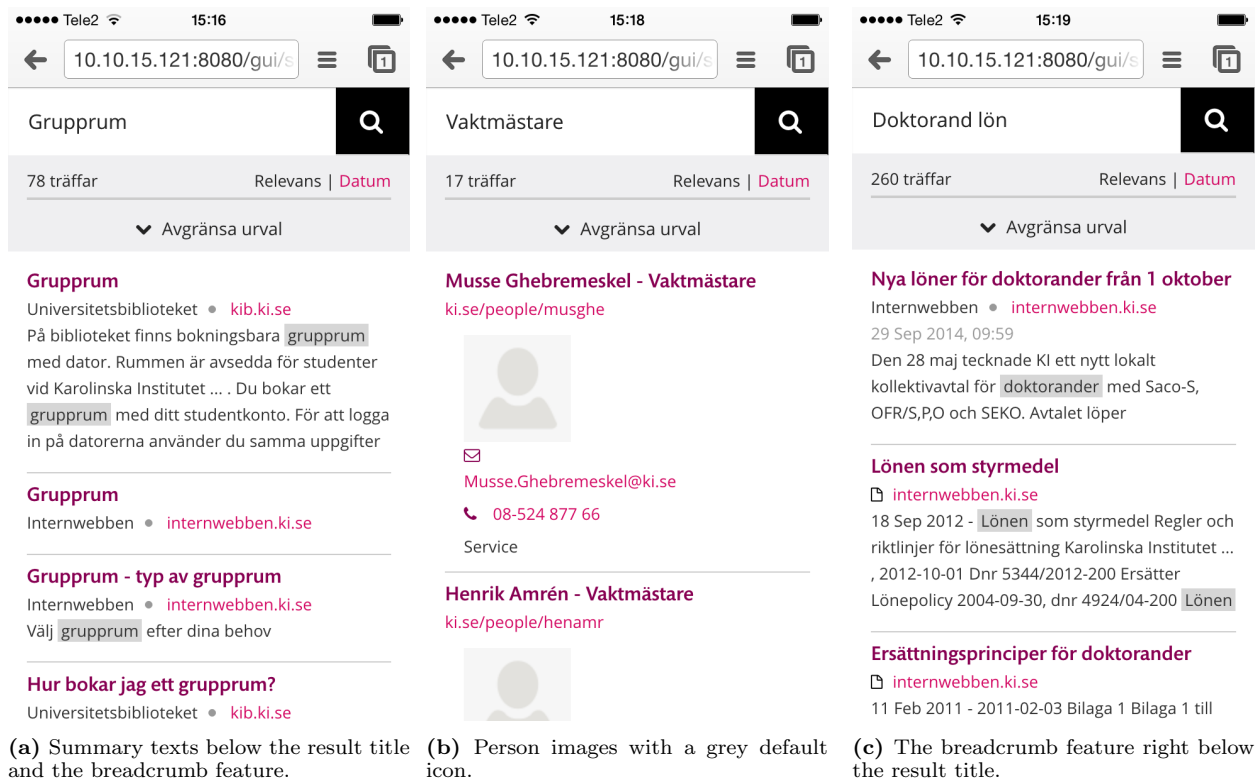


Figure 2: The original GUI in swedish.

the most from the other SERP cells' layouts is the one for persons (seen in Figure 2b) . In the structure of this cell the image of the person occupies about half of the space of the cell's area. Because of this, it would be interesting to know if the images have any efficiency impact on the user.

2.3.3 Categories

In the study by Dumais et al. [12], which was explained in section 2.1 Different types of desktop SERP layouts, the authors compared search results presented in a relevance sorted list to a category system. Their experiments showed that categories were more efficient than the relevance sorted list in all of the quantitative study cases and were also the most preferred when looking to the qualitative studies. In their experiments they tested to systematically eliminate contextual content from the category interface such as result titles and summary texts. The qualitative results of this prior work were also supported by the findings of Heimonen and Käki [15] where a categorised mobile GUI was preferred to a relevance list search layout. However, both of these prior studies implemented their respective solutions in a desktop layout and in early generation (2007) of a smartphone (Nokia 6680). Therefore it would be of interest to test this feature's impact in a current smartphone web GUI.

2.3.4 Breadcrumb

This feature is a navigational support tool to help the user to understand his or her position on the website in relation to the rest of the website. The *breadcrumb* is a textual representation of the website's structure [22]. It usually starts with a *breadcrumb item* (a text string) and is followed by a

separator (usually ">"). According to Nudelman [8] breadcrumbs can be divided into two main types; *historical* and *hierarchical*. Historical breadcrumbs, seen in Figure 3 displays the trail of a user's browsing pattern on a website - from the initial location to the current.

Home > Mobile devices > Smartphone > Apple > iPhone
Home > Brands > Apple > iPhone

Figure 3: The historical breadcrumb structure which is also a trail. The trail consists of different browsing patterns from the initial location to the current one.

Home > Cell phones > Apple > iPhone > Models
 iPhone6
 iPhone5s
 Skins
 Gadgets

Figure 4: The look-ahead breadcrumb. An hierarchical breadcrumb structure but with an added drop-down menu.

Instead of showing where the user has been the hierarchical breadcrumb displays the user's position in website's architecture (hierarchy). Furthermore, the hierarchical breadcrumb shows different layers within the website where the users can go - helping them to navigate through and between different sections of the website. A previous study by Blustein et al. [23] compared a version of the hierarchical breadcrumb called *look-ahead* breadcrumb to the standard version. This new type of breadcrumb, seen in Figure 4, has the same structure as the regular hierarchical breadcrumb

with an addition of a drop-down list containing categories from the same tier. From the study, the look-ahead breadcrumb proved to be more effective than the regular version.

The breadcrumb implemented in the KI web GUI has a hierarchical structure, which tells the user where in the website's structure he or she can find the source of the result or find similar web pages.

2.4 Variables of effectiveness and efficiency

Prior work [5, 12] which reported on experiments within the area of search GUIs measured mostly the same effectiveness and efficiency variables, which were completion time, number of clicks and eye tracking data (gaze time and gaze pattern). The dependent variables that will be measured in this thesis, quantitative as well as qualitative, are explained below.

In the book *Handbook of Usability Testing* [24] the authors states the following task measurements:

- Accuracy
 - Numbers of errors
 - * Omission, missing something
 - * Commission, doing something unnecessary
 - Requiring hints
 - Points of hesitation
- Time
 - Time of completion
 - * Mean
 - * Standard deviation

When measuring the efficiency the previous studies [5, 12] indicated that most of the variables from the book were used with the added measurement of number of clicks. For example the experimental user study of Xu et al. [17] evaluated the efficiency of their prototype through time of completion and the effectiveness of number of clicks, satisfaction feedback score and answer quality. However, these measurement variables are not set in stone but rather applied to suit the experiment. For example in the study by Heimonen and Käiki [15] the accuracy measurements were made up of the variables: relevant research selections and qualified search speed. In the study by Nilsson et al. [25] the effectiveness was defined as the performance of achieving a goal (success rate) and efficiency included the resources (time, clicks) used on the way to the goal, interpreted from the interpretation of the standardization of usability guidance [26].

In a study by Xu et al. [17] qualitative variables were also measured in the form of answers from *pre-study* and *post-study questionnaires* and quotations from *think-aloud*. Which of these variables that were used in this thesis is explained in the 3 Method section.

3. METHOD

A combined study of experiments, complementing think-aloud¹ and semi-structured interviews was performed on students from KI, in a controlled setting and in Swedish. The experiments were different versions of the KI smartphone web user interface; an original version and an implemented version. Before the experiments were created scientifically based hypotheses emerged, which are further explained in 3.2 Hypotheses. The measured variables of the experiments were also set before the implementation and were finally as in the table below, Table 1:

Quantitative	Qualitative
<i>Time of completion</i>	<i>Free text answers</i>
<i>Number of hints</i>	<i>Comments and quotations</i>
<i>Success rate of tasks</i>	
<i>Number of clicks (taps)</i>	

Table 1: The desired measured variables.

During the procedure, the participating students were randomly split into three sample groups and were asked to perform search tasks (in Swedish) on one of the implemented smartphone GUI designs as well as the existing KI design. The tasks consisted of nine (three for each group) questions which were developed by researching prior work and consulting the KI contact. Short interviews were held before and after the experimental tasks.

3.1 Participants

Initially the target group was first-year students at KI, because older students use another system (called Ping Pong) for getting information about their courses and schedules. The target group was changed to KI students from all grades due to the lack of participation, by the first recruitment try. Eventually a number of 30 persons (18 women and 12 men) participated in the study and were given individual test sessions which took place at KI in Solna, Sweden, over the period of three days. In a short pre-study questionnaire (shown in the test protocol in Appendix A) the students were asked to rate on a rating scale between 1-5, how often they used a web browser in the smartphone and how often they used the KI search solution. All the participants were very experienced in navigating in the web browser on a smartphone (4.8/5.0), but prior experience with the KI search solution was low (2.0/5.0). The previously mentioned sample groups were as follows:

- Group A >> No summary text GUI vs. Original GUI
- Group B >> No person images GUI vs. Original GUI
- Group C >> Categories GUI vs. Original GUI

3.2 Hypotheses

To help investigating the main research question, experiments were executed with the following hypotheses. The features which these hypotheses were based on are explained in section 2.3 Features of interest, FOI.

¹An interview method explained in Rubin and Chisnell book *Handbook of Usability Testing* [24].

H_A

You get a significant improvement in efficiency and effectiveness of the KI SRWUI by removing summary texts from the search results i.e. implementing the "Summary text" feature.

This hypothesis was based on prior work [12, 19] which concluded that displaying summary texts were more effective in a DWUI. In contrast to this conclusion the author of this thesis assumes the other since the experiments will be conducted only on smartphones.

H_B

You get a significant improvement in efficiency and effectiveness of the KI SRWUI by not displaying images of persons i.e. implementing the "Person images" feature.

The previous study of Gossen et al. [21] was an underlying factor for this hypothesis, where the authors could not conclude that displaying images were more effective.

H_C

You get significant improvement in efficiency and effectiveness of the KI SRWUI by removing the breadcrumb feature and replace it with categories.

This hypothesis was based on the findings in Heimonen and Käkis study which could show an improving effect by categorising SERPs [15]. The *breadcrumb* feature is explained more below in 3.4.1 Substituting the breadcrumb feature.

3.3 Experiment tasks

Experimental tasks were performed in some previous studies [12, 15]; and were based on information seeking and their goals were for the user to reach a SERP that validated a certain information need. The goals and tasks, for this thesis, were created from consulting with the KI client contact in order to clarify what the users most likely would search for.

When the tasks were created they followed the suggested results of Raneburger et al. [14], to have a well pre-defined scenario. In order to refine and accomplish the tasks they were pilot tested by Computer Science students, in two iterations. The first version of tasks were created to contain the investigated FOI when searching as well as in the goal SERP. The first version of tasks proved to take too much attention from the feature, as the pilot testers were more focused on completing the actual task. The second version followed the previous study by Gossen et al. [21], where two of the tasks were navigational based and one was informational based. The final version of tasks resulted in a number of three tasks per group (nine in total), with two of them being navigational and one informational. The list below contains the tasks for each group and the informational task is marked with a star (*).

Group A. No summary text GUI

1. Find an address to a group study room at KI.
2. Find where you can heat up your food on campus.
3. What is the loan period for other books than course literature?*

Group B. No person images GUI

1. Find a caretaker who does not have a phone number in their profile.
2. Which head of communication manages external communication?*
3. Find contact information to a librarian.

Group C. Breadcrumb to category GUI

1. How much salary does a PhD have at KI? Find a news article about it.*
2. Find the event for the next lunch lecture with a professor.
3. Find the event for the next *Coffee Hour*.

The tasks were also designed to be searched for in Swedish, because there is a larger amount of data in this language and it is also the default language of the website.

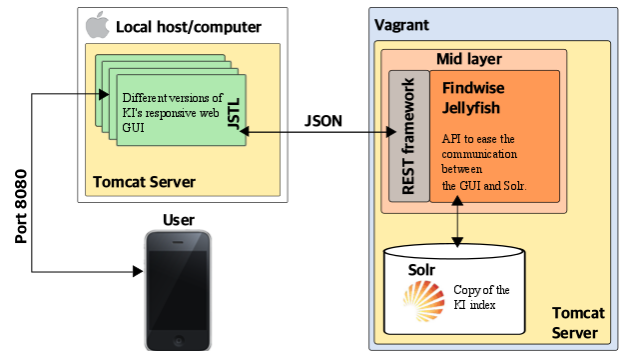


Figure 5: A structural overview of the technical architecture.

3.4 Implementation of features of interest

Before implementing the FOIs measurement variables and hypotheses were decided, which were explained in the 3 Method section and in the section 3.2 Hypotheses. Then, conceptual sketches were developed on paper and transferred into the wireframing software Balsamiq². The categories feature was not implemented like the other two by deleting the feature. Instead categories substituted a feature in the KI solution called *breadcrumb*, which is explained in 3.4.1 Substituting the breadcrumb feature. From the designs the high-fidelity versions were developed with the same tools that Findwise used in their development of the KI search solution. Findwise helped to set up a virtual environment on the test computer³ using the develop environment configure tool Vagrant⁴. The implementations of the FOIs can be seen in Figure 6.

²<http://balsamiq.com> (accessed 2014-11-26)

³MacBook Air '2012

⁴<https://www.vagrantup.com/> (accessed 2014-11-26)

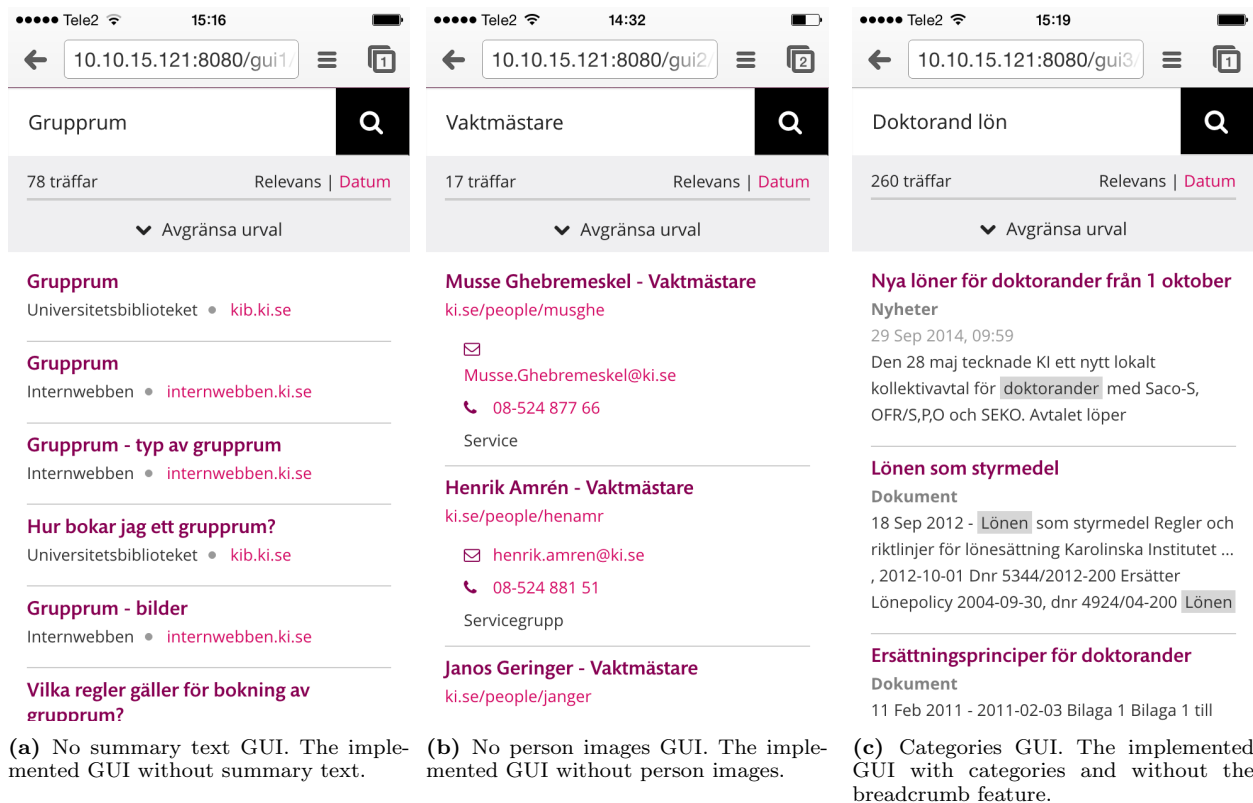


Figure 6: The developed alternative smartphone GUI designs.

A brief walkthrough of the architecture (seen in Figure 5): on the client side the user sends a search query to the Tomcat server on the computer, which hosts the different GUI versions. The different GUI versions were implemented with a tag library JSTL⁵. The web GUI communicates with the server side (which is hosted in a virtual environment) through JSON⁶. The JSON is then interpreted by a server framework called REST⁷ to fit for Findwise's Jellyfish framework⁸. Jellyfish then sends the query to the search platform Solr⁹ where a copy of all the parsed documents and web pages of KI can be accessed, the so called *index*.

3.4.1 Substituting the breadcrumb feature

A noticeable feature in the KI search solution is the so called *breadcrumb* which is placed right below the result title and can contain three parameters, see Figure 2c. The parameters being the organisation, the organisation's hyperlink and hyperlink to where the web page can be found. The purpose of this feature (also explained in 2.3.4 Breadcrumb) is to help the users in their navigation by revealing the result's structure. Since the feature contains some non-intuitive parameters (the first and second) it would be interesting to replace them with only a category. Space is saved by doing this but categorical sectioning has also proved to be effective

⁵<https://jstl.java.net/> (accessed 2014-11-26)

⁶<http://www.json.org/> (accessed 2014-11-26)

⁷<http://restclient.org/> (accessed 2014-11-26)

⁸<http://www.findwise.com/services/glossary/jellyfish-component> (accessed 2014-11-26)

⁹<http://lucene.apache.org/solr/> (accessed 2014-11-26)

in the findings of Dumais et al. and Heimonen et al.[12, 15]. The breadcrumb feature was substituted by a bold category text.

3.5 Procedure

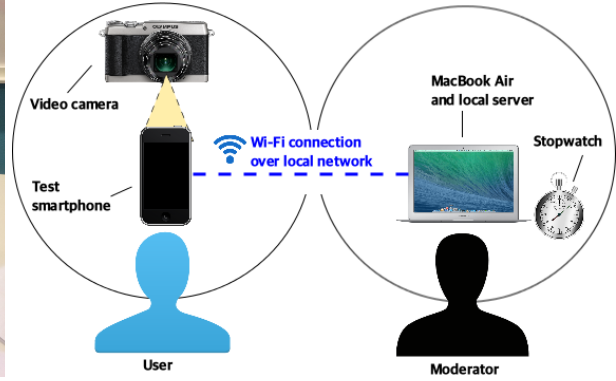
Before the start of each experiment session, the participants were asked if they wanted to use their own smartphone or an iPhone5s or a Samsung Google Galaxy Nexus. However, there were complications with the Wi-Fi network at the lab venue (seen in Figure 7a) and the participants mostly did not use their own smartphones (6 of 30 did), because of the session time limit of 15 minutes. The user connected as a client to one of the SRWUI versions on the local host, a MacBook Air computer. An overview of the experimental setting and the experimental venue can be seen in Figure 7.

First, the short pre-study questionnaire was performed as explained in the section 3.1 Participants. Then, the participant was assigned a group (A-C) randomly and was instructed to complete three search tasks of that group, with a time limit of two minutes per task. The participant was informed about the time limits of the tasks (2min) and the whole experiment (15 min). The same tasks were performed on the implemented test design as well as the original design. The participants were also asked to perform think-aloud during the experiment. After each task the participants were asked to rate the difficulty of it, on a Likert scale (1-5).

During the experiments, speech and video were recorded of the smartphone screen and the hand motions of the participants. When the tasks were completed a semi-structured



(a) Picture of the test environment and the setting.



(b) Overview of the setting.

Figure 7: The experimental setting.

interview followed. Before leaving, the participants were rewarded with a cinema ticket which was sponsored by KI.

In order to counteract the familiarity of tasks, since the participants were assigned the same tasks for both designs, the order of the designs were reversed for half of the participants in each group.

The time was measured for each task and the stopwatch started when the search input field was empty and the task was acknowledged by the participant. The time stopped when the participant was done, or if two minutes passed. Hints were also given if the participant clicked too far away, which is three pages away from the SERP.

Time, hints and difficulty and other user expressions and actions, were all filled in a test protocol, see Appendix A.

4. RESULTS

In this section the results of the experimental study are presented. It starts off by analysing the measured quantitative data from the users' performance during the experiments, which will answer how effective and efficient the different features were.

When measuring, the sample groups were the independent variables in all the experiments and the dependent variables were the ones listed in section 2.4 Variables of effectiveness and efficiency. Wilcoxon's signed rank test (tested at $p < 0.05$) and Pearson's correlation coefficient (PCC) test were conducted on the quantitative data. The Wilcoxon's signed rank test was used because of the small number of samples (30) and because they were paired as well as dependent. The PCC test is a statistical method that checks if two or more variables are related to each other. In statistics the PCC is a value between -1 to +1 where the minimum and maximum values are perfect correlations and 0 meaning that there are no correlations at all. The PCC test was used to support findings from the Wilcoxon's signed rank test.

Finally, qualitative data from the post-questionnaire and the think-aloud comments during the experiments, will be presented. This data was analysed by gathering all the quotes from transcriptions of the video recordings and notes from the post-study questionnaires. The qualitative data was also analysed, within groups, with a deductive approach to find similarities and differences. These qualitative findings answered which kinds of features that were important for a smartphone responsive web user interface.

4.1 Effectiveness

To measure effectiveness a scoring system (Table 2), similar to the one in the study by Gossen et al. [21], was made based on two variables. The two variables being whether the participant reached a solution and the amount of required help in form of hints.

Score	Explanation
4	The user successfully solved the task without any help.
3	The user successfully solved the task with one hint.
2	The user successfully solved the task with two or more hints.
1	The user did not solve the task.

Table 2: Rating of the participants' task performance.

In all of the experiment sessions but four (4) the participants successfully solved all of their assigned tasks. The failing participants all belonged to sample group A - the one comparing search results with and without summary texts. The task that the participants failed at was the same task for everyone and it was task 1 of the alternative SRWUI of A. In two out of four of these failed tasks the participants did not solve the task for neither of the test SRWUIs.

The Wilcoxon's signed rank test showed no significant difference of the participants' success scores between the test design and the original design in any of the groups, as can be seen in Table 3.

Group	A	B	C
p -value	0.155	0.477	0.475

Table 3: The Wilcoxon’s signed rank test between success scores.

Looking at the ranks of the Wilcoxon’s signed rank test (in Table 4) of the success scores there is a very small (6 of 30) positive rank advantage in group A, meaning that the original design was slightly more successful. In group B there was a small advantage for the alternative design of B. In group C there was no notable significance at all.

Original - implemented	A0 - A	B0 - B	C0 - C
Negative ranks	2	7	6
Positive ranks	6	4	7
Ties	22	19	17

Table 4: Ranks of the Wilcoxon’s signed rank test between success scores. The original designs A0, B0 and C0 are subtracted with the implemented designs A, B and C.

As the results from the success scores do not show a significant effectiveness improvement, in any of the groups, the alternative hypotheses (H_A , H_B and H_C) as well as the corresponding null hypotheses must be retained.

4.2 Efficiency

The efficiency variables that were measured were; the completion time of the tasks and the number of taps executed during the experiments. The completion time was measured during the experiment and the number of taps were counted thoroughly by watching the video recordings.

4.2.1 Completion time

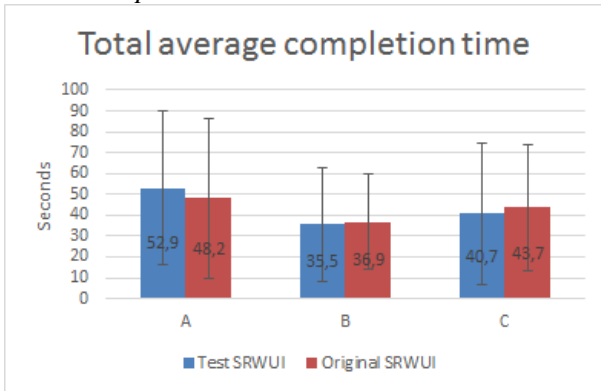


Figure 8: Average time of completion with standard deviations for group A, B and C.

Comparing the completion times between the test designs and the original design, just by looking at Figure 8, small time differences can be observed. The participants were slightly faster when they had access to the summary text, in other words the original GUI of A was more efficient time wise, which agrees with the studies of Dumais et al. and Clarke et al. [12, 19], but disagrees with the alternative hypothesis H_A . For the other groups (B and C) the participants’ times were faster when they were using the alternative GUIs, thus agreeing with the alternative hypotheses H_B and H_C . However, the Wilcoxon’s signed rank test showed no

significant difference in completion time between the original design and the test designs seen in Table 5.

Group	A	B	C
p -value	0.627	0.577	0.624

Table 5: Results of the Wilcoxon’s signed rank test between completion times of the original and test designs.

The difference in average completion time between the different groups depends on the difficulty of the task. Shown in the individual graph of group A, in Appendix B, the first task is the one that has the highest completion time by far. The first task of group A was also the only one that was unsuccessful according to the previous finding in section 4.1 Effectiveness. The PCC, seen in Table 6, also revealed that there was a positive correlation; the completion time increased as the tasks were rated more difficult - which made the correlation evident.

Group	A	B	C
PCC	0.669	0.747	0.759

Table 6: Correlation between completion time and the rated task difficulty.

Just by studying the completion time there was no strong (significant different) evidence to be found that either of the test designs were improving the efficiency.

4.2.2 Number of taps

When measuring the tap count Walmsley’s definition of a tap was used [27], where actions such as swipe, flick and typing on the keyboard were not counted.

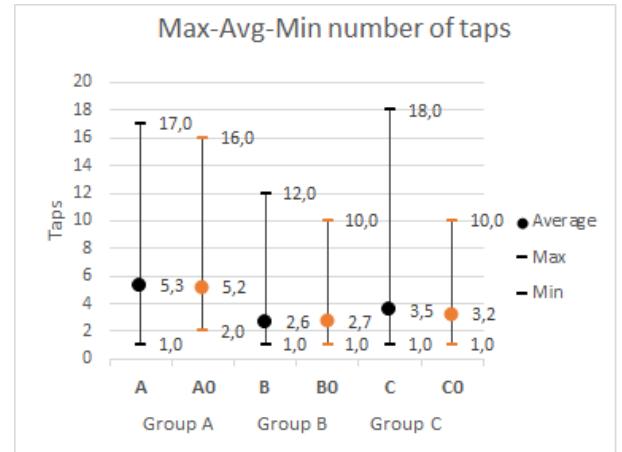


Figure 9: The average and maximum and minimum number of taps. The original GUIs are A0, B0, C0 and implemented GUIs are A, B, C.

The average number of taps, seen in Figure 9, did not have a significant difference between the test designs and the original design. The Wilcoxon’s signed rank test also showed that there was no significant difference, seen in the Table 7.

Group	A	B	C
<i>p</i> -value	0.830	0.339	0.902

Table 7: The Wilcoxon’s signed rank test between number of taps.

However, there was a positive correlation between the number of taps and the task difficulty, seen in Table 8 - when the task was difficult the number of taps increased. This also explained why the maximum number of taps were at least three times higher than the average. The higher number of taps represented the participants that had a hard time during the tasks.

Group	A	A0	B	B0	C	C0
PCC	0.564	0.711	0.819	0.458	0.821	0.613

Table 8: Correlation between number of taps and the rated task difficulty. Where the original designs are A0, B0 and C0 and the implemented designs A, B and C.

The PCC test also showed that there was a strong positive correlation between the number of taps and the completion time of the tasks, observed in Table 9. The efficiency results in this thesis follows the traditional findings of user performance of search interfaces [28].

Group	A	A0	B	B0	C	C0
PCC	0.903	0.929	0.863	0.852	0.784	0.845

Table 9: Correlation between completion time and the number of taps. Where the original designs are A0, B0 and C0 and the implemented designs A, B and C.

The results of the two efficiency measurements, completion time and number of taps, did not prove any significant efficiency improvement, for either of the implemented SRWUIs. Therefore the alternative hypotheses (H_A , H_B and H_C) as well as the corresponding null hypotheses must be retained.

4.3 Post-study questionnaires

To get a clearer perspective of the participants’ experiences with the different smartphone GUI versions, this section will analyse the feedback from the post-study questionnaires, seen in Table 10, and the think-aloud comments from the experimental sessions.

Post-study questionnaire questions
Q1: Did you detect any difference between the two designs?
Q2: Did you prefer a particular design?
Q3: Did you notice anything else that you thought were good or being annoying?
Q4: After telling the difference to the participant. Did you notice this? What do you prefer?

Table 10: The post-study questionnaire translated from the test protocol in Appendix A.

All of the questionnaire answers and comments from the sessions were extracted from the transcription of the video

recordings and from the notes of the test protocols (in Appendix A). The qualitative data was analysed with a deductive approach within groups. The deductive analysis approach were performed with the focus on detecting similarities and differences, of the participants’ opinions about the different smartphone GUI versions.

4.3.1 Group A

There were only three of ten participants that spotted the difference between the designs, when question Q1 was asked.

Through the feedback of Q2 it was also evident that familiarity of the tasks affected the participant’s decision. Three of the participants preferred the second tested design, because they knew what to search for. The quote: *“The other was better because you knew what to look for and you got a fast response.”*, highlights an issue that was common in all the groups - that fast response and a less effort spent is highly satisfactory, which also agrees with the results of Xu et al.[17].

After revealing the difference between the designs, three of ten participants argued that they preferred alternative GUI of A and the reason being that the SERP was easier overviewed. However, two of them preferred no summary texts if the search result titles were informative enough with the following reason: *“It is easier to see more results at the same time. If the answer is in the title it is better with the first (the alternative GUI).”*

The rest of the participants preferred the original design of A as it was more informative, which agrees with findings of Clarke et al. and Dumais et al.[12, 19]. A specific suggestion from a participant, who also preferred to have summary texts said: *“...It would be nice with a ‘more-function’ if you search longer texts.”* The user wanted the “more”-function to hide and reveal longer summary text snippets when the result was insufficient. This function is actually something that was considered after the pilot tests instead of implementing the a GUI version without the person image feature. This was discarded because it was too similar to the “Summary text” feature. Another function that two of the participants observed was the highlighting of search hits which was appreciated.

4.3.2 Group B

None of the participants in the group were able to pin-point the FOI - the omission of person images. An explanation for this was that the existing profile pictures in the SERP of the original design of B were actually not linked to a person image and was now presented as a grey picture icon (seen in Figure 6b). After Q4 was prompted the majority thought that they would have noticed a difference if there were actual portraits of people in the search results.

The familiarity effect was evident for this group as eight participants preferred the later design tested to the first as it was faster and easier to search in. One of the quotes being: *“The other one because I knew what to search for.”*

After explaining the difference there were three participants that preferred the implemented design of B. A participant said: *“If there is no real image, it is not needed. I only look*

for the name and contact information when I search.” This was the main reason why these participants preferred the implementation - that the contact information was the most important attribute to look for. However, six out of ten participants thought that an implementation of the feature was situation dependent. One of the participants quoted: *“It depends if you have seen the person before. But I would not prefer the image if I am searching for a text. It is more important with facts.”* which tells that the use for person images depends on if the participant had seen the person they searched for or not.

Another feature that was highlighted in the feedback from Q3 was the relevance of the search results: *“You were forced to scroll because the results were not at the top. Person profiles should be at the top if you search for them.”*, which was not satisfactory. The participants encountered and clicked other types of results before person profiles popped up on the screen, which gave non-satisfactory answers. This phenomena of pressing results displayed at top has also been found in prior studies [3, 4], which implies that relevance is an important feature. This was not investigated in this thesis by recommendation from the supervisor at Findwise, since it is a huge component of the search solution.

During the experiments most of the participants used the existing smartphone devices because of the network complication. However, there was no major negative affect to be observed in the results. A participant said *“I got used to the smartphone.”* and this was the only participant that expressed that using an unfamiliar device had an affect on him or her. In addition, this particular participant did not own a smartphone, but instead a tablet, and the user did not explain how this affected him or her.

4.3.3 Group C

In group C only one participant detected the feature of categories instead of the breadcrumb feature. One explanation for that was that this feature occupied the least amount of space out of the three FOIs and that it is also substituted instead of removed. The fact that the breadcrumb was not a space demanding feature was mentioned by half of the participant. Another explanation that was evident by viewing the video recordings, was that the participants were mostly just focusing on the task and not paying attention to what happened with the GUI layout. This occurred in the other groups as well, but only expressed in the post-study questionnaire of this group: *“No, I did not detect any difference. I was too focused on finding what I searched for.”*

Four of the participants were also affected by the task familiarity and preferred the second one with the reason being speed and knowing the task beforehand. As one participant expressed: *“The second was easier as I knew what words to search for. I also knew that writing ‘ki’ was not good because I got too many hits.”*

A distinct validation, eight out of ten, for the implemented C design arose after showing the difference. One of the two participants that did not explicitly preferred the alternative design of C said: *“A combination of categories and breadcrumb is good, as you are able to see where the result is coming from. It is particularly good on a large search en-*

gine.” These two participants either thought it was situational based or that a combination of the categories and the breadcrumb would be the best choice. In addition, neither of the participants knew of the buzzword *breadcrumb* and this was the feature that needed further explanation compared to the other two.

In the feedback from Q3 two participants brought up the sorting of dates being in reverse order for task 2 and task 3 (tasks about events), and that it was not preferable. Two others also wanted to see an auto complete and spellcheck features, which was disabled in the version of the GUI acquired from Findwise.

5. DISCUSSION

The main purpose of this thesis was to investigate what kind of features that makes a search solution effective, by performing experimental studies on the KI search solution in a smartphone environment. The experimental studies were a combination of experiments with think-aloud and short interviews. These studies compared the effectiveness and efficiency variables of three implemented smartphone web GUIs, against the released product. By comparing the implemented designs in pairs with the released product it was shown by using statistical methods, that the empirical quantitative data, did not show any significant improvements in effectiveness or efficiency. The qualitative data from the post-study questionnaire feedback however highlighted a few different aspect of an effective SRWUI.

5.1 The results

Observed from the completion times in the quantitative results group A agreed that the original web GUI design was more efficient than the test design without result summary texts. These results suggests that summary texts are in fact efficient features when searching on a smartphone. The findings are also supported by the results of the prior work of Dumais et al. and Clarke et al. [12, 19]. This result differs from the other two experiment groups as the implemented designs for group B and C would be more efficient than the original, according to Figure 6. However, the differences were very small and as the statistical test did not show a significant difference it is not entirely sure to assume that the implemented version is better than the original. It is also worth mentioning that, as reported in 4.3.1 Group A, only three of ten participants spotted the difference between the designs which could explain why the feedback from Q2 was inconsistent. Inconsistent in the sense of participants preferring neither of the designs or that the differing feature was something else than the FOI. Detecting other noticeable features also appeared in the other two groups but the participants in group A were the ones who spoke about it the most. The author believes that an explanation to why this was most occurring in group A is that the summary text appears the same in the KI SRWUI, as in other search engines; below the result titles.

By analysing the users’ answers from the post-study questionnaire, especially Q4, together with the quantitative results, the users’ perception of the alternative GUI of C becomes clearer. A high number (8 of 10) preferred the use of categories, after revealing the differences between the two GUI versions. The two participants that did not totally

agree with this feature suggestion did not disagree either, but rather wanted a combination of both of the tested GUI designs. The users' perspectives strongly approves that categories are an important feature for streamlining search solutions, which supports the prior work of Dumais et al. [12]. The results were still too indecisive to be entirely certain, regarding efficiency and effectiveness for this feature. However, the completion time slightly suggests that a category is more efficient than the breadcrumb feature.

The performance results for group B were disambiguous as the efficiency results were suggesting different designs to be the most efficient. Likewise there was also an uncertainty in the qualitative results which design that was more preferable, since six of ten participants did not prefer any design as they thought it depended on the situation. An explanation for this result was that this was also the only group where none of the participants recognised the GUI difference. Therefore it is clear why the result did not show a significant improvement in efficiency or effectiveness, as the change in the GUI was not detected.

Another notable difference that was observed in the completion time's graph (Figure 8) were the actual differences between the groups' completion times. An explanation to this was that the completion times were in correlation with the task difficulties. This result strongly suggested that this was the case as the completion times of group A were the highest and that its participants were the only to fail a task. As tasks probably were strongly related to time measurements it is important to spend time when creating them, which is also recommended by Rubin and Chisnell [24]. The results of the experiments in this thesis also were a proof of that the tasks were not well balanced even after two pilot tests, between the groups. Because of this, it was very likely that the tasks and the experiments were the factor that affected the results the most in terms of not observing any statistically significant results.

To get the participants more familiar with the tasks and the tested GUI some preparatory test could have been made to balance out the variance of the quantitative results. A study of Deshpande et al. [29] showed that the users' task completion effectiveness and efficiency improved after conducting preparatory tests of GUIs before the actual experiment task. In the case of this master thesis the suggestion would have been to let the users search with the KI SRWUI for a while before starting with the experiments. The outcome of this would hopefully follow the results of Deshpande et al. [29] and even out the participants' level of experience with the KI GUI and as a result lower the variance.

Another approach would be to make the group dividing process different. Instead of assigning the participants to a group randomly they could have been divided into three different groups, depending on the results from the pre-study questionnaire (regarding smartphone search experience). The theory of splitting the groups according to prior experience is that within the groups the participants are equally used to the test program environment. If this theory also was executed with the participants experimenting on all the different GUI versions, connections about how the participants' experience level suits different features could

have been drawn. These connections, between participants' experience and the features are something that the current thesis can not conclude from its results.

5.2 Focus on quantitative data - a suitable method?

By choosing experiments as the main method in this thesis a lot of time was spent on designing and planning the experiments. As mentioned before, the tasks were a major part of this study, because they demand a lot of planning [24]. In retrospect, before starting the experiment creating process, a pre-study with KI students would have been a better way to go than generating tasks from related work and previous research. As shown in the results the tasks were probably a bit hard as there were only a few cases where the participants actually discovered the tested feature. This was also somewhat expected since the participants of pilot tests were Computer Science students. However, because of time constraints there was not enough room for perfecting the task process. The time constraint depended on development environmental setup problems, which were conflicting software versions and operating systems.

Something that is important to acknowledge is also the variables that could not be affected during the experimental design phase. The main one being the delimitation of not removing or adding more than one feature. This unwritten delimitation was implemented because the analysis process would have been hard if not impossible to execute otherwise. The differences between the SRWUIs quantitative data would have been disambiguous with a smartphone web GUI of multiple features. However, a GUI design with a lot of changes would probably have increased the participants' results of detecting a difference between the different design suggestions. This would probably get significant results in a pure explorative study within the search web GUI area.

When conducting the experiments a possible risk was that it would be hard to find 30 subjects for the experimental study. This was why it was decided that there would be a short time limit (two minutes) during the experiment tasks. The time limit was also something that constrained the tasks' depth and their difficulty. The fact that the participants knew that a time limit existed before the experiment started, could have induced stress. This could also have affected the participants negatively by making them overlook certain aspects of the layout. Even if this was not measured it is still likely to have been a factor to why so few of the participants did notice the difference between the two design versions.

Another limitation of the method was the quantitative data analysis method. This was the dependent non-parametric Wilcoxon's signed rank test and it had been used before in Gossen et al.'s study [21], for similar kinds of statistical analysis. A statistical method that was considered before the Wilcoxon's signed rank test was the Mann-Whitney U test. However, further investigation [30] of the Mann-Whitney U test showed that it did not suit the type of data, as it compares two independent samples rather than paired samples. It is believable that another experimental approach such as the one in the study by Trattner et al. [28] would perhaps have produced significant results. In this prior study the participants tested all the different GUI versions. This

would however demand a change of the whole experimental design. Otherwise if there was no way to change the design another approach would be to increase the number of subjects in the study to get a large sample size, which would boost the statistical power. By using this, other statistical methods which have been observed in most of the prior work [5, 12, 28] would be feasible, such as ANOVA. However, within the thesis' time limit of 20 weeks an experimental study (planned to be executed during 2 weeks) with 30 participants which generated both quantitative and qualitative results were something that the author should be and is proud of.

Even if this thesis focuses on the quantitative results the qualitative results delivered clearer opinions about the GUI versions' differences. Perhaps a better approach for this thesis would have been to focus on user feedback and satisfaction, not the other way around, since satisfaction is a deciding factor in GUI development [8]. That aspect is philosophical for now but can be investigated in future work.

5.3 Implications for KI, Findwise and smartphone search solutions

The only real change that is suggestable is to apply categories to the KI SERP. The GUI change by itself is small but would be rather time consuming for Findwise and costly for KI, if not timed well with a software deployment, according to the author's supervisor at Findwise. This is because the search solution is a responsive website that is implemented for desktop and smartphone. The change of this feature only concerns the smartphone layout as of now, but a professional implementation would result in changes in more than one place. This would however provide a more intuitive search results page layout where the user could understand all the featuring elements, since none of the participants understood the purpose of a breadcrumb feature. However, by implementing categories it is important to contemplate; how much Findwise and KI actually would profit from this versus how much resources they would be spending. This change, switching breadcrumb to categories, actually did not show any revolutionising improvements to the KI GUI. Therefore, the author believes that the category feature by itself will not be profitable for either of the organisations. Nevertheless it is believable that if several more suggestions, of small improving features, were to be implemented the profit value would be higher. The drawback of implementing and changing several features in an already existing product is that the risk of complications will be higher. Also, an implementation of this sort may affect the larger and more complex components and there could also be complications between the different implemented features.

The findings in this paper, which implies that summary texts and categories are efficient features, still agrees with prior work [5, 12, 15] and to some extent the idea of saving space [8]. However in prior research, there has been a lack of pointing out concrete features which contributes to an effective and efficient search GUI. This paper will then contribute to the area of smartphone search GUIs by pointing out that informational value, in the form of summary text, is a key feature to an efficient and intuitive presentation of a search result page. Categories showed in older work to be efficient in both mobile GUIs and desktop GUIs, but this

thesis' results hopefully proved that it is viable for modern smartphone responsive web GUIs as well.

5.4 Future work

Besides of focusing on user satisfaction other features that are interesting to investigate in terms of effectiveness and efficiency are the ones from the post-study questionnaires, highlighting of search queries and *more*-function. It would also be interesting to investigate a desktop view as well, because most of the related work found around GUI efficiency and effectiveness [5, 12, 17] are conducting experiments on search GUIs suited for desktop. One interesting approach would be to explore the efficiency and effectiveness of a search GUI in desktop mode and then compare these results to the smartphone's. Furthermore, to implement these results into an optimised search GUI would be another interesting issue for future work, as the results from this current thesis were not significant enough to do so. However, the consequences of an implementation of this size would be rather time consuming for Findwise. The consequences would be as described in the section above 5.3 Implications for KI, Findwise and smartphone search solutions.

6. CONCLUSION

In this thesis three implemented GUI versions of Karolinska Institutet's search solution were compared to the original version, through an experimental study with the focus on effectiveness and efficiency with the research question:

In a search solution which kind of features are important for making an effective smartphone responsive web user interface?

The experiments revealed no statistically significant improvement of effectiveness or efficiency, for any implemented GUI. However, the qualitative results showed that summary text and categorisation were important in a search web GUI, which increased the user's performance. The feedback also revealed that images of persons are preferable when they are known to the one who searches.

7. ACKNOWLEDGEMENTS

The author would like to thank Johan Holmström, the supervisor at Findwise, for the assistance during the whole project with practical expertise as well as helping out with the contact to KI. In addition, the author would like to thank Maria Sjögren for assisting with experiment venue and student relations at KI, as well as serving as the main contact at KI. Also, the author would like to thank Leonard Saers at Findwise for helping to configure the test environment. Finally a thanks to Jonas Moll, supervisor at the Royal Institute of Technology, for his academic assistance and insightful comments.

8. REFERENCES

- [1] J. Cho and S. Roy, "Impact of search engines on page popularity," in *Proceedings of the 13th international conference on World Wide Web*, pp. 20–29, ACM, 2004.
- [2] Z. Guan and E. Cutrell, "An eye tracking study of the effect of target rank on web search," in *Proceedings of*

- the SIGCHI conference on Human factors in computing systems, pp. 417–420, ACM, 2007.
- [3] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, “Accurately interpreting clickthrough data as implicit feedback,” in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 154–161, ACM, 2005.
 - [4] A. Aula, P. Majaranta, and K.-J. Räihä, “Eye-tracking reveals the personal styles for search result evaluation,” in *Human-Computer Interaction-INTERACT 2005*, pp. 1058–1061, Springer, 2005.
 - [5] Y. Kammerer and P. Gerjets, “How the interface design influences users’ spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface,” in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 299–306, ACM, 2010.
 - [6] Accenture, “Mobile web watch 2012.” <http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Mobile-Web-Watch-Internet-Usage-Survey-2012.pdf>, 2012. (accessed 2014-10-24).
 - [7] E. Marcotte, “Responsive web design.” <http://alistapart.com/article/responsive-web-design>, 2010. (accessed 2014-10-23).
 - [8] G. Nudelman, *Designing Search: UX Strategies for eCommerce Success*. Wiley Publishing Inc, 2011.
 - [9] L. Wroblewski, *Mobile First. A Book Apart*, 2011.
 - [10] A. Ghazarian, “The pros and cons of responsive web design vs. mobile website vs. native app.” <http://designmodo.com/responsive-design-vs-mobile-website-vs-app/>, 2010. (accessed 2014-10-23).
 - [11] M. Bryson, “How common are seo problems with responsive web design?.” <http://searchengineland.com/how-common-are-seo-problems-with-responsive-web-design-152672>, 2013. (accessed 2014-10-24).
 - [12] S. Dumais, E. Cutrell, and H. Chen, “Optimizing search by showing results in context,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 277–284, ACM, 2001.
 - [13] R. Weisman and J. Bar-Ilan, “Intranet search patterns in a complex organization—the hybrid information model,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 277–284, AIS Electronic Library, 2010.
 - [14] D. Ranenburg, D. Alonso-Ríos, R. Popp, H. Kaindl, and J. Falb, “A user study with guis tailored for smartphones,” in *Human-Computer Interaction-INTERACT 2013*, pp. 505–512, Springer, 2013.
 - [15] T. Heimonen and M. Käki, “Mobile findex: supporting mobile web search with automatic result categories,” in *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pp. 397–404, ACM, 2007.
 - [16] S. Jones, M. Jones, G. Marsden, D. Patel, and A. Cockburn, “An evaluation of integrated zooming and scrolling on small screens,” *International Journal of Human-Computer Studies*, vol. 63, no. 3, pp. 271–303, 2005.
 - [17] S. Xu, T. Jin, and F. Lau, “A new visual search interface for web browsing,” in *Proceedings of the second ACM international conference on web search and data mining*, pp. 152–161, ACM, 2009.
 - [18] G. Schmiedl, M. Seidl, and K. Temper, “Mobile phone web browsing: a study on usage and usability of the mobile web,” in *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, p. 70, ACM, 2009.
 - [19] C. L. Clarke, E. Agichtein, S. Dumais, and R. W. White, “The influence of caption features on clickthrough patterns in web search,” in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 135–142, ACM, 2007.
 - [20] L. Granka, M. Feusner, and L. Lorigo, “Eye monitoring in online search,” in *Passive eye monitoring*, pp. 347–372, Springer, 2008.
 - [21] T. Gossen, J. Höbel, and A. Nürnberger, “Usability and perception of young users and adults on targeted web search engines,” in *Proceedings of the 5th Information Interaction in Context Symposium*, pp. 18–27, ACM, 2014.
 - [22] B. L. Rogers and B. Chaparro, “Breadcrumb navigation: Further investigation of usage,” *Usability News*, vol. 5, no. 2, pp. 1–7, 2003.
 - [23] J. Blustein, I. Ahmed, and K. Instone, “An evaluation of look-ahead breadcrumbs for the www,” in *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, pp. 202–204, ACM, 2005.
 - [24] J. Rubin and D. Chisnell, *Handbook of Usability Testing*. Wiley Publishing Inc, 2008.
 - [25] E. G. Nilsson and A. Følstad, “Effectiveness and efficiency as conflicting requirements in designing emergency mission reporting,” in *I-UxSED*, pp. 20–25, 2012.
 - [26] E. DIN, “9241-11. ergonomic requirements for office work with visual display terminals (vdts)—part 11: Guidance on usability,” *International Organization for Standardization*, 1998.
 - [27] W. Walmsley, “Taps and swipes: Intuition vs. machine learning in ux design.” <http://minuum.com/taps-and-swipes/>, 2014. (accessed 2014-12-04).
 - [28] C. Trattner, Y.-l. Lin, D. Parra, Z. Yue, W. Real, and P. Brusilovsky, “Evaluating tag-based information access in image collections,” in *Proceedings of the 23rd ACM conference on Hypertext and social media*, pp. 113–122, ACM, 2012.
 - [29] Y. Deshpande, S. Bhattacharya, and P. Yammiyavar, “A study of the impact of task complexity and interface design on e-learning task adaptations,” in *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction*, pp. 19–27, ACM, 2013.
 - [30] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.

APPENDIX

A. TEST PROTOCOL

Testprotokoll

Namn: _____

För-intervju:

Hur ofta använder du ki.se:s sökmotor?	<input type="checkbox"/> varje dag	<input type="checkbox"/> >1/vecka	<input type="checkbox"/> 1/vecka	<input type="checkbox"/> 1/månad	<input type="checkbox"/> sällan
Hur ofta använder du webbläsaren i mobilen?	<input type="checkbox"/> varje dag	<input type="checkbox"/> >1/vecka	<input type="checkbox"/> 1/vecka	<input type="checkbox"/> 1/månad	<input type="checkbox"/> sällan
Hur ofta använder du en sökmotor när du söker?	<input type="checkbox"/> varje gång	<input type="checkbox"/> oftast	<input type="checkbox"/> hälften	<input type="checkbox"/> ibland	<input type="checkbox"/> aldrig

TEST1: A0 A00 B0 B00 C0 C00 **Fråga hur svår uppgifterna var!**

Uppg.	Tid	Hints	Svårighet	Kommentar
1				
2				
3				

TEST2: A0 A00 B0 B00 C0 C00 **Fråga hur svår uppgifterna var!**

Uppg.	Tid	Hints	Svårighet	Kommentar
1				
2				
3				

Efter-intervju:

- Upptäckte du någon skillnad mellan designerna?
- Var det någon design som du föredrog?
- Var det något annat du som du hakade upp dig på/lade märke till?
- Skillnaden var detta <feature>. Lade du märke till det? Går det att framhäva bättre?

B. COMPLETION TIMES PER TASK

