2018

CBS
COPENHAGEN
BUSINESS SCHOOL
HANDELSHØJSKOLEN

# Designing higher performing e-commerce search engines

## Design objectives from an evaluation of existing e-commerce search engines with users

MASTER THESIS IN BUSINESS ADMINISTRATION AND INFORMATION SYSTEMS (E-BUSINESS)

**EMIL HERMANSSON, EMHE15AD@STUDENT.CBS.DK**

Supervisor: Torkil Clemmensen
Date of submission: 2018-05-15
Pages: 80
Characters: (181.931)

# ABSTRACT

With the aim of formulating design science objectives for increasing performance of e-commerce search engines, the present research used a mixed methods approach to perform an extensive evaluation of current e-commerce search engine solutions, with actual e-commerce users. Divided into two phases, the study started, by the help of users following a think aloud protocol with different e-commerce related search tasks, with exploring and identifying e-commerce search engines, user behaviours and problems. By combining theory from the e-commerce and information retrieval domains, and the findings from the first phase, a framework for performance evaluation was then developed. In the second phase, more users evaluated the performance of three e-commerce search engines, using the produced framework. A quantitative analysis of the results found that the search engines Google.se and Prisjakt.se performed significantly better in four out of seven of the performance measures, than the search engine Elgiganten.se. Using Google.se, Prisjakt.se, user behaviours and the identified problems as a foundation, an analysis of the findings from both phases, combined with theory of what is feasible in search, allowed the researcher to develop recommendations, in the form of ten different design objectives for increasing the performance of e-commerce search engines.

The design objectives propose e-commerce search engines to (1) choose what products to index and make sure their information is complete, (2) allow users to filter and sort the categorized products already at the results listing, (3) allow users to use common natural words (e.g. "budget" or "lightweight") by helping in translating such words to product attributes, (4) carefully consider how much influence advertisers should have on the ads, as ads easily could cause relevance problems, (5) clearly show the destination of ads, in order not to confuse users, (6) implement optional personalization that shows its benefits for the users, (7) use the historical interests of users to make the search results more relevant, (8) use referral queries of other search engines, to understand and adjust the content to the users' information need even before they type a query, (9) implement a browser add-on that complements existing search engines with e-commerce related information and (10) implement a browser add-on that tracks the users' information needs between search engines, and adjusts the content thereafter. Given these research objectives, developers of e-commerce search engines should be able to increase the performance of their search engines.

# Contents

## List of tables



## List of figures



## List of abbreviations

IR: Information retrieval
IIR: Interactive information retrieval
SERP: Search engine results page

# 1. INTRODUCTION

This thesis evaluates current e-commerce search engines with users, in order to produce recommendations on how performance could be increased for e-commerce search. The result is a set of managerial recommendations in the form of design objectives for new entrants and practitioners in the industry of developing e-commerce search engines. The following section will describe the background and motivation for this research and introduce the research questions.

## 1.1. Background

The newspapers' headlines are telling a clear message, e-commerce keeps breaking new records worldwide. In the last 20 years e-commerce has grown from being negligible to approximately a tenth of the total retail sales worldwide (eMarketer.com, 2017). In USA we read, *"Cyber Monday becomes largest online shopping day in US history"*, increasing by 16.8 percent from last year totalling the enormous amount of $6.59 billion for a single day (CNBC.com, 2017). In China, Alibaba is even worse with Single's day sales of amassing *"/…/$25.3 billion, doubling 2016 Black Friday and Cyber Monday sales combined"* (Theverge.com, 2017). E-commerce has come to stay.

Transaction efficiency (i.e. keeping transaction costs low) is considered a key driver to the e-commerce development, where the search cost (i.e. cost of time for the user to find information/products) needs to be kept low for efficient transactions to take place (Amit & Zott, 2001). In Sweden alone, there are now more than 5000 e-commerce stores to choose from (Prisjakt.nu, 2017). Searching through all their products would likely prove unfeasible for a single user, this is why search engines have come to play a critical part for the success of e-commerce. Search engines take many forms; the internal search engines searching through a specific website (e.g. e-commerce store), the general-purpose search engines (e.g. Google), but also the specific e-commerce search engines (e.g. Prisjakt.se), often including price comparisons, reviews and product information, which have found an important role in organizing products online.

Though search engine market shares are hard to measure exactly, estimations suggest Google is holding between 76 % and 92 % (NetApplications.com, 2017; StatCounter.com, 2017) of the worldwide share (in July 2017). Therefore, there is no question that Google is a major player in the search engine market.

Research has also shown that many users use general purpose search engines for e-commerce search (Jansen & Spink, 2006). To strengthen its position in e-commerce search, Google introduced its own shopping comparison search engine Froogle already in 2003 to compete with the specific e-commerce search engines available. However, in recent years Google has further increased its efforts in counter the competition from specific e-commerce search engines, by placing Google Shopping (i.e. a newer version of Froogle) directly into the Google search results. This move by Google, has been very controversial from a competition perspective, due to the favourable position Google Shopping gets in Google search. A search listing otherwise dominated by fierce competition between e-commerce stores, various websites and other specific e-commerce search engines. As a consequence, The European Commission has threatened Google with fines of €2.42 billion for abusing its *"/…/dominance as search engine by giving illegal advantage to own comparison shopping service"* (European Commission - Press release, 2017). E-commerce search is currently a very hot topic.

## 1.2. Motivation

Theory and current research in information retrieval (IR, i.e. the academic sub-discipline of computer science that includes the designing of search engines) suggest that search is an extensive process, involving multitasking and multiple search episodes (Spink, Ozmutlu, & Ozmutlu, 2002). Some users tend to work with opening many tabs and windows, then switching between these (Obendorf, Weinreich, & Herder, 2007). Users also use many different information platforms in their search process (Ho, Lin, & Chen, 2012) and they often transition between different search engines with the same search task, coming from a general-purpose search engine (e.g. Google) and ending up searching internally in a specific website (Ortiz-Cordova, Yang, & Jansen, 2015). Jansen and Molina (2006) investigated the effectiveness of search engines for retrieving e-commerce links using a three-grade relevance scale (relevant, partially relevant and non-relevant) classification of links. Their findings suggested that the e-commerce links of general purpose search engines were on average partially relevant, and those of specific e-commerce search engines just slightly better (Jansen & Molina, 2006).

In economic research on consumer search, established literature have tended to show the opposite, that the search of users is fairly limited (De los Santos, Hortaçsu, & Wildenbeest, 2012; Koulayev, 2014). However, these studies have had major limitations (e.g. limited tracking of users, analysis of a limited industry) and a recent study by Blake, Nosko, and Tadelis (2016) could therefore shed new light on the comprehensiveness of consumer search, using data from the e-commerce giant eBay.com. The users in their study make on average 36 searches before a purchase, in a search process that takes days, sometimes weeks, and where many searches never end in an actual purchase. As pointed out in the introduction: a key to successful e-

commerce is lower transaction costs (i.e. a higher transaction efficiency) which includes search being efficient (i.e. search costs being low). Therefore, it is clearly a business problem when users have to conduct extensive searching before a purchase.

As initiated by the researcher's doubt in the current performance of e-commerce search, combined with the objective of design-science research which is to *"develop technology-based solutions to important and relevant business problems."* (Hevner, March, Park, & Ram, 2004, p. 83) this thesis will produce recommendations for increasing e-commerce search's performance. The recommendations aim to answer the first two activities of the design science research methodology (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), namely (1) *Problem identification and motivation* and (2) *Defining the objectives for a solution*. The problem identification and motivation of performance in e-commerce search is already described in the introduction, but will be further investigated to find their underlaying cause. Defining the objectives for a solution requires a deeper understanding of the problem together with knowledge regarding what is feasible, including knowledge about *"current solutions, if any, and their efficacy"* (Peffers et al., 2007, p. 55).

When designing new IT artefacts, a key to success is starting out by identifying problems and opportunities in the actual application environment (Hevner & Chatterjee, 2010). The problems and opportunities are together with existing artefacts, theories and analogies/metaphors used as inspiration for creative design activities (Iivari, 2007). The overall goal of the current research is therefore to make the recommendations by examining and evaluating existing artefacts with real users. The recommendations should work as design objectives for a future e-commerce search engine artefact. The focus of the recommendations is first for possibly new entrants in the market, secondly for established players. The thesis takes a design science perspective, rather than an economic market perspective, which implies that the focus will be the actual IT artefacts (e.g. the search engines and their user interactions) rather than the competition and market situation behind. In order to succeed, the research will be guided by the following research questions and research objectives.

## 1.3. Research question

The research question for the current research consists of an overall question to be answered by the design objectives and three subordinate questions to steer the answering of the overall question. The overall question follows as:

*How should search engines be designed to perform better when solving e-commerce users' information needs?*

The subordinate questions follow as:

1. *How do e-commerce users find information for e-commerce related information needs?*
2. *What problems do e-commerce users experience in e-commerce search engines?*
3. *Are there differences in the relative performance of common search engines when used to find e-commerce products?*

## 1.4. Research objectives

In order to address the research question as well as the subordinate research questions, the following research objectives are used to steer the research:

1. Identify search engines used by e-commerce users.
2. Identify and describe problems and search behaviours in the search engines from a user perspective.
3. Construct an evaluation framework that can measure the performance of e-commerce search engines.
4. Evaluate the performance of a search engines sample with real users.
5. Compare the performance and the problems across the different e-commerce search engines.
6. Make recommendations of design objectives for search engine developing businesses that can improve performance and address the previously identified problems.

## 1.5. Scope

To limit the scope, the current research is performed within the Swedish e-commerce landscape and uses only Swedish e-commerce users. The choice of Sweden relates to accessibility[1] and that e-commerce is very well spread in Sweden[2]. To further limit the scope, the present research focuses on e-commerce with physical products, not with services.

[1] The author is located in Sweden, having access to Swedish users.
[2] Sweden had an 81 % proportion of individuals who had purchased online within the last 12 months. The second highest proportion after the United Kingdom (82 %) in the European Union (avg. 57 %) in 2017 (Eurostat, 2017).

# 2. THEORETICAL FRAMEWORK

This section introduces the theoretical concepts that will be used in the thesis.

## 2.1. Design objectives in design science

As pointed out in the introduction, the goal of this thesis is to make recommendations in the form of design objectives for the development of an e-commerce search engine. In order to do so, understanding how to make design objectives is crucial. Peffers et al. (2007, p. 55) describe the design objectives as qualitative descriptions *"/…/of how a new artifact is expected to support solutions to problems not hitherto addressed"* or descriptions in quantitative terms of how *"/…/a desirable solution would be better than current ones/…/"*. Peffers et al. (2007, p. 55) further suggests that *"The objectives should be inferred rationally from the problem specification."*. This implies that, as the overall problem of the thesis is specified as low performance (e.g. search requires many queries and is time consuming for users) in current e-commerce search, the objectives should address this problem, guided by the overall research question.

Peffers et al. (2007, p.55) further suggest that then resources needed for defining the design objectives are the *"/…/knowledge of the state of problems and current solutions, if any, and their efficacy."*. For this thesis, this requirement is achieved through extensive evaluation of existing search engines with users. At last, Peffers et al. (2007, p.55) also points out that the design objectives should be using *"/…/knowledge of what is possible and feasible."* This last criterion is considered as the thesis combines the results of the user evaluations with the theoretical body of e-commerce and information retrieval.

## 2.2. Information retrieval and search engines

Historically, *information retrieval (IR)* has been a very broad concept which refers to the storing and accessing of information of all kinds (Baeza-Yates & Ribeiro-Neto, 2011). For this thesis information retrieval will be put in the light of e-commerce and therefore relate to the storing and accessing of e-commerce information, such as products and product information.

### 2.2.1 Search, query, information need and document

*Searching* refers to the process in which the user of an information retrieval system formulates a query that usually consists of a set of words that describes the information he or she is looking for (Baeza-Yates & Ribeiro-Neto, 2011). A *query* can be defined as *"what the user conveys to the computer in an attempt to communicate the information need"* (Manning & Raghavan, 2008, p. 5). This is the input the user gives the information retrieval system. The *information need* is another central concept in information retrieval and refers to the topic that the user aims to increase knowledge about when using the information retrieval system (Manning & Raghavan, 2008). The term *document* has a central role in information retrieval, which refers to the material the user aims to find by utilising information retrieval (Manning & Raghavan, 2008). To adapt the information retrieval definition of *documents* to fit the e-commerce context, a *document* in the following study will refer to a web page (product page) containing information about a specific product. This product page is often a part of the actual e-commerce transaction, as the users commonly use the product page to get information about it and to place the product in the shopping cart.

### 2.2.2 Search Engines

A search engine (also called web search engine) is an information retrieval system for finding information on the web. This information can be web pages, images, files or more specific entities such as companies, persons or products (as often in e-commerce). In the following sections, the two dimensions and some concepts regarding search engines will be presented.

#### 2.2.2.1 Dimension: General purpose - Specific purpose

Search engines can be differentiated by their dimension of specification, as either specific focus search engines or general search engines. The focus can take many different approaches. As an example, Teixeira Lopes & Ribeiro (2011) conducted an evaluation of web search engines in health information retrieval, where they used four general web search engines (Bing, Google, Sapo and Yahoo!) and three health-specific search engines (MedlinePlus, SapoSau´de and WebMD). Jansen & Molina (2006) studied effectiveness of e-commerce link retrieval in five major search engines and distinguished between *general purpose web search engines* and *e-commerce search engines*. The Specific purpose search engines are in some literature referred to as *vertical search engines* (Zhou, Cummins, Lalmas, & Jose, 2013).

#### 2.2.2.2 Dimension: Internal - External

Another distinction of search engines can be made by the internal-external dimension (Ortiz-Cordova & Jansen, 2014; Ortiz-Cordova et al., 2015). Internal search refers to *"/.../one or more queries submitted to a site's specific search service in order to find information that is contained on that site."* (Ortiz-Cordova et al., 2015, p. 719) and

external search is the *"/../capability and the action of searching using a general purpose search engine/../"* (Ortiz-Cordova et al., 2015, p. 719).

However, as pointed out by Ortiz-Cordova et al. (2015), the definition of what is internal or external search is not permanent, but instead depending on the context of the research. Therefore, to better fit the context of e-commerce search, the definition of external search for this research is adapted to also include specific search engines, in addition to general purpose search engines.

Though these dimensions relate to the actions and capabilities of the user in their search episodes (Ortiz-Cordova et al., 2015), the dimensions are still very relevant for categorising the actual search engines in the e-commerce domain. As the final information need in e-commerce search is to conduct a transaction, a general-purpose search engine such as Google, is defined as an external search engine (i.e. users will not conduct their transaction with Google itself). An internal search engine is therefore the search engine on the website which sells the actual product (i.e. on the e-commerce store's website where the transaction is conducted). A store (or e-commerce store) will onwards be used to refer to an online e-commerce store, from which the user can buy the actual product.

#### 2.2.2.3 Search engine results page (SERP)

The Search engine results page (SERP or short: search results page) refers to the page were the search results are presented. It is commonly divided into different sections, such as the "organic" or the "sponsored". Organic refers to being produced algorithmically (Ortiz-Cordova et al., 2015). An example is *organic traffic* which is defined as *"visits referred by a major search engine based on relevance listings rather than ads"* (Ortiz-Cordova & Jansen, 2014, p. 1346) The opposite, the so called "sponsored search" is defined as *"targeted, relevance-based advertisements that are displayed alongside major search engine results (e.g., Google AdWords)"* (Ortiz-Cordova & Jansen, 2014, p. 1346). Organic or sponsored can also be used to describe the listings, links etc. found on the search results page. An example of Google.se is presented in Figure 1, where the sponsored listings are found in the green and red rectangles, the green one showing the Google Shopping results; while the red one shows the Google AdWords results. The blue rectangle shows the organic results, which continues downwards out of the figure.

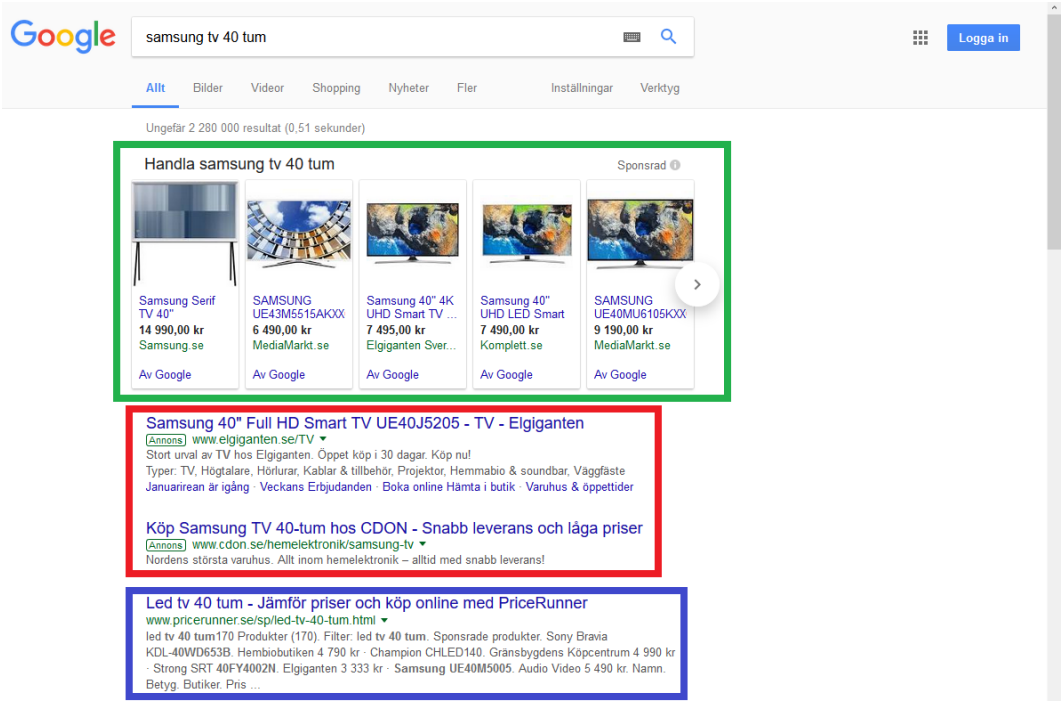



**FIGURE 1 - SEARCH ENGINE RESULTS PAGE OF GOOGLE.SE**

#### 2.2.2.4 Collections and coverage

Different search engines can contain different *collections* of documents. A collection of documents is the group of documents that the search engine perform retrieval from (Baeza-Yates & Ribeiro-Neto, 2011; Manning & Raghavan, 2008). In this thesis, the search engines will be discussed in terms of their different *coverage*, where coverage refers to the extent a search engine's collection covers what is currently being discussed (e.g. webpages, products etc.). The term *indexed* refers to something being a part of the collection (or synonymy index) of a search engine.

#### 2.2.2.5 Information quality, accuracy and completeness

Information systems are commonly described by their information quality (DeLone & McLean, 1992). The information quality refers to the content that the information system produces and it includes the accuracy and completeness of the information (DeLone & McLean, 2003).

## 2.3. E-commerce search

Most of the current research on search engines do not specify focus on e-commerce search engines. The studies that do take an e-commerce perspective on search mainly do so by considering the retrieval of e-commerce related content from general purpose search engines, rather than retrieval in search engines specifically designed for e-commerce information. Spink, Jansen, Wolfram, & Saracevic (2002) started out early by noticing an e-commerce related queries growth in the general-purpose search engine EXCITE's

query logs. Simultaneously Broder (2002) developed a taxonomy of queries by using user surveys and queries logs from AltaVista (i.e. another general purpose search engine). Already then, Broder (2002)s results suggested ~8 % of the total account of searches was made for the purpose of buying something on the internet (i.e. e-commerce). He, Meng, Yu, & Wu, (2003) Peng, Meng, He, & Yu (2004); Yu, He, Meng, & Peng (2004) utilized e-commerce search engines (ESEs) as a central input for the construction of e-commerce metasearch engines (EMSEs) (i.e. a search engine allowing the user to search at many other search engines simultaneously), though without providing a clear definition of the concept itself. They rather describe it as search engines with e-commerce sites as sources (He et al., 2003), where the searchable content are products; and provided an example, the e-commerce website Barnesandnoble.com which sells books (C. Yu et al., 2004).

### 2.3.1 Defining E-commerce search engines

Though e-commerce search is examined in previous studies (e.g. Broder, 2002; Spink, Jansen, et al., 2002), a clear definition of a specific e-commerce search engine is first seen in Jansen & Molina (2006). They define *e-commerce search engines* as search engines specifically designed for retrieving e-commerce information, such as product information, comparisons, prices and reviews (Jansen & Molina, 2006). As suggested in the previous sections, e-commerce search is not limited to just search within specifically designed e-commerce search engines, but it's also a phenomena occurring greatly in general purpose search engines (e.g. (Broder, 2002; Jansen, Booth, & Spink, 2008; Spink, Jansen, et al., 2002)).

As the overall research question of the current research addresses search engines in a general perspective, where e-commerce is included through the information need of the user, the definition of e-commerce search engines should be very wide, as *any search engine that allows for the search of e-commerce products.* This implies that general purpose search engines such as Google, DuckDuckGo (i.e. privacy respecting profile) and Baidu (i.e. major Chinese search engine) are included in the definition; but not specific search engines for other topics than e-commerce, such as PubMed (i.e. medical publications focus search) or Mocavo (i.e. family history search focus).

## 2.4. User behaviours in search

This section addresses theoretical concepts on user behaviours in search, as user behaviours are included in the first of the subordinate research questions, and also believed to be relevant for developing design objectives for practice. The first subsection addresses user behaviours when formulating queries and different information retrieval approaches for helping the users with this. The next two sections address user's behaviours when transferring between search engines, and how to search without queries.

### 2.4.1 Ambiguous queries

Previous research suggests that users on average perform poorly when formulating queries in web search engines. They tend to make queries short (2-3 words) (Hearst, 2009; Spink, Jansen, et al., 2002); while few (6-7%) use operators (e.g. AND, OR, -, + ), and when they do, half of the users use them incorrectly (Jansen, Spink, & Saracevic, 2000). The result is what is often referred to as *ambiguous queries*, which currently is and has been a common problem in IR for long (Cronen-Townsend & Croft, 2002; Krovetz & Croft, 1992; Song et al., 2007). Simply put, ambiguous queries are queries that are hard for the search engine to understand (i.e. there are formal technical definitions on when this occur, but such provide little value here). The queries can be ambiguous in different ways and there are different information retrieval approaches to counter ambiguity, some of which are described in the following four sections.

#### 2.4.1.1 Long-term and short-term interests

One example of an ambiguous query is the query "java", as "java" can refer to either the island Java in Indonesia or the popular programming language Java (Singh & Sharma, 2016). The user's information need probably concerns one of these topics, but for a search engine receiving the simple query of "java", it will be hard to guess one of them. Singh and Sharma (2016) suggest an approach to handling this kind of ambiguous queries, which involves using personalization. Personalization (further explained in section 2.6) in search refers to adjusting the search results depending on the preferences of the individual user (Dou, Song, Wen, & Yuan, 2009). Different solutions to this problem have been proposed, such as combining the ordinary search rankings with user's long-term and short-term interests, to better understand the user's information need (Matthijs & Radlinski, 2011; Singh & Sharma, 2016). Another approach is to use preferences of similar users in order to make the search engine rankings more relevant (Singh & Sharma, 2016).

#### 2.4.1.2 Query tagging and modifiers

Ambiguous queries can also occur as it is difficult for the search engine to interpret the meaning of words. One example of this, is the e-commerce user searching for "designer handbags", which often returns results where the word "designer" is included in the product title or description for sales boosting reasons, rather than the handbag actually being a designer handbag (Gollapudi, Ieong, & Kannan, 2012). Adding the time aspect further complicates the search engines task, as a handbag considered "designer" a year ago, could be completely out of fashion today (Gollapudi et al., 2012). In order to address the problem, solutions have been suggested that divide the different keywords found in a query and tag each of them to product attributes and categories (Li, Wang, & Acero, 2009; Sarkas, Paparizos, & Tsaparas, 2010). An example could be the query "silver canon digital camera" in which the search engine would map the query to category

"cameras", subcategory "digital" and filter the cameras with the attributes colour "silver" and brand "canon".

Gollapudi et al. (2012) further develop this approach by suggesting a model that takes historical search data into consideration when designing the modifiers (i.e. the pointers of query words to product attributes, e.g. silver → colour). Their solution is based on the assumption that users who have previously searched for "designer handbags" are likely to also have browsed and clicked at such products in the search results. Compared to the original results, their approach of modifiers is preferred by the users in 87 % of the cases (Gollapudi et al., 2012).

#### 2.4.1.3 Diversity, search suggestions and user feedback

Agrawal, Gollapudi, Halverson and Ieong (2009) suggest an approach to the ambiguous query problem by including diversity into the rankings of the results, in addition to the traditional ranking of relevance. The logic behind this is to minimize the risk that the user finds no relevant result at all. Given the Java example from section 2.4.1.1, the search engine might consider the probability that the user is looking for the Indonesian island Java as 70 % and for the programming language Java as 30 %, thereby returning results only about the Indonesian island as it is more probable that this is the topic that the user is looking for. 70 % of the users will be satisfied but 30 % will be very dissatisfied with the results. The better solution is instead to use taxonomies (i.e. the search engine understanding that "Java" could refer to the two different topics) and consider the diversity of the results (not only the relevance of each individual document) when making the ranking, implying that results about the programming language Java will also be included, though to smaller extent. The users looking for the island Java are likely a little less satisfied, but the users looking for the programming language Java will not be completely dissatisfied anymore, which leads to an overall higher satisfaction with the results.

The above mentioned approach can also be combined with search suggestions and user feedback. J. Yu, Mohan, Putthividhya and Wong (2014) suggests such an approach for e-commerce where users of ambiguous queries are suggested results from the different topics and given the options of *"see more items like this"* for each of the products. By this approach, the users' feedback on which topic they are looking for can further help guiding the search engine towards the users' information need.

#### 2.4.1.4 Spelling errors

Another type of queries that is hard for search engines to understand are those who contain spelling errors. Research suggests that 10-15 % of all queries committed to search engines contain errors (Cucerzan & Brill, 2004), a percentage which is even higher (21 %) for long tail queries (e.g. more rare queries) (Broder et al.,

2009). There are different approaches such as offline correction and online correction of spelling errors (Duan & Hsu, 2011). The offline provide spelling corrections after the search query has been submitted to the search engine, while the online spelling correction produce spelling correction already as the user is typing the query into the search engine (Duan & Hsu, 2011).

### 2.4.2 External to internal search

As pointed out in the Motivation (section 1.2), studies have found that search tends to be an extensive process (Spink, Ozmutlu, et al., 2002) spanning over multiple information platforms (Ho et al., 2012). A more recent study (Ortiz-Cordova & Jansen, 2014; Ortiz-Cordova et al., 2015) has focused on the transition of users between different search engines, more specifically from an external to an internal search engine. The study used a music entertainment website allowing the users to *"/…/play songs on demand, watch music videos, view song lyrics, look up artist information and biographies/…/"* (Ortiz-Cordova et al., 2015, p. 722). By recording search queries on the internal search engine and referral queries (i.e. the query entered in the external website before entering the internal) from external website, it allowed the researchers to track the users and their intent between external and internal search. They identified six different user site searching strategies, which can support websites in advertising and internal search (Ortiz-Cordova et al., 2015).

### 2.4.3 Browsing in search

Baeza-Yates & Ribeiro-Neto (2011) points out that not all searches start with a query. *Browsing,* or *navigating*, refers to the process in which the user utilizes an information structure to view the available information, and in a sequence, refine their view by scanning and selecting from its contents. This is usually the case when the user can't determine a precise query, since its interest is weakly defined or very wide. An information structure could for example be a web page with hyperlinks (Baeza-Yates & Ribeiro-Neto, 2011).

## 2.5. Evaluations in information retrieval and e-commerce

Evaluation in general refers to "*the process of determining the merit or worth or value of something; or the product of that process* ." (Scriven, 1981, p. 53). In information retrieval this refers to measuring how well an information retrieval system meets the information needs of the users (Baeza-Yates & Ribeiro-Neto, 2011). As the third subordinate research question addresses performance evaluation of the search engines, different approaches for such will be presented in the following sections.

### 2.5.1 Classic information retrieval evaluation

The classic approach towards evaluation of IR systems is to test the system on a standardised corpus of documents in a lab environment and then calculate performance measures, such as precision and recall, according to specific definitions. By controlling the user variables and search tasks, this experimental design can test the indexing and search components of the IR system (Su, 2003a). This classical IR evaluation is often referred to as a system-centred approach, where the questions asked relate directly to the performance of a certain system: "/…/*does this system retrieve relevant documents?"* (Kelly, 2009, p. 3). The relevance of this approach on evaluation has been questioned by many researchers (Kelly, 2009; Su, 2003a; Wang & Forgionne, 2008), as it does not take into account how real life end-users use the system.

### 2.5.2 Modern information retrieval evaluation

To contrast the classic system-centred approach on evaluations, more current studies tend to put a larger emphasis on the interactive nature of information retrieval (e.g. Dumais et al., 2016; Su, 2003b; Teixeira Lopes & Ribeiro, 2011; Vaughan, 2004), where the full interaction between the system and the user is important, rather than just the system itself. The classic approach of evaluations in laboratory environments is also clashing with the user-centred approach, which focuses on studying information retrieval in actual real life conditions (Borlund, 2003). Judgements about a system is often very subjective and a classic IR assessor might not make the same judgement as the actual users, thereby making the classic measures to asses performance of a system, less meaningful (Kelly, 2009). Real-life environments and conditions based studies are often referred to as naturalistic studies, contrary to laboratory studies (Kelly, 2009). The advantage of such is that it's less likely that the research design or laboratory setting will bias the participants behaviours (Kelly, 2009).

## 2.6. A framework for evaluating e-commerce search engines performance

Through the literature review, lots of literature on evaluation of e-commerce and information retrieval has been found separately. But a clear framework for evaluation of e-commerce search engines specifically has not been found. One of the most e-commerce focused search engine evaluation frameworks is found in Jansen and Molina (2006), who evaluate the relevance of e-commerce links in web search engines. Though relevance is one important metric in search, it is not the only one, and hence the following section will present a wider framework for evaluations of search engines in e-commerce. The concepts in the framework have been selected from the e-commerce and information retrieval domain respectively, due to their relevance as metrics for evaluations in the combined domain. An illustration of the framework is available in Figure 2 and its concepts are explained in the following sections.

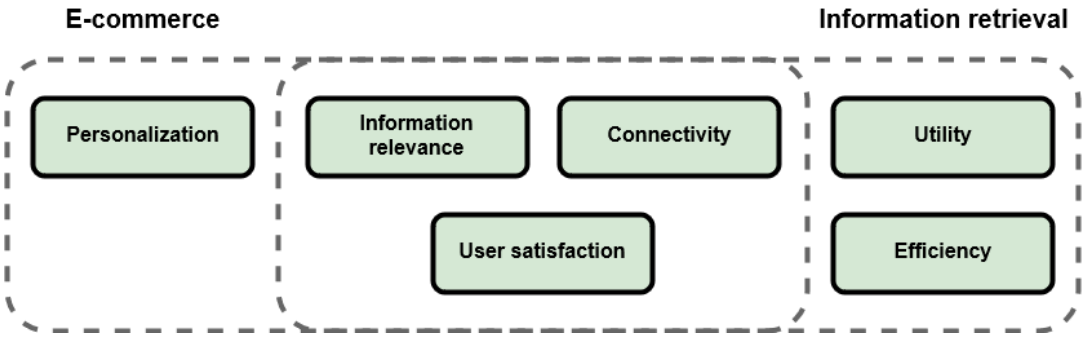



**FIGURE 2 - THE DEVELOPED E-COMMERCE SEARCH EVALUATION FRAMEWORK, ITS CONCEPTS AND OVERLAP WITH IN THE DOMAINS OF E-COMMERCE AND IR.**

### 2.6.1 Information relevance

Information relevance is a central concept for evaluations in both information retrieval and in e-commerce. In e-commerce it's frequently used as one determinant of the quality of information that a system can deliver (Cao, Zhang, & Seydel, 2005; DeLone & McLean, 2004; Huang & Benyoucef, 2013; Jaiswal, Niraj, & Venugopal, 2010; Liu & Arnett, 2000; van der Merwe & Bekker, 2003). In information retrieval, relevance is central to the information retrieval problem, described as: *"the primary goal of an IR system is to retrieve all the documents that are relevant to a user query while retrieving as few non-relevant documents as possible"* by (Baeza-Yates & Ribeiro-Neto, 2011, p. 4). Relevance is further used as a performance measure of search engines (Ali & Beg, 2011; Kelly, 2009; Su, 2003a) in many search engine evaluations (Bilal, 2012; Hawking, Craswell, Bailey, & Griffiths, 2001; Spink, 2002; Su, 2003b). Relevant information is further pointed out as a part of cultivation, which increases customer loyalty in e-commerce, leading to improve the desired outcomes of word-of-mouth and the willingness to pay more (Srinivasan et al., 2002).

### 2.6.2 User satisfaction

User satisfaction is an established and common concept for measuring the success of Information Systems in general (DeLone & McLean, 1992; Zviran, Glezer, & Avni, 2006). It's been pointed out as *"/…/probably the most widely used single measure of I/S success."* (DeLone & McLean, 1992, p. 69). The user satisfaction measure has also been suggested as well fit for evaluations in e-commerce (DeLone & McLean, 2004; Zviran et al., 2006). In information retrieval evaluations, user satisfaction is also commonly used (Kelly, 2009; Su, 2003a). With the aim of finding the best evaluation measure(s) for interactive information retrieval, Su, (1992) concluded that the second single best evaluation measure of information retrieval performance was user satisfaction.

User satisfaction affects the intention of a user to use the information system, at the same time as the actual use of the information system will affect the user satisfaction. It's aimed to cover the full user experience (DeLone & McLean, 2004), making it a great measure due to its wideness.

### 2.6.3 Connectivity

Connectivity has been pointed out in both e-commerce (Susser & Ariga, 2006; van der Merwe & Bekker, 2003) and search engines (Dong & Su, 1997; Su, 2003a) as an important element of web sites. Connectivity can be measured as how well the links in a search engine are working (Su, 2003a). If the links in an e-commerce search engine don't work, it doesn't matter how relevant rankings the search engine itself provides, or how fast it performs, it won't be useful for the users anyway.

### 2.6.4 Personalization

Personalization is a common concept in e-commerce (DeLone & McLean, 2004; Molla & Licker, 2001; Zviran et al., 2006). It is, just as relevance, a part of the information quality of e-commerce (DeLone & McLean, 2004). Personalization refers to *"the tailoring of products and purchase experience to the tastes of individual consumers based upon their personal and preference information."* (Chellappa & Sin, 2005, p. 181). This includes the customization of information and offerings to the specific user (Chiou, Lin, & Perng, 2010).

The implementation of personalization can be approached in numerous ways. One approach is to use click-stream data to track the user activities on the website (Cao et al., 2005). For personalization in search the use of user preferences (short-term and long-term interests) has been proposed to modify the rankings of search results to better match the user (Singh & Sharma, 2016) (section 2.4.1.1).

As personalization requires the collection of information about the user, one issue of personalization is the trade-off between value of personalization and concern for privacy. This can affect the users' willingness to share information about themselves online. However, privacy is not absolute, and if enough benefits are added, users might abandon some of their privacy (Awad & Krishnan, 2006; Chellappa & Sin, 2005). Information transparency features refer to features that provide knowledge of processing procedures and the collected information that a company has about its users. Awad & Krishnan (2006) suggest that the users that desire information transparency are less likely to engage in personalization than the users that do not desire information transparency. Hence, they recommend managers to focus on satisfying users that are more willing to take part in personalization, ignoring the minority of privacy sensitive users (Awad & Krishnan, 2006).

### 2.6.5 Utility

Utility measures has been used previously in information retrieval (Kelly, 2009; Su, 2003a; Wang & Forgionne, 2008). According to Su (1992), *value of search results as a whole* is the best single measure of information retrieval performance. It measures a search engine's ability to deliver results at a level that meets the user's need (Su, 2003a). Consequently, this measure differs from relevance measures (e.g. precision), as it focuses on what subjective level the search engine delivers results that meet the user's need (i.e. the matching between results and user need), rather than the specific proportion of relevant links. Hence, if there is a search engine frequently delivering only one relevant result, but one that solves the user's information need, it can still provide the user more value than a search engine delivering a larger proportion of relevant or partly relevant results, but with none completely solving the information need.

### 2.6.6 Efficiency

Efficiency measures have previously been used in information retrieval (Su, 2003a). It is related to the time the user needs to complete the different decision-making steps along the information retrieval process. This includes the time of phrasing search queries, as well as the time needed to judge the relevant documents (Wang & Forgionne, 2008).

# 3. METHODOLOGY

## 3.1. Paradigmatic assumptions

This research takes its philosophical stand in the pragmatic philosophy of science. The choice of pragmatism relates to the nature of the overall research question in this study. As the research question is very practical in nature and aims to produce design objectives for practice, it aligns with the pragmatic view of science, in which the aim of producing *"/…/practical solutions that inform future practice"* is central (Saunders, Lewis, & Thornhill, 2016, p. 143). The advantage, but also criticism (Goles & Hirschheim, 2000), of pragmatism in relation to other philosophies of science, is its flexibility on ontology[3] and epistemology[4] (Johnson & Onwuegbuzie, 2004). Instead of a strict stand on ontology and epistemology, which can limit the researcher, a pragmatist researcher should focus on what is most important for answering the research question (Tashakkori & Teddlie, 2010).

As such, the ontological stance of pragmatism is that actions and change are viewed as the fundamentals of reality (Goldkuhl, 2012). Actions and change are also what constitute acceptable knowledge, which in epistemological terms implies, that concepts are only considered relevant when they can be used to support action or change in practice (Goldkuhl, 2012; Saunders et al., 2016). Hence, for this thesis, the focus will be to produce design objectives that could change and potentially disrupt e-commerce search in its foundation. As such, when later valuing and discussing the findings, priority will be given to findings that are likely able to cause action, and that could provide major practical implications for the increase of performance.

### 3.1.1 Multi-paradigmatic research

The implications of the flexibility of pragmatism for this thesis also mean that different methods can be used to answer the research questions. The methods can come from different paradigms and use different underlying assumptions, leading to a multi-paradigmatic research (Saunders et al., 2016). The first two subordinate research questions aim to qualitatively explore and identify social phenomenon (i.e. user behaviours and problems). This calls for a qualitative approach commonly used with an interpretivist

[3] Ontology refers to the assumptions the researcher makes about the nature of reality (Saunders et al., 2016).
[4] Epistemology refers to the assumptions the researcher makes about what constitutes acceptable knowledge. E.g. What is knowledge and how can it be communicated it to others? (Saunders et al., 2016).

philosophy. For the interpretivist researcher, there is no universal reality and users experience and interpret the same situation differently, leading to new worldviews and interpretations as acceptable knowledge (Saunders et al., 2016). The third subordinate research question aims to measure the performance of the different search engines. This calls for a quantitative approach, which is commonly combined with a positivistic philosophy. For the positivistic researcher, there is only one reality and knowledge is law-like generalisations and measurable facts (Saunders et al., 2016). As the aims of these subordinate research questions differ substantially, the strength of the overall pragmatic assumptions is therefore that they can be accompanied by interpretivist and positivist assumptions for the qualitative and quantitative subordinate research questions respectively.

It is the researcher's opinion that each of these philosophical stances alone, are less likely to provide findings that can support strong design objectives, than if used together. If the research focus would be purely positivistic, the simplified and generalisable findings produced would likely miss a lot of potential problems that individual users experience, as the individual problems are not significant on an aggregated level. Similarly, in a pure interpretivist approach, the findings could prove too complex and too influenced by the researcher to be usable. Therefore, pragmatism is the great middle way, that allows focusing on the practical implications that the different findings could have.

## 3.2. Approach to theory development

This thesis takes an inductive approach to theory development. This as it aims to first collect data by exploring and evaluating e-commerce search engines, in order to then produce theory in the form of recommendations for e-commerce search engine developers. This contrasts the deductive theory building, which first develops hypotheses using existing theory, which are later tested in order to falsify or verify existing theory (Saunders et al., 2016). The advantage of induction is that, contrary to deduction, it's a less structed research approach, which allows for alternative explanations of phenomena, not covered in specific hypotheses (Saunders et al., 2016).

## 3.3. Research design

The research design follows the overall purpose of evaluative research. The purpose is therefore to *"/…/find out how well something works."* (Saunders et al., 2016, p. 176), which for the present research is e-commerce search and its search engines. The research takes a sequential double phase research design, using mixed methods as illustrated in Figure 3.

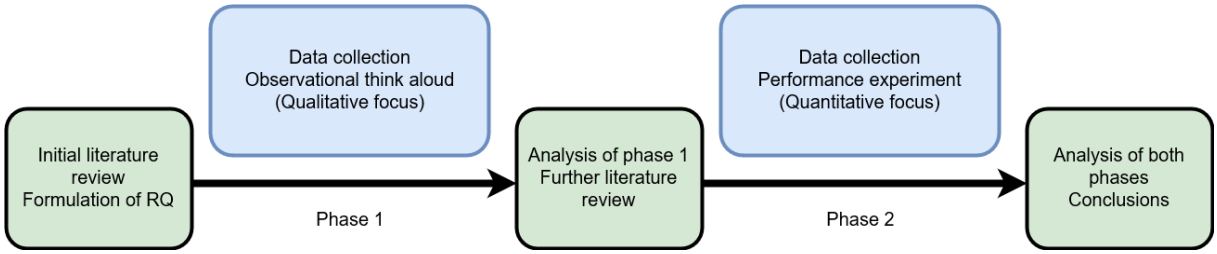



**FIGURE 3 - OVERVIEW OF THE SEQUENTIAL DOUBLE PHASE RESEARCH DESIGN**

This sequential approach is appropriate as the findings of the first phase can be used to inform the second phase (Venkatesh, Brown, & Bala, 2013), which wouldn't be possible in a concurrent design where all the data collection takes place at the same time (phase) (Saunders et al., 2016). The advantage of the double phase design is also flexibility of starting out broadly by clarifying the understanding of the phenomenon before narrowing down the research focus (Saunders et al., 2016).

As design science objectives are aimed to provide solutions to problems and opportunities identified in the actual application environment (Hevner & Chatterjee, 2010), a key is to first understand this context. For this thesis, the overall evaluative purpose is therefore combined with an explorative purpose in the first phase, to make sure that the environment of e-commerce search is properly understood. The implication of the design for the research was further that the findings of the first phase were used to inform the second phase, by the selection of appropriate criteria (i.e. measures) and sample (i.e. search engines) for the performance evaluation. The preliminary analysis (analysis of phase 1) further helps narrowing the research to a focus of certain concepts (i.e. external to internal search, advertising, specific search, etc.).

### 3.3.1 Methodological choice – mixed methods

The methodological choice is to use a mixed method approach, with both qualitative and quantitative methods. This works well with the underlying pragmatic assumptions of the research, which sees mixed methods as beneficial and supplementing when relevant for answering the research question (Saunders et al., 2016; Tashakkori & Teddlie, 2010; Venkatesh et al., 2013). The mix of methods also provide many benefits, such as initiation and triangulation for the research (Bryman, 2006). The mixed methods provided initiation in the beginning of the current sequential research by producing a context of the research, aiding the refinement of the research question, as well as prioritizing the questionnaire measures and sampling the search engines (also pointed out in section 3.3).

The choice of mixed methods can also be related to the exploratory and evaluative nature of the research questions and research design. As the first phase of the research aimed to explore e-commerce search,

highly unstructured methods were chosen. The observational think aloud is viewed as unstructured as the researcher has little control over how the user behaves, and the space of possible observations is consequently very wide (van Someren, Barnard, & Sandberg, 1994). In the second phase, the experiment was instead conducted using a structured methodology for the search engines performance to be tested. In the following two sections, the methods will be shortly introduced.

#### 3.3.1.1 Phase 1 methods: Observational think aloud

The choice of methods for the first phase, is a combination of an observational study, think aloud study and interviews. The combination is that a user is given some tasks to solve, and while doing so, the user must verbalise his or her thoughts out loud. The researcher observes the user while he or she is solving the tasks and thinks out loud. When the user finishes his or her tasks, the researcher can ask questions through a short debriefing interview.

The strength of the think aloud method is its flexibility, as it can be used to collect data about a wide variety of phenomena (e.g. usability of systems (Holzinger, 2005; Nielsen, 1993), information processing (Eveland & Dunwoody, 2000) or health care research (Lundgrén-Laine & Salanterä, 2010)). Advantages of think aloud include very rich data, user comments including clear quotes, simultaneously as collection of user preferences and performance (Holzinger, 2005; Nielsen, 1993). Some disadvantages of the think aloud are performance measurement limitations (Holzinger, 2005; Nielsen, 1993), distraction from the task (e.g. users might be distracted by talking when they are performing the tasks), high time consumption and intrusion for the user (Holzinger, 2005).

The implications for this thesis is that the think aloud method can be easily used in the e-commerce search context, without modifications, and that it can potentially generate very rich data, which is desirable for the explorative aim of the first phase and two first subordinate research questions. The method's lack of performance measurement is judged to be a negligible problem, as the second phase of the study instead has this aim. The aim of the first phase is explorative, where identification of problems and behaviours is the objective.

The reason for mixing the methods in the first phase comes from the interpretivist assumptions. Saunders et al. (2016, p. 141) points out the challenge as: *"The challenge for the interpretivist is to enter the social world of the research participants and understand that world from their point of view."* It's the researcher's opinion that the implication of combining these methods (i.e. observation, think aloud and interviewing), is that he can indeed come deeper into the social world of the research participant. For example, if using only observations, the researcher could identify problems that users experience, but he wouldn't be able to

understand their opinions on the problems or the motives of their behaviours to counter them. On the other hand, if only using the think aloud method without observation, it's likely that the researcher could miss problems that the users do not address themselves.

#### 3.3.1.2 Phase 2 method: Performance evaluation experiment

For the second phase, an experimental method is chosen to evaluate the performance of the different search engines. Experiments are commonly used in search engine evaluations, where different systems or different versions of the same system are evaluated using a set of outcome measures, such as performance (Kelly, 2009). The experiment works by having a set of independent variables (e.g. the systems being used by the user) and dependent variables (e.g. performance measures). By manipulating the independent variable (i.e. changing system), the changes in the dependent variables are measured, and then statistically tested, in order to find differences in performance between systems (Kelly, 2009).

For this thesis, the choice of the experimental method implies that the research isn't limited to identifying problems and behaviours in the search engines (i.e. focus in first phase), but also a more general assessment of in which dimensions (e.g. performance measures) the search engines differ. This is important because in a comparison of many search engines, it is not necessarily the search engine with the most identified problems that is the worst performing search engine. This has implications for the development of design objectives, as the point of departure for building new improved solutions, is well performing artefacts, rather than poorly performing artefacts.

### 3.3.2 Cross-sectional time horizon

The time horizon for the research project is cross-sectional. Cross-sectional implies that the research project is undertaken at one point in time and not as a comparison over time (i.e. longitudinal) (Saunders et al., 2016). Though the research design is sequential, and the data collection is not done at the exact same moment, it's still considered cross-sectional as there is no comparison over time and the data collection points are relatively close in time.

The reasons for using a cross sectional time horizon are (1) the limited time frame available for a master thesis and (2) the nature of the research. From a pragmatist view, truth, meaning, and knowledge are not fixed, but instead constantly changing over time (Johnson & Onwuegbuzie, 2004). This suggests that what is truth and knowledge today, might not be true tomorrow. For this thesis, this implies that the design objectives produced can quickly become irrelevant, and that the discovered results should not be seen as any everlasting law-like generalizations (as in a positivistic study). The design objectives are answers to the

research question for the moment, due to the fact that the research question isn't asking for a comparison over time.

## 3.4. Sampling of e-commerce users

The participants of the study (both phases) were recruited through advertisements in student-related Facebook groups with Swedish profiles (e.g. Facebook groups with discussion mainly in the Swedish language, not international students' Facebook groups), as well as through physical advertisements placed on university campuses in Sweden. The advertisements asked specifically for people with experience within e-commerce, and included a link (or short URL) to the study and its instructions, which were located online. This approach is a non-probability volunteer self-selection sampling, which was appropriate for the first phase as of its explanatory aim (Saunders et al., 2016), but can be limiting to the generalisability of the inferential statistics in the second phase (Kelly, 2009).

The reason for using a non-probability volunteer sampling relates to the problems of defining the elements of the targeted population (i.e. all e-commerce users). Likely, there is not an existing list of Swedish e-commerce users, but even if there was and access to that list was available, it would likely be very hard to gain access to each of the elements (i.e. the users) (Kelly, 2009). Therefore generalizations of the quantitative findings should be done with caution (Kelly, 2009).

## 3.5. Research setting

As pointed out in section 2.5.2, though the classic approach to IR evaluations uses a laboratory setting for isolating the variables, it will not likely represent the real life behaviours of users (Kelly, 2009). Therefore, a naturalistic setting is chosen for this research. As Hevner and Chatterjee (2010, p. 17) points out that *"Good design science research often begins by identifying and representing opportunities and problems in an actual application environment."*, this is an argument for the naturalistic setting, as it is believed to be closer to the actual application environment than a laboratory setting is. For both phases, the research participants were therefore asked to be in an environment where they usually are when conducting e-commerce.

## 3.6. Search tasks

Information needs for a study are usually described through search tasks. It's often difficult to artificially generate information needs so that these correspond to the user's actual information needs, as these are considered very subjective, existing only in the user's head. The tasks can be too specific or too general (Kelly, 2009). This can be a source of validity issues, as if the tasks do not correspond to reality, the results from using the tasks might not correspond to reality. The search tasks in this research were therefore

developed backwards from the use of actual e-commerce queries extracted from a query log, adopted from Jansen and Molina (2006). This is a common approach previously used in IR (Kelly, 2009).

The following are examples of queries adopted from Jansen and Molina (2006, p. 1094-1095): "For sale Yamaha xt600", "Food under $5.00", "Cigarettes for $12 carton", "Seacat boat prices", "Nokia 2160 hands free kit". It was found that many queries included brands, models, prices and location. The search tasks used in both phases were therefore developed to contain such elements. A full list of the search tasks can be found in appendix 1 and appendix 2.

The first phase of the study aimed to explore users' approaches for searching e-commerce products. To keep a wide perspective of users' search strategies, the search tasks containing the information need were therefore not limited to a specific industry. Instead the search tasks for the first phase were designed to contain diverse tasks from ten different industries, identified through a survey on physical product based e-commerce in Sweden (HUI Research, 2017).

## 3.7. First phase data collection

The data collection in the first phase was conducted through notetaking, audio and video recording of the think aloud session, and the debriefing interview.

### 3.7.1 Procedures

Phase 1 utilized task-based participant observation through video recording and think aloud protocol.

Upon contact on the advertised research task, a time slot of one and a half hour was scheduled between the researcher and the participant at their convenience. The participants were instructed to be ready at a place they usually conducted e-commerce, using a personal computer of their own choice (i.e. no mobile phones) with a stable internet connection. There were no time requirements for the tasks, so each of the participants could complete the tasks at their own pace. The choice of allowing the participants their preferred equipment, place and time was to *"/…/ensure that the test reflected real-world conditions/.../"* (Jansen & Molina, 2006, p. 1085).

At the time, the researcher and the participant connected using the video-calling software Skype. The participants shared their screens with the researcher and the participants' screens and the conversations between the participant and the researcher were recorded using the video and audio recording software

Open Broadcaster Software[5]. The participants were then explained the think aloud protocol method and given a small training exercise where they had to think aloud. Then the participants were instructed to follow a list of search tasks available through a website. One task at a time was shown to the user, in order not to distract them from the present task.

During the tasks, the researcher was observing the participant, taking notes as well as prompting the participant for thoughts when needed to keep the think aloud speaking flowing. Upon finishing the tasks, each participant made a short debriefing interview to answer questions the researcher had about the session (see section 3.12.1.1).

## 3.8. Second phase data collection

### 3.8.1 Procedures

Upon contact through the link in the advertisement for the research, each participant was taken to a webpage with information about the research project, such as expected time consumption and examples of tasks. Accepting to participate in the research project took the user to a questionnaire containing the tasks and the instructions. Each user first had to fill out a pre-task questionnaire with some background info. Afterwards, instructions for each search task were presented on a separate page, where the answers to that given task were also reported. After completing the tasks, another page of the questionnaire contained post-task questions.

### 3.8.2 Controls

A repeated-measures design was used (Montgomery, 2012), where each of the research participants did complete one different search task for each of the search engines.

**TABLE 1 – COUNTERBALANCED DESIGN OF PHASE 2**

| | Task type | | |
|---|---|---|---|
| **Subject group** | **Task 1** | **Task 2** | **Task 3** |
| **SG1** | Google | Prisjakt | Elgiganten |
| **SG2** | Google | Elgiganten | Prisjakt |
| **SG3** | Prisjakt | Elgiganten | Google |
| **SG4** | Prisjakt | Google | Elgiganten |
| **SG5** | Elgiganten | Google | Prisjakt |
| **SG6** | Elgiganten | Prisjakt | Google |

In order to reduce order effects such as learning and fatigue, counterbalancing was used to control and rotate the search engines in different orders. Each search engine was given each task and position (in

[5] Available at http://obsproject.com

relation to the other search engines) an equal amount of times, as seen in Table 1. The participants were randomly divided between the different subject groups.

### 3.8.3 Sampling of search engines for second phase

#### 3.8.3.1 Selection of industry

For the second phase of the study (the performance evaluation), the consumer electronics industry was selected, due to its large volume of e-commerce sales. Consumer electronics is currently the largest industry in Swedish physical product based e-commerce in terms of turnover (HUI Research, 2017). The e-commerce part of the industry had a turnover of 12.2 billion SEK during 2016 which accounted for 26 % of the overall sales of consumer electronics in Sweden (HUI Research, 2017). The industry was further chosen as a result of the phase 1 analysis, where the electronics industry task involved relatively much of searching, in all three dimensions of search engines (i.e. dimensions from Figure 4).

#### 3.8.3.2 Selection of search engines

For the second phase, a non-probability sample of search engines was used. It was collected using purposive sampling, where each search engine was selected on the basis of (1) their fit within the selection criteria, (2) populational popularity (i.e. Alexa.com ranking), (3) their popularity in phase 1 and (4) their fit for possible contributions in answering the research questions. Alexa.com (Alexa) provides popularity rankings (i.e. Alexa traffic ranks, by pageviews and daily visitors) for web sites on the Internet, located in different regions. The rankings are not perfect due to measurement difficulties, but Alexa is one of the most prominent rankings of websites' popularity and commonly used as a ranking reference in previous research on search engines (e.g. Jansen & Molina, 2006; Ortiz-Cordova et al., 2015; Teixeira Lopes & Ribeiro, 2011).

Purposive sampling was used as it is beneficial for small samples where the researcher can use his/her intuition on which cases that are likely to help answering the research question, rather than being based on statistical generalisability (Saunders et al., 2016). An example of the advantage of purposive sampling in small samples, could be the choice of Google. Google could for example be excluded due to chance in probability sampling, even though Google was the most frequently used search engine in both Alexa's ranking and in the first phase (77 % of the queries made, see Figure 5). Excluding Google from the sample would thereby likely have affected the quality of the findings negatively, as the research question aims on understanding e-commerce search in an overall perspective, where Google was shown to be a major part.

The sample selection is based on the two dimensions of search engines from previous studies (Jansen & Molina, 2006; Ortiz-Cordova et al., 2015; Teixeira Lopes & Ribeiro, 2011) as also pointed out in sections

2.2.2.1 and 2.2.2.2. The dimensions are external vs. internal search and general vs. specific search, which is illustrated in Figure 4, together with the selected search engines Google.se, Prisjakt.nu and Elgiganten.se.

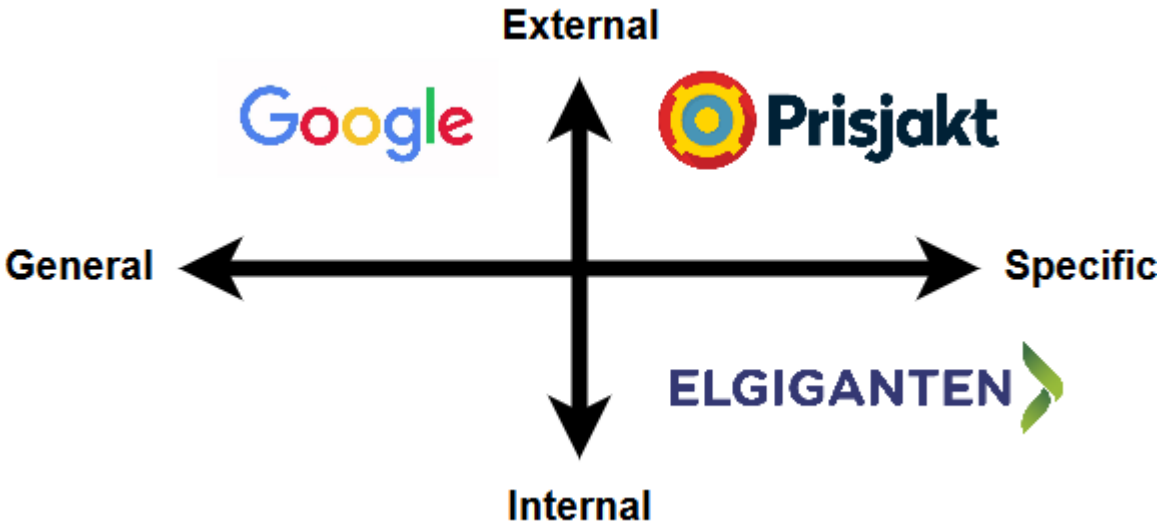



FIGURE 4 - SEARCH ENGINE DIMENSIONS INCLUDING THE EVALUATION SEARCH ENGINES

**Google.se – External general search**

Google.se (*from now on just Google*) is the most frequently used website in Sweden (Alexa.com, 2017) and also the most used website and search engine in phase 1, accounting for 77 % of the overall searches made (Figure 5). Statcounter.com (2017) suggests that Google had a 94 % search engine market share in Sweden 2016.

**Prisjakt.nu – External specific search**

Prisjakt.nu (*from now on just Prisjakt*) is the highest ranked price comparison shopping site in Sweden, ranking 32 out of all websites in Sweden according to Alexa.com (2017). A similar service is Pricerunner.se, ranked 142. It was also used, but less than Prisjakt in phase 1.

**Elgiganten.se – Internal specific search**

Elgiganten.se (*from now on just Elgiganten*) is an e-commerce store (i.e. seller of actual goods). The company with the same name also operates physical stores in Sweden. The search engine in the Elgiganten website is an internal specific search engine. Elgiganten ranks 87 out of all websites in Sweden on Alexa traffic rank (Alexa.com, 2017), which is prominent for being an e-commerce store. Potentially similar search engines in terms of internal e-commerce search with an electronics assortment is Blocket.se (11), Tradera.com (38) Microsoft.com (47), Apple.com (69) and Aliexpress.com (73), but as these were not, or very little, used in phase 1, Elgiganten is prioritized.

**Missing general internal search engine**

As shown in Figure 4, no general internal search engine was used. This relates to the definitions of general purpose search engines as well as internal search engines. General purpose implies that the search engine indexing of all kinds of webpages. Internal search implies only indexing of the pages at the website itself. For a search engine to be considered both internal and general, it should therefore include all kinds of documents, but only from the website itself. This research hasn't come across any of such search engines, and thereby the general purpose internal search dimension is left out. An example of a very wide (almost general) internal search engine in e-commerce could for example be Amazon.com, as it contains a very wide assortment. However, the search engine at Amazon.com does still only contain e-commerce related results, making it a specific search engine rather than a general search engine (such as Google).

### 3.8.4 Instruments and measurement

The following section presents the instruments and measurements used in the second phase performance evaluation.

#### 3.8.4.1 Relevance and precision

As pointed out by Kelly (2009), classic IR-performance measures are not always meaningful to real users. A measure that makes its assessment from the retrieval of the 1000 first results (documents), wouldn't be useful for real users, as they would never consider looking through such a high amount of results. As the current study aims to evaluate web search engines, where the collection of webpages as well as the results often count in millions, it wouldn't be feasible to have the users evaluate each individual result from the results list. As such, Su (2003a) uses a precision measure based on the 20 first results in order to address this issue. Still, the 20 first results is a high amount, considering that studies suggest that users in most searches (80 %), only view the first page of results (where it's often just 10 results) (Jansen et al., 2000; Silverstein, Marais, Henzinger, & Moricz, 1999). The relevance measure of precision in this study will therefore be limited to examine the first 10 results on each search engine results page (i.e. approximately the first page), which has also recently been used in other studies (e.g. Lewandowski, 2015).

Different scales have been used for grading (measure) relevance. Three-point scales (Bilal, 2012; Chu & Rosenthal, 2003; Jansen & Molina, 2006; Su, 2003a) and four-point scales (Gordon & Pathak, 1999) have been common. In this study, the three-point scale is adopted with the grades: (1) Relevant, (2) Partially relevant and (3) Not relevant. The definitions of the grades were adopted in full from (Su, 2003a)[6] as:

[6] The definitions were translated into Swedish and included in the questionnaire provided for the evaluation.

RELEVANT (R): Any item which, on the basis of the information it provides, is related to your information need or problem, even if it is already familiar to you.

PARTIALLY RELEVANT (PR): Any item which, on the basis of the information it provides, is only somewhat or in part related to your information need or problem, even if it is already familiar to you.

NOT RELEVANT (NR): Any item which, on the basis of the information it provides, is not at all related to your information need or problem.

From the relevance gradings, two measures of precision were constructed in accordance with Su (2003a). Precision ratio 1 (PR1) as the proportion of relevant and partially relevant hits out of the first 10 hits (R+PR/10). Precision ratio 2 (PR2) as the proportion of relevant hits out of the first 10 hits (R/10).

#### 3.8.4.2 User satisfaction

Though there are many different user satisfaction measures available, the ones used by Su (2003a) were adopted, with a small modification. Su (2003a) presents eight different measures for measuring the user satisfaction, but only two of them were adopted, as the questionnaire needed to be kept short to avoid fatigue. The measure scale of *"Overall success of an engine in providing help for the user's information need or problem"* (Su, 2003a, p. 1182) was chosen. It was measured using a 7-point Likert scale, using the Swedish translations of "extremely unsatisfactory" to "extremely satisfactory." Also, the measure how well the search engine helped the user to save time was used on a 7-point scale from *1: no time saving at all*, to *7: saving a lot of time.* The time saving aspect has also been pointed out as important in the e-commerce context (DeLone & McLean, 2003)

#### 3.8.4.3 Efficiency

Su (2003a) suggested two different measures of efficiency, namely search time and search strategy. Search time is measured as the time taken from start to completion of search in each of the search engines. Search strategy refers to the number of queries committed to the search engine for solving the search task (Su, 2003a). Due to technical difficulties in measuring search time on the participant's computer, this study adopted the second measure and the study participant was given a field in the questionnaire to input the number of queries as an integer.

#### 3.8.4.4 Utility

The utility concept uses the measure, *value of search results as a whole,* adopted from (Su, 2003a). It was measured using a 7-point Likert scale, using the Swedish translations of "extremely unsatisfactory" to

"extremely satisfactory." As pointed out in section 2.6.5, *value of search results as a whole* is the best single measure of information retrieval performance.

#### 3.8.4.5 Personalization

The measurement of personalization was adapted from Chellappa and Sin (2005). It was measured using a 7-point scale, using the Swedish translations of "strongly disagree" to "strongly agree" on how well the search engine implemented personalization, for each of the search engines. In addition, an open question was included where the participant was asked to elaborate on their view about personalization, whether it was positive or negative. This question was included to collect potentially privacy related opinions about personalization.

### 3.8.5 The questionnaire

The questionnaire for the second phase performance evaluation was divided into three parts: (1) non-performance characteristics (2) the search tasks for each of the search engines and (3) post-task general and comparative questions. The following section describes each of the parts.

#### 3.8.5.1 Non-performance characteristics

A few non-performance characteristics were collected in order to get a better understanding of the research participants. The collected attributes were: year of birth, sex, occupation and e-commerce purchase frequency.

#### 3.8.5.2 Search tasks

The search tasks were given the users for one search engine at a time to reduce distraction. During each search task the users had to provide answers for the relevance and efficiency measures. The users also provided their final search query. The search tasks are available in the appendix 2.

#### 3.8.5.1 Post task general and comparative questions

The last part of the questionnaire included the utility, user satisfaction and personalization measures. In addition, users were also asked to suggest their preferred e-commerce search engine and a motivation for this. Users were also asked to mention advantages and disadvantages of each of the search engines in the study. The use of open questions at the end, allowed for triangulation of the findings of phase 1.

## 3.9. Qualitative analysis methods

Qualitative research is commonly associated with an interpretivist philosophy as of the subjective and socially constructed meanings of the research participant that the researcher wants to understand (Saunders et al., 2016). The data in qualitative research is therefore likely to be more complex and variating than quantitative data, which is more standardised (Saunders et al., 2016). The data collection in the qualitative research therefore required the researcher to actively classify the unstructured data into categories, and the data analysis was conducted with conceptualisation (Saunders et al., 2016).

A problem of the exploratory aim and an inductive approach in the first phase, was the risk of failing to examine the data properly during the exploration (Saunders et al., 2016). Therefore, a main focus was to interpret and asses the data recursively as it was collected. This was done through the examination of the notes from each of the user' sessions before the next one started.

During this process, existing theory also helped by providing an initial analytical framework (Saunders et al., 2016). The framework, which consisted of loosely structured components from theory of search and e-commerce (i.e. search engine dimensions and user search strategies) worked as a starting point in the think-aloud and observational data collection, and was combined and developed with the emerging data. As such, though the initial data collection was inductive in nature, it also included elements of deduction (Saunders et al., 2016).

### 3.9.1 Thematic analysis

The analysis of the data from the first phase was done using thematic analysis, which works by identifying patterns in the data (e.g. behaviours, actions, thoughts) and then combining them into sub-themes and themes (Aronson, 1995). By doing so, the data was put in a context, as it would likely be meaningless if viewed alone (Aronson, 1995). An important part of this is coding, which relates to labelling the different actions, behaviours, believes, events, ideas, interactions and more, with a code that describes its meaning (e.g. summarizes it) (Saunders et al., 2016).

The advantage of thematic analysis is its flexibility, due to its detachment from any specific philosophical position or approach to theory development (Saunders et al., 2016), which allowed it to work well with the pragmatic assumptions of the current research, focusing on practical implications. Thematic analysis and open coding further fit the exploratory aim of the first phase, as it allowed the researcher to go back and forth with the data as new themes emerged, not limiting to the initial framework (Saunders et al., 2016). The

coding was also possible to adjust to the research question, rather than following specific rules, common in other analytic techniques (Saunders et al., 2016).

## 3.10. Quantitative analysis methods

This section presents the quantitative analysis which follows positivistic assumptions using statistical testing to analyse the quantitative data.

### 3.10.1 Statistical testing

In order to judge the system's impact on the performance variables, the results need to be statistically tested. The independent variables (IV:s) are the different systems being evaluated, and the dependent variables (DV:s) are the different performance measures that the systems affect. By variating the systems (IV:s) that the user uses, the study expects to note differences in the performance measures (DV:s).

#### 3.10.1.1 Parametric testing

The first step in deciding which tests to use was to detect the distributions of the collected data. Many statistical tests for comparing results between two or more groups assume the data to be normally distributed (e.g. ANOVA and t-tests). Those statistics are referred to as parametric statistics, while statistics not assuming that the data is normally distributed, are referred to as non-parametric statistics (Kelly, 2009). To test for normality, the data was tested using the Kolmogorov-Smirnov test and the Shapiro-Wilk test in SPSS (Saunders et al., 2016).

#### 3.10.1.2 Choice of statistical test – Kruskal-Wallis with post-hoc tests

To test for differences in one variable between two different groups, the common parametric test is the t-test (more exactly the independent groups t-test) (Kelly, 2009; Saunders et al., 2016). However, in the parametric testing both tests (Kolmogorov-Smirnov and Shapiro-Wilk) indicated that the data was non-parametric. As of this, the tests for evaluating differences between the groups (i.e. search engines groups) in the data had also to be non-parametric. For this, a Mann–Whitney U Test, which is a non-parametric test also commonly used in IR (Kelly, 2009), can be used (Kelly, 2009; Saunders et al., 2016). The problem with the Mann–Whitney U Test is that it only tests the difference between two groups. As the study uses three different groups (i.e. search engines), a Kruskal-Wallis test was instead used, as this allows for testing between three or more groups in non-parametric data (Kelly, 2009). Another test commonly used for testing between three or more groups is the one-way ANOVA (Kelly, 2009; Saunders et al., 2016). The reason for using a Kruskal-Wallis test, instead of the one-way ANOVA, is that the one-way ANOVA

assumes the data is either parametric or that the number of cases in each group are more than 30 and the variance difference between the groups is small (Saunders et al., 2016). The cases in each group are 26.

Furthermore, Saunders et al. (2016) differentiates between non-directional and directional hypothesises. In the Kruskal-Wallis test, a non-directional hypothesis (i.e. two-sided or two-tailed) (IBM Corporation, 2016) is used, as the aim is to test for any differences in performance between the search engines, rather than testing for a specific search engine having a higher performance than the others (Saunders et al., 2016).

**The Kruskal Wallis hypotheses in SPSS** (IBM Corporation, 2016):

- **Null hypothesis ($H_0$):** The distributions of all the groups are the same.
- **Alternative Hypothesis ($H_A$):** At least one of the groups is different.

#### 3.10.1.3 Pairwise comparisons

As noted by the hypotheses in the previous section, the Kruskal-Wallis test will only provide a decision regarding the significance of the null hypothesis (i.e. $H_0$: The distributions of all the groups are the same). If the decision is to reject the null hypothesis, this implies that the alternative hypothesis is accepted and thus that there is a difference for at least one of the groups. However, the Kruskal Wallis test does not indicate for which of the groups there is a difference.

For those variables in which Kruskal Wallis rejected the null hypothesis ($H_0$), pairwise comparisons were therefore conducted in order to asses for which systems there was a significant difference in performance (IBM Corporation, 2016).

#### 3.10.1.1 Likert scales as interval data

There are different opinions on whether Likert scales should be treated as ordinal or interval data. Technically the Likert scale data is ordinal level, as it allows for identifying differences between variables (i.e. System A is better than System B) but it does not describe how much the differences between the systems are. However, in IR, Likert scales are still commonly used as interval level data in order to allow for more sophisticated analysis (Kelly, 2009), which was also the case in the current research.

## 3.11. Role of the researcher and ethical considerations

Axiology refers to the role of values and ethics in the research (Saunders et al., 2016). In terms of axiology, the pragmatist researcher acknowledges that his values and beliefs are a part of the research, and that these values both initiates and drives the research forward (Saunders et al., 2016).

### 3.11.1 Ethical considerations

As the research uses humans as participants, there are ethical considerations to be made. The researcher was eager to record the think aloud sessions made in phase 1. Before doing so, the users were first required to consent to the recording of their screen and speech. All the participating users gave their consent for recording. The users were asked for consent before the recording started, and as the recording started they were also asked a second time to confirm their understanding of the recording taking place. Before consenting, they were also given information about the processing of the recordings. This included: (1) recordings were to be kept only by the researcher for the specific research project and were to be deleted directly afterwards; and (2) all personal identifiable information was to be anonymized before publishing the thesis.

## 3.12. Validity, reliability and limitations

In quantitative research, validity refers to (1) if the research measures actually measure what they are intended to, (2) if the analysis of results and relationships are accurate and (3) if the research is generalisable to a wider context? (Saunders et al., 2016). Reliability in quantitative research refers to replication and consistency, that is, if another researcher would adopt the same research design, would he get the same results? (Saunders et al., 2016). As qualitative research builds on other assumptions (often interpretive instead of positivistic) validity and reliability should be seen in the light of these assumptions.

### 3.12.1 Qualitative data

#### 3.12.1.1 Audio and video recording of observations

The observations of the think aloud sessions were audio and video recorded. The advantage of video recording is that it creates a permanent record, which is not possible with regular observations (Saunders et al., 2016). This enabled the researcher to view the think aloud sessions multiple times, replaying events that were found particularly interesting. This implied that the coding was possible to do more accurately, as observations were verified, thereby increasing the reliability of the findings.

#### 3.12.1.2 Validation

The qualitative data was validated by two forms of validation in order to strengthen its validity: (1) debriefing interviews with participant validation and (2) triangulation of the findings. As soon as the participant finished their think aloud sessions in phase 1, a short (5-15 minutes) debriefing interview took place. The debriefing interview was semi-structured, following open questions and questions from the notes of the researcher, which allowed the users to elaborate further on their actions, as well as validating observations made by the researcher. This partial use of this participant validation was conducted to

increase the validity of the findings (Saunders et al., 2016), as the participants provided confirmations and corrections on the researcher's observation notes.

Triangulation refers to using different sources/methods of data to confirm the validity of the data (Saunders et al., 2016). The sequential double phase design with mixed methods allowed for triangulation of the findings, which increases the validity of the results, given similar findings of the different methods (Bryman, 2006; Saunders et al., 2016). An example is that the quantitative results from the second phase were compared to the qualitative results of both the first and the second phase, in order to triangulate the results, but also to interpret (explain) the quantitative results.

#### 3.12.1.3 Researcher bias and observer effects

Researcher bias refers to the researcher biasing the recording of answers from the users, with his own views and beliefs. Though a pragmatist researcher lets his values and beliefs drive the research forward, the aim is still that user's behaviours and problems shouldn't be biased by those predetermined values and beliefs of the researcher. The researcher therefore tried to act as neutral as possible to the answers given by the participants, in order not to influence the participants with his own views.

There is also a risk that the research participants will try to give answers that they believe the researcher would like to hear (Saunders et al., 2016). For this, the researcher was clearly pointing out in the beginning of each think aloud session, that there were no right or wrong answers to the questions, and that the users' actual thoughts were of greatest importance.

### 3.12.2 Quantitative data

#### 3.12.2.1 Participant error and bias

Participant error and bias are threats to the validity of the research (Saunders et al., 2016). To reduce the risk of participant error, the questionnaire in phase 2 was kept short to avoid fatigue.

#### 3.12.2.2 Reliability of instruments

As pointed out by Kelly (2009), in research based on self-reported data where the researcher is the instrument, reliability issues can be complicated. A good way to minimize the reliability issues is to use instruments with established reliability. This is the reason for this thesis to adopt the established instruments of Su (2003a, 2003b), which have also been referred to by (Jansen & Molina, 2006; Kelly, 2009; Wang & Forgionne, 2008) and used as a basis in other studies (Campbell & Ash, 2006).

#### 3.12.2.3 Pilot testing of questionnaire

In order to reduce the risk of distributing an ambiguous questionnaire to a large number of users, a pilot test of the questionnaire was conducted with a handful of acquaintances. This allowed for feedback on the formulations in the questionnaire to be incorporated before sending out the questionnaire to the larger group in the regular study. The pilot testers also helped with estimating the time consumption of the questionnaire, which allowed for adjustment (removal) of some questions from the questionnaire, to keep the questionnaire short and avoid participants' fatigue.

#### 3.12.2.4 User sampling limitations

From a statistical point of view, when making inferences about a population given results from a sample, the sample needs to be representative of the population. As pointed out in section 3.4 the users were sampled using volunteer self-selection sampling aimed at students. This was due to resource constraints, as recruitment from the whole population of e-commerce users (if possible) would have been much more resource intensive. Section 5.1 indicates that the sample consists of mostly young employed males. As such the sample is not representative of the population, which limits the generalisability of the findings.

Most of the study participants were employed, in spite of targeting a students sample. This likely relates to the fact that many members of the students' Facebook groups where the researcher advertised the study, had already graduated. Some of the study participants recruited from these groups, had also previous relationships with the researcher, as members of these communities commonly help each other on research projects. This further biases the sample and limits the generalisability of the findings.

# 4. FIRST PHASE FINDINGS

The following subsections present the findings of the first phase.

## 4.1. Observations and interviews

In the think aloud search tasks, a total of 178 searches were observed in the eight video recorded user sessions, each following the ten different search tasks (appendix 1). In the following section, a brief overview of the quantitative data extracted from these observations will be presented.

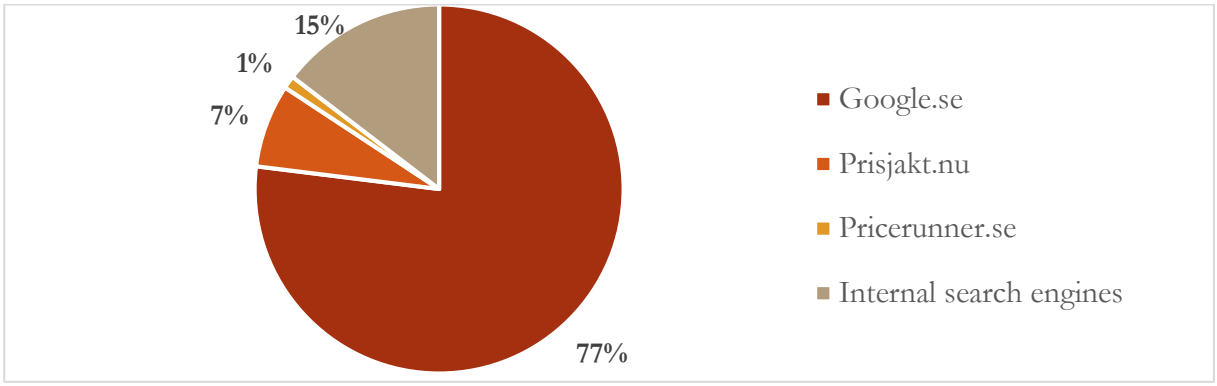



**FIGURE 5 - OVERVIEW OF SEARCH ENGINES USED, AND THEIR SHARES OF TOTAL QUERIES MADE.**

It's important to point out that Figure 5 doesn't show the time spent on each website, but only the percentage of the total number of queries made on each domain. The users spent most of their time in other websites than Google in order to solve their information needs, but in these other websites it was less common to use search (queries). This topic will be further elaborated in the discussion.

## 4.2. Users' approaches to search

The following section presents findings that relates to how users search in e-commerce. It includes strategies, techniques and web search functionality that the users employed to solve their information needs. The section starts with summarizing specific findings from each of the three dimensions of search, and then follows with more general findings.

### 4.2.1 The three dimensions

#### 4.2.1.1 A single general-purpose search engine: Google

As pointed out by Figure 5, Google dominated as the preferred search engine in e-commerce. Out of all queries (searches) made, 77% percent (137 out of 178 queries) was made using Google. In the video recordings, Google was used for all general-purpose searches. No other general-purpose search engine was used by any of the users to complete the search tasks.

#### 4.2.1.2 External specific search: Prisjakt and Pricerunner

This section uses e-commerce as the specific purpose of the search. Prisjakt.se and Pricerunner.se were the only external e-commerce specific search engines used. Those combined accounted for 8 % of all searches made in the video recorded user search tasks, as shown in Figure 5. Prisjakt was searched 13 times while Pricerunner was searched two times (number of queries made in each search engine). However, observing how the users used Prisjakt and Pricerunner, it became clear that the users tended to use them to a larger extent than the search query data suggests, as of browsing (section 4.2.2). Pricerunner was visited by five of the users and Prisjakt was also visited by five users. Two users used Pricerunner but not Prisjakt and two other users used Prisjakt but not Pricerunner. Only one user, user 2, neither used Pricerunner nor Prisjakt at all.

#### 4.2.1.3 Internal search engines

As seen from Figure 5, 15 % of the queries were used in internal search engines (i.e. at the website selling the actual product). The internal search engines were spread over 18 different domains (e-commerce stores). The most frequently used were Adlibris.se and Elgiganten.se, both receiving three queries each. There were also substantial differences in how well the search engines were working.

### 4.2.2 Browsing

In the external specific search and in the internal search, users commonly either (1) entered directly to the product page from an external general search engine (Google) or (2) used the navigational features of the website (such as categories, filters and sorting) to find the relevant products. This in contrary to using the search queries in the search engines. The second behaviour of users navigating themselves to a product page is identified as a browsing behaviour. This behaviour was also common in external to internal search (section 4.2.3).

### 4.2.3 External to internal search

The recordings of user search tasks showed that the transition from external to internal search was present. In total, it was observed 15 times that a user transitioned from external to internal search and continued with their search. The behaviour was observed for all users, except user 8.

An even more common behaviour was that users transitioned from an external search engine to an internal search engine, but instead of using the search function in the internal website, they used browsing (i.e. navigating with categories/filters/sorting) to solve their information need (i.e. find a relevant product). This browsing behaviour occurred more frequently than searching, for all users but one, who instead used searching and browsing an equal amount of times in external to internal search.

#### 4.2.3.1 Internal to external search

The observations also showed that users sometimes transitioned from internal to external search. This was often the case when the internal search returned no or just a few satisfying results, and the user would then go to the external search engine (i.e. Google) and make the same or a similar query. One behaviour was to include the name of the website the user intended to reach and combine it with the name of the product the user was searching for. Some examples are user 2 with the query "apotea nasal spray" and user 4 with the query "emporio armani watch pricerunner".

Two motives for this behaviour were identified: (1) to compare e-commerce stores selling the same product, (2) to use Google for searching the store's website, as the user doubts the performance of the internal search engine on the store's website.

An example of the second case is given by user 7 when he is searching for *"toner i natten"* at Prisjakt and receives no results. This is the same problem of coverage (i.e. Prisjakt do not include books) that also user 5 experiences in section 4.3.4. User 7's approach to the problem is to go to Google and make the search there too, including both the book title and the keyword "Prisjakt". He believes the Prisjakt search engine sometimes performs poorly, and that Google in these cases is more likely to find the product, even though it is actually located at Prisjakt website. When he does not find the book on Prisjakt nor through Google, he concludes: *"Prisjakt seems to have a poor search function here. I don't find any link"*

### 4.2.4 External modification – Refunder browser add-on

An unexpected but interesting finding is the Refunder web browser add-on. Refunder is a company providing cash-back to e-commerce customers, when they purchase goods and services from more than 700 online stores (Refunder.se, n.d.). User 4 is a user of this service and she has an add-on installed in her browser that helps her track her purchases and suggests the levels of cashbacks earned through the purchases on different sites. Figure 6, illustrates this by "FÅ 7.5% ÅTERBÄRING på alla köp från APOTEK HJÄRTAT" which translates to "Receive 7.5% cash-back on all purchases from Apotek hjärtat", where "Apotek hjärtat" is the name of a pharmacy. The interesting finding for e-commerce search is that the add-on also integrates itself into the search results page. It shows the cash-back levels for different e-commerce shops directly with their entry in the organic search listings, as marked with yellow in Figure 10.

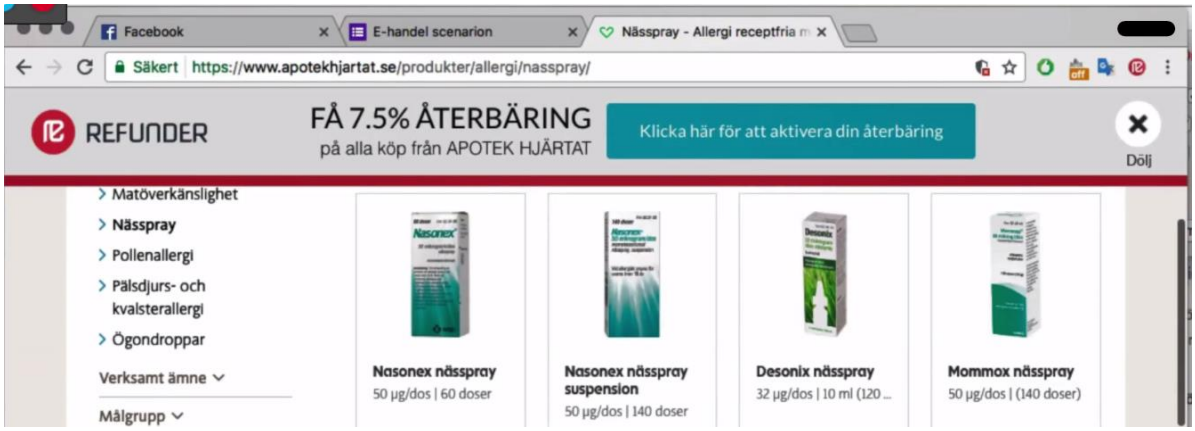



**FIGURE 6 - REFUNDER WEB BROWSER ADD-ON INDICATING CASHBACK LEVEL, SCREENSHOT FROM VIDEO RECORDING OF USER 4**

### 4.2.5 User experience influence choice of e-commerce stores

Another theme emerging from the think out loud was that most users preferred to visit websites or companies that they are already familiar with and trust. This implied that the result entries from such companies on the search results page, was more likely to be visited than results belonging to a website unknown to the user or difficult to interpret. The theme can be exemplified by reflections made by user 7 in search task 3, when he has just received the Google search results page for the query *"nasal spray delivery"*:

User 7 (thinks out loud): "*First, I look to see what pages I recognize. If it is, like medicines, then I would like to buy it at some place I know about.*" In the search results there are ads from Apotea.se (online pharmacy store) and Happygreen.se (online health store), and organic results from Apotea.se and Apoteket.se (online and offline pharmacy store). The rest of the results are non-visible. User 7 continues by explaining: "*I don't know what*

*Happy Green is, it sounds a little suspicious. But Apoteket, as an example, I do know what it is, and Apotea as well. There I would probably have clicked."* And so User 7 does, clicking on the organic result of Apoteket.se.

In many cases the user would go directly to the domain of the e-commerce store instead of using a search engine. For example, when user 4 starts out in search task 4, she navigates directly to Elgiganten and Cdon.se to start searching for dishwashing machines. She suggests that Elgiganten has good function for comparing different products and that Cdon.se usually has good prices.

### 4.2.6 Organization, reviews and comparison in external specific search

Though it was pointed out in the previous section that the user interaction with Prisjakt and Pricerunner is sometimes problematic, the results also show that the users do like using these sites when searching for products in e-commerce. The main advantages are their price comparisons, categories and filters for organizing products. Reviews are also mentioned as an advantage. Users believe they get more comprehensive information when searching with those external e-commerce specific search engines, as they include many different e-commerce stores and lots of product data about each product. This was particularly important for more expensive purchases, as the dishwasher in search task 4, where all users except user 2 used either Pricerunner or Prisjakt to find product information. User 8 exemplifies this thinking on Adlibris where he finds a book for 54 SEK. He says *"I believe 54 kr is a little too little/…/"*, implying that 54 SEK is already such a small price that there is really no need for price comparisons.

#### 4.2.6.1 Price comparisons

User 7 suggests that if he thinks the same product will be sold by many stores (such as books), he will use a website like Prisjakt or Pricerunner to compare its price. User 5 agrees as he suggests that: *"I want as low price as possible. There is no reason for paying too much as one will get the same product regardless of where one purchases it",* when he searches Pricerunner for the book in search task 2. User 1 also agrees, by describing that when she finds a product she likes, she will use Prisjakt.se or Pricerunner.se to see where the product is sold to the lowest price.

#### 4.2.6.2 Product reviews

User 1 uses Pricerunner to find product reviews, which she believes are important when choosing a dishwasher in search task 2, as she specifically don't want the dishwasher to be very loud. She reads the reviews to see what other people say about the noise level. She further suggests that she would be willing to pay more for the dishwasher if it has good reviews in general.

#### 4.2.6.3 Product organization

As previously mentioned, a common approach to finding products in the external specific search engines is to use categories and filters. Prisjakt.se makes this by showing icons to illustrate the contents of each category when the users browse the website (Figure 7). In each of the categories Prisjakt and Pricerunner offers a wide selection of filters to allow the user to restrict the list into a more manageable amount. The filters available depend on the category, as different products hold different attributes. The filter was for example used by user 5, user 8 and user 9 to restrict the measurements of dishwashing machines in search task 4. Sorting was also commonly used by the users, to arrange the products according to the highest reviews or the lowest price.

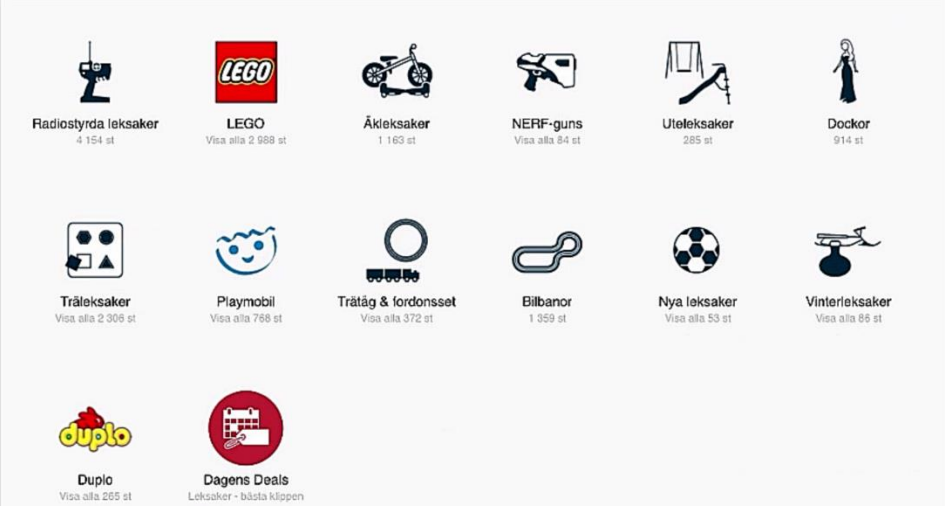



**FIGURE 7 - EXAMPLE OF PRODUCT CATEGORIES AT PRISJAKT.NU FROM VIDEO RECORDING OF USER 8**

## 4.3. Identified problems

The following sub-sections present the problems that were identified in the user sessions.

### 4.3.1 Ambiguous query: Lloyds pharmacy and Lloyds bank

The findings from the first phase show problems with ambiguous queries. An example is found in Figure 8, where the SERP of the Google search with query "Lloyds" is shown. The user was trying to reach Lloyds pharmacy. The initial search query, before obtaining the result displayed on Figure 8, was *"nasal spray online" ("Näs别spray online")*. The user was also logged in with her Google account (marked out with red for anonymization) in the Google website during both searches. The search results in Figure 8 are completely focused on Lloyds Bank, rather than on Lloyds Pharmacy (Lloyds Apotek), which is a major pharmacy in the Swedish market, with 79 pharmacies in 40 different cities and e-commerce (only for prescribed medicines) (LloydsApotek.se, n.d.). Further, it should be noted that the user was physically located in

Sweden at the time of the search, using no virtual private network service. She was using English language, instead of Swedish, with her Google account. To resolve the problem, the user reformulated her query to "Lloyd's pharmacy" *("Lloyds apotek")* and did then receive relevant results.

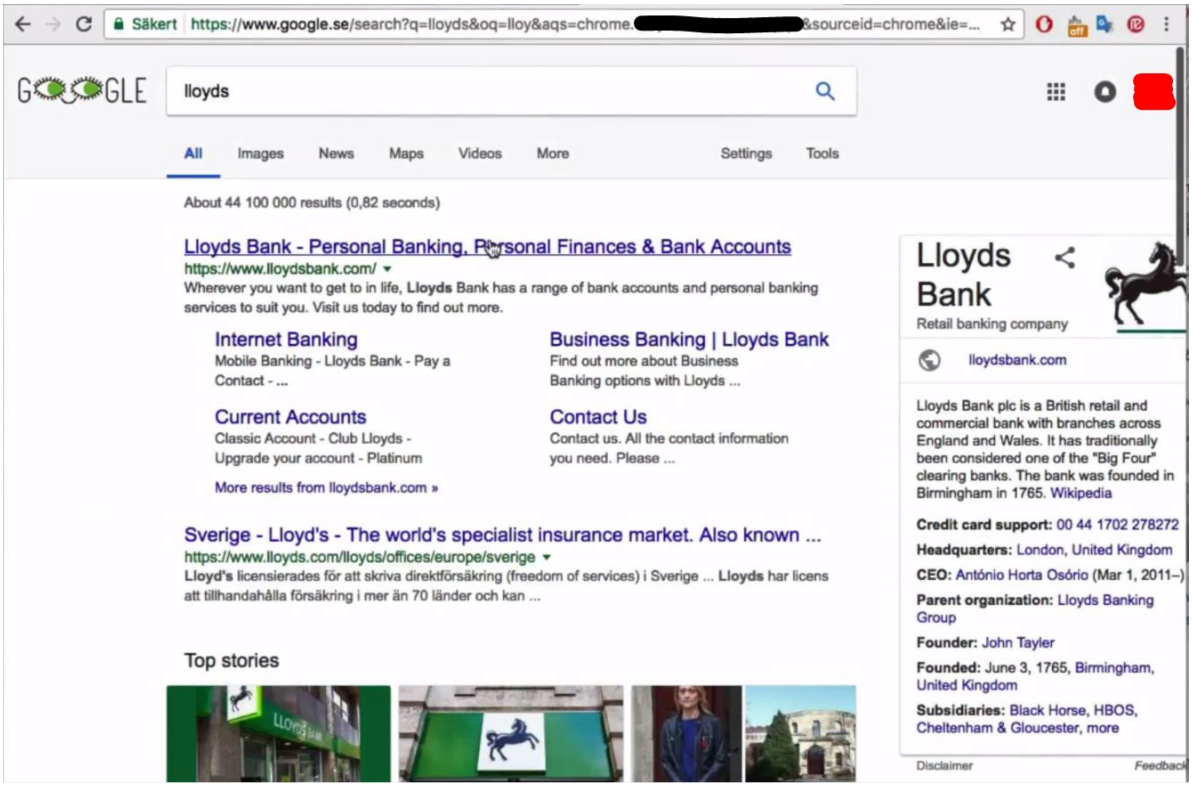



**FIGURE 8 - GOOGLE MISINTERPRETING THE INFORMATION NEED OF THE USER, SCREENSHOT FROM VIDEO RECORDING OF USER 4**

### 4.3.2 Product category as ambiguous query at Elgiganten

Another observation is made about query ambiguity at Elgiganten in Figure 9. Though user 2 uses a search query containing the exact match of the product name "bänkdiskmaskin" ("counter dishwasher"), the search engine's suggestions only present two relevant matches out of ten listed results, (1) the "Electrolux bänkdiskmaskin ESF2200DW" and (2) the category ("Kategori") "Bänkdiskmaskin". All the remaining results are results with product names containing the word "bench" ("bänk").

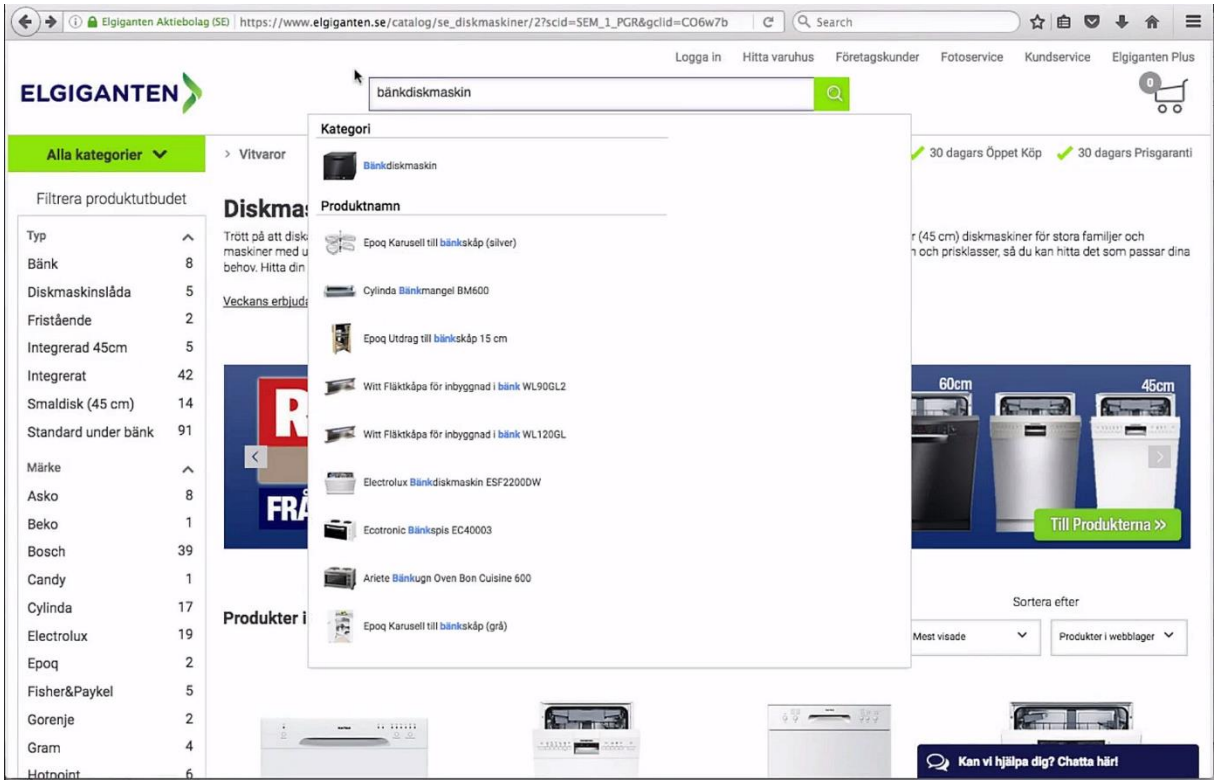



**FIGURE 9 - SCREENSHOT OF ELGIGANTEN.SE SEARCH FROM VIDEO RECORDING OF USER 2**

### 4.3.3 Google ads relevance problem

As previously suggested, Google was the completely dominating search engine in the study. Given its dominance, observations about its advertisement system for the search engine, Google AdWords, were made. This will be the focus of the following section, though sponsored search results listings are also found in other search engines. All participants made think out loud comments regarding the ads (the Google AdWords advertisements will from here on be referred to as just "ads"), ads that can be found in the red rectangle in Figure 10. The findings suggest that the users did have overwhelmingly negative opinions about the ads, and that they avoid looking or clicking on them. The following example is transcribed from the video recording (user 4):

> User (thinks out loud): *"I would never click on these ad-things, one doesn't even look at them."*
> Researcher probes: *"How do you think about that?"*
> User: *"Then you don't have pages good enough to get there without paying for it."*

User 5 describes his dislike towards Google ads and admits that he in general avoids using them. However, if he already knows what he wants, then he can use them. He says, that if he is looking for something specific, like a product or a website, and it is presented on the first row on the search results page as an ad,

he might as well click on it, rather than scrolling down and look through the organic results. This is especially common when he is using his phone, where a smaller part of the search results is visible than when using his laptop.

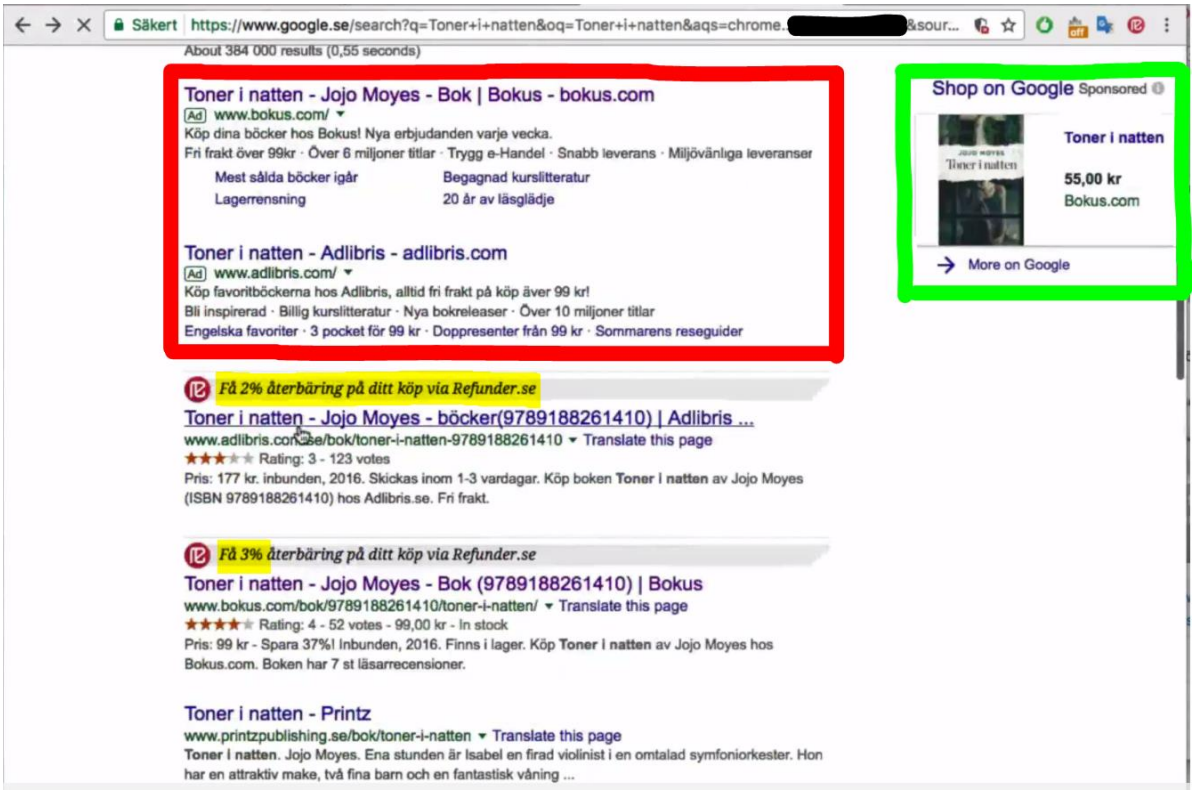



**FIGURE 10 - SCREENSHOT OF THE SEARCH RESULTS PAGE FROM THE RECORDED VIDEO OF USER 4.**

User 3 agrees that he also avoids using Google ads. He points out that the links are often less relevant than those in the organic search results listings. However, when he is searching for a watch in search task 10, he searches for "Rolex" (a manufacturer of luxurious watches), clicks on the top ad from Rolex.com in the search results listings, and speaks out: *"They have money, so they can pay a little to Google"*.

#### 4.3.3.1 Problem 1: Different listings for ambiguous queries

A problem of relevance in Google Ads that was observed, refers to the different listings Google makes for ambiguous queries. An example is found with user 5, in Figure 11. He uses the search query *"pricerunner Toner i natten"*, where *"Toner i natten"* is the exact title of the book being used as the information need in search task 2. The issue of ad relevance can be observed by comparing the sponsored listing (two top results) with the organic results (the two bottom results). The Swedish word "toner" have double meanings, "toner" as in the book title refers to *tone* (characteristics of a sound), "toner" as in the ads refers to *toner cartridge* (a container of powder mixture for printers). Hence, Google succeeds in presenting relevant results

in the organic listing (both organic results refer to the book) but fails in presenting relevant results in the sponsored listing (both results refer to the printer cartridge). Note: A third ad was cut out from the screenshot (Figure 11) to conserve space. This ad did also refer to printer cartridges.

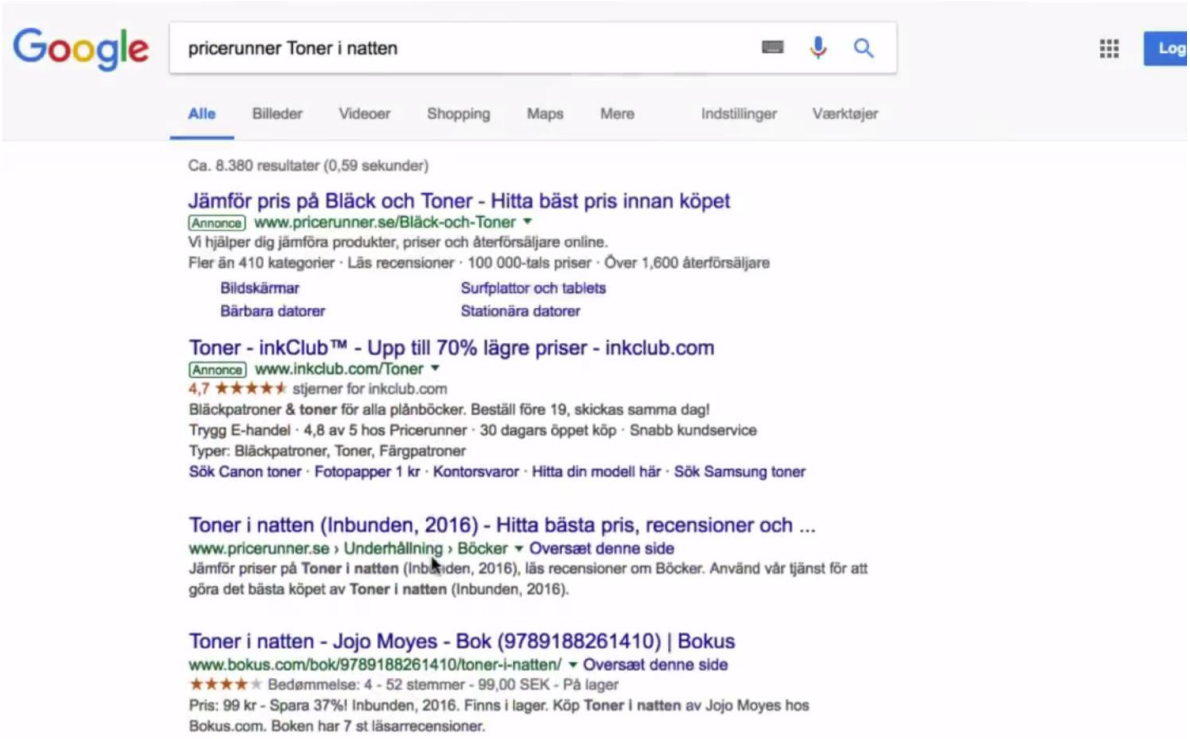



**FIGURE 11 - MODIFIED SCREENSHOT FROM THE RECORDED VIDEO OF USER 5. (MODIFICATIONS: REMOVED PRIVACY WARNING AND SPONSORED RESULT NO. 2)**

The same query as used in Figure 11, but with "pricerunner" cut out, is also used by user 4 in Figure 10. In this search results page, both the ad listings and the organic listings refer correctly to the book, and not to printer cartridges. This though the ambiguous word "toner" is used in both queries (Figure 10 and Figure 11).

#### 4.3.3.2 Problem 2: Ad content

Another observation of ad irrelevance relates to the content (message) of the ad itself. This problem can be illustrated by comparing the content of the ad from Adlibris (Figure 10) with the content of the organic result of Adlibris (Figure 10). It's clear that the content differ, though the result (and its link[7]) refer to the same product page on Adlibris.se. However, due to the content difference, users expressed confusion about

[7] As most users avoid using the sponsored links, it was not possible to observe the destination of the link of the ad from Figure 10. Instead, when the observational session had ended, to better understand the ad, the researcher used the same query as the user in Google and tried the link. It showed that the link in the ad had the same destination as the organic result (the product page of the requested book).

the expected destination of the ad. The organic result clearly suggests that the link destination is the actual product page. This as the destination (the green URL, located under the title of the ad), is perceived as specific when it includes the product name and ISBN (product id): www.adlibris.com/se/bok/toner-i-natten-9789188261410 (Figure 10). The destination (the green url) in the ad is showing only the general website: www.adlibris.com. In Figure 10, the ad's phrase "Buy the favourite books at Adlibris, always free shipping for purchases over 99 kr!" describes the shipping terms of Adlibris in general, while the organic result phrase describes attributes and terms for the book, such as price, binding, year of publication, shipping speed and ISBN. In addition, the organic result includes a rating of the book from previous users.

Some users reflected on this difference and commented that they believe they are more likely to end up on the product page if they use the organic link rather than the sponsored link (the ad). This is a reason for preference in organic results rather than sponsored results.

### 4.3.4 Information accuracy and coverage in specific external search

The problem with coverage could be exemplified with user 7's attempt to search for the book in search task 2 using its ISBN number in Prisjakt, without receiving any results. He also makes a separate query with the book's title, "Toner i natten", but receives only results that he concludes not to be relevant. The user switches to Pricerunner and with the book title as a query, he gets a well-done product page showing the book cover, title, publication date, language, binding and prices from at least five different stores and their links (The user didn't continue further in the price listings). As he returns to the Prisjakt tab, he admits that he is *"A little irritated that I didn't find it here"* referring to Prisjakt and the book, as Prisjakt didn't find the book there.

#### 4.3.4.1 Problem with information accuracy

User 3 points out the price history feature of Prisjakt as valuable, as he can know if the current price is a favourable price. However, as he uses Prisjakt to track the prices on a pair of shoes for search task 1, he sometimes doubts the accuracy of the prices:

> *"Then I saw that the price has been as low as 450kr during the end of the summer in 2016. I do believe that it might be incorrect, as another time when I was online looking, the Prisjakt prices wasn't always correct. It was actually more expensive on the actual website"*

User 3 refers to "the actual website" as the website where the goods are sold (the e-commerce store).

#### 4.3.4.2 Behaviours for countering the information accuracy and coverage problems

The users also have strategies for dealing with the information accuracy and information coverage problems. As an example, User 5 suggests that he usually checks with both Pricerunner and Prisjakt in case one of them lacks some information. User 5: *"Maybe I'm strange but I believe that they sometimes don't get all companies in both of the websites. So, I always try to double check"* where "companies" refer to the stores selling the product, and "both websites" refer to Pricerunner and Prisjakt. A similar strategy is employed by user 7 and user 5 as they switch between Prisjakt and Pricerunner when the information needed is not found.

# 5. SECOND PHASE FINDINGS

The following subsections present the findings of the second phase.

## 5.1. Non-performance characteristics

In this section the non-performance characteristics (i.e. the background) of the participants are presented. In Figure 12, the year of birth among the users in the second phase are shown. The users are born between 1997 and 1988, and the most common year of birth was 1992. Out of the 26 users participating in the second phase, eight were females (31 %) and 18 were males (69 %).

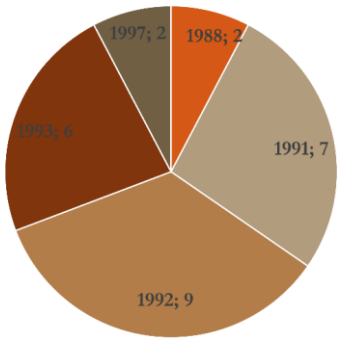



**FIGURE 12 - YEAR OF BIRTH OF USERS**

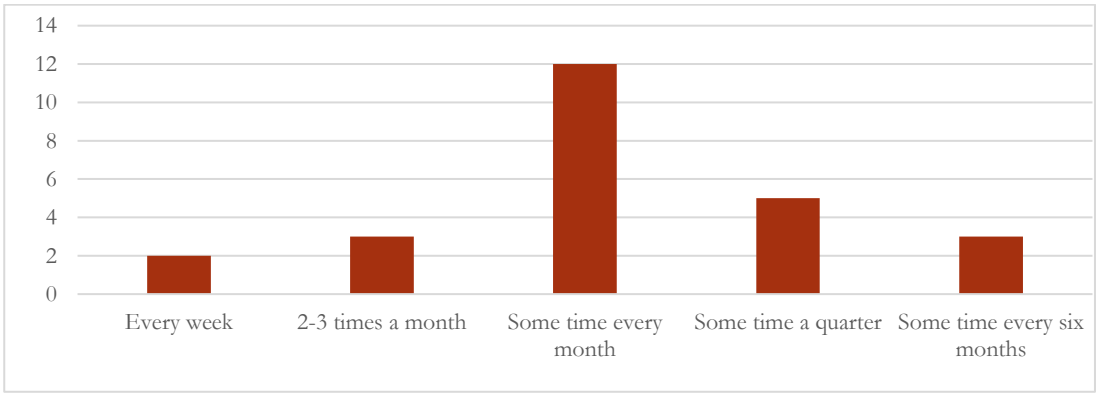



**FIGURE 13 - E-COMMERCE EXPERIENCE OF USERS**

Figure 13 shows the frequency of conducting e-commerce among the users in the second phase. 80 % of the users purchase goods online at least some time every month. The most common occupation among the

users was being an employee (76 %), followed by being a student (16 %). One of the users was working at home with the household, and one user was self-employed.

## 5.2. Parametric testing

The parametric testing was conducted using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Each of the tests on each of the variables gave significant results ($p < 0.05$), indicating that the collected data on the variables was not normally distributed.

## 5.3. Performance measures

The following subsection present the analysis of the performance data from the second phase.

### 5.3.1 Relevance

The study used relevance measures in the form of two precision ratios, Precision ratio #1 and Precision ratio #2. In Table 2, an overview of the precision ratios for the different search engines is presented.

**TABLE 2 - OVERVIEW OF RELEVANCE MEASURES**

| Measure | Mean | Median | SD |
|---|---|---|---|
| **Precision ratio #1 (%)** | | | |
| Google | 91,54 | 100 | 12,23 |
| Prisjakt | 88,08 | 90 | 12,34 |
| Elgiganten | 78,08 | 80 | 22,98 |
| **Precision ratio #2 (%)** | | | |
| Google | 61,15 | 60 | 27,18 |
| Prisjakt | 54,62 | 50 | 25,18 |
| Elgiganten | 45,00 | 40 | 26,87 |

Both precision ratios indicate that Google has the highest relevance, followed by Prisjakt and finally Elgiganten. The ratios were also statistically tested using the Kruskal-Wallis test, for which precision ratio #1 showed a significant difference (p-value = 0,037) for at least one of the groups at the 5 % level. The pairwise comparisons found a significant difference between Elgiganten and Google at the 5 % level. No significant difference was found in the other pairwise comparisons for precision ratio #1.

For Precision ratio #2, the Kruskal-Wallis test did not find any significant difference for any of the groups at the 5 % level. An extra Kruskal-Wallis test was conducted at the 10 % level, and showed a significant

difference (p-value = 0,065) for at least one of the groups. The pairwise comparisons found the difference between Elgiganten and Google significant (p-value = 0,062) at the 10 % level.

### 5.3.2 Efficiency

Efficiency was measured using the search strategy measure, operationalized as the number of queries committed to the search engine for each search task. The lowest average of queries was submitted to Google (mean=2.15, median=1, SD=2.15), followed by Prisjakt (mean=2.58, median=2, SD=2.04) and most queries were submitted to Elgiganten (mean=3.31, median=3, SD=2.77). The Kruskal-Wallis test showed no significance (p-value=0.111) neither at the 5 % level nor at the 10 % level and so, no pairwise comparisons were conducted.

### 5.3.3 Utility

Utility was measured using the *value of search results as a whole* on a Likert scale. Prisjakt got the highest scores (mean=5.85, median=6, SD=1.01), followed by Google (mean=5.27, median=5, SD=1.25) where higher rating is better. Elgiganten got the lowest score (mean=3.35, median=3, SD=1.47). The differences were also tested using the Kruskal-Wallis test, which found significant differences (p-value=0.000) in at least one group at the 5% level. The pairwise comparisons found the difference between Prisjakt and Elgiganten (p-value=0.000), as well as between Google and Elgiganten (p-value=0.000) to be significant at the 5 % level. No significant difference was found between Google and Prisjakt.

### 5.3.4 Personalization

The degree of personalization was measured using a 7-point scale. Google (mean=4.85, median=5, SD=1.71) got the highest scores, followed by Prisjakt (mean=3.85, median=4, SD=2.07) and at last Elgiganten (mean=2.54, median=2, SD=1.45). The Kruskal-Wallis test for differences between groups showed a significant difference (p-value= 0.000) for at least one group at the 5% level. The pairwise comparisons revealed a significant difference (p-value 0.000) between Elgiganten and Google at the 5 % level and a significant difference (p-value=0.058) between Elgiganten and Prisjakt at the 10 % level.

#### 5.3.4.1 User opinions on personalization

In addition to the user judgement on levels of personalization for the search engines, the users were also asked about their opinion regarding personalization. The user responses were categorised as either clearly positive, clearly negative or other. The findings suggest that eight users were clearly positive, four users were clearly negative and the remaining 14 users fell in the category other (i.e. those that don't have a clear opinion or no opinion at all).

From the opinions, the themes that emerged were that many users view personalization positively in terms of search performance (nine users, e.g. *"Positive for finding more relevant results.", "Of course, positive that search engines try to deliver optimal results depending on my behaviour online"*), but also in conjunction with concerns and risks. Most frequently in terms of privacy concerns (five users, e.g. *"Negative in terms of monitoring is a little scary"*, *"However, it's not fun to feel supervised as information about what I'm doing online is stored and thus creates a profile about who I am and what I do.")* and some mentions the risk that one might "live in a bubble" (four users, e.g. *"The negative, however, is that you may live in a bubble or just get served in the same style, pattern, idea, etc. and does not get the opportunity to discover new things", "It can be smooth but unpleasant that it adjusts the results without showing certain parts."*).

In addition, four users pointed out that the performance of personalization today is poor according to their experiences (e.g. *"Not so important, never feel like it works or the results are so customized to me"*). Two users also mentioned retargeting marketing (i.e. marketing that follows the user through many different websites because of something the user has viewed) as a disturbing form of personalization *(e.g. "Good in connection with the search, but unrewarding when targeted advertisements persist for a long period of time after specific searches").*

### 5.3.5 User satisfaction

The two user satisfaction measures; overall success and time saving were both measured using Likert scales. Table 3 shows an overview of the user satisfaction measures.

**TABLE 3 - OVERVIEW OF USER SATISFACTION MEASURES**

| Measure | Mean | Median | SD |
|---|---|---|---|
| **Time saving** | | | |
| Google | 5.50 | 5 | 1.17 |
| Prisjakt | 6.08 | 6 | 0.98 |
| Elgiganten | 3.23 | 3 | 1.56 |
| **Overall success** | | | |
| Google | 5.77 | 6 | 0.95 |
| Prisjakt | 5.73 | 6 | 1.04 |
| Elgiganten | 3.35 | 3 | 1.79 |

The findings indicate that Google got the highest scores on both time saving and user satisfaction, while Elgiganten had the lowest score on both. The statistical testing showed a significant difference for at least one of the groups at the 5 % level for both time saving (p-value=0.000) and overall success (p-value=0.000)

according to the Kruskal-Wallis test. The pairwise comparisons showed a significant difference between Elgiganten and Google (p-value=0.000), as well as between Elgiganten and Prisjakt (p-value=0.000) for time saving at the 5 % level. Furthermore, the pairwise comparisons also showed a significant difference between Elgiganten and Google (p-value=0.000), as well as between Elgiganten and Prisjakt (p-value=0.000) for overall success at the 5 % level. There were no significant differences between Google and Prisjakt for neither time saving or overall success.

### 5.3.6 Overview of performance findings

The following section contains a summary of the statistical analysis of the quantitative data, as seen in Table 4.

**TABLE 4 - OVERVIEW OF PERFORMANCE FINDINGS**

| **Measure** | **Kruskal Wallis sig.** | **Pairwise comparisons adj. significance** | |
|---|---|---|---|
| **Precision ratio #1** ***(Relevance)*** | 0,037 ($H_0$ Rejected) | Elgiganten - Google | 0,032 |
| | | Elgiganten - Prisjakt | 0,445 |
| | | Google - Prisjakt | 0,801 |
| **Precision ratio #2** ***(Relevance)*** | 0,065 ($H_0$ Retained) (rejected at 10% level) | Elgiganten - Google | 0,062* |
| | | Elgiganten - Prisjakt | 0,453 |
| | | Google - Prisjakt | 1,000 |
| **Search strategy** ***(Efficiency)*** | 0,111 ($H_0$ Retained) | | |
| **Value of search results as a whole** ***(Utility)*** | 0,000 ($H_0$ Rejected) | Elgiganten - Google | 0,000 |
| | | Elgiganten - Prisjakt | 0,000 |
| | | Google - Prisjakt | 0,474 |
| **Personalization** | 0,000 ($H_0$ Rejected) | Elgiganten - Google | 0,000 |
| | | Elgiganten - Prisjakt | 0,058* |
| | | Google - Prisjakt | 0,191 |
| **Time saving** ***(User satisfaction)*** | 0,000 ($H_0$ Rejected) | Elgiganten - Google | 0,000 |
| | | Elgiganten - Prisjakt | 0,000 |
| | | Google - Prisjakt | 0,471 |
| **Overall Success** ***(User satisfaction)*** | 0,000 ($H_0$ Rejected) | Elgiganten - Google | 0,000 |
| | | Elgiganten - Prisjakt | 0,000 |
| | | Google - Prisjakt | 1,000 |

* Significant at 10 % level, but not at 5 % level.

### 5.3.7 Preferred e-commerce search engine

In response to the question *"If you were allowed to choose only one search engine to use when you conduct e-commerce, which one would you choose? Why?",* the results of the preferred search engine can be found in Figure 14. The results show that half of the users (13 users) preferred Google and the other half (13 users) preferred Prisjakt. No users preferred Elgiganten.

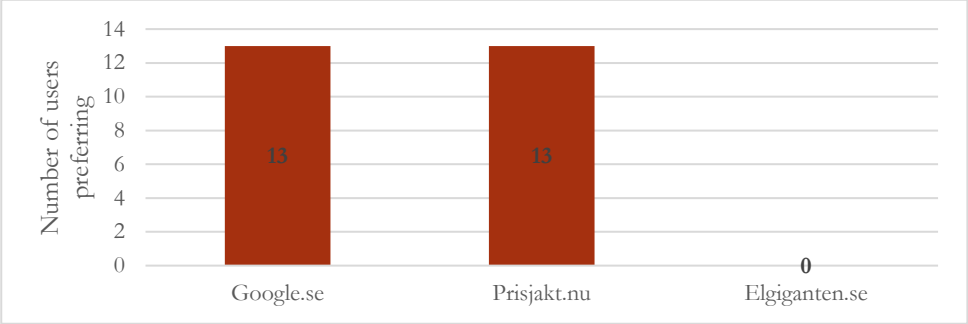



**FIGURE 14 - NUMBER OF USERS PREFERRING EACH OF THE SEARCH ENGINES FOR E-COMMERCE SEARCH**

The preferred search engine question also asked for the participants' motivation to their choice, which 16 of the users supplied. The most common reasons for using each of the search engines are listed below in Table 5 and Table 6. Elgiganten is left out as no user preferred it and thereby didn't provide any motivation. The number in parenthesis refer to the number of users mentioning the motive.

**TABLE 5 - MOTIVES FOR USING GOOGLE AS PREFERRED E-COMMERCE SEARCH ENGINE**

| **Google motives** | **Example formulations** |
|---|---|
| Coverage (4) | *"Google, where everything is gathered"*<br>*"Google allows me to enter many sites, Prisjakt only gives me Swedish sites"* |
| Habit (2) | *"I'm used to how Google is working/…/"*<br>*"Google, by pure habit"* |
| Precision (1) | *"Google, more accurate and detailed search"* |
| More detailed (1) | *"Google, more accurate and detailed search"* |
| Usability (1) | *"Google is unbelievably user friendly"* |
| Google shopping (1) | *"I like that it now has proposals of products and their prices above the search results"* |

**TABLE 6 - MOTIVES FOR USING PRISJAKT AS PREFERRED E-COMMERCE SEARCH ENGINE**

| **Prisjakt motives** | **Example formulations** |
|---|---|
| Product data (4) | *"Prisjakt as it has price comparisons and information with many details."* |
| Filtering (4) | *"Prisjakt, as it allows one to easy filter on price, features and reviews"* |
| Comparisons (4) | *"/…/ easy to overview and compare prices between sources"*<br>*"Prisjakt as it has price comparisons and information with many details."* |
| Relevancy (3) | *"Relevant hits from many stores"*<br>*""/…/ and more relevant results than the other ways"* |
| Coverage (2) | *"largest range of products"* |
| Reviews (2) | *"/…/ there is more advanced filtering and reviews"* |
| E-commerce focus (1) | *"It is more e-commerce oriented than Google"* |
| Usability (1) | *"Clearly the most user friendly"* |

### 5.3.8 Advantages and disadvantages of the search engines

#### 5.3.8.1 Google

**Advantages**

The three most mentioned advantages of Google are its (1) coverage, (2) relevance and (3) overall performance on general queries. The coverage relates to Google's ability to index everything everywhere. Users mentioned that indexing websites outside Sweden was an advantage (e.g. *"Also comes with many different results and sources - even foreign which is good"*). In addition, a user suggests that websites don't need to *"signup themselves to be visible there"*, (i.e. websites do not need to sign up to Google but get automatically included).

Relevance relates to the finding that several users mean that an advantage of Google is that many of its results are relevant (e.g. "*Relevant answers but many ads.", "Most commonly find the search results that are most relevant at the beginning."*). Some users connect the characteristics of coverage and relevance to broad queries, suggesting that the performance is high on broad general queries, but not as good on specific queries (e.g. *"Extremely wide and with many results but difficult to specify when searching for a very specific product"*, *"Difficult to immerse and find just the kind of product you want"*.)

**Disadvantages**

The users mentioned three times more disadvantages than advantages of Google. The three most common disadvantages mentioned about Google are (1) ads, (2) inadequate price information (3) poor specific search. Each of the areas can be further divided into smaller problems. Starting with Google's ads, many users found them to be irrelevant, poorly separated from organic results and that Google has too many of

them. This is supported by the following user quotes. A user suggests that Google contains *"Many purchased ads that are profiled as search results but do not match what you are looking for"*. Another user believes that it *"can be difficult to separate the ads from the independent results, which is a disadvantage"*. Other users point out the large number of ads *(e.g. "Relevant answers but a lot of ads"*).

The disadvantage of inadequate price information, included the lack of price comparisons in Google, which implies that the user has to enter each result to compare prices (e.g. *"May be difficult to compare prices when you have to click on each page", "The downside was that I had to go into each website individually to compare prices etc."*). The problem of price comparisons is also connected to the problems of ads, which one user describes as. *"The disadvantages are that many are sponsored and may not necessarily show the most affordable results"*. Google is great at answering *"what"* to buy but not *"where"* to buy it. One of the users explains it like this *"Google is a great way to get advice on what to buy, but it's not as easy to find where the product is cheapest"*.

The third disadvantage of Google relates to what users describe as its lack in specific searching (e.g. *"Difficult to immerse and find just the kind of product you want.", "/…/difficult to specify when searching for a very specific product"*). The lack of specific searching can be connected to Google's absence of filters (e.g. *"/…/And then there are no filters."*). One user points out that there are no possibility filtering out only products, as a disadvantage (e.g. *"Google has a lot of different sources, unfortunately, you will not only get products when you search"*) Not all users agree on Google's lack in specific searching. As an example, one user believes that Google is both broad and specific, as Google gives *"/…/a wide search while you can easily target more specifically when you find something you want."*

Furthermore, another user mentioned forced personalization as one disadvantage of Google. This as they track you and save your information in order to put *"blinders"* on your searches (i.e. a pair of small leather screens attached to a horse's bridle to prevent it seeing sideways and behind), implying that personalization disallows the user from seeing the whole picture within a search.

#### 5.3.8.2 Prisjakt

**Advantages**

The three main advantages of Prisjakt are (1) product info (2) coverage and (3) filters and sorting. Many users say that the product information at Prisjakt is its main advantage. This includes its price comparisons (e.g. *"If you already know what product you want, then the price comparison is very good"*, *"You can easily see who sells the product at the cheapest price"*) but also its overall detailed information about products (e.g. *"Comparative price and detailed information", "Advantage of being easy to see price and more information about products quickly"*). Reviews, ratings and product pictures are also part of the product information and considered an advantage (e.g. "*The advantage is that in the result you can see ratings on the respective TV"*).

Coverage is also considered high for Prisjakt (e.g. *"'Large selection, easy to compare prices"*) pointed out specifically for products (e.g. *"Everything is in Prisjakt and you can easily see who sells at the cheapest price*", *"Shows only products, and feels like it shows all of them."*). However, coverage is a point of conflict in the user opinions, as many believe the coverage of e.g. prices isn't always complete and accurate.

The third advantage, filters and sorting (also referred to as specific search, e.g. *"Advantages: advanced filtering"*, *"It was difficult to search for a price, but it is easy to filter by price after the search. Would rather use the filter function than the search function."*) does partly overlap with coverage, as users point out that it is specifically coverage on products (i.e. filters away other results) that is Prisjakt's strength (e.g. *"Prisjakt is good because you only get products, but you need to use the filters to make it work well"*). Another part of filters and sorting is Prisjakt's categorization of products, which many users point out as an advantage.

**Disadvantages**

The main disadvantages that users experience with Prisjakt are that (1) product information is inaccurate or incomplete and (2) coverage. There isn't a clear third problem, but rather a cluster of other problems. As pointed out in the previous section, a main advantage with Prisjakt is its comprehensive product info, which is also the source to its largest disadvantage, that the information is inaccurate. The users mention many different ways in which the product information is inaccurate or incomplete, but the major ones are price and stock information (e.g. *"Not always completely up to date with current prices and inventory"*, *"sometimes shows misleading prices, for example, there may only be one store selling at the advertised price, and that store you do not trust", "Bad that the price isn't always correct"*). One user finds the product ratings of Prisjakt untrustworthy (e.g. "*I do not really trust Prisjakt's ratings*"), while another user points out that *"some products have no ratings and reviews"*.

Coverage is also pointed out as a disadvantage by some users (e.g. *"There are no endless results.", "All search results are not available, but it's convenient to search there if the price is crucial."*.), which not only applies to products but also to prices (e.g. *"However, I'm unsure whether they compare all companies' prices"*)

Other disadvantages pointed out by the users are that Prisjakt do not allow misspellings in the search, it only exactly matches the search queries (e.g. *"My experience is that one cannot make spelling errors in their search engine"*). The search results listings do only show twelve products, which is too few. The search results listings also display many of what seems like the same product, with minor differences. This is hard to identify from the search results listings (e.g. *"The same product appears several times with differences that are not always easy for the customer to spot"*).

#### 5.3.8.3 Elgiganten

**Advantages**

Users point out the possibility to order directly at the website as an advantage (e.g. *"You can shop directly on the page",*). Some users also point out that Elgiganten provides good filters (e.g. *"Clear results and good filtering to sort the results", "Good ability to weed out irrelevant results, as the search field didn't accept specific queries"*), but at the same time some users emphasizes the opposite. There are also views that Elgiganten's results are presented in a clear way, with high relevance (e.g. *"Advantages: Relevant Results"*, but some users have the opposite opinion.

**Disadvantages**

The most frequently mentioned disadvantages with Elgiganten's search engine are that (1) it only includes Elgiganten's products and (2) it performs poorly on specific searches. The first problem (e.g. *"It is limited to Elgiganten's assortment"*) relates to the user's need for comparing prices within e-commerce, which Elgiganten lacks (e.g. *"The disadvantage is that you want to be able to compare between different stores in a smooth way, which obviously isn't possible directly on a dealer's side.", "I would never shop on a page without comparing prices with competitors. Of course, I cannot do that from Elgiganten's webpage.")* The second is the quality of the search engine. Users find the relevance to be low (e.g. *"Hard to find relevant results with only keywords"*, *"One-sided search function that does not always match the relevance"*), absence of spelling correction (e.g. *"Also, it cannot generate search results if you spell errors. ex. "samsun tv 40" shows nothing but "samsung tv 40" shows search results with products."*) and no hits if queries are too specific (e.g. *"It was enough for some predetermined search parameters, but when you're looking for something more specific, you will not get any hit"*). Also, some users experienced the sorting/filtering to be poor (e.g. *"Difficult to search with, as the sorting of the hits was poor").*

In addition, users mentioned inadequate product information for some products as a disadvantage (e.g. *"The information about the products differ a lot - sometimes inadequate"*).

## 5.4. Identified problems

### 5.4.1 Specific search

The comments from users on their experiences from each of the search engines, shows problems when formulating queries. An example that many users point out is the problem of formulating specific product attributes in the search query at Prisjakt and Elgiganten. Specific queries are identified as queries that are more specific in nature (in terms of product attributes etc.). The following queries from the second phase are examples of more specific queries *("vacuum cleaner Electrolux energy class C cheap", "Samsung 40 full HD smart*

*tv", "laptop thin and lightweight"*). The following queries are examples of more general queries (*"vacuum cleaner", "Samsung tv 40", "laptop"*).

The following are examples of user comments on problems in specific search with Prisjakt:

*"Could not search for, for example, energy class C, then no results were found. You had to go in and read more about each vacuum cleaner to see if it met my requirements, which reduces the rating."*

*"Without filter, Prisjakt's search function and its results are not good at all."*

The following are user comments on problems in specific search with Elgiganten*:*

*"I could not use price requirements in the search function"*

*"Searched for the phrase "laptop" and then it showed results including tablets that I do not consider relevant to what is considered a laptop.*

*"All vacuum cleaners showed up. Thereby no directly specified search results"*

# 6. PRELIMINARY DISCUSSION

## 6.1. Identified e-commerce search engines

As seen in Figure 5, the dominant search engine is Google, accounting for 77 % of the searches. This is not an unexpected result as it was already pointed out in the introduction that Google has a very dominant position in the search engine industry, and that users commonly use general search engines, such as Google for their e-commerce queries.

Prisjakt was also identified as an important search engine for e-commerce, accounting for 7 % of the queries. The rest of the identified search engines can be found in appendix 3.

## 6.2. Implications for second phase and the construction of an evaluation framework

### 6.2.1 Connectivity

Connectivity was used in the previous literature as the proportion of valid links out of all links tested by a user in a search engine (Dong & Su, 1997; Su, 2003a). High validity of links helped the user to save time. Results indicated that between 11 and 18 percent of the links were broken (Su, 2003b), and thus connectivity was an important measure of a search engine's performance. Already during phase 1, connectivity was in the spotlight through the observations. The observational findings of phase 1 only revealed one broken link in approximately nine hours of e-commerce searching by the different users.

Following the outlined philosophical assumptions of pragmatism (section 3.1), where findings are prioritized giving their potential to provide change and action, this small amount of broken links does not motivate the use of connectivity as measure for the second phase. It is not likely to provide major implications on practice if measured, and should therefore be excluded to give priority for other measures. The previous high proportion of broken links, likely refers to search engines being less advanced in detecting broken links 15 years ago when Su (2003b)'s study was conducted.

### 6.2.2 Personalization

The results of phase 1 indicate problems with detecting the user information need in search, also in Google. The example of the confusion between Lloyd's Pharmacy and Lloyd's Bank in section 4.3.1, clearly illustrates the issue. As described in the theoretical framework section 2.4.1.1, one approach to dealing with this problem is to use short-term interests (e.g. browsing history) of the user when the search engine makes the rankings. As the search engine was aware that the user had just been searching with the query "nasal spray online" before entering the ambiguous query "lloyds", it shouldn't be difficult for the search engine to figure out that the user was looking for Lloyd's Pharmacy and not for the Lloyd's Bank. This is example is very practical and an indication that Google.se either ignores the user's short-term interests (section 2.4.1.1) in the rankings, or that Google's performance on using the user's short-term interests (section 2.4.1.1) is poor.

As seen from Figure 8, the initial results page for the query "lloyds" offers only results related to Lloyd's Bank and Insurance, no results on Lloyd's Pharmacy. Though the query "lloyds" is clearly ambiguous, there is no diversity or search suggestions (section 2.4.1.3) available. This example is very practical and an indication that Google does not properly implement the diversity or search suggestions features (section 2.4.1.3).

To conclude, for the first phase personalization of search is very interesting and could potentially have large practical implications (section 3.1) on the performance of e-commerce search (i.e. the research question). Personalization should therefore be included for further investigation in the performance evaluation (i.e. phase 2) and potentially extended. Though the questionnaire is intended to be short, and to focus on collecting quantitative data, an open question should be included for better understanding the users' attitudes towards personalization and its performance, as those attitudes could be important for producing well informed design objectives.

# 7. FINAL DISCUSSION

## 7.1. Comparison of search engine performance

This section addresses the question of a possible existence of "*differences in the relative performance of common search engines when used to find e-commerce products?"*. The quantitative findings suggest that there are relative differences between different search engines. As shown in section 5.3, Google scores best out of the three search engines in all measures, except for in *utility* and in the user satisfaction measure *time saving,* where Prisjakt had the best scores. Elgiganten got the worst scores in all measures. Not all differences were significant, but in Table 4 in section 5.3.6, it can be observed that Google performs significantly better than Elgiganten in the measures: *precision ratio #1, value of search results as a whole, personalization, time saving and overall success*. Prisjakt performs significantly better than Elgiganten in the measures: *value of search results as a whole, time saving and overall success*.

As previously pointed out, the study uses mixed methods to triangulate and to broaden the understanding of the complex interaction between the users and the search engines. Therefore, the quantitative findings of each of the performance measures, will in the following sections be discussed in relation to the qualitative findings and identified problems.

### 7.1.1 Relevance

As suggested, there is a significant difference in relevance between the results of Google and Elgiganten for precision ratio #1 (i.e. including partly relevant hits). The users evaluate between 78 % and 91,5 % of the hits in the search engines as relevant, which is relatively high (in Su (2003b) the search engines scored between 51 % and 63 %). For precision ratio #2 (i.e. including only relevant hits), the scores are instead between 45 % and 61%, which can seem relatively high in relation to the scores between 30 % and 37 % in Su (2003b). However, it's still poor performance when only around half of the retrieved results are relevant. Likely, it takes extra effort from the user, as he or she has to evaluate each result at the SERP.

In terms of the qualitative findings, the Elgiganten problem on dishwasher results (section 4.3.2) and the Google ads relevance problem (section 4.3.3) supports the problem of relevance found in the quantitative results.

### 7.1.2 Efficiency

The efficiency measure (search strategy) didn't provide any significant differences between the search engines. This is likely due to the high variance between the results, as the number of search queries made can fluctuate significantly between different search episodes and users.

### 7.1.3 Utility

As suggested, there are significant differences between Google and Elgiganten as well as between Prisjakt and Elgiganten, were Elgiganten had significantly lower scores of utility in both cases. One of the major problems with Elgiganten, pointed out by users (section 5.3.8.3), was the lack of price comparisons on the website. Consumer electronic products, as TVs, are often sold by many e-commerce stores, and users are interested to find the store with the lowest price. Using a specific internal search engine, will not allow this comparison, as stores rarely list their competitors' prices. Even if the products are relevant for the user, many will dismiss them as they also need to know that prices are good, to find them of good utility.

### 7.1.4 User satisfaction

There were two different user satisfaction measures, *time saving* and *user's judgement of overall success*. For both the measures Google and Prisjakt scored significantly better than Elgiganten, but with no significant difference between Google and Prisjakt. In terms of time saving (i.e. whether or not the search engine helps the user save time), it was shown that many users in the first phase browsed the categories instead of using the search interface on the website, which could be a potential explanation to the low score (i.e. browsing being more time consuming than searching). In terms of overall success, the lack of price comparisons and poor performance on specific searches (section 5.3.8.3) likely contributed to the low overall success rating.

### 7.1.5 Personalization

In terms of personalization, the findings showed that users believed that Google and Prisjakt[8] personalized their services significantly more than Elgiganten, but with no significant difference between Google and Prisjakt. The topic of personalization is different compared to the previous measures, as a higher degree of

[8] Significant only at the 10 % level, not at the 5 % level.

personalization - ceteris paribus - doesn't necessarily mean that the service is better. Support for this view is found in the theory (section 2.6.4) and in the findings in section 5.3.4, that users have concerns of privacy.

### 7.1.6 Preferred e-commerce search engine

The question of preferred search engine for e-commerce showed a clear preference for Google and Prisjakt, where each option got 50 % of the preference. Elgiganten wasn't preferred by any user. Analysing the arguments for using Prisjakt and Google compared to Elgiganten, it is clear that the coverage and comparison of products from many competing stores is missing in an internal search engine like Elgiganten, and that these are two major reasons for preferring a search engine.

### 7.1.7 Implications of the comparison of search engine performance

Summarizing the discussion from section 7.1.1 to section 7.1.6, the answers provide a coherent picture of Google and Prisjakt as the search engines of higher performance in the study. The further discussion will therefore focus on how to design an e-commerce search engine for increased performance with Google and Prisjakt as a departure point.

## 7.2. Design objectives from the user search behaviours and the identified problems

The sub-sections of this section will discuss the empirical findings of the first two subordinate research questions, in relation to existing theory, to arrive at suggestions for design objectives that can answer the overall research question.

### 7.2.1 Comparisons of advantages and problems of Google and Prisjakt

To design a novel search engine for e-commerce, the researcher will in this section compare the advantages, disadvantages and problems between Google and Prisjakt. The results can help the design objectives to build on what the users perceive as valuable in an e-commerce search engine, at the same time pointing out the problems.

### 7.2.2 Coverage and information accuracy

Comparing the findings of Google (section 5.3.8.1) and Prisjakt (section 5.3.8.2), a difference between the search engines exsits in the users' opinions regarding coverage. Coverage is found to be an unambiguous advantage of Google, but for Prisjakt the coverage is seen both as an advantage and as a problem. The users agree that Google includes everything in its search engine, while Prisjakt only includes products. Some users

suggest Prisjakt has great coverage in terms of e-commerce products, which likely relates to Prisjakt's specific focus on e-commerce. However, Prisjakt doesn't include all kind of products (e.g. books are absent, section 4.2.3.1 and 4.3.4) in its search engine, and it doesn't communicate clear enough which products users should be able to find there.

There are similar problems in the coverage of stores in the price listings of Prisjakt (section 4.3.4.2). Users complain about prices not being accurate (section 4.3.4.1), and that they have to check with multiple price comparison sites (Prisjakt and Pricerunner in section 4.3.4.2) as they doubt the completeness of prices (stores) in the price comparison.

### 7.2.3 Specific search abilities

#### 7.2.3.1 Specific search with filters, categories and sorting

Specific search was found to be a major problem with Google as well as with Elgiganten, but an advantage of Prisjakt. As users connected the specific search to the filter capabilities of the search engines (See section 5.3.7 and 5.3.8), filter and sorting is likely a part of the answer to improving specific search. An example is the following user's opinion on Google: *"Difficult to immerse and find just the kind of product you want.", "/…/difficult to specify when searching for a very specific product"*. The response can further be connected to Google Shopping. Though Google Shopping is commonly shown directly in the SERP of Google, it does often include only a single product (see green box in Figure 10), and when including multiple products, no filter or sorting of the products is possible (e.g. Figure 1).

It should be noted that if users enter from the SERP of Google (e.g. Figure 1) to the specific SERP of Google Shopping, there are some filter and sorting alternatives available. However, none of the users did so during phase 1, so it didn't help. For the development of an e-commerce search engine, an important design objective is consequently to allow users to conduct very specific search, using filters, sorting and categories already in the initial SERP. A potential approach could be to include the filtering options already when the user types the query.

#### 7.2.3.2 Specific search with specific queries and ambiguous queries

The second approach to specific search is where the user wants to find a specific result, given only the query, without filters, sorting or categories. This requires more functionality in the search engine, as users often expect search engines to solve queries in a more complex way than just matching the exact product name. The second phase findings (section 5.4.1) indicated poor performance for specific queries in Elgiganten and Prisjakt.

As the findings indicate that users like to use natural language when formulating queries, it would be desirable for an e-commerce search engine to better process natural language queries. As pointed out by the literature (section 2.4.1), such queries are often difficult for search engines to understand and as such referred to as ambiguous queries. For allowing the search engine to understand words that users commonly use, tags and modifiers (section 2.4.1.2) can be used to map certain words to product attributes. This would help users, as they to a larger degree could use their own words (i.e. "lightweight" and "budget"), instead of only filters or sorting to find products.

#### 7.2.3.3 Spelling correction

The absence of correction of spelling errors was found as a disadvantage of Prisjakt and Elgiganten, but not for Google (section 5.3.8). Phase 1 instead indicated that Google provides spelling correction or suggestions if it suspects that the user has a typo in the query. As pointed out in section 2.4.1.4, the problem of spelling errors can be as common as in 10 % of all searches. Having the user correcting 10 % of all searches implies extra search time for the user, which is opposite to the goal of providing the users an efficient e-commerce search experience. Given this, it's remarkable that Prisjakt and Elgiganten don't implement better spelling corrections in their search engines (section 5.3.8.2). Design objectives for a future search engine should therefore include spelling error countermeasures, such as spelling corrections.

### 7.2.4 The problem of the relevance and ads trade-off

The approach to advertising differs between the search engines. While the internal search engine Elgiganten doesn't include ads, both Prisjakt and Google do so. There is however a clear difference between the two external search engines; while no users provide criticism over the Prisjakt ads (advertisements), many users have strong negative opinions about the ads in Google, in the first phase (section 4.3.3) as well as in the second phase of the study (section 5.3.8.1). Ads is the number one disadvantage of Google (section 5.3.8.1). The user response in section 4.3.3 illustrates this well, as the user suggests that the reason why e-commerce websites advertise on Google is the lack of quality of their websites (the user uses the term "pages" relating to web pages, i.e. the pieces of a website). Hence, advertisement in Google can be perceived as a sign of poor store quality by some users.

**Why are there problems with Google's advertisements but not Prisjakt's advertisements?**

There are similarities between Google and Prisjakt. It is free for stores to be listed in the search results of both Google and Prisjakt, and stores have the possibility to buy sponsored links with advantages in both search engines (Google.com, n.d.-c; Prisjakt.nu, n.d.).

However, the differences refer to the advertisements (1) placement and (2) content. In Google, the sponsored links provide a more attractive placement in the SERP than the organic links, as pointed out in section 2.2.2.3 and as seen in the green and red areas of Figure 1. This is different from Prisjakt where there are no organic links, just organic price listings. Prices from all participating websites (stores) are listed in the same price comparison list, no matter if the store advertise (pay) or not (screenshot available in appendix 4). The advantages that the advertising store gets at Prisjakt are (1) logo and links (i.e. non-sponsored websites only get their name and price, no link in the results listing), (2) small promotional text in the results listing (e.g. stores commonly listing their top three selling points) and (3) statistics. Paying for sponsored links does not give a more attractive placement in the results list on Prisjakt. The implications of the differences are referred to as the ranking problem and the content problem, which will be explained in the following two sections.

#### 7.2.4.1 The ranking problem

Compared to Google, Prisjakt allows the users to sort the results on whatever ranking they prefer, like store's price, reviews or shipping speed, without considering if the store pays for advertisement or not. Google on the other hand organizes their advertisements through their Google AdWords program (section 2.2.2.3), in which the ads position (rank) is determined by an auction which takes a combination of the stores bid, the ad's relevance for the search and user, and some additional variables (Google.com, n.d.-b). The implication is that ads can claim a top result as long as they pay enough.

Digging in to the microeconomic motives of Google's approach to ads are outside the scope of this thesis, though the maximization of income per search can be a potential argument for a bid-based approach. However, from a user perspective, the highest paying ad is not necessarily the most relevant ad. So even though Google suggests that it takes more variables (e.g. the ads' relevance) than just the bid into account when ranking the ads, the findings (i.e. user opinions in section 4.3.3 and section 5.3.8.1) suggest that this is far from sufficient. Given two competing stores selling an identical product, the company placing a higher ad bid is - ceteris paribus - more likely to charge a higher price on the customer, to cover for the cost of the ad. As users are price sensitive (section 4.2.6.1), a higher price for the same product is - ceteris paribus - considered less relevant by the customers.

It can also be the case that the advertiser unawarely has selected keywords that causes the mismatch. This as the advertisers themselves can decide what search query keywords they like their ads to be triggered for (Google.com, n.d.-a). In the end, it is therefore the search engine's decision how much influence the advertiser should be given over what queries the advertisements should be visible for. The problem of advertiser's influence is further discussed in the next section.

#### 7.2.4.2 The content and advertisers' influence problem

The content problem relates to the problem of poor ad messages pointed out in section 4.3.3.2. As suggested in the findings, when users doubt if the ad links to a specific product page, or to the website in general, they will likely consider it irrelevant and avoid it.

This problem also relates to the users' preference on previous experiences with stores, where users prefer buying from stores that they've already positive experiences from (section 4.2.5). When the user has such experiences, it becomes even more relevant with product information rather than common store selling points in the ad, as the user at that point probably already is aware of shipping speed, year in business (i.e. reliability), etc (section 4.3.3.2). Personalization (further addressed in section 7.2.5) can then be employed to identify and adjust the ad content for previous customers.

#### 7.2.4.3 Further support

Further support for these theories are the findings that Google provides highly relevant results in the organic results. One example is the problem of the search query *"pricerunner Toner i natten"* for the book in section 4.3.3.1, which includes the ambiguous word "toner". Even though Google understands that the user is looking for a book (i.e. return relevant organic results), it will still show less relevant ads of toner cartridges. (Reminder: *"toner"* in Swedish can translate to both tones and toner cartridges for printers).

Though the exact reasons for Google's choice of ranking of ads is unknown, a potential explanation could be that InkClub (the second advertiser on SERP) and Pricerunner have placed very high bids on the search terms "Toner" and "Pricerunner", and that the relevance part of Google's auction-algorithm isn't enough to compensate for this. This observation provides support for the (1) placement problem of ads previously addressed.

As it's also been pointed out by some users (section 5.3.8.1) that the separation between sponsored and organic search on Google is poor, it is likely to further extend the relevance issues, as users face problems determining what is ads and what is not.

Given the different relevance problems of Google ads, a trade-off between advertiser influence (both through bids and through content) and relevant ads has been identified, where a high degree of advertiser influence is likely to lead to a lower degree of relevance of the ads. Though it is the advertiser that writes the ad content, by choosing what keywords to bid on, poor advertiser choices also affect the search engine in the form of low relevance.

#### 7.2.4.4 Implications for Advertisers

Regarding implications for search engine developers controlling the content of ads (also for advertisers), one important lesson is that considering a more product-oriented content of the ad could likely deter less users from clicking on it, as pointed out in *Ad content* in section 4.3.3.2. As the generic content of *"Toner i natten – Jojo Moyes – Book | bokus.com, buy your books at Bokus. New offers every week! Free shipping over 99kr – More than 8 miljon titels – Safe e-commerce/…/"* of the Bokus.com ad in Figure 10, confuses the user about the actual destination of its link.

### 7.2.5 Personalization and ambiguous queries

The findings indicate that many users see a value in personalization (section 5.3.4.1). As pointed out in the theoretical underpinnings (section 2.4.1.1 and section 2.6.4), personalization can also be used for search engines and e-commerce. The following two sections discuss different problems of the findings, and provide design objectives as solutions for these.

#### 7.2.5.1 Personalization with opt-out

From the findings in phase 2 (section 5.3.4.1), certain objections to personalization were also made. It was mainly two problems (1) privacy concerns and (2) "living in a bubble". From the findings, the privacy concerns mainly relate to the users being *"supervised", "monitored"* etc, which feels scary. As suggested by the literature (section 2.6.4), the privacy fears can be overcome if enough benefits are added to the service. The second problem of "living in a bubble", is described by the users as the risk that they will miss out on discovering new things, as content could be outside their regular search patterns, and as such ignored by the search engine.

Though Awad & Krishnan (2006) recommend managers to ignore the privacy sensitive minority, the findings (section 5.3.4.1) shows a substantial part (five of the users) care about these issues, and therefore information transparency should not be completely ignored. The recommendation to the identified privacy problems is therefore that the personalization features of the search engine should be easy to turn on and off. The privacy concerned users can then choose to use the search engine without personalization, at the same time as the ability to easily turn personalization on and off also can help in showing the benefit of personalization, as it is more transparent for the user how the search results differ with or with or personalization. As such, even privacy concerned users might see the value of personalization and reconsider their choices.

#### 7.2.5.2 Use long-term interest (LTI) and short-term interest (STI)

The findings also show that current search engines perform poorly in using user's interest when making results. One example is the ambiguous query problem of Lloyd's Bank in section 4.3.1, where Google fails to recognize the information need of the user. The ambiguity of the query is that "lloyds" can relate to different things and that it is hard for the search engine to interpret (just as in the example of "Java" in section 2.4.1.1). As seen in the example, Google presents the user with a SERP full of Lloyd's Bank, even though the query that the user previously entered in Google was *"nasal spray online"*, which likely would indicate that the information need is to find Lloyd's Pharmacy rather than Lloyd's Bank. As pointed out in the findings (section 4.3.1), the user is logged with her Google account, which could potentially help Google recognizing (track) the short-term interest of the user. Still, the displayed Lloyd's Bank results are irrelevant to the user. The currently discussed findings indicate that Google fails to use the short-term interest of the user, as described in section 2.4.1.1. The findings also show that Google fails to suggest diverse search suggestions (section 2.4.1.3), as all the results at the SERP refers to Lloyd's Bank (Figure 8).

The finding in section 4.2.5 indicates that many users prefer buying from e-commerce stores that they have previously experienced positively. An understanding of users' long-term interests of e-commerce stores, could therefore help e-commerce search engines provide more relevant results. Design objectives should therefore include personalization with long-term interest (LTI) and short-term interest (STI) as well as diverse search suggestions (section 2.4.1.3) to perform better with ambiguous queries.

### 7.2.6 External to internal search

The findings (section 4.2.3) indicate that the concept of external to internal search is present in e-commerce search, just as web search in general search. The findings also suggest that the concept of external to internal search could be wider than previously thought. This as search transitions do not only occur from external search engines to internal search engines (external to internal search), but also from internal search engines to external search engines (internal to external search). As pointed out in section 4.3.4, a motive for this could be that the user believes that the external search engine will perform better in searching at the website of the internal search engine, than the actual internal search engine. This is a clear failure for the internal search engine.

#### 7.2.6.1 Browsing

The findings (section 4.2.3) suggest that the majority of the transitions from external to internal search is performed through the use of browsing (navigation), instead of queries on the internal website. If the definition of search is adapted to also include browsing, this would imply that the concept of external to

internal search is approximately much more common than previously known, at least in e-commerce search. A possible reason for its omission is that the methods (query log analysis) used by previous research (Ortiz-Cordova & Jansen, 2014; Ortiz-Cordova et al., 2015) does not capture navigation, as it only takes the search queries into consideration; and search queries do not contain the navigational steps taken by users when browsing the websites.

For practitioners in e-commerce and search, this implies that the theory of external to internal search is potentially much more important than previous studies have shown, which has a variety of implications. For example, the design and content of landing pages (i.e. the pages that users enter an (internal) website to from an external search engine), search keywords and targeting advertising can be better tailored around the users' needs. Internal search engines can use the referral search query from the external search engine of the incoming user to personalize the landing page, but also for delivering better results if the user continue to search with the internal search engine.

In terms of external to internal search, the design objectives of search engines should include tracking of user's information needs, as the users transition between search engines. This could be achieved using referral queries, as suggested by Ortiz-Cordova et al. (2015). In regards to browsing, this study hasn't identified any opportunities for tracking such (except for the browser add-on in section 7.2.7). This is therefore a recommendation for future research (section 9.1.1).

### 7.2.7 Browser add-on

The finding that one user has a dedicated browser add-on that generate discount (section 4.2.4) is interesting also for e-commerce search. One alternative solution to developing a web-based e-commerce search engine from scratch, would be to develop an e-commerce search browser addon. There are many advantages of such a solution, which will be discussed in the following sub-sections.

#### 7.2.7.1 Modifying existing search: prices, reviews, availability, filters

This shows that users prefer Google and Prisjakt when searching in e-commerce (section 5.3.7) but for different reasons. The advantages of Google are its coverage, relevance and overall high performance on general queries, while for Prisjakt the product info, coverage, filters and sorting (section 5.3.8.1) are the advantages. As seen in yellow in Figure 10, a browser add-on can modify the SERP with complementing information. This implies that the information that users suggest that Google is missing (e.g. price comparisons and specific search by filters, section 5.3.8.1) could be added to the SERP of Google by a browser add-on. That would likely increase efficiency of e-commerce search, as users would be able to

directly compare prices and to see reviews, availability and other product information directly in the SERP, thereby decreasing the number of different websites and queries the user needs to find the desired information. Allowing the browser add-on to filter the results would further help the users, as the results could be limited to products only. As the browser add-on is separated from a specific website, it can also provide similar information in many internal search engines, or at any website (such as stores) where it identifies e-commerce products.

#### 7.2.7.2 Tracking between external and internal search at different websites

Another advantage of using a browser add-on for e-commerce search refers to external to internal search (section 7.2.6). As it has been identified in both literature (section 2.4.2) and by the findings (section 4.2.3), that transitions between different search engines are common, the advantage of using a browser add-on in e-commerce search is that such an add-on can potentially track users queries and information needs between websites better than a single web-based search engine can. Keeping track of queries and browsing behaviours from multiple search engines, would allow the add-on to better understand the user's information need, and thereby provide more accurate product recommendations.

# 8. DESIGN OBJECTIVES FOR E-COMMERCE SEARCH ENIGES

The following sections present the managerial recommendations produced by this thesis, answering the overall research question of *"How should search engines be designed to perform better when solving e-commerce users' information needs?"*. As pointed out in section 2.1, the design objectives will describe solutions (qualitatively or quantitatively) for performance problems in e-commerce search, by using insights from user evaluations combined with existing theory from search and e-commerce fields to motivate the choices. Each design objective is therefore accompanied with a summary of the corresponding problem identified in the application domain. The recommendations are made to be used as design objectives for the development of a novel e-commerce search engine.

## 8.1. Coverage and information accuracy

**Problem 1:** Users search for products not covered (i.e. indexed) by the e-commerce search engines. However, when they do search for products covered, the information about the products are inaccurate or incomplete.

**Design objective 1:** Decide what products to cover in the search engine, and make sure to clearly communicate that to the users. This could be done by providing coverage information next to the search input. If users still search for a product type not indexed by the e-commerce search engine (e.g. books in Prisjakt), the search engine should return an error message in the search results, clearly indicating that such products are not available in the search engine. Also make sure the information about each of the products is accurate and complete. Failing to do this will likely cause irritation from the users, as they cannot find what they expect.

## 8.2. Specific search

**Problem 2:** Users experience problems when trying to find specific products (e.g. products with certain attributes) in e-commerce, as search engines are not implementing categories, filters or sorting.

**Design objective 2:** Implement categories, filters or sorting already in the SERP of e-commerce search engines. This allow the users to search with browsing when they prefer so, or when their queries are not returning results specific enough, increasing the performance of the search engine.

**Problem 3:** Users experience problems in expressing their information need when trying to find specific products (e.g. products with certain attributes), as search engines cannot interpret natural language properly, users get irrelevant or no results. The problems also include misspelled words.
**Design Objective 3:** Implement tags and modifiers in e-commerce search engines, to tag common e-commerce related query key words (i.e. "lightweight" and "budget") to common product attributes (i.e. 25% of computer with lowest weight and price lower than 5000 SEK). Also, implement spelling correction to help the users correct misspelled words. Together this will increase the search performance for users in e-commerce, as they can use their own language in queries and they are not limited to use filters and categories when specifying attributes for products.

## 8.3. Advertiser influence and relevance trade-off in ads

**Problem 4:** In some search engines, advertisers influence the advertisements through both content and bidding, which can affect the relevance of the sponsored search negatively. Simultaneously, users in e-commerce often prefer e-commerce specific rankings such as price, review ratings etc.
**Design objective 4:** When designing search engines for e-commerce, consider limiting the advertiser's influence and prioritize advertising that doesn't affect or disturb the organic listings. Failing to do so will likely annoy users.

**Problem 5:** In some search engines, advertising messages sometimes confuse users, as the ads destination (i.e. e-commerce store's product page or general home page) isn't clear.
**Design objective 5:** When designing search engines for e-commerce, make sure that the ads clearly shows the destination to which they lead. This could be done by including the URL, in the ad or by using a complete breadcrumb trail (e.g. store name > category name > product name) to make it clearer for the user that they will arrive at the product page when clicking on the ad. This would make the ads more relevant for users, as one element of ambiguity is removed. The design objective could also be used by copy writers, writing ads that signalises a clear link destination.

## 8.4. Transparent personalization

**Problem 6:** Some users see privacy issues and risks of being isolated from new things when they are subject to personalization.

**Design objective 6:** Implement personalization that users themselves can activate or deactivate, to make the personalization benefit more transparent. This option could encourage even users objecting to personalization, to use the e-commerce search engines, as the benefits are better understood.

**Problem 7:** Current search engines fail to recognize user interests, and thereby providing a less personalized experience and a more time-consuming search.

**Design objective 7:** By using long-term interest of the users, search engine could better track the users' preferences on e-commerce stores and thereby increase their rankings in search results. By using short-term interest of the user, search engines can more accurately process ambiguous queries, as the search engines will have more details on the user's information need. Together this would likely increase the efficiency of e-commerce search, as users would have to look at fewer irrelevant results (i.e. stores that they do not intend buying from), and use fewer queries to find desired answers to their information needs.

## 8.5. External to internal search

**Problem 8:** Internal and external engines fail to recognize user' information need when users arrive to the search engine from another external search engine (i.e. the tracking of users' information needs during search engine transitions works poorly).

**Design objective 8:** Implement the use of referral queries, to track users arriving from other search engines. The referral queries can sometime be extracted from the referral URL, thereby allowing the search engine to understand what the user has been previously searching for. This is likely to help increase the performance of the search engines, as the they can adjust the results to fit the user without having the user repeating the query.

## 8.6. E-commerce search browser add-on

**Problem 9:** Search engines are missing e-commerce relevant information and do not allow the users to filter out products.

**Design objective 9:** Design an e-commerce search web browser add-on that complements the original search engine information with comprehensive e-commerce related information, such as prices, reviews, availability, etc. Also allow the add-on to filter and sort results from major web-based search engines. By providing users with accurate product information through a browser add-on, their search can be made

more efficient, as they don't have to spend time visiting several different websites to collect the relevant information.

**Problem 10:** In their search process, users visit many different search engines and websites, to repeatedly express their information need, either through search queries or through browsing.
**Design objective 10:** Design an e-commerce search web browser add-on that tracks and supports the user across different websites, providing the user with relevant information for the information need (purchase) the user is currently interested in. By collecting information on the user across websites, the e-commerce search browser add-on could propose more accurately relevant products than individual search engines at each website.

# 9. CONCLUSION

The purpose of this thesis has been to develop design objectives for a novel e-commerce search engine, by user evaluations of current solutions in the e-commerce search area. Doing so, the first phase of the thesis started with exploring and identifying the current solutions, their problems and user behaviours. An evaluation framework was then developed by combining theory of e-commerce and information retrieval with the findings of the first phase. For the second phase, the evaluation framework was then used to evaluate the performance of three popular e-commerce search engines. In addition to evaluating the performance, qualitative user feedback was also collected to triangulate the first phase findings. The evaluation found significant differences in performance between the search engines in five out of seven measures, where the search engines Google and Prisjakt performed better than Elgiganten in most measures. Together with the identified user behaviours and problems, these findings answered the three subordinate research questions of the thesis, and met the research objectives.

By discussing the findings of performance, user behaviours and identified problems in relation to previous literature, a set of managerial recommendations in the form of ten design objectives were produced, answering the overall research question. The design objectives can all be implemented to produce action and change in the e-commerce search area, which has been a cornerstone for the underlying pragmatic assumptions of this thesis.

## 9.1. Recommendations for further research

### 9.1.1 External to internal search with browsing

As this research contributes with a wider perspective on external to internal search, where a majority of the searches in internal search engines are done through browsing instead of search queries, this opens an opportunity for further research on the topic, with new methods that could also capture browsing (i.e. clickstream data combined with query logs instead of exclusively query logs).

As a cornerstone of the research in external to internal search (or any research of transitions between search engines), is the possibility of following the information need of the user between search engines. Further

research needs to explore methods for understanding how historical browsing behaviours could be used to complement search queries in search engines. An approach could be to use a browser add-on as discussed in section 7.2.7, as this could help tracking information needs between websites. The research could potentially provide implications in a variety of fields, such as advertising (e.g. keyword section based on browser behaviour), ranking (i.e. browsing behaviour influencing search engine result rankings) content creation and indexing (browsing behaviour influencing on which content to produce, or index for search engines/websites).

### 9.1.2 Relevance and search engine revenue models in e-commerce

Future research should address the relevance problem of Google AdWords for e-commerce. An approach could be evaluating the precision of the organic listings and sponsored listings separately, to confirm differences and to better define areas of improvement. Making the ads more relevant would likely enhance users' attitude towards them, and the users' usage of them. As e-commerce search differs from general web search, by the available sorting of results (price, review ratings, shipping duration etc.), the payment models for advertisement needs to be further researched and understood. Even minor relevance improvements would likely have major financial (practical) implications for the search engines and their advertisers, as users could potentially use the ads more frequently. Further research should also address the maximization of the user perceived relevance of ads as well as the revenue of the ads, as those through an auction model could be related.

### 9.1.3 Ad content

Further research should investigate how changing the content of specific product ads, to contain product data (like organic results) instead of generic store selling points, could affect the click rate of the ad, as discussed in section 7.2.4.4. This could be particularly useful for returning customers, already aware of the selling points of the particular e-commerce store and consequently, more interested in the product information. Furthermore, the approach could potentially be combined with personalization, so that the ad adjusts its content given the viewing user's history with the advertiser.

# BIBLIOGRAPHY


Agrawal, R., Gollapudi, S., Halverson, A., & Ieong, S. (2009). Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09* (p. 5). https://doi.org/10.1145/1498759.1498766

Alexa.com. (2017). Top Sites in Sweden. Retrieved October 2, 2017, from https://www.alexa.com/topsites/countries/SE

Ali, R., & Beg, M. M. S. (2011). An overview of Web search evaluation methods. *Computers and Electrical Engineering*. https://doi.org/10.1016/j.compeleceng.2011.10.005

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, *22*(6–7), 493–520.

Aronson, J. (1995). A Pragmatic View of Thematic Analysis. *J. The Qualitative Report*, *2*(21), 1–3. https://doi.org/10.4135/9781446214565.n17

Awad, N. F., & Krishnan, M. S. (2006). The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly*, *30*(1), 13–28. https://doi.org/10.2307/25148715

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd ed.). Essex CM20 2JE, England: Pearson Education Limited.

Bilal, D. (2012). Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Yahoo!, Bing, Yahoo! Kids, and ask Kids. *Journal of the American Society for Information Science and Technology*, *63*(9), 1879–1896. https://doi.org/10.1002/asi.22675

Blake, T., Nosko, C., & Tadelis, S. (2016). *Returns to Consumer Search: Evidence from eBay*. *NBER Working Paper Series* (Vol. 22302). https://doi.org/10.1017/CBO9781107415324.004

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*. https://doi.org/10.1002/asi.10286

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, *36*(2), 3. https://doi.org/10.1145/792550.792552

Broder, A., Ciccolo, P., Gabrilovich, E., Josifovski, V., Metzler, D., Riedel, L., & Yuan, J. (2009). Online expansion of rare queries for sponsored search. In *Proceedings of the 18th international conference on World wide web - WWW '09* (p. 511). https://doi.org/10.1145/1526709.1526778

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97–113. https://doi.org/10.1177/1468794106058877

Campbell, R., & Ash, J. (2006). An evaluation of five bedside information products using a user-centered, task-oriented approach. *Journal of the Medical Library Association : JMLA*, *94*(4), 435–41, e206-7. https://doi.org/10.1089/tmj.2011.0119

Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: an empirical examination. *Industrial Management & Data Systems*, *105*(5), 645–661. https://doi.org/10.1108/02635570510600000

Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management*, *6*(2–3), 181–202.

https://doi.org/10.1007/s10799-005-5879-y
Chiou, W.-C., Lin, C.-C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management*, *47*(5–6), 282–290. https://doi.org/10.1016/j.im.2010.06.002
Chu, H., & Rosenthal, M. (2003). Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology. *American Society for Information Science (ASIS) 1996*, 1–15.
Cronen-Townsend, S., & Croft, W. B. (2002). Quantifying query ambiguity. In *Proceedings of the second international conference on Human Language Technology Research* - (p. 104). https://doi.org/10.3115/1289189.1289266
Cucerzan, S., & Brill, E. (2004). Spelling correction as an iterative process that exploits the collective knowledge of web users. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 293–300). Retrieved from https://www.microsoft.com/en-us/research/uploads/prod/2004/07/Cucerzan.pdf
De los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior. *The American Economic Review*, *102*(6), 2955–2980. https://doi.org/10.1257/aer.102.6.2955
DeLone, W. H., & McLean, E. R. (1992). Information System Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60–65. https://doi.org/10.1287/isre.3.1.60
DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success : A Ten-Year Update. *Journal of Management Information Systems*, *19*(4), 9–30. https://doi.org/10.1073/pnas.0914199107
DeLone, W. H., & McLean, E. R. (2004). Measuring e-Commerce Success: Applying the DeLone & McLean Information Systems Success Model. *International Journal of Electronic Commerce*, *9*(1), 31–47. https://doi.org/1086-4415/2004
Dong, X., & Su, L. T. (1997). Search Engines On The World Wide Web And Information Retrieval From The Internet: A Review And Evaluation. *Online and CD-Rom Review*, *21*(2), 67–82. https://doi.org/10.1108/eb024608
Dou, Z., Song, R., Wen, J.-R., & Yuan, X. (2009). Evaluating the Effectiveness of Personalized Web Search. *IEEE Transactions on Knowledge and Data Engineering*, *21*(8), 1178–1190. https://doi.org/10.1109/TKDE.2008.172
Duan, H., & Hsu, B.-J. (Paul). (2011). Online spelling correction for query completion. In *Proceedings of the 20th international conference on World wide web - WWW '11* (p. 117). https://doi.org/10.1145/1963405.1963425
Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. (2016). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *ACM SIGIR Forum*, *49*(2), 28–35. https://doi.org/10.1145/2888422.2888425
eMarketer.com. (2017). Worldwide Retail and Ecommerce Sales: eMarketer's Estimates for 2016–2021. Retrieved November 15, 2017, from https://www.emarketer.com/Report/Worldwide-Retail-Ecommerce-Sales-eMarketers-Estimates-20162021/2002090
European Commission - Press release. (2017). Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service. Retrieved November 20, 2017, from http://europa.eu/rapid/press-release_IP-17-1784_en.htm
Eurostat. (2017). 68 % of internet users in the EU shopped online in 2017. Retrieved December 31, 2017, from http://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce_statistics_for_individuals
Eveland, W. P., & Dunwoody, S. (2000). Examining Information Processing on the World Wide

Web Using Think Aloud Protocols. *Media Psychology*, *2*(3), 219–244. https://doi.org/10.1207/S1532785XMEP0203_2
Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, *21*(2), 135–146. https://doi.org/10.1057/ejis.2011.54
Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead…long live the paradigm: the legacy of Burrell and Morgan. *Omega*, *28*(3), 249–268. https://doi.org/10.1016/S0305-0483(99)00042-0
Gollapudi, S., Ieong, S., & Kannan, A. (2012). Structured query reformulations in commerce search. In *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12* (p. 1890). New York, New York, USA: ACM Press. https://doi.org/10.1145/2396761.2398538
Google.com. (n.d.-a). About keyword matching options. Retrieved February 3, 2018, from https://support.google.com/adwords/answer/2497836?hl=en
Google.com. (n.d.-b). Ad Rank. Retrieved February 16, 2018, from https://support.google.com/adwords/answer/1752122
Google.com. (n.d.-c). Get your website on Google. Retrieved January 27, 2018, from https://support.google.com/webmasters/answer/34397?hl=en
Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: The retrieval effectiveness of search engines. *Information Processing and Management*, *35*(2), 141–180. https://doi.org/10.1016/S0306-4573(98)00041-7
Hawking, D., Craswell, N., Bailey, P., & Griffiths, K. (2001). Measuring Search Engine Quality. *Information Retrieval*, *4*(1), 33–59. https://doi.org/10.1023/A:1011468107287
He, H., Meng, W., Yu, C., & Wu, Z. (2003). Wise-integrator: an automatic integrator of web search interfaces for e-commerce. *Proceedings of the 29th VLDB Conference*, *29*, 357–368. https://doi.org/http://dx.doi.org/10.1016/B978-012722442-8/50039-2
Hearst, M. A. (2009). *Search User Interfaces* (1st ed.). Cambridge University Press.
Hevner, A. R., & Chatterjee, S. (2010). *Design Research in Information Systems Theory and Practice*. *Integrated Series in Information Systems Volume 22*. https://doi.org/10.1007/978-1-4419-5653-8
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625
Ho, C. I., Lin, M. H., & Chen, H. M. (2012). Web users' behavioural patterns of tourism information search: From online to offline. *Tourism Management*, *33*(6), 1468–1482. https://doi.org/10.1016/j.tourman.2012.01.016
Holzinger, a. (2005). Usability engineering methods for software developers. *Communications of the ACM*, *48*(1), 71–74. https://doi.org/10.1145/1039539.1039541
Huang, Z., & Benyoucef, M. (2013). From e-commerce to social commerce: A close look at design features. *Electronic Commerce Research and Applications*, *12*(4), 246–259. https://doi.org/10.1016/j.elerap.2012.12.003
HUI Research. (2017). E-barometern Årsrapport 2016. Retrieved August 12, 2017, from http://www.hui.se/statistik-rapporter/index-och-barometrar/e-barometern
IBM Corporation. (2016). IBM SPSS Statistics 24 Algorithms. IBM Corporation. Retrieved from ftp://ftp.software.ibm.com/software/analytics/spss/documentation/statistics/24.0/en/client/Manuals/IBM_SPSS_Statistics_Algorithms.pdf
Iivari, J. (2007). A Paradigmatic Analysis of Information Systems As a Design Science. *Scandinavian Journal of Information Systems*, *19*(2), 39–64.
Jaiswal, A. K., Niraj, R., & Venugopal, P. (2010). Context-general and Context-specific

Determinants of Online Satisfaction and Loyalty for Commerce and Content Sites. *Journal of Interactive Marketing*, *24*(3), 222–238. https://doi.org/10.1016/j.intmar.2010.04.003
Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management*, *44*(3), 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015
Jansen, B. J., & Molina, P. R. (2006). The effectiveness of Web search engines for retrieving relevant ecommerce links. *Information Processing and Management*, *42*(4), 1075–1098. https://doi.org/10.1016/j.ipm.2005.09.003
Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*. https://doi.org/10.1016/j.ipm.2004.10.007
Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, *36*(2), 207–227. https://doi.org/10.1016/S0306-4573(99)00056-4
Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.2307/3700093
Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, *3*(1—2), 1–224. https://doi.org/10.1561/1500000012
Koulayev, S. (2014). Search for differentiated products: Identification and estimation. *RAND Journal of Economics*, *45*(3), 553–575. https://doi.org/10.1111/1756-2171.12062
Krovetz, R., & Croft, W. B. (1992). Lexical ambiguity and information retrieval. *ACM Transactions on Information Systems*, *10*(2), 115–141. https://doi.org/10.1145/146802.146810
Lewandowski, D. (2015). Evaluating the retrieval effectiveness of web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*, *66*(9), 1763–1775. https://doi.org/10.1002/asi.23304
Li, X., Wang, Y.-Y., & Acero, A. (2009). Extracting structured information from user queries with semi-supervised conditional random fields. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '09*, 572. https://doi.org/10.1145/1571941.1572039
Liu, C., & Arnett, K. P. (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & Management*, *38*(1), 23–33. https://doi.org/10.1016/S0378-7206(00)00049-5
LloydsApotek.se. (n.d.). LloydsApotek. Retrieved November 1, 2017, from https://www.lloydsapotek.se/om-oss/om-oss
Lundgrén-Laine, H., & Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative Health Research*, *20*(4), 565–575. https://doi.org/10.1177/1049732309354278
Manning, C. D., & Raghavan, P. (2008). *An Introduction to Information Retrieval* (1st ed.). New York: Cambridge University Press.
Matthijs, N., & Radlinski, F. (2011). Personalizing web search using long term browsing history. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 25–34). https://doi.org/10.1145/1935826.1935840
Molla, A., & Licker, P. S. (2001). E-commerce systems success: An attempt to extend and respecify the Delone and Maclean model of IS success. *Journal of Electronic Commerce Research*, *2*(4), 131–141. https://doi.org/10.1.1.92.6900
Montgomery, D. C. (2012). *Design and Analysis of Experiments* (8th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

NetApplications.com. (2017). Search Engine Market Share. Retrieved November 12, 2017, from https://www.netmarketshare.com/search-engine-market-share.aspx
Nielsen, J. (1993). *Usability Engineering*. London: Academic Press, Inc.
Obendorf, H., Weinreich, H., & Herder, E. (2007). Web page revisitation revisited: implications of a long-term click-stream study of browser usage. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 597–606. https://doi.org/10.1145/1240624.1240719
Ortiz-Cordova, A., & Jansen, B. J. (2014). Linking External and Internal Search: Investigating the Site Searching Patterns of Referred Searchers. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (pp. 1345–1350). https://doi.org/10.1145/2559206.2581199
Ortiz-Cordova, A., Yang, Y., & Jansen, B. J. (2015). External to internal search: Associating searching on search engines with searching on sites. *Information Processing and Management*, *51*(5), 718–736. https://doi.org/10.1016/j.ipm.2015.06.009
Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302
Peng, Q., Meng, W., He, H., & Yu, C. (2004). Clustering E-Commerce Search Engines. In *WWW Alt. '04 Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 416–417). https://doi.org/http://doi.acm.org/10.1145/1013367.1013503
Prisjakt.nu. (n.d.). Vanliga frågor - FAQ. Retrieved January 27, 2018, from https://www.prisjakt.nu/info.php?t=about_faq
Prisjakt.nu. (2017). Jämför priser från 5 240 butiker. Retrieved October 7, 2017, from https://www.prisjakt.nu
Refunder.se. (n.d.). Om Refunder. Retrieved December 12, 2017, from https://www.refunder.se/foretagsinformation
Sarkas, N., Paparizos, S., & Tsaparas, P. (2010). Structured annotations of web queries. In *Proceedings of the 2010 international conference on Management of data - SIGMOD '10* (pp. 771–782). https://doi.org/10.1145/1807167.1807251
Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students* (7th ed.). Essex CM20 2JE England: Pearson Education Limited.
Scriven, M. (1981). *Evaluation Thesaurus (Third Edition)*. Pt. Reyes California 94956: Edgepiess.
Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, *33*(1), 6–12. https://doi.org/10.1145/331403.331405
Singh, A., & Sharma, A. (2016). An Approach for Personalized Web Information Retrieval using Modified PageRank Method. In *Proceedings of the International Conference on Advances in Information Communication Technology & Computing - AICTC '16* (pp. 1–5). New York, New York, USA: ACM Press. https://doi.org/10.1145/2979779.2979881
Song, R., Song, R., Luo, Z., Luo, Z., Wen, J.-R., Wen, J.-R., … Hon, H.-W. (2007). Identifying ambiguous queries in web search. *Proceedings of the 16th International Conference on World Wide Web - WWW '07*, 1169. https://doi.org/10.1145/1242572.1242749
Spink, A. (2002). A User-Centered Approach to Evaluating Human Interaction with Web Search Engines: An Exploratory Study. *Information Processing & Management*, *38*(3), 401–426. https://doi.org/10.1016/S0306-4573(01)00036-X
Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From E-sex to e-commerce: Web search changes. *Computer*, *35*(3), 107–109. https://doi.org/10.1109/2.989940
Spink, A., Ozmutlu, H. C., & Ozmutlu, S. (2002). Multitasking information seeking and searching processes. *Journal of the American Society for Information Science and Technology*, *53*(8), 639–652. https://doi.org/10.1002/asi.10124

Statcounter.com. (2017). Search Engine Market Share Sweden Jan - Dec 2016. Retrieved May 5, 2017, from http://gs.statcounter.com/search-engine-market-share/all/sweden/2016
StatCounter.com. (2017). Search Engine Market Share Worldwide Jan - Dec 2016. Retrieved October 15, 2017, from http://gs.statcounter.com/search-engine-market-share/all/worldwide/2016
Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing & Management*, *28*(4), 503–516. https://doi.org/10.1016/0306-4573(92)90007-M
Su, L. T. (2003a). A comprehensive and systematic model of user evaluation of web search engines: I. Theory and background. *Journal of the American Society for Information Science and Technology*, *54*(13), 1175–1192. https://doi.org/10.1002/asi.10303
Su, L. T. (2003b). A comprehensive and systematic model of user evaluation of web search engines: II. An evaluation by undergraduates. *Journal of the American Society for Information Science and Technology*, *54*(13), 1193–1223. https://doi.org/10.1002/asi.10334
Susser, B., & Ariga, T. (2006). Teaching e-commerce Web page evaluation and design: a pilot study using tourism destination sites. *Computers and Education*, *47*(4), 399–413. https://doi.org/10.1016/j.compedu.2004.11.006
Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed.). Thousand Oaks: SAGE Publications, Inc.
Teixeira Lopes, C., & Ribeiro, C. (2011). Comparative evaluation of web search engines in health information retrieval. *Online Information Review*, *35*(6), 869–892. https://doi.org/10.1108/14684521111193175
Theverge.com. (2017). Alibaba's Singles' Day sale amassed $25.3 billion, doubling 2016 Black Friday and Cyber Monday sales combined. Retrieved November 30, 2017, from https://www.theverge.com/2017/11/11/16637362/alibaba-singles-day-2017-sales-record-online-shopping-black-friday-cyber-monday
van der Merwe, R., & Bekker, J. (2003). A framework and methodology for evaluating e-commerce Web sites. *Internet Research*, *13*(5), 330–341. https://doi.org/10.1108/10662240310501612
van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. *Department of Social Science Informatics, University of Amsterdam*. https://doi.org/10.1016/0306-4573(95)90031-4
Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing and Management*, *40*(4), 677–691. https://doi.org/10.1016/S0306-4573(03)00043-8
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21–54. https://doi.org/10.25300/MISQ/2013/37.1.02
Wang, Y. D., & Forgionne, G. (2008). Testing a decision-theoretic approach to the evaluation of information retrieval systems. *Journal of Information Science*, *34*(6), 861–876. https://doi.org/10.1177/0165551508091308
Yu, C., He, H., Meng, W., & Peng, Q. (2004). WISE-cluster: clustering e-commerce search engines automatically. In *Proceedings of the 6th annual ACM international workshop on Web information and data management - WIDM '04* (p. 104). New York, New York, USA: ACM Press. https://doi.org/10.1145/1031453.1031473
Yu, J., Mohan, S., Putthividhya, D. (Pew), & Wong, W.-K. (2014). Latent dirichlet allocation based diversified retrieval for e-commerce search. In *Proceedings of the 7th ACM international conference on Web search and data mining - WSDM '14* (pp. 463–472). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556195.2556215
Zhou, K., Cummins, R., Lalmas, M., & Jose, J. M. (2013). Which Vertical Search Engines are

Relevant ? *Proceedings WWW 2013*, 1557–1567. https://doi.org/10.1145/2488388.2488524
Zviran, M., Glezer, C., & Avni, I. (2006). User satisfaction from commercial web sites: The effect of design and use. *Information and Management*, *43*(2), 157–178. https://doi.org/10.1016/j.im.2005.04.002

# APPENDICES

## Appendix 1 – Search tasks first phase

### Task 1 – Clothes/Shoes

You need a new pair of summer shoes. Can you find a pair that you would like to buy? Answer with a link to the product page. Remember to think aloud!

### Task 2 – Books/Media

You have been recommended a book by a friend that you would like to buy. The title is "Toner i natten" by the author Jojo Moyes. Where would you buy it? Answer with a link to the product page. Remember to think aloud!

### Task 3 – Beauty/Health care

You have borrowed a friend's apartment in Stockholm, but unfortunately cached such a bad cold that you can't go outside. You would like to get a nasal spray delivered to your apartment as soon as possible. The location is Kungsholmsgatan 6, 112 27 Stockholm. How would you do? Answer with a link to the product page of the nasal spray and the shipping method of your choice. Remember to think aloud!

### Task 4 – Consumer electronics

Your counter dishwasher has broken down and you need to buy a new one. In order for it to fit in your kitchen, it cannot be higher than 44 cm or wider than 55 cm. You would also like the dishwasher to have a fast run programme, which should take 40 minutes or less to run. Where would you buy it? Answer with a link to the product page. Remember to think aloud!

### Task 5 – Baby/Children

A relative of yours is turning two years old. Can you find a toy that you can give as birthday gift to him/her? Answer with a link to the product page. Remember to think aloud!

### Task 6 – Sports/Leisure

You have decided to start to play badminton. See if you can find a badminton racket that would suit you! Answer with a link to the product page. Remember to think aloud!

### Task 7 – Interior/Furniture

You have found a picture of a chair that you would really like to buy. Where can you find and order it? Answer with a link to the product page. Remember to think aloud!

### Task 8 – Groceries

You have decided that doing groceries takes way too much of your time, so instead you would like to order a subscription for a weekly delivery of groceries to your home. Your home address is: Erik Dahlbergsgatan 60, 411 31 Göteborg. From where will you order the groceries? Answer with a link to the product page. Remember to think aloud!

### Task 9 – Car/Boat/Motorbike

You need new wipers for the windshield of your car, a Volvo v70 from 2006. Can you find some that fits? Answer with a link to the product page. Remember to think aloud!

### Task 10 – Watches/Jewellery

You need a new watch; can you find one that matches your preferences? Answer with a link to the product page. Remember to think aloud!

# Appendix 2 – Search tasks second phase

### Task 1 – Computer

You need to buy a new computer. You want it to be a laptop that is easy to travel with (e.g. not too large or heavy). Use {SEARCH ENGINE}to find a computer that suits your needs!

### Task 2 – Vacuum cleaner

It is time for you to buy a new vacuum cleaner. You would like one of the brand Electrolux with energy class C or better. Your budget is max 2000 SEK. Use {SEARCH ENGINE}to find a vacuum cleaner that suits your needs!

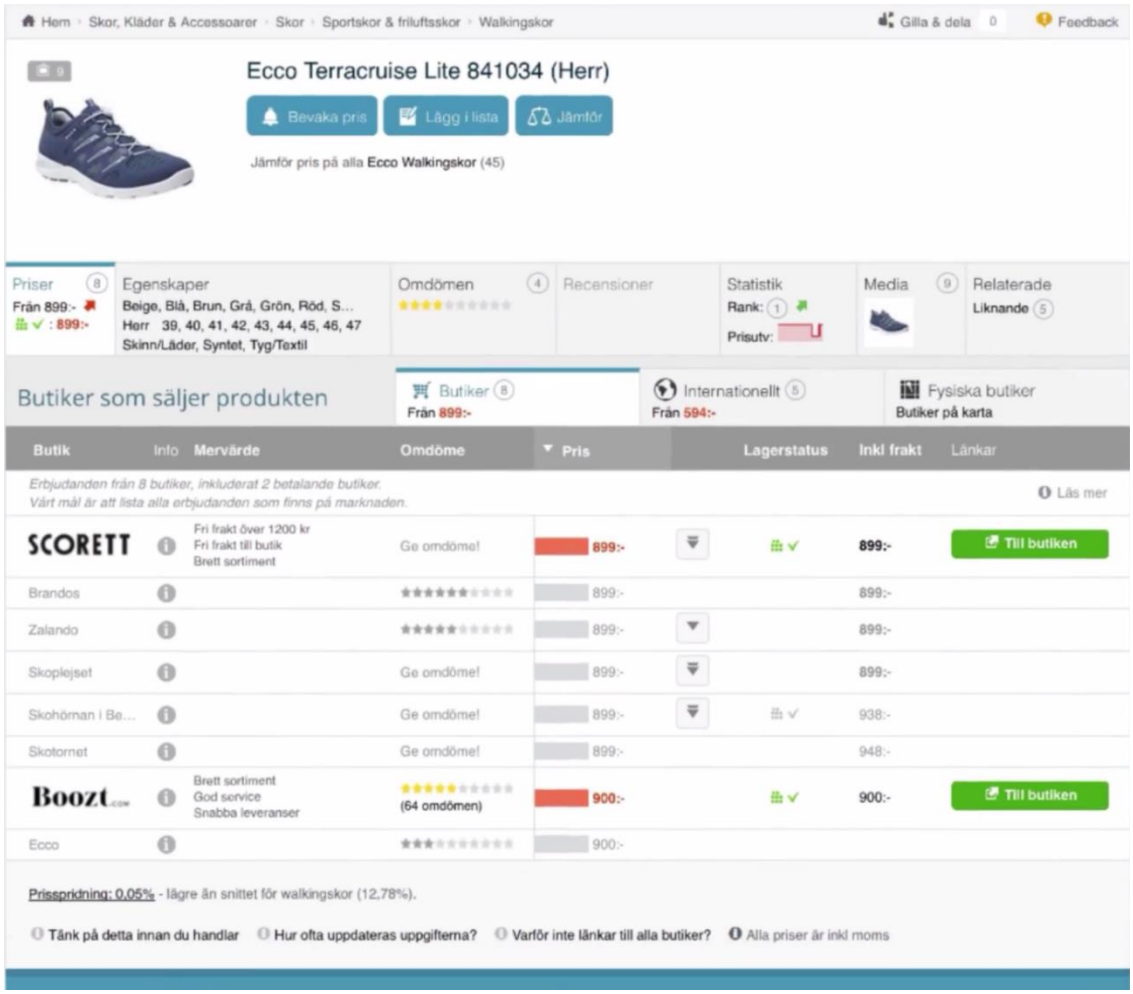
### Task 3 – TV

Your TV has broken down. Can you find a new one that seem good? It should be 40 inches and of the brand Samsung. You budget is max 6000 SEK. Use {SEARCH ENGINE}to find a TV that suits your needs!

# Appendix 3 – First phase number of queries per search engine

| **Domain** | **Number of queries** |
|---|---|
| Google.se | 137 |
| Prisjakt.nu | 13 |
| Adlibris.se | 3 |
| Elgiganten.se | 3 |
| Pricerunner.se | 2 |
| Ikea.com | 2 |
| Apotea.se | 2 |
| Biltema.se | 2 |
| Studentkortet.se | 2 |
| Apotekhjartat.se | 1 |
| Amazon.co.uk | 1 |
| Cdon.se | 1 |
| Lekmer.se | 1 |
| Furniturebox.se | 1 |
| Blocket.se | 1 |
| Svenssons.se | 1 |
| Rum21.se | 1 |
| Bokus.com | 1 |
| Xxl.se | 1 |
| stadium.se | 1 |
| Zalando.se | 1 |
| **Sum:** | **178** |

# Appendix 4 – Prisjakt price comparison listing



# Appendix 5 – Overview of performance measures for second phase

| | Criteria group | This criteria group evaluates/measures | Part | Adapted*/adopted from: |
|---|---|---|---|---|
| **Relevance measures** | | | | |
| | Precision ratio #1 or PR1 | Proportion of R and PR hits retrieved among first 10 hits from an engine | Search task | *(Su, 2003a) |
| | Precision ratio #2 or PR2 | Proportion of R hits among first 10 hits from an engine | Search task | *(Su, 2003a) |
| | | | | |
| **Efficiency measure** | | | | |
| | Search strategy | Number of search queries submitted by a user in searching for his or her problem on an engine | Search task | (Su, 2003a) |
| | | | | |
| **Utility measure** | | | | |
| | Value of search results as a whole | Usefulness of search results as a whole to the user for meeting the need or resolving the problem | Post-search task | (Su, 2003a) |
| | | | | |
| **User satisfaction measures** | | | | |
| | Time saving | Time spent by using an engine to find information, 1: no time saving at all, 7: saving a lot of time | Post-search task | (Su, 2003a) |
| | User's judgment of overall success | Overall success of an engine in providing help for the user's information need or problem | Post-search task | (Su, 2003a) |
| | | | | |
| **Personalization measure** | | | | |
| | Personalization | How well the search engines acquire users' personal preferences and personalize the services and products for them | Post-search task | (Chellappa & Sin, 2005) |
| | | | | |