Deciphering search ranking credibility and quality: an exploratory analysis

Author: Karlijn Pots
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT- Due to the use of search engines as primary source of information and increased registered data online, search engines such as Google, can collect, store, and exploit tremendous amounts of personal data. Search rankings, displayed by Google, are generated and personalized based on an online identity. However; personal data can be inaccurate or irrelevant due to the misrepresentation of an individual. Furthermore, the process of ranking based on registered data, also referred as *Back-end Googlization*, can alter and manipulate the individual's perception of what information is available. Both of these scenarios can create credibility discussions. Therefore, this study examines the impact of personalization of search rankings on the credibility perceived by the search engine users. An experiment is conducted to identify the differences between top *k* lists of two categories of search engines; [1] a search engine that tracks personal registered data (e.g., Google) and [2] a search engine that does not track personal registered data (e.g., DuckDuckGo). Subsequent, the perceived credibility is determined by a survey, which contains a set of credibility constructs. The experimental results show that there is little to no overlap between the search rankings of Google and DuckDuckGo. Thus, the top *k* lists of both search engines are significantly different. Furthermore, the negative correlation between the personalization of the search rankings and the perceived credibility score indicates that personalization has a negative impact on the perceived credibility of a search engine. This results in the fact that organic services, such as DuckDuckGo, are perceived as more fair and trustworthy to perform search queries due to its anonymous platform. The results of this study offer a basis for the synthesis of different insights on the discursive mechanisms of search engines and the (political) platforms developed by search engines.

Keywords- Online identity, self-presentation theory, credibility, Back-end Googlization

1. INTRODUCTION

Once, the Internet days were characterized by the cartoon illustrated by Peter Steiner (1993) named "on the Internet nobody knows you're dog." The cartoon symbolizes the understanding that protocols did not force Internet users to identify or register themselves, and it was probably safe to assume that individual's Internet behaviour did not reveal much about their real-world personas. However, anonymity has almost disappeared and new online identities are constructed. Online identity refers to the sum of registered data of an individual that is available online (Papacharissi, 2002). One reason for this phenomenon is due to ubiquitous presence of mobile devices. Social network profiles, online purchases, registered preferences, and even email registration reflect elements of an online identity. Every time an individual sends out a tweet, logs a status or posts a picture individuals are adding data to their online identity (Papacharissi, 2002). Google is an example of a corporation that uses registered data to create and refine an individual's profile; what they are doing and what they will do next. Google is a prediction and recommender search engine. Google gathers data to create a unique universe of information for each individual (Pariser, 2011). Hannak *et al.* (2013) and Pariser (2011) state that personalization of the search rankings may result in *Filter Bubble* effects, "where users are only given results that the personalization algorithm thinks they want, while other, potentially important, results remain hidden" (Hannak et al. 2013, p. 1). With the study of Powers (2011) the *Filter Bubble* effects can also be evaluated as a type II error or false negative; indicating that the relevant documents are not retrieved (p, 3).

Prior studies featuring eye-tracking technology show that individuals scan the search rankings in the order in which they appear, then fixate and click on the results that rank highest (Bilal & Kirby, 2002; Pan *et al.,* 2007; Rogers, 2013). Furthermore, large search engines retrieve over millions of search engine results, with little overlap on identical queries. (Bharat & Broder, 1998; Bar-Ilan, 2005; Lawrence & Giles, 1999). Therefore, search rankings become very important to retrieve a significant proportion of relevant documents (i.e., recall) for the search engine user (Powers, 2011).

An important study that features search rankings is the research of Epstein and Robertson (2015). The researchers discuss the impact of biased rankings on United States 2016 presidential election. They state that search engine users are influenced by biases in search rankings, even when the search engine users know that the search ranking results are biased (Epstein & Robertson, 2015 p. 4519). The researchers conjecture that the search engine is a powerful means that can be abused, where specific demographic group can be accessed. Which makes detection and regulation very difficult, and ultimately the search ranking bias can pose significant threats to the democratic system, when individuals' votes can be influenced due to the biased information displayed by Google. Research from Kay *et al.* (2015) state that "the information individuals access affects their understanding of the world around them and the decisions they make; biased information can affect both how individuals treat others and how they evaluate their own choices or opportunities" (p. 1). Thus, the algorithms used by Google can ultimately alter the way individual's encounter information and the perception of what information is available, by presenting inferior results to what they would otherwise find if all results were displayed organically (i.e., without personalization).

As mentioned, self-registered data can influence search rankings, this data can also be inaccurate or misrepresented, due to the fact that individuals manipulate, reinvent, or only reveal certain aspects of their identity in online communities (Marwick, 2013; Papacharissi, 2002). Hence, search rankings can be biased and dissatisfy users by offering inaccurate and irrelevant matches.

So, based on the plethora of online information, the different algorithms used to display personal rankings, and coupled with the heavy reliance of Internet users on the search engines, it can be stated that the overall credibility of search engine rankings can be called into question. The scenarios mentioned are some of the facts that shape the concept of the "ranking bias", also referred as *Back-end Googlization* (Rogers, 2013). This process refers to the algorithm that recommends sources hierarchically, based on signals; everything from what you shared, from what you were browsing, and to what you have searched before (Nguyen, Hui, Harper, Terveen, & Konstan, 2014; Pariser, 2011).

Previous work

Multiple researches have examined the differences between search rankings from different search engines (Bar-Ilan, Keenoy, Yaari, & Levene, 2007; Vaughan, 2004). Bar-Ilan *et al.* (2007) and Vaughan (2004) both conducted user studies to compare search engines. Users submitted rankings for a set of terms that where compared with search engine rankings. Both studies concluded that there were significant differences (e.g., order) between competing search engines, however, neither study examined the concept of personalization to explain these differences. Hannak *et al.* (2013) researched the personalization of web searches, but did only include browser history and geolocation. Other scholars (Kang, 2010; Meyer, 1988; Thorson, Vraga, & Ekdale, 2010) researched the credibility of blogs and newspapers, but excluded search engines research. Therefore, this study focuses on search engine research and aims to investigate the impact of personalization of search rankings on the credibility perceived by the search engine users, and whether these outcomes differentiate from search rankings of a search engine that does not track registered data (e.g., DuckDuckGo).

Therefore, *how do personalized search rankings affect the credibility perceived by search engine users?* is the central research question that this study will seek to answer. In addition, the following sub questions are formulated in order to guide the data collection process.

- How do profile based search engines influence search rankings?
- Are there differences in the top *k* lists of different search engines?

A feature that distinguishes this study from previous work is that this study includes four metrics, instead of one, to measure the differences between search engines. Prior studies (Bar-Ilan *et al.,* 2007; Vaughan, 2004) included one or two metrics to compare and validate the experimental results. However, some of the results were not conclusive and could not be validated. To extent the validity and the comparability of the results, this study uses four metrics to compare the differences between top *k* lists. Furthermore, the study includes a self-constructed survey to include the concept of personalization to elaborate the experimental results. The practical contribution of this research is found in the development of creating awareness among individuals of the importance and nuances of personalized rankings.

2. Theoretical framework

Registered data can alter and influence the search rankings displayed by Google, this data can be self-presented by the Internet users. (Rui & Stefanone, 2013). Therefore, the first section of the theoretical framework contains and elaborates the theory of self-presentation, which allows a theoretical lens to explain the customization of the online identity of an individual. The second section will discuss the two search engines researched in this study. The final sections will elaborate the concept of credibility and the process of *Back-end Googlization*.

2.1. Self-presentation theory

Multiple studies show that there is a tendency of individuals to create online personae's that differentiate from offline identities; due to the *self-presentation* online (Marwick, 2013; Papacharissi, 2002). The process of creating an online identity is through customization. When creating a social network profile, registering an email address or purchasing a product individuals use a variety of information to present themselves (e.g., preferences, interests) (Papacharissi, 2010). These items become symbolic markers of an online identity (Papacharissi, 2010). In an online setting, individuals can engage in an environment that can be controlled by themselves, where the symbolic markers can be conveyed, reinvented, and manipulated to an ideal identity (Papacharissi, 2010; Rui & Stefanone, 2013). The distance between "presenter" and "audience" is greater online, which makes it easy to conceal aspects of the offline self. Therefore, an online identity is often not the same as a real-world identity (i.e., offline identity), because online presented characteristics may differ from the characteristics presented in the physical world.[1]

Misrepresented symbolic markers and characteristics can influence the search rankings, because the search engine algorithm only validates the online presented symbolic markers and characteristics. Thus, the search rankings displayed will not match the offline identity of an individual. These biases may ultimately affect the credibility of the search engine rankings.

2.2. Search engines

According to Pan *et al.* (2007), search engines "act as an information intermediary that facilitates the information seeking process" (p, 3). Despite the popularity of search engines, users often are not aware of how they perform and know little about the implications of the algorithms used. The search engines in this study are divided into two categories; [1] search engines that track personal registered data (e.g., Google, Yahoo) and [2] search engines that do not track personal registered data (e.g., DuckDuckGo, Ixquick).

2.2.1. Google

When using Google, search terms entered by the search engine user are sent to the specific site an individual clicked on (via the HTTP referrer header); this is also referred as a search leakage, where personal information regarding a search query is shared with other databases.[2] Google keeps a record of what a browser has searched and collects data including account information, date, time, and information about the computer (e.g., IP address).

An algorithm is a series of (mathematical) steps that governs the flows of information on which the search engine user depends (Gillespie, 2014). Google's algorithm searches data from the Google index and uses a search term entered as input. This process is, however, not limited or static, due to the fact that the search rankings displayed are used as new input for the algorithm. For instance, if a search engine user clicks on a link to a specific website this information is collected and stored by Google, so that the algorithm can learn and improve the information displayed in future search rankings. This shows that algorithms are very complex and exits out of a lot of signals and metrics. The position of a website is determined by the number of links to the page from other sites (i.e., inbound links) and the number of links from the page to other sites (i.e., outbound links). Google's PageRank is one of the metrics that also assesses inbound and outbound links (Brin & Page, 2012). However, it quickly became clear that PageRank is very easy to manipulate. Individuals and corporations who have the opportunity, time, and the money to invest in search engine optimization (i.e., SEO) could severely affect the search rankings. Therefore, since 2013,

[1] http://www.internetsociety.org/

[2] https://duckduckgo.com/privacy

Google has chosen not to update and share this information anymore in an attempt to make their search rankings harder to manipulate. PageRank is the most revealing and critical metric that governs a domain's ability to determine a page's relevance or importance. PageRank is given as follows:

$$\mathrm{PR(A)} = (1 - \mathrm{d}) + \mathrm{d}\left(\frac{\mathrm{PR(T1)}}{\mathrm{C(T1)}} + \cdots + \frac{\mathrm{PR(Tn)}}{\mathrm{C(Tn)}}\right) =$$

PageRank (PR) is determined for each page individually and includes a scale from 0 to 10. 0 indicates that the specific page is not listed at all. To determine the PageRank for a specific page, for instance for Page A, all the pages linked to page A need to be identified. "Every page has a notion of its own self-importance. That's "PR(T1)" for the first page on the web all the way up to "PR(Tn)" for the last page" (Rogers, 2002).

"C(T1) refers to the number of outbound links on page 1 and C(Tn) refers to the number of outbound links for n, and so on for all other pages" (Rogers, 2002). Outbound links are the links pointing to other websites from Page A. "Thus, if page A has an inbound link from page N, the scare of the vote page A will get is "PR(Tn)/C(Tn)" (Rogers, 2002). This deviation needs to be done for all other pages linking to page A (e.g., T2, T3) (Langville & Meyer, 2011). "All these fractions of votes are added together. To stop the other pages from having too much influence, this total vote is "damped down" by multiplying it by 0.85 (the damping factor d)" (Rogers, 2002). The d is the normalized damping factor, thus ranging between 0 and 1. According to Brin and Page (2012) "the d damping factor is the probability at each page that a search engine user will get bored and request another random page." (p. 4), thus, the probability that a search engine user will continue searching. Because the PageRank theory beholds that a search engine user, who is randomly searching, surfing, and clicking, will eventually stop searching (Brin & Page, 2012; Page, Brin, Motwani, & Winograd, 1999; Richardson & Domingos, 2001). The PageRank algorithm can also be revised as:

$$\text{PageRank of a site} = \sum \frac{\text{PageRank of inbouding links}}{\text{Number of links on that page}}$$

Inbounding links are the links pointing to Page A, also known as backlinks. In the last few years Google has published a lot of changes in their algorithm, for instance, Google Panda, Google Penguin, and Google Hummingbird. Google Panda focuses on "the quality of user's browsing experiences. It aimed to punish low-quality, poor content sites, while promoting high quality, rich content sites" (Adams, 2013, p. 26). Google Penguin is "designed to penalize pages that have been scamming Google for rankings through webscam tactics" (Canon, 2011, p. 1). The update is aimed at decreasing search rankings that violate Google's guidelines by using spamdexing. Google Hummingbird was aimed at making interactions more human; so, the search engine was more capable of understanding concepts and correlations between search queries. Google Hummingbird offers "an opportunity to create content that is more focused on a search engine user's true intent" (Lin & Yazdanifard, 2014, p. 52). According to Lin and Yazdanifard (2014), such updates are influencing the current marketing strategies (e.g., social media, content marketing). "With the fact that Hummingbird presents the opportunity to create content that is more focused on a searcher's true intent, websites, pages or blogs of businesses should contain more liable and related content that the search engine users are looking for" (p, 52). This update is an extension of Google's Knowledge Graph, which was used to gather information from a wide variety of sources (Lin & Yazdanifard, 2014).

2.2.2. DuckDuckGo

DuckDuckGo is an example of a hybrid search engine, which does not track personal registered data. The search engine emphasizes on protecting the search engines user's privacy and avoiding *Filter Bubble effects*. DuckDuckGo prevents search leakages by not collecting, storing or sending data. DuckDuckGo developed a highly evolved contextual library for intuiting the search engine user's intent by receiving data from upstream search partners (e.g., Bing and Yahoo). DuckDuckGo focuses on directories operated by humans to evaluate and extent the "Instant Answers"; developed in the DuckDuckHack platform. This allows DuckDuckGo to place high-quality answers above organic results and advertisement. However, DuckDuckGo also has its own crawler bot and added an intelligence layer that removes spam and re-orders the provided results to optimize the quality of search rankings. To improve specific features (e.g., misspellings) the search engines saves searches. These searches are saved in a non-personal identifiable way, because the search engine does not store IP addresses or unique User agent strings.

2.3. Credibility

In 1986, communication researchers pointed out the lack of theoretical clarification of the credibility construct in the media. Subsequent, Gaziano and McGrath (1986) developed a credibility construct containing twelve items to measure credibility, which was later revised by Meyer (1988) into a construct containing five items; [1] fairness, [2] concerned with your interests, [3] bias, [4] exactness, and [5] trust. Nowadays, researchers refer to credibility as the objective as well as subjective parts of the believability of data and its source (Kang, 2010; Flanagin & Metzger, 2008). Credibility is often associated with concepts such as trustworthiness, reliability, and accuracy (Self, 1996). Kiousis (2001) made the distinction between two dimensions of credibility; [1] source credibility and [2] medium credibility. Source credibility focuses on "how different communicator characteristics influence the processing of messages", such as fairness, trustworthiness, and reliability (Kiousis, 2001, p. 382). The second dimension, medium credibility, refers to "credibility of the channel through which content is delivered" (Kiousis, 2001, p. 382). Metzger, Flanagin, Eyal, Lemus, and McCann (2003) later added a third dimension; the content credibility, referring to perceived credibility of the communicated message itself, such as accuracy and relevance. Prior research (Kiousis, 2001; Kohring & Matthes, 2007) conducted with the revised scales resulted in five constructs that consistently emerge in research; [1] trustworthiness, [2] reliability, [3] motivation to earn money, [4] accuracy, and [5] fairness. The study by Thorson *et al.* (2010) incorporated these constructs into a scale to measure blog credibility comprised of six semantic differentials: [1] fair/unfair, [2] relevance/irrelevance, [3] reliable/unreliable, [4] accurate/inaccurate, and [5] trustworthy/untrustworthy, and [6] balanced/imbalanced. This study draws on a combination of Meyer's (1988) and Thorson *et al.* (2010) credibility constructs, including [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, and [5] accuracy. A summary of the descriptions of the constructs is provided in *table 2.1* below.

Constructs	*Description of construct*	*Source*
Trustworthiness	The confidence that data will be handled competently, and that the corporation will not behave opportunistically Disinclined to deceive Being honest	Dinev & Hart, 2005, p. 64 Wilson, 1983, p.15 Hilligoss, & Rieh, 2008, p. 2
Reliability	The ability of a service to perform its required functions under stated conditions for a specified period of time	Ran, 2003, p. 7 Geraci, Katki, McMonegal, Meyer, Lane, Wilson & Springsteel, 1991
Fairness	Separation of fact and opinion	Gaziano & McGrath, 1986, p. 454
User based relevance	The aboutness, usefulness, usability, or utility of information objects in relation to the fulfillment of goals, interests, work tasks, or problematic situations intrinsic to the user	Borlund, 2003, p. 915
Accuracy	Exactness and correctness, which refers to the correspondence of a specification with the real needs of the user	Zowghi & Gervasi, 2003, p. 9

Table 2.1. Summary of the descriptions for the credibility constructs: [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, and [5] accuracy.

“Relevance is dynamic and changes as time progresses” (Taylor, 2012, p. 137) and Mizzaro (1998) addresses this problem as the inconsistent use of relevance as a concept. There are different relevance theories applicable depending on the type of research field (Mizzaro, 1998). According to Borlund (2003), “system or algorithmic relevance is the most common and clearest definition of relevance, and is applied in traditional evaluation of information retrieval systems” (p. 914). System or algorithmic relevance refers to “the relation between the query (terms) and the collection of information objects expressed by the retrieved information object(s)” (p. 915). However, this study focuses on the relevance concerned with the aboutness and appropriateness of a retrieved information object, and focuses on the degrees of intellectual interpretations carried out by search engine users (Borlund, 2003). Therefore, user based relevance is used in this study.

The most common evaluation criteria for relevance are [1] precision and [2] recall. According to Powers (2011), precision “denotes the proportion of Predicted Positive cases that are correctly Real Positives” (p. 38); the fraction of retrieved documents that is relevant. Recall is defined as “the proportion of Real Positive cases that are correctly Predicted Positive” (p. 38); the fraction of relevant documents that are retrieved.

2.4. Back-end Googlization

Search engines respond to a query with a ranked list of top ten documents (i.e., top *k* list) in their default setting. The search ranking reflects the search engines’ estimated relevance of web pages to the query. Three parties influence the search rankings; [1] web authors, who put their website online, [2] the search engines, and [3] the users of search engines, who enter various search queries (Pan *et al.*, 2007).

According to Rogers (2013) Google’s model relies on registrational interactivity, where user’s preferences and activity are registered, stored and employed, to serve tailored results. Prior to this model, search engine rankings returned the same information for all users at any given time. Luca *et al.* (2015) found three, sometimes overlapping, assumptions; [1] the search engine technology is not neutral, but contains features that favour some values over others, [2] the algorithms used do not apply objective criteria in generating rankings, and finally [3] the favouritism of other websites in the search rankings presented (Cothey, 2004; Van Couvering, 2010). These assumptions feature the ranking bias (i.e., *Back-end Googlization*) (Rogers, 2013), which refers to the propensity to rank certain sites higher than other sites, given that both websites are included in its index, based on registered data.

As mentioned, the study of Epstein and Robertson (2015) elaborates the impact of biased rankings on United States 2016 presidential election. The researchers portray that voters are subjected to a wide variety of factors, but are influenced by manipulated search rankings. When the algorithm, that ranks election-related information, favours one candidate over another, competing candidates would have no means of compensating for the bias. The researchers call this construction the *search engine manipulation effect*. “Voters are relatively powerless when trying to resist sources of influence they cannot see. They are not conscious of the fact that they are being manipulated” (Epstein & Robertson, 2015, p. 4520). Even when search engine users are aware of the fact that the search rankings are personalized (Epstein & Robertson, 2015, p. 4519). The researchers expect that this phenomenon will only expand over time, because the attention of individuals, in general, shifts from traditional sources of information towards the mobile sources.

Furthermore, Hindman, Tsioutsiouliklis, and Johnson (2003) researched the domination of politics by heavy linked websites. Their results showed that in every category, a few large sites dominate linkage patterns, creating a *Googlearchy* of “a small number of heavily-linked sites receiving more links than the rest of the sites combined—effectively dominating the community they are a part of” (Hindman *et al.*, 2003, p. 6). Höchstötter and Lewandowski (2009) reviewed the composition of search rankings of five separate engines. They concluded that Google and Yahoo favored their own subsidiaries; with Google consistently returning far more YouTube links than other searches. Furthermore, according to Purcell, Brenner, and Rainie (2012) search engine users perceive personalization as a bad thing, “because it may limit the information you get online and what search results you see” (p. 19).

3. Research design

3.1. Sample

Google is a popular search engine across different demographic groups, whom all have different information strategies and success rates (Bilal & Kirby, 2002). This study uses purposeful sampling to select a sample of N=46 young adults, sampled at the University of Twente, because research showed that young adults are often more successful in reviewing search rankings (Jansen & Spink, 2004). Individuals are labelled as young adults when they are aged 18 to 35 (Petry, 2002). Prior studies concluded that age was found to be very important when interpreting user search engine results (Purcell *et al.,* 2012, Van Couvering, 2010). Search engine users who are younger perceived certain events differently than search engine users who are older. Therefore, age is an important control variable in this study.

Furthermore, the participants need to use Google as primary search engine and need to have a high familiarity with Google's interface. The requests to participate are shared through e-mail correspondence and through personal approach. The research design of this study follows a systematic pattern. First, the participants conduct the experiment and after that the same sample of participants fills out a survey regarding the perceived credibility of Google and DuckDuckGo (see Appendix A for the survey).

3.2. Experiment design

To measure the similarity between top *k* lists of the two search engines and to discuss the impact of personalization, an experiment is conducted. The experiment also follows a systematic pattern; first, a sample of *x* participants is selected that execute *q* identical search terms, displayed in *table 3.1*, on their own computer for Google and DuckDuckGo. This process is monitored by the researcher. Subsequent, all search rankings for both search engines are saved and compared by the researcher to determine differences between top *k* lists. Furthermore, the rankings are also tested for stability.

After the experiment, the same sample of participants will fill in a self-constructed survey, which includes eleven questions regarding the perceived credibility (see Appendix A for the survey).

To create a valid data set, a set of screening criteria is used; the participants are requested to conduct the experiment on their own computer and not to delete their history and cookies, one month prior to the experiment, to make sure Google can offer personalized search rankings.

It is important to select identical search terms that have both breadth and impact, since there is no data available of what type of search terms are personalized. The selected search terms are retrieved from Google Trends, which publishes the most popular search terms. The search terms represent three categories; [1] business and politics, [2] economics, and [3] entertainment (see *table 3.1*). As Pariser (2011) elaborated, personalization of political related search terms raises some of the most contentious issues. This statement is also confirmed by Hannak *et al.* (2013) where "highest personalization for queries related to political issues, news, and local businesses." (Hannak *et al.,* 2013, p.2).

Top search terms

Category	Term	q_no
Business & Politics	Brexit	q1
Economics	Student loans	q2
Entertainment	Music	q3

Table 3.1. Top search terms for the three experimental categories (q1, q2, and q3): [1] business and politics, [2] economics, and [3] entertainment; identified by Google Trends

3.3. Data collection

The credibility of search engines rankings is determined by a survey and a quality assessment. The survey will determine the credibility of the search engine perceived by search engine users through human judgments. The first section contains three demographic questions, to distinguish differences between participants. The second step in constructing an overall credibility score is conducting an analysis of the five constructs to determine the participant's attitude towards search engines, gathered through the survey data. This analysis is based on a combination of abbreviated scales created by Meyer (1988) and Thorson *et al.* (2010) (see Appendix A for the survey). Unless otherwise noted, the constructs are measured on a five point Likert-type scale. Due to the high degree of subjectivity, this study also conducts a quality assessment. This quality assessment contains a set of quality indicators for the search engines; [1] up-to-dateness, [2] design, [3] instant search, [4] quality of search rankings, [5] privacy, and [6] response time (see Appendix B for the summary of descriptions). This analysis is based on a combination of revised scales proposed by Loiacono, Watson, and Goodhue (2002), Parasuraman, Zeithaml, and Malhotra (2005), and Yoo and Donthu (2001).

3.4. Data analysis

As mentioned, the overlap on identical queries is relatively low. (Bharat & Broder, 1998; Bar-Ilan, 2005; Lawrence & Giles, 1999). Thus, to measure and compare the distance between Google's and DuckDuckGo's permutations, which is the act of arranging documents into the same order, four different metrics are used. First the Jaccard Index (i.e., Jaccard similarity coefficient), which compares the similarity of documents. A Jaccard Index of 0 represents no overlap, while 1 indicates the lists contain the same rankings. (Hannak *et al.,* 2013). Jaccard similarity coefficient is defined as:

$$\mathrm{JS}(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B}{|A| + |B| - |A \cap B|}$$

A ∩ B refers to the search rankings displayed in list A and B and A ∪ B refers to the search results displayed in list A or B, where $0 \leq J\,(A,B) \leq 1$. The Jaccard distance, which measures the dissimilarity between sets, is obtained by the formula stated below. A higher value indicates a larger dissimilarity**.**

$$d_J(A,B) = 1 - \mathrm{JS}(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

The Spearman correlation coefficient is also computed to identify differences between Google and DuckDuckGo. The Spearman correlation coefficient is applied to ranked lists of *n* items, where the search rankings are valued between 1 and n. The Spearman correlation coefficient, r_s, can value from +1 to -1. An r_s of +1 indicates a complete agreement, 0 indicates no correlation between the two rankings, and -1 indicates a complete disagreement. The formula of the Spearman correlation coefficient, with no tied ranks, is stated below, where d_i =difference in paired ranks and *n* = number of cases. No ties are applicable due to the fact that there is no possibility that an identical search result with be replicated in same top *k* list.

$$\rho = 1 - \frac{6\sum d\frac{2}{i}}{n(n^2-1)}$$

The last metric is the extended metric of the Spearman footrule. The Spearman's footrule is computed as $F(\sigma 1, \sigma 2) = \sum | \sigma_1(i) - \sigma_2(i))$, when two lists are identical *F* is zero (Bar-Ilan, 2005; Bar-Ilan, Keenoy, Yaari, & Levene, 2007). In the case when two lists are not identical, an arbitrary placement will be assigned. This placement is (*k*+1) for all documents not appearing in the list (Fagin *et al*., 2003). The formula stated below is the extension of the Spearman footrule:

$$F^{(k+1)}(\tau_{1,}\tau_2) = 2(k - z)(k+1) + \sum_{i \epsilon Z} |(\tau_1(i) - \tau_2(i)| - \sum_{i \epsilon Z} \tau_1(i) - \sum_{i \epsilon Z} \tau_2(i)$$

Z is the set of overlapping documents; s is the set of documents only appearing in the first list, and *t* is the set of documents only appearing in the second list. The probability that two lists of documents are identical is very small, thus the non-overlapping documents have a tremendous effect on the outcome (Bar-Ilan *et al.*, 2006). Therefore, the formula was normalized, so that the *F* will be valued between 0 and 1, independent of the overlap. For k=10 the normalization factor is 110 (Fagin *et al.*, 2003; Bar-Ilan *et al.,* 2006). For *k*=10 the normalization factor is 110. The normalized formula is referred as *G* measure.

$$F^{51} = 1 - \frac{F^{(k+1)}}{k(k+1)}$$

Note that *F* is a distance measurement, therefore, the smaller the value the more similar the rankings are. For all search terms, the average correlation and standard deviation between the search engines are computed based on the overlapping URL's. Furthermore, to test whether DuckDuckGo indeed displays the search rankings organically and returns the same rankings for all search engine users, when entering an identical search term; the non-parametric Friedman test is used. The Friedman test is used to test for differences between samples when the dependent variable being measured is ordinal.

This study includes an elaboration of the researched credibility constructs, identified by Meyer (1988) and Thorson *et al.* (2010); [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, and [5] accuracy. All variables are tested for normality, with the Shapiro-Wilk Test. When the significance value is greater than 0.05, it can be concluded that the variable for this particular subset of individuals is normally distributed. This test is explicitly useful for small sample sizes. To measure whether the credibility scores of both search engines are statistically different, a paired-samples *t* test is chosen to compare the mean scores of both search engines. To measure whether there is a negative association with the personalized search engine (dichotomous variable) and perceived credibility (continuous variable) a Point Biserial Correlation test is computed (Jaccard & Becker, 2002).
Furthermore, due to potential effects of other variables on the dependent variable, this study will also control for age, gender, usage frequency, and educational level. Demographic data is also obtained through the survey listed in Appendix A. The survey data set has a relative high reliability with a Cronbach's alpha =.797 with N=9. The research is anonymously conducted and the participants are informed of the anonymity beforehand.

4. Results

The research results will be elaborated within this chapter. First the demographic characteristics of the sample will be briefly discussed. Subsequently, the experimental results will be elaborated. Section three includes the credibility results, these results are divided into six parts; [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, [5] accuracy, and [6] the concluding credibility results. The quality assessment will be briefly mentioned in the final section.

4.1. Demographic characteristics

The sample for the survey and the experiment consists out of 22 male and 24 female participants. Aged between 18 and 31, averaged 23 years (*s* = 2.650) (see Appendix D for the descriptive statistics). The sample consists out of 15 undergraduate students, 21 graduate students and 10 participants with a completed master's degree.

4.2. Experimental results

The results of the first metric, the Jaccard similarity coefficient (see *table 4.1),* show that almost all JS (*A, B)* coefficients are significantly small. Therefore, it can be concluded that the top *k* lists of DuckDuckGo and Google are highly dissimilar (M = .193, SD = .029). The high average dissimilarity($d_J(A,B)$ = .807, SD = .029) value indicates that both search engines display significantly different links, factors, and information.

Furthermore, the r_s coefficients for q2 and q3 are highly negative, indicating an almost complete disagreement between the top *k* lists.

To extent the current body of literature, the F^{51} is also computed, which is very useful in the case when two lists are not identical. The results are normalized, thus, the value ranges between 0 and 1. The F^{51} values, listed in *table 4.1*, also confirm that Google's top *k* lists and DuckDuckGo's top *k* lists are highly dissimilar, averaged F^{51} =.836, SD = .014. Note that, *F* is a distance measurement, thus, the smaller the value, the more similar the top *k* lists are. All metrics are equivalent, thus, calculate the same results; however, the different methodologies of the metrics help to compare and validate the results more thoroughly.

Permutation distance DuckDuckGo and Google

q_no	Query	$JS(A,B)$	$d_J(A,B)$	ρ	F^{51}
q1	Brexit	.160	.840	-.321	.845
q2	Student loans	.205	.795	-.806	.843
q3	Music	.214	.786	-.914	.819
AVG		.193	.807	-.681	.836
SD		.029	.029	.316	.014

Table 4.1. Search engine permutation distance (DuckDuckGo and Google) for the three experimental categories (q1, q2 and q3): Jaccard's Index, Jaccard's distance, Spearman correlation coefficient, and the extended Spearman footrule.

The χ^2 values in *table 4.2* show that for q2 and q3 no significant differences are found; the lists are completely identical (see *table 4.2*). The value for q1 also shows no significant difference between the top *k* lists, however, the Chi-square value, $\chi^2(45)$ = 16.132, *p* = 1.00, indicates that there are differences within the top *k* lists, when entering the search term ‘Brexit’. This result is unexpected, because DuckDuckGo states that the search engine should return the same rankings for all search engine users, when entering an identical search term. *Figure 4.1* and *Figure 4.2* are examples of dissimilarities in the search rankings. The figures display two DuckDuckGo’s top *k* lists that are re-ranked and even missing from previous search ranking results.

Stability DuckDuckGo (non-parametric Friedman test)

q_no	Term	df	$\chi^2(45)$	Sig. (2-tailed)	
q1	Brexit	45	16.132	1.00	Insig.
q2	Student loans	45	-	-	-
q3	Music	45	-	-	-

Table 4.2. DuckDuckGo’s stability values for search term q1, q2, and q3; non-parametric Friedman test; SPSS outcome

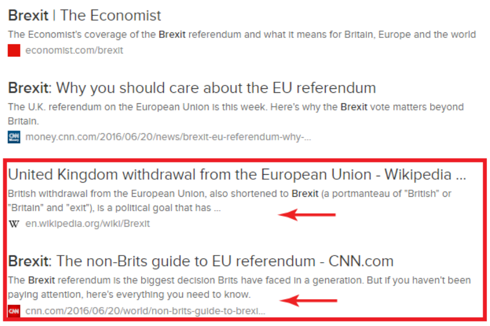

Figure 4.1. Participant #12 DuckDuckGo search rankings for search query ‘Brexit’ (q1)

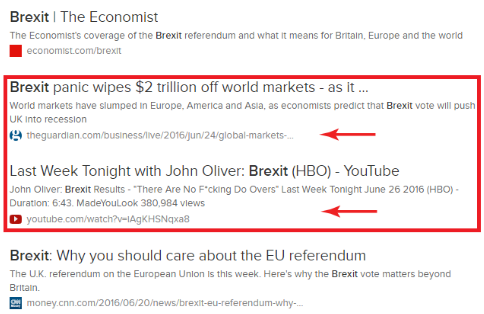

Figure 4.2. Participant #18 DuckDuckGo search rankings for search query ‘Brexit’ (q1)

Table 4.3 displays the χ^2 values for Google. The Friedman test shows that there is a statistically significant difference in Google’s top *k* lists for each of the search terms (q1, q2 and q3), which indicates that all participants received a different top *k* list. This is expected, due to the fact that Google personalizes the search rankings based on the online identity of the specific search engine user.

Stability Google (non-parametric Friedman test)

q_no	Term	df	$\chi^2(2)$	Sig. (2-tailed)	
q1	Brexit	45	150.967	.000	Sig.
q2	Student loans	45	161.984	.000	Sig.
q3	Music	45	139.311	.000	Sig.

Table 4.3. Google’s stability values for the three experimental categories (q1, q2, and q3); non-parametric Friedman test; SPSS outcome

4.3. Credibility results

This study includes an elaboration of the researched credibility constructs, identified by Meyer (1988) and Thorson *et al.* (2010); [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, and [5] accuracy. As mentioned, prior research showed that age is an important variable when interpreting search engine results. Therefore, the correlation between the five credibility constructs and age is measured separately. All credibility constructs showed a violation of normality, linearity or homoscedasticity, therefore, a Spearman’s rho test is used to find inter-correlations between these variables.

4.3.1. Participants’ usage frequency

Figure 4.3 shows the differences between the usage of DuckDuckGo and Google. Prominent is that only 2% of the participant uses DuckDuckGo very often and 50% of the participants rarely uses DuckDuckGo’s interface. Whereas 46% of the participants use Google very often and only 4% participants use Google occasionally. Thus, it can be concluded that Google is used more frequently than DuckDuckGo. This result corresponds with the study by Purcell *et al.* (2012), where they conclude that Google is the search engine of choice, preferred by 83% of the search engine users.

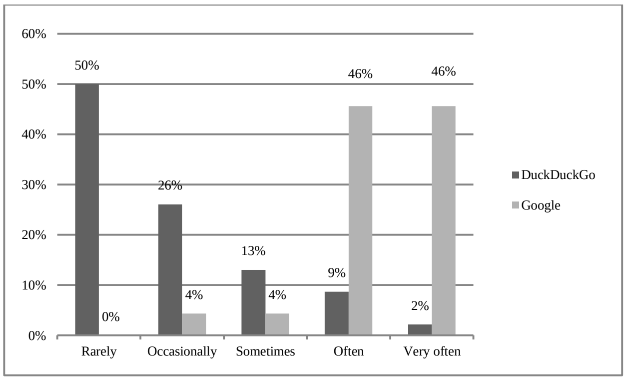

Figure 4.3. Participants’ usage frequency of Google and DuckDuckGo; percentages

4.3.2. Trustworthiness

Trustworthiness refers to “the confidence that data will be handled competently, and that the corporation will not behave opportunistically” (Dinev & Hart, 2005, p. 64). *Figure 4.4* illustrates that no extreme number of participants perceive the search engines as ‘untrustworthy’ or ‘very trustworthy’.

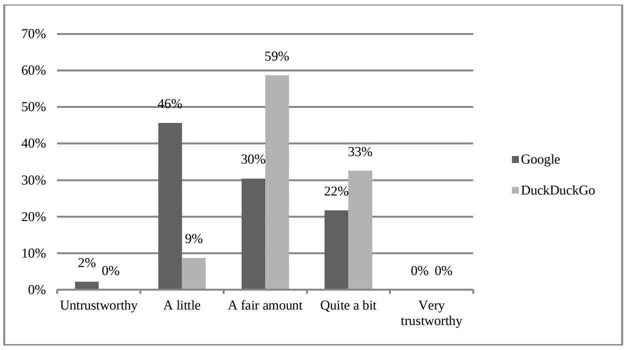

Figure 4.4. *Participants'* perceived trustworthiness results of Google and DuckDuckGo; percentages

The results listed in *table 4.4* indicate that the participants (N=46) rated Google, averaged 2.717 (*s* =.834) as less trustworthy than DuckDuckGo, averaged 3.239 (*s* = .603), even though the user frequency of DuckDuckGo is relatively low.

Descriptive Statistics

	N	Min	Max	Mean	Std.
Google	46	1.00	4.00	2.717	.834
DDG	46	2.00	4.00	3.239	.603
N	46				

Table 4.4. Descriptive statistics concerning the participants' perceived trustworthiness for Google and DuckDuckGo search rankings; SPSS outcome

4.3.3. Reliability

Reliability refers to "the ability of a service to perform its required functions under stated conditions for a specified period of time" (Ran, 2003, p. 7). *Figure 4.5* shows that Google is perceived as more reliable than DuckDuckGo. Furthermore, no participant perceived DuckDuckGo or Google as unreliable or as very reliable.

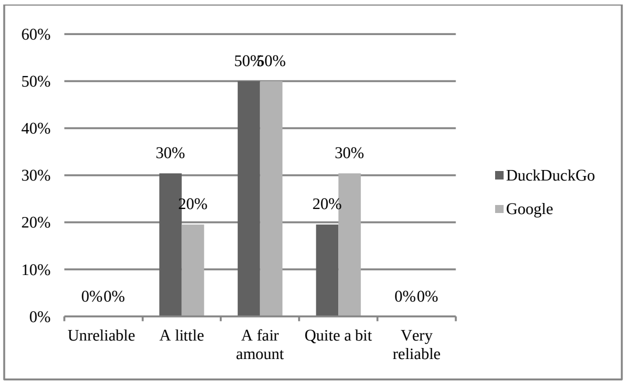

Figure 4.5. *Participants'* perceived reliability results; percentages

The results listed in *table 4.5* indicate that the participants (N=46) rated Google, averaged 3.109 (*s* =.706) as more reliable than DuckDuckGo, averaged 2.891 (*s* = .706).

Descriptive Statistics

	N	Min	Max	Mean	Std.
Google	46	2.00	4.00	3.109	.706
DDG	46	2.00	4.00	2.891	.706
N	46				

Table 4.5. Descriptive statistics concerning participants' perceived reliability for Google and DuckDuckGo search rankings; SPSS outcome

4.3.4. Fairness

Fairness refers, according to Gaziano and McGrath (1986), to the separation of fact and opinion (p. 454). *Figure 4.6* illustrates that 15% of the participants perceive Google as opinionated and no participant perceived Google as factual. Furthermore, no participants perceive DuckDuckGo as opinionated and 4% of the participants perceive DuckDuckGo as factual**.**

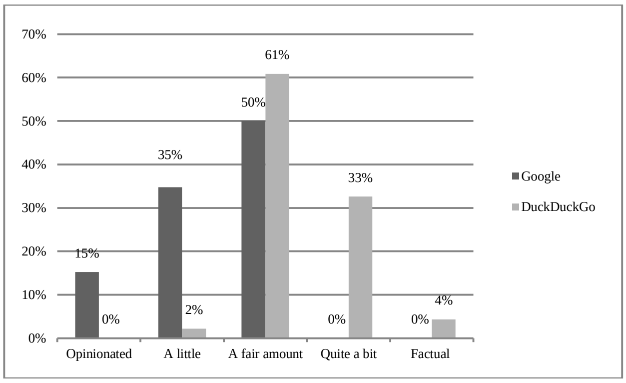

Figure 4.6. *Participants'* perceived fairness results; percentages

Table 4.6 illustrates that Google, averaged 2.348 (*s* = .737) is perceived as more opinionated than DuckDuckGo, averaged 3.391 (*s* = .614).

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Google	46	1.00	3.00	2.348	.737
DDG	46	2.00	5.00	3.391	.614
N	46				

Table 4.6. Descriptive statistics concerning participants' perceived fairness for Google and DuckDuckGo search rankings; SPSS outcome

Furthermore, *Figure 4.7* depicts an interesting example of search rankings generated by Google when entering the search term 'student loans' (q2). The election-related information regarding Hilary Clinton's United states 2016 presidential campaign is ranked in the top *k* lists of Google, whereas, no information regarding Donald Trump is displayed. This top *k* list could be an example of the *search engine manipulation effect*, elaborated by Epstein and Robertson (2015), where individuals receive and perceive specific election related information based on their own content and interests.

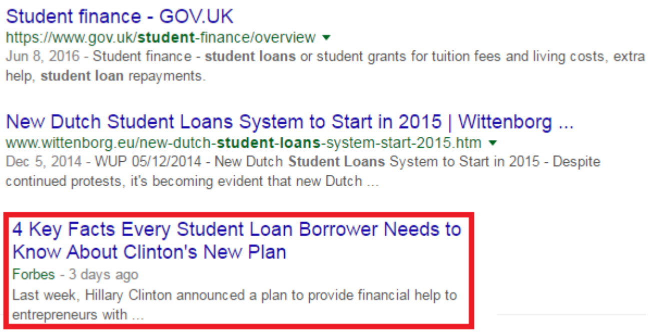

Figure 4.7. Participant #26 Google search rankings for search query 'Student loans' (q2)

4.3.5. *Relevance*

Relevance in user generated content refers to "the aboutness, usefulness, usability, or utility of information objects in relation to the fulfillment of goals, interests, work tasks, or problematic situations intrinsic to the user" (Borlund, 2003, p. 914). *Figure 4.8* illustrates that 57% of the participants perceive Google's search rankings as "quite relevant", whereas 39% of the participants assign DuckDuckGo's search rankings in this category.

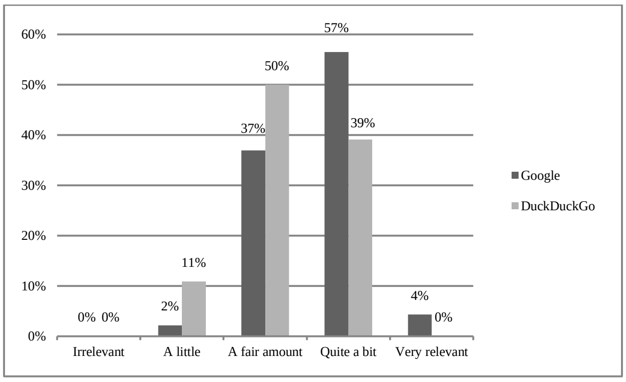

Figure 4.8. Participants' perceived relevance results; percentages

Table 4.7 illustrates that Google, averaged 3.630 (s = .610), is perceived as more relevant than DuckDuckGo, averaged 3.283 (s = .655).

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Google	46	2.00	5.00	3.630	.610
DDG	46	2.00	4.00	3.283	.655
N	46				

Table 4.7. Descriptive statistics concerning participants' perceived relevance for Google and DuckDuckGo search rankings; SPSS outcome

Figure 4.8 and *figure 4.9* depict documents listed by Google after entering the search term music (q3). Google's search rankings displayed several artistical suggestions (e.g., Rihanna, 'The Voice' artists). The top k lists documented by DuckDuckGo only listed general websites (e.g., YouTube, Apple Music), without inviting suggestions. The symbolic markers of an online identity (e.g., preferences, interests) contribute to the individual collection of registered data. With this content Google can generate unique and personalized search rankings, as displayed in *Figure 4.8* and *Figure 4.9*.

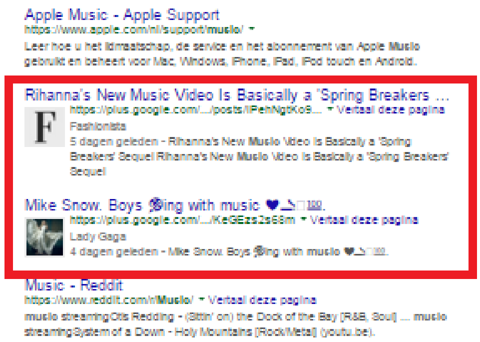

Figure 4.8. Participant #18 Google search rankings for search query 'Music' (q3)

Figure 4.9. Participant #5 Google search rankings for search query 'Music' (q3)

4.3.6. *Accuracy*

As mentioned Zowghi and Gervasi (2003) define accuracy as "the correspondence of a specification with the real needs of the user" (p, 9). *Figure 4.10* illustrates that 46% of the participants perceive Google's search rankings as 'quite a bit' accurate, whereas 30% of the participants ranked DuckDuckGo's search rankings in this category. These results are comparable with the research of Purcell *et al.* (2012), where 45% of the participants state that search engines are accurate most of the time.

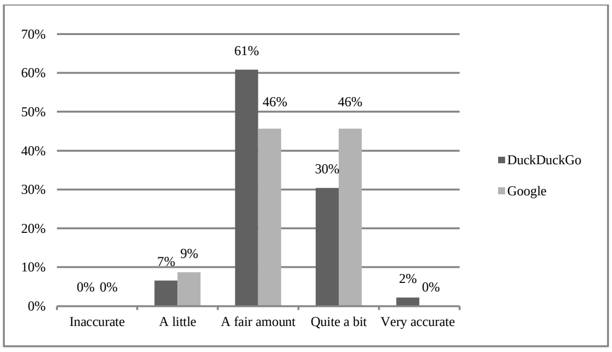

Figure 4.10. Participants' perceived accuracy results; percentages

Table 4.8 shows that Google's search rankings, averaged 3.370 (s= .645), are perceived as more accurate than DuckDuckGo's search rankings, averaged 3.283 (s = .621).

Descriptive Statistics

	N	Min	Max	Mean	Std.
Google	46	2.00	4.00	3.370	.645
DDG	46	2.00	5.00	3.283	.621
N	46				

Table 4.8. Descriptive statistics concerning participants' perceived accuracy for Google and DuckDuckGo search rankings; SPSS outcome

4.3.7. Credibility results

Due to the means of both credibility scores, listed in *table 4.9,* and the direction of the *t*-value, T= -2.166, it can be concluded that there is a statistically significant difference between the perceived credibility of Google's search rankings, averaged 3.035 (s = .481) and the perceived credibility of DuckDuckGo's search rankings, averaged 3.217 (s = .279), t (45) = -2.166, $p < .036$

Descriptive Statistics

	N	Min	Max	Mean	Std.
Google	46	1.80	3.80	3.035	.481
DDG	46	2.40	4.00	3.217	.279
N	46				

Table 4.10. Descriptive statistics concerning participants' average credibility score; SPSS outcome

Because personalization is a dichotomous variable and the perceived credibility score is a continuous variable, a Point Biserial Correlation test is used to determine the relationship between personalization (independent variable) and perceived credibility (dependent variable) (see *table 4.11).* Results show that there is a negative correlation between personalization and the perceived credibility score, which is statistically significant, r_{bp} (45) = -.228, $p < .029$; indicating that personalization of the search rankings could result in a lower credibility score.

As mentioned by several scholars, credibility consists out of five constructs; [1] trustworthiness, [2] reliability, [3] fairness, [4] relevance, and [5] accuracy. *Table 4.12* shows the calculated weighted average of the five credibility constructs of all participants. Google receives an average score of 3.630 for their relevance; indicating that Google's search rankings are perceived between the 'a fair amount' and 'quite a bit'. DuckDuckGo receives their highest score for the fairness construct, averaged 3.391, whereas Google receives the lowest score for the fairness construct, averaged 2.348; the participants perceive Google as "a little bit" fair. DuckDuckGo scores the lowest on reliability, averaged 2.891; between "a fair amount" and "quite a bite" reliable.

Paired Samples Test

		Paired Differences					T	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Average_Google Average_DDG	-.18261	.57167	.08429	-.35237	-.01284	-2.166	45	.036

Table 4.9. Differences between the participants' average perceived credibility results; SPSS outcome

Credibility (Point-Biserial Correlation)

Independent variable	Dependent variable	df	Point-Biseral Correlation	Sig.(2-tailed)	Confidence int.	
Personalization	Perceived credibility score	45	-.228	.029**	95%	Significant

* p <.05, two-tailed; * * p <.01, two-tailed

Table 4.11. Correlation between personalization and the participants' average perceived credibility score; SPSS outcome

Credibility constructs		
Construct	Google	DuckDuckGo
Trustworthiness	2.717	3.239
Reliability	3.109	2.891
Fairness	2.348	3.391
Relevance	3.630	3.283
Accuracy	3.370	3.283

Table 4.12. Average of the credibility constructs for the five credibility constructs, for both search engines; survey outcome.

4.4. Control variables

Due to the potential external effects of other variables, this study has also controlled the averaged credibility score for age, gender, usage frequency, and educational level. A linear regression analysis with the control variables is executed. Dummy variables are created for the variables educational level (categorical), usage frequency (categorical), and gender (dichotomous) to execute the regression analysis. *Table 4.13* (see Appendix C for the model summary) displays the results of the regression analysis. Model 1, including all control variables, shows a R-squared value of .312, which indicates that 31% of the variance in the average credibility score can be explained by the control variables [1] age, [2] gender, [3] usage frequency, and [4] educational level. Model 2, including all control variables and the predictor variable (i.e., personalization), shows an R-squared value of .495, which indicates that 50% of the variance in the average credibility score can be explained by the control variables and the predictor variable. The R-square change value of .183, displayed in Model 2, indicates that the independent variable, personalization, explains 18% of the variance in the average credibility score, even when the effects of age, educational level, usage frequency, and gender are controlled for. *Table 4.14* the Anova table, indicates that the model as a whole is a significant predictor of the average credibility score, $p < .000$. *Table 4.15*¸ the coefficient outcome shows that only two variables contribute to the significance of the model; age and personalization (independent variable). Gender, educational level, and usage frequency do not make a unique, significant contribution to the model. The standardized coefficients of both contributing variables show that personalization contributes the most to the model.

The variable age has a unique contribution to the model and, therefore, the correlation between age and the five credibility constructs is measured separately (see Appendix E for the correlation table). Statistics show that for Google there are three significant negative correlations; trustworthiness, fairness, and accuracy. Thus, the older the participants are the lower the perceived score for these variables. Purcell *et al.* (2012) also stated in their study, that younger search engine users tend to perceive search engines as more fair. 79% of the participants aged between 18 and 29 years state that search engines are fair and unbiased (Purcell *et al.*, 2012).

Furthermore, for DuckDuckGo only two significant correlations are found; trustworthiness and relevance. The correlation between trustworthiness and age is negative, $r_s (45) = -.666$, $p < .000$, similar as Google. Thus, the older the participants are the lower the perceived score for trustworthiness. The positive correlation between relevance and age, $r_s (45) = .523$, $p < .000$, indicates that older participants perceive DuckDuckGo's search rankings as more relevant than younger participants.

4.5. Quality assessment

Due to the subjective nature of the survey, a quality assessment of both search engines is conducted. A summary of the quality assessment is provided in *table 4.17*. The quality assessment consists of six factors; [1] up-to-dateness, [2] design, [3] instant search, [4] quality of search rankings, [5] privacy, and [6] response time. It is possible to insert a timeframe of at least one hour and sort by date in Google's search rankings. The sortation by date is also possible in DuckDuckGo features; however, the search engine provides a timeframe of one day. Thus, it can be concluded that Google is more up-to-date than DuckDuckGo. *Figure 4.10* illustrates this factor. Both encircled search results have a marked date of before the Brexit referendum, however, these top *k* lists returned weeks after the actual referendum. Thus, it is peculiar that DuckDuckGo's top *k* lists still display older links to websites containing information which may be irrelevant to the search engine user.

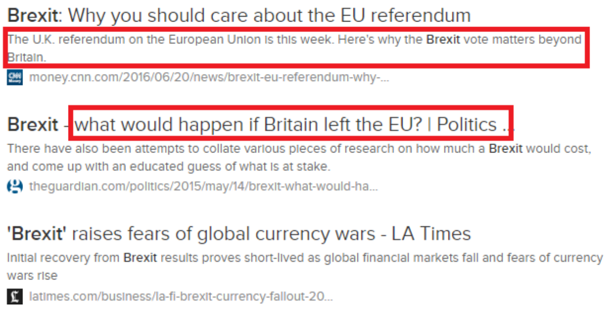

Figure 4.10. Participant #12 DuckDuckGo search rankings for search query 'Brexit' (q1)

DuckDuckGo offers an option to tailor the search engine to individual preferences, histories, and ways of shopping. Google, however, provides a clean design with features to change the content and specifications of the search rankings. Furthermore, both search engines consist of a computer arithmetic sequence. An arithmetic sequence is a list of numbers in which the difference between consecutive terms is constant. The difference between consecutive terms is always the same, regardless of the starting number.

The quality of the search rankings is based on the display and ordering of search rankings. DuckDuckGo state that the search engine displays the search rankings results organically (i.e., without personalization), whereas Google's search rankings are personalized based on an individual's online identity. As mentioned, the personalized search rankings could result in *Filter Bubble effects*, "where users are only given results that the personalization algorithm thinks they want (while other, potentially important, results remain hidden)" (Hannak et al. 2013, p. 1). Based on the study of Powers (2011) *Filter Bubble effects* can also be evaluated as a type II error or false negative. When an individual misrepresent themselves online Google can display less relevant documents based on the online identity due to the lack of correlation with the offline identity. Therefore, the proportion of relevant documents that are retrieved (i.e., recall) is very low. However, when the symbolic markers and characteristics presented online correlate with the offline identity of the search engine user personalization can enhance the relevance of the search rankings displayed by Google. This enhancement can indicate that the amount of actual retrieved relevant documents, displayed by DuckDuckGo (i.e., recall), could be significantly lower than Google's recall. However, recall is difficult to determine, because it is difficult to know exactly how many relevant documents exist in a database (Powers, 2011).

This study did conclude that certain identical search terms display different DuckDuckGo top *k* lists, even providing certain unique and personal suggestions. Therefore, it can be questioned whether DuckDuckGo generates organic and unbiased search rankings. The quality factor also contributes to the privacy factor; Google is a profile oriented search engine that focuses on the provision of search rankings personalized by (self) registered data. Thus, Google is focusing on interconnection. Whereas, DuckDuckGo is a search engine that offers an anonymous environment, where search engine users often perceive a higher level of privacy, because the search engine does not collect, store, and employ personal identifiable data.

The final factor collects information about the response time, the time to get a response after entering a search term. Google has a lower response time than DuckDuckGo. When entering an identical search term, Google provides the search rankings after an average of 0.38 seconds, whereas DuckDuckGo provides the search rankings after an average of 1.53 seconds.

Quality assessment

Factors	Google	DuckDuckGo
Up-to-dateness	Add timeframe of at least one hour and sort by date	Add timeframe of at least one day and sort by date
Design	Clean design without customization options	Customization of the looks and behaves (e.g., font, stop favicons from displaying)
Instant search	Computer arithmetic	Computer arithmetic
Quality of search rankings	Type II error	Organic display of results
Privacy	Profile oriented (crawler-based search engine)	Hybrid search engine
Response time	0.38 seconds	1.53 seconds

Table 4.17. Quality assessment of both search engines divided into six factors: [1] up-to-dateness, [2] design, [3] instant search, [4] quality of the search rankings, [5] privacy, and [6] response time

5. Discussion

Currently, several books concerning SEO and consultancy regarding search engine manipulation are very popular (Dover & Dafforn, 2011; Fleischner, 2009; Halavais, 2013; Jones, 2008; Langville & Meyer, 2011; Ledford, 2015). With this increasing demand for SEO strategies, tactics, and the manipulation of search rankings it is important to be aware of a search engine's discursive mechanisms (e.g., *Back-end Googlization*, *Googlearchy, search engine manipulation effect*) that have been addressed in this study.

This study uses four metrics to measure the differences between Google and DuckDuckGo. The results of these metrics are conclusive and indicate that there are extensive differences between Google's and DuckDuckGo's top *k* lists; there is little or no overlap.

As expected, the stability measurements for most DuckDuckGo terms show that there are no significant differences between DuckDuckGo's top *k* lists. However, the smaller Chi-square value for the search term 'Brexit' (q1) shows that there are several differences in the top *k* lists for q1. This is an unexpected result, because DuckDuckGo states that identical search terms should return the same search rankings for every participant, regardless of browser history, online activities, and cookies. Thus, it can be questioned if DuckDuckGo's search rankings are indeed organic (i.e., without personalization) and unbiased. Therefore, extensive research regarding the organic structure of DuckDuckGo would be interesting. The stability values for Google showed that there are significant differences between Google's top *k* lists, which is likely due to the personalization of the search rankings; search engine users receive and perceive information based on their online identity, thus, search rankings can be highly different depending on the search engine user's preferences.

As mentioned, Luca *et al.* (2015) listed three assumptions. The first assumption discusses the neutrality of search engine technology (Cothey, 2004; Van Couvering, 2010). The survey results show that Google was perceived as less neutral than DuckDuckGo. Furthermore, Google's high permutation distance values for 'Brexit' (q1) indicate that different search engine users are subjected to a wide variety of different factors and links containing different types of information and opinions. This could ultimately influence the opinion of the search engine user regarding certain events, such as the Brexit or even the United States 2016 presidential election. When an algorithm, which ranks political-related information, favours one campaign over another it could ultimately lead to a biased vote. Subsequent, when entering the term 'student loans' (q2) several top *k* lists of Google showed United States 2016 presidential election-related information regarding Hilary Clinton's campaign, whereas, no information regarding Donald Trump was displayed.

In both situations, the *search engine manipulation effect* (Epstein and Robertson, 2015) could be very prominent. Because the British as well as the American voters receive different top *k* lists from Google based on registered personal data and could be influenced by biases in these top *k* lists, even when they are aware that these search rankings are personalized.

The second assumption elaborates the application of objective criteria in generating search rankings (Cothey, 2004; Van Couvering, 2010). As mentioned, Google is a predication and recommender search engine, which gathers data to create a unique profile for the specific search engine user. However, when the self-presented data is manipulated or reinvented by the search engine user, Google could fail to present objective, relevant, and accurate data based on the online identity of the search engine user.

The final assumption discusses the favouritism of other websites in the search rankings presented (Cothey, 2004; Van Couvering, 2010). Höchstötter and Lewandowski (2009) concluded that Google favours their own subsidiaries compared to other search engines, thus, the cooperation displays fair more YouTube links than, for instance, DuckDuckGo. Höchstötter and Lewandowski's (2009) conclusion correlates with the outcome of this study, because Google always places own subsidiaries in the same ranking order; all top *k* lists displayed by Google, when entering 'music' (q3), listed YouTube as the first result. As mentioned, "the information individuals access affects their understanding of the world around them and the decisions they make; biased information can affect both how individuals treat others and how they evaluate their own choices or opportunities" (Kay *et al.*, 2015, p, 1). Thus, the algorithms used by Google can ultimately alter the way individual's encounter information and the perception of what information is available, by presenting inferior results to what they would otherwise find if all results were displayed organically (i.e., without personalization).

The second part of the study includes a survey. The survey incorporates all identified credibility constructs. Prior research (Thorson *et al*., 2010) concluded that Google is considered as more trustworthy than traditional media (e.g., newspapers) for news, because the search engine does not generate own content. However, prior studies did not research whether the level of perceived trust varies between search engines, where DuckDuckGo "vows to protect the search engine's user's privacy by not saving and utilizing registered data", and is perceived as more trustworthy than Google.[3]

Purcell *et al.* (2012) state in their research that almost 66% of the search engine users perceive search engines in general as a fair and unbiased source of information. This statement contradicts with the results in this study, where relatively low fairness scores are generated. The participants evaluate Google as more opinionated than DuckDuckGo. An explanation for these results can be found in the personalization of Google's search rankings. "The search engine users are given the results that the algorithm thinks they want" (Hannak et al. 2013, p. 1). Thus, certain particular opinionated articles or webpages can be ranked in the top *k* lists due to Google's collection of an online identity.

Furthermore, the participants in this study evaluate Google's search rankings as more relevant, due to the higher degree of aboutness, usefulness, usability, or utility of information objects in relation to the fulfillment of goals, interests, work tasks, or problematic situations intrinsic to the search engine user (Borlund, 2003). Lewandowski (2011) also stated that Google shows significantly more personalized (commercial) results than other search engines (e.g., Yahoo, Microsoft) and the search engine users are aware of this bias. However, this does not affect the perceived relevance of the search rankings, which remains very high.

The results show that the participants evaluated DuckDuckGo as more credible based on two out of the five constructs; [1] trustworthiness and [2] fairness. The outcome is remarkable due to the fact that 45% of the participants rarely use DuckDuckGo. An explanation for the higher level of trust is the statement of DuckDuckGo illustrating that the search engine does not save and utilize registered data, in comparison with Google.[4]

A regression analysis shows that age has a unique and significant contribution to the overall credibility score. Indicating that the older the participant is the lower the perceived credibility scores are. However, the sample can be broadened with older individuals to validate these results more thoroughly.

The results of the experiment, survey, and quality assessment show that personalization provides a lot of benefits for the search engine user (e.g., up-to-dateness, speed, relevance, accuracy), however, personalization also develops an environment where search engine users perceive the service as less trustworthy and more opinionated. Resulting in the fact that organic services, such as DuckDuckGo, are perceived as a fair and more trustworthy choice of platform to perform search queries, due to its anonymous environment and the protection of the privacy of the search engine user. Pariser (2011) and Hannak *et al* (2013) already stated that search rankings regarding political issues, news, and local businesses are most personalized and could raise some of the most contentious issues. Constructions such as *Back-end Googlization, Googlearchy*, and the *search engine manipulation effect* will not cease to exist and in addition with the (political) examples mentioned in this study it shows that these constructions need to be examined more closely to understand the (political) consequences of the information age. Similarly, how the media pays attention to the framing methods used by politicians, the search engine users and scholars should be aware of the discursive constructions of the algorithms created by search engines. by search engines.

The research results are not representative of all search engine users, due to the significant small sample size. Thus, following this study, the proposed methodology can be used to establish more confirmatory research with a broader sample of on-going search engine users. Subsequent, Jansen and Spink (2004) state that search engine users in the U.S. engage in different search techniques than search engine users in Europe. Therefore, it would also be interesting to replicate the study in different countries and using a larger set of priory selected search terms, to offer more generalizable results. Next to that, both the web and the search ranking algorithms constantly change, thus, these studies should be carried out periodically.

6. Conclusion

The goal of this study was to explore how the personalization of search rankings affects the credibility perceived by a search engine user. This exploratory research delivered some interesting findings and offers a good foundation for further research on search engine optimization and search ranking biases.

The present study advances the current body of research by including more metrics to validate the data and to include personalization as a variable to elaborate differences between search engines. The study could serve as an alert for search engine users in marketing literature as well as in information management, offering a visualization of the personalized search rankings and manipulation effects in political campaigns.

ACKNOWLEDGEMENTS

I would like to thank the supervisors of my thesis: Dr Fons Wijnhoven and PhD candidate Anna Priante for their insightful comments and guidance, but also for the encouragement to specify my research into a smaller perspective.

References

Adams, R. L. (2013). *SEO Black Book: A Guide to the Search Engine Optimization Industry's Secrets*. CreateSpace Independent Publishing Platform.

Bar-Ilan, J., Mat-Hassan, M., & Levene, M. (2006). Methods for comparing rankings of search engine results. *Computer networks*, *50*(10), 1448-1463.

Bar-Ilan, J., Keenoy, K., Yaari, E., & Levene, M. (2007). User rankings of search engine results. *Journal of the American Society for Information Science and Technology*, *58*(9), 1254-1266.

Bharat, K., & Broder, A. (1998). A technique for measuring the relative size and overlap of public web search engines. *Computer Networks and ISDN Systems*, *30*(1), 379-388.

[3] https://duckduckgo.com/privacy

[4] https://duckduckgo.com/privacy

Bilal, D., & Kirby, J. (2002). Differences and similarities in information seeking: children and adults as Web users. *Information processing & management*, *38*(5), 649-670.

Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, *56*(18), 3825-3833.

Canon, C. (2011). SEO After the Panda Update: How Google Rewards Original, Relevant Online Content.

Cothey, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology*, *55*(14), 1228-1238.

Dinev, T., & Hart, P. (2005). Internet privacy concerns and social awareness as determinants of intention to transact. *International Journal of Electronic Commerce*, *10*(2), 7-29.

Dover, D., & Dafforn, E. (2011). *Search engine optimization (SEO) secrets*. Wiley publishing.

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), 4512-4521.

Fagin, R., Kumar, R., & Sivakumar, D. (2003). Comparing top k lists. *SIAM Journal on Discrete Mathematics*, *17*(1), 134-160.

Fillmore, M. (2012). Google Penguin Update Continues Assault on Black Hat SEO.

Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. *Digital media, youth, and credibility*, 5-27.

Fleischner, M. H. (2009). *SEO made simple: Strategies for dominating the world's largest search engine*. SEO Made Simple.

Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism and Mass Communication Quarterly*, *63*(3), 451.

Geraci, A., Katki, F., McMonegal, L., Meyer, B., Lane, J., Wilson, P., ... & Springsteel, F. (1991). *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*. IEEE Press.

Gillespie, T. (2014). 9 The Relevance of Algorithms. *Media technologies: Essays on communication, materiality, and society*, 167.

Google Trends. (2015). Consulted on 22th of April 2016. http://www.google.com/trends/topcharts?hl=nl#date=2015&geo

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web*, 527-538.

Halavais, A. (2013). *Search engine society*. John Wiley & Sons.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003). Googlearchy: How a few heavily-linked sites dominate politics on the web. *Inannual meeting of the Midwest Political Science Association* (pp. 1-33).

Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, *44*(4), 1467-1484.

Höchstötter, N., & Lewandowski, D. (2009). What users see–Structures in search engine results pages. *Information Sciences*, *179*(12), 1796-1812.

Jaccard, J., & Becker, M. A. (2002). *Statistics for the behavioral sciences*. Wadsworth Publishing Company.

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, *42*(1), 248-263.

Jones, K. B. (2008). *Search Engine Optimization: Your visual blueprint for effective Internet marketing*. John Wiley & Sons.

Kang, M. (2010). Measuring social media credibility: A study on a Measure of Blog Credibility. *Institute for Public Relations*, 59-68.

Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3819-3828). ACM.

Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication & Society*, *4*(4), 381-403.

Kohring, M., & Matthes, J. (2007). Trust in news media development and validation of a multidimensional scale. *Communication research*, *34*(2), 231-252.

Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press.

Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, *400*(6740), 107-107.

Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In *Proceedings of the 2011 iConference* (pp. 452-458). ACM.

Lin, C. O. Y., & Yazdanifard, R. (2014). How Google's new algorithm, Hummingbird, promotes content and inbound marketing. *American Journal of Industrial and Business Management*, *2014*.

Ledford, J. L. (2015). *Search Engine Optimization Bible*. John Wiley & Sons.

Loiacono, E. T., Watson, R. T., & Goodhue, D. L. (2002). WebQual: A measure of website quality. *Marketing theory and applications*, *13*(3), 432-438.

Luca, M., Wu, T., Couvidat, S., Frank, D., & Seltzer, W. (2015). Does Google Content Degrade Google Search? *Experimental Evidence. Experimental Evidence. Harvard Business School*, 16-035.

Marwick, A. (2013). Online identity. Hartley, J. Burgess, J. & Bruns, A.(eds), A.

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication yearbook*, *27*, 293-336.

Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism & Mass Communication Quarterly*, *65*(3), 567-574.

Mizzaro, S. (1998). How many relevances in information retrieval?. *Interacting with computers*, *10*(3), 303-320.

Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web* (pp. 677-686). ACM.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer‐Mediated Communication*, *12*(3), 801-823.

Papacharissi, Z., 2002. The presentation of self in virtual life: Characteristics of personal home pages. *Journalism and Mass Communication Quarterly*, *79*(3), pp.643–660.

Papacharissi, Z. (Ed.). (2010). A networked self: Identity, community, and culture on social network sites. Routledge.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). ES-QUAL a multiple-item scale for assessing electronic service quality. *Journal of service research*, *7*(3), 213-233.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Petry, N. M. (2002). A comparison of young, middle-aged, and older adult treatment-seeking pathological gamblers. *The Gerontologist*, *42*(1), 92-99.

Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Purcell, K., Brenner, J., & Rainie, L. (2012). Search engine use 2012.

Richardson, M., & Domingos, P. M. (2001). The Intelligent surfer: Probabilistic Combination of Link and Content Information in PageRank. In *NIPS* (pp. 1441-1448).

Rogers, I. (2002). The Google Pagerank algorithm and how it works.

Rogers, R. (2013). *Digital methods.* MIT press.

Ran, S. (2003). A model for web services discovery with QoS. *ACM Sigecom exchanges*, *4*(1), 1-10.

Rui, J., & Stefanone, M. A. (2013). Strategic self-presentation online: A cross-cultural study. *Computers in Human Behavior*, *29*(1), 110-118.

Self, C. C. (1996). Credibility. *An integrated approach to communication theory and research*, *1*, 421-441.

Steiner, P. (1993, July 5). On the Internet, nobody knows you're a dog. The New York Times.

Taylor, A. (2012). User relevance criteria choices and the information search process. *Information Processing & Management*, *48*(1), 136-153.

Thorson, K., Vraga, E., & Ekdale, B. (2010). Credibility in context: How uncivil online commentary affects news credibility. *Mass Communication and Society*, *13*(3), 289-313.

Van Couvering, E. (2010). Search engine bias: The structuration of traffic on the World-Wide Web (Doctoral dissertation, The London School of Economics and Political Science (LSE)

Vaughan, L., & Wu, G. (2004). Links to commercial websites as a source of business information. *Scientometrics*, *60*(3), 487-496.

Wilson, P. (1983). Second-hand knowledge: An inquiry into cognitive authority.

Yoo, B., & Donthu, N. (2001). Developing a scale to measure the perceived quality of an Internet shopping site (SITEQUAL). *Quarterly journal of electronic commerce*, *2*(1), 31-45.

Zowghi, D., & Gervasi, V. (2003). On the interplay between consistency, completeness, and correctness in requirements evolution. *Information and Software technology*, *45*(14), 993-1009.

Appendix A

Part I

What is your gender?	- Male - Female
What is your age?	
What is your highest level of education?	- Undergraduate - Graduate - Completed master - Completed PHD

Part II

1. **How much do you use Google?**
   Rarely Occasionally Sometimes Often Very often
2. **How much do you use DuckDuckGo?**
   Rarely Occasionally Sometimes Often Very often
3. **Do you perceive Google as trustworthy? (i.e., honest, disinclined to deceive)**
   Untrustworthy A Little A fair amount Quite a bit Very trustworthy
4. **Do you perceive DuckDuckGo as trustworthy? (i.e., honest, disinclined to deceive)**
   Untrustworthy A Little A fair amount Quite a bit Very trustworthy
5. **Do you perceive Google as reliable (i.e., perform without failure)?**
   Unreliable A Little A fair amount Quite a bit Very reliable
6. **Do you perceive DuckDuckGo as reliable (i.e., perform without failure)?**
   Unreliable A Little A fair amount Quite a bit Very reliable
7. **How relevant are the search rankings presented by Google to the search term inserted (i.e., useful for the fulfillment of goals, interests, work tasks, or problematic situations)?**
   Irrelevant A Little A fair amount Quite a bit Very relevant
8. **How relevant are the search rankings presented by DuckDuckGo to the search term inserted (i.e., useful for the fulfillment of goals, interests, work tasks, or problematic situations)?**
   Irrelevant A Little A fair amount Quite a bit Very relevant
9. **Do you perceive Google's search rankings as accurate (i.e., exact, correct)?**
   Inaccurate A Little A fair amount Quite a bit Very accurate
10. **Do you perceive DuckDuckGo's search rankings as accurate (i.e., exact, correct)?**
    Inaccurate A Little A fair amount Quite a bit Very accurate
11. **Do you perceive Google's search rankings as fair (i.e., without opinions, judgments)?**
    Opinionated A Little A fair amount Quite a bit Factual
12. **Do you perceive DuckDuckGo's search rankings as fair (i.e., without opinions, judgments)?**
    Opinionated A Little A fair amount Quite a bit Factual

Appendix B

Part III Quality assessment

Constructs	*Description of Concept*	*Source*
Up-to-dateness	The extent it reflects the latest information or changes	Parasuraman *et al.*, 2005
Customization/personalization	How much and how easily search engine can be tailored to individual customers' preferences, histories, and ways of shopping	Parasuraman *et al.*, 2005
Quality of search rankings	The extent to which the search engine promises are fulfilled.	Parasuraman *et al.*, 2005
Instant search	The ease and speed of accessing and using the search engine	Loiacono *et al.*, 2002 Parasuraman *et al.*, 2005
Privacy	The degree to which the search engine is safe and protects customer information.	Loiacono *et al.*, 2002 Yoo & Donthu, 2001
Response Time	Time to get a response after entering a search query or an interaction with a search engine	Parasuraman *et al.*, 2005

Appendix C

Descriptive statistics of the demographics of the sample; SPSS outcome

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.
Age	46	18,00	31,00	22.956	2.650
Educational level	46	1	3	1.89	0.737
Gender	46	1	2	1.50	.506
Valid N (listwise)	46				

Table 4.1: Descriptive statistics of the demographics of the sample; [1] age, [2] educational level, and [3] gender; SPSS outcome

Appendix D

Regression analysis for the control variables

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.558[a]	.312	.223	.38451	.312	3.524	9	70	.001
2	.703[b]	.495	.422	.33183	.183	24.988	1	69	.000

a. Predictors: (Constant), Usage frequency, Educational level, Gender, Age
b. Predictors: (Constant), Usage frequency, Educational level, Gender, Age, Personalization

Table 4.13: Model Summary of the multiple regression analysis to control for external effects for [1] age, [2] gender, [3] usage frequency, and [4] educational level; SPSS outcome

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.689	8	.521	3.524	.001[b]
	Residual	10.349	71	.148		
	Total	15.038	79			
2	Regression	7.440	9	.744	6.757	.000[c]
	Residual	7.598	70	.110		
	Total	15.038	91			

a. Dependent Variable: Average_credibility score
b. Predictors: (Constant), Usage frequency, Educational level, Gender, Age
c. Predictors: (Constant), Usage frequency, Educational level, Gender, Age, Personalization

Table 4.14: Anova table of the multiple regression analysis to control for external effects for [1] age, [2] gender, [3] usage frequency, and [4] educational level; SPSS outcome

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.803	571		8.416	.000
	Educational d1 graduate	-.078	.019	-.470	-4.038	.000
	Educational d2 Master	-.047	.136	-.049	-.347	.730
	Age	-.167	.117	-.193	-1.422	.159
	Gender d1	.089	.096	.103	.930	.355
	Usage d1	.170	.133	.164	1.282	.204
	Usage d2	.138	.122	.144	1.134	.261
	Usage d3	.001	.168	.000	.004	.997
	Usage d4	.195	.144	.166	1.358	.179
2	(Constant)	5.640	.520		10.849	.000
	Educational d1 graduate	-.085	.017	-.514	-5.100	000
	Educational d2 Master	-.028	.118	-.029	-.236	.814
	Age	-.191	.101	-.221	-1.887	.063
	Gender d1	.065	.083	.075	.782	.437
	Usage d1	-.426	.165	-.410	-2.586	.012
	Usage d2	-.379	.147	-.395	-2.572	.012
	Usage d3	-.174	.149	-.121	-1.169	.246
	Usage d4	.087	.126	.074	.687	.494
	Personalization	-.614	.122	-.709	-5.032	.000

a. Dependent Variable: Average_combined

Table 4.15: Coefficient table of the multiple regression analysis to control for external effects for [1] age, [2] gender, [3] usage frequency, and [4] educational level; SPSS outcome

Appendix E

Spearman Correlation for variable age

Correlation age (Spearman Rho)

Independent variable	Dependent variable	df	r_s	Sig.(2-tailed)	Confidence int.	
Age	Trustworthiness Google	45	-.491	.001**	95%	Significant
Age	Trustworthiness DDG	45	-.666	.000**	95%	Significant
Age	Reliability Google	45	-.192	.202**	95%	Insignificant
Age	Reliability DDG	45	.106	.483**	95%	Insignificant
Age	Fairness Google	45	-.431	.003**	95%	Significant
Age	Fairness DDG	45	-.282	.058**	95%	Insignificant
Age	Relevance Google	45	-.672	.000**	95%	Significant
Age	Relevance DDG	45	.523	.000**	95%	Significant
Age	Accuracy Google	45	-.519	.000**	95%	Significant
Age	Accuracy DDG	45	.287	.053*	95%	Insignificant

Table 4.16: Correlations between age and the five credibility constructs [1] trustworthiness, [2] reliability, [3] fairness [4] relevance, and [5] accuracy; SPSS outcome