Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DEPARTMENT INFORMATION

Bachelorarbeit

Die Praxis der Relevanzbeurteilung von Google-Nutzern auf dem Prüfstand: eine Eye-Tracking-Studie

vorgelegt von

Sebastian Schultheiß

Studiengang: Bibliotheks- und Informationsmanagement

erster Prüfer: Prof. Dr. Dirk Lewandowski
zweiter Prüfer: Sebastian Sünkler, M.A.

Hamburg, August 2016

Abstract

Diese Bachelorarbeit geht der Frage nach, ob Suchmaschinennutzer der durch Google vorgegebenen Trefferreihenfolge mehr vertrauen als ihren eigenen Relevanzurteilen. Dafür wurde eine Studie aus dem Jahr 2007 repliziert. Kernbestandteil beider Arbeiten ist ein Eye-Tracking-Experiment. Dabei wurden den Probanden ohne deren Kenntnis Google-Ergebnisseiten in zum Teil manipulierter Form präsentiert, womit Rechercheaufgaben zu lösen waren. Währenddessen wurde das Blick- und Klickverhalten aufgezeichnet. Zusätzlich sollte die Relevanz der betrachteten Suchergebnisse bewertet werden. Alle erhobenen Daten wurden später statistisch ausgewertet.

Die US-amerikanischen Teilnehmer der zu replizierenden Studie wiesen ein großes Vertrauen Google gegenüber auf. Dies offenbarte sich darin, dass sie sich in ihrem Blick- und Klickverhalten eher von der Position als von der selbst bewerteten Relevanz eines Google-Ergebnisses leiten ließen. Im wiederholten Experiment dieser Abschlussarbeit standen Hamburger Studierende als Probanden zur Verfügung. Sie unterlagen aufgrund der manipulierten Ergebnislisten wie auch die Teilnehmer der Referenzarbeit einer gewissen Verunsicherung. Dies äußerte sich unter anderem durch ein längeres Verweilen auf den manipulierten Google-Ergebnisseiten. In ihrer Entscheidung, einen der Treffer anzuklicken, war jedoch der Einfluss der Treffer-Relevanz größer als der Rang innerhalb der Ergebnisliste. Daraus kann auf ein emanzipiertes Rechercheverhalten der Probanden des vorliegenden Experiments geschlossen werden. Dadurch wird die Forschungsfrage der vorliegenden Arbeit verneint, die lautet, ob sich die Erkenntnisse der Referenzstudie im zeitlichen und geografischen Kontext auf heutige, deutschsprachige Suchmaschinennutzer übertragen lassen.

Keywords

Suchmaschine, Google, SERP, Suchergebnisseite, Suchmaschinennutzer, Rechercheverhalten, Relevanzbeurteilung, Eye-Tracking, Experiment, Replikation

Inhaltsverzeichnis

Abstract ..... **II**

Inhaltsverzeichnis ..... **III**

Abbildungsverzeichnis ..... **V**

Tabellenverzeichnis ..... **VI**

1 Einleitung **1**

1.1 Motivation 2

1.2 Aufbau der Arbeit 3

2 Stand der Forschung **4**

2.1 Eye-Tracking-Studien mit Webseiten- oder Suchmaschinen-Bezug 4

2.2 Studien zum Rechercheverhalten der Suchmaschinennutzer 7

3 Forschungsfrage und Hypothesen **9**

3.1 Forschungsfrage 9

3.2 Hypothesen 9

4 Methodik **11**

4.1 Eye-Tracking-Methodik 11

4.2 Studiendesign des Eye-Tracking-Experiments 13

4.2.1 Zielsetzung 13

4.2.2 Versuchsdurchführung 14

4.2.3 Probanden 16

4.2.4 Rechercheaufgaben 17

4.2.5 Manipulation der Suchmaschinen-Ergebnisseite (SERP) 19

4.2.6 Explizite Relevanzbeurteilungen von Snippets und Webseiten 20

4.2.7 Technische Umsetzung 21

4.3 Auswertung der Eye-Tracking-Daten 24

5 Ergebnisse des Experiments **27**

5.1 Analyseebene ‘SERP’ 27

5.1.1 Betrachtungsdauer auf den SERPs 27

5.1.2 Anzahl der Fixationen auf den SERPs 30

5.1.3 Anzahl der angeklickten Snippets auf den SERPs 32

5.1.4 Vergleich der Ergebnisse aus 5.1 mit der Hypothese H1 33
5.2 Analyseebene ‘Snippet’ 34
5.2.1 Blick- und Klickverhalten auf den Snippets 35
5.2.2 Statistische Auswertungen zu den Fixationen auf den Snippets 38
5.2.3 Vergleich der Ergebnisse aus 5.2 mit den Hypothesen H2 und H3 39
5.3 Analyseebene ‘Relevanz’ 40
5.3.1 Statistische Auswertung der Relevanzurteile per gemischter Modelle 41
5.3.2 Vergleich der Ergebnisse aus 5.3 mit der Hypothese H4 43
5.4 Analyseebene ‘Vergleich’ 43
5.4.1 Vergleich der Ergebnisse aus 5.4 mit der Hypothese H5 46

6 Diskussion 47
6.1 Fazit 48
6.2 Grenzen der Studie 49
6.3 Zukünftige Forschungsvorhaben 50

7 Literaturverzeichnis 51

Anhang 1: Beigabe (CD) ..... A

Anhang 2: Suchergebnisseiten zu einer Rechercheaufgabe ..... B
Anhang 2.1 SERP “normal“ C
Anhang 2.2 SERP “vertauscht“ ..... D
Anhang 2.3 SERP “umgekehrt“ ..... E

Anhang 3: Dokumente zur Durchführung der Experimente ..... F
Anhang 3.1 Testleitfaden ..... F
Anhang 3.2 Datenschutzvereinbarung L
Anhang 3.3 Einverständniserklärung ..... M

Eidesstattliche Erklärung 55

Abbildungsverzeichnis

Abbildung 1: Goldenes Dreieck (Hotchkiss et al. 2005, S. 7) 5
Abbildung 2: SERP mit AOIs in Tobii Studio 25
Abbildung 3: Blick- und Klickverhalten SERP "normal" 35
Abbildung 4: Blick- und Klickverhalten SERP "vertauscht" 36
Abbildung 5: Blick- und Klickverhalten SERP "umgekehrt" 37
Abbildung 6: Blick- und Klickverhalten SERP "umgekehrt" (Pan et al. 2007, S. 814) 45
Abbildung 7: Beispiel SERP "normal" C
Abbildung 8: Beispiel SERP "vertauscht" ..... D
Abbildung 9: Beispiel SERP "umgekehrt" ..... E

Tabellenverzeichnis

Tabelle 1: Navigationsorientierte Suchanfragen des Eye-Tracking-Experiments 18
Tabelle 2: Informationsorientierte Suchanfragen des Eye-Tracking-Experiments 18
Tabelle 3: Deskriptive Statistiken zur Betrachtungsdauer einer SERP pro Rechercheaufgabe 28
Tabelle 4: Post-Hoc-Test nach Bonferroni: Betrachtungsdauer SERP 29
Tabelle 5: Deskriptive Statistiken zur Fixationsanzahl auf einer SERP pro Rechercheaufgabe 30
Tabelle 6: Test auf Homogenität der Varianzen 31
Tabelle 7: Post-Hoc-Test nach Games-Howell: Fixationsanzahl auf einer SERP pro Rechercheaufgabe 31
Tabelle 8: Deskriptive Statistiken zur Klickhäufigkeit auf Snippets pro Rechercheaufgabe 32
Tabelle 9: Post-Hoc-Test nach Games-Howell: Klickhäufigkeit auf Snippets pro Rechercheaufgabe 33
Tabelle 10: Lineares gemischtes Modell: Blickverhalten 42
Tabelle 11: Lineares gemischtes Modell: Klickverhalten 42

1 Einleitung

Suchmaschinen sind die heutzutage am häufigsten genutzten Anwendungen des Internets (Frees, Koch 2015, S. 372). Die Treffersortierung beeinflusst dabei entscheidend die Sichtbarkeit eines Unternehmens, die direkten Einfluss auf dessen Erfolg oder Misserfolg ausübt. Auch der Suchmaschinennutzer selbst steht ständig vor der Herausforderung, aus der immensen Menge an Webseiten die für die eigenen Bedürfnisse passendste herauszusuchen – es sei denn, er vertraut den Suchmaschinen. Dadurch tritt er seine individuelle Relevanzbeurteilung der Treffer an Konzerne wie Google ab und bevorzugt unabhängig von der eigentlichen Relevanz der präsentierten Webseiten tendenziell die erstplatzierten Ergebnisse.

Die Studie “In Google We Trust: Users’ Decisions on Rank, Position, and Relevance” von Pan, Hembrooke, Joachims, Lorigo, Gay und Granka (2007) kam zu ebensolchen Ergebnissen. Dafür untersuchten die Autoren, wie die Probanden im Klick- und Blickverhalten reagieren, wenn die ersten zehn Treffer der Google-Ergebnisliste nicht wie herkömmlich nach Relevanz sortiert, sondern komplett umgekehrt (*Treffer eins wird zu Treffer zehn usw.*) oder vertauscht (*Treffer eins mit Treffer zwei*) vorlagen. Dabei konnten sie ermitteln, dass die Probanden Google gegenüber ein großes Vertrauen aufwiesen. Dies äußerte sich darin, dass sie in der Entscheidung, ein Suchergebnis auszuwählen, die Position der Relevanz eines Treffers vorzogen.

Die gewonnenen Erkenntnisse sind nicht nur aufgrund des hohen Vertrauens der Nutzer in die Relevanzsortierung der Suchmaschinen als kritisch zu erachten. Denn auch bei irrelevanten Treffern zu Beginn der SERP (*Search Engine Results Page*) wird die Trefferqualität der manipulierten Ergebnisliste offenbar nicht ausreichend hinterfragt. SERPs sind von Suchmaschinen erstellte HTML-Seiten, die die Ergebnisse einer Suchanfrage darstellen (Lewandowski 2015, S. 125).

Ebenso führt solch ein Rechercheverhalten, bei dem die ohnehin schon populären und auf der SERP weit vorne stehenden Ergebnisse bevorzugt werden, zu einer weiteren Problematik.

Denn diese Einschränkung seitens der Suchmaschinennutzer schließt Dokumente aus, die zwar weniger populär sind, aber durchaus von Relevanz sein können. Ihr niedriger Rang nimmt solchen Treffern oftmals die Chance, in das Augenmerk potentieller Webseitenbesucher zu geraten.

Zielsetzung der vorliegenden Arbeit ist es, die Studie von Pan et al. (2007) zu replizieren. Anhand der Ergebnisse soll ermittelt werden, ob die Probanden beider Studien ähnliche Verhaltensmuster aufweisen.

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Gleichwohl gelten sämtliche Personenbezeichnungen für beiderlei Geschlecht.

1.1 Motivation

Die Motivation zum Verfassen dieser Arbeit fußt auf zwei Säulen, die als fachlich und persönlich bezeichnet werden können. In ihrer Verbindung begründen sie das Interesse an dem gewählten Thema. Im Sinne einer besseren Lesbarkeit wird dafür die Ichform verwendet.

Kurse und Projekte mit Suchmaschinen-Bezug innerhalb des Studiums ermöglichten mir den Einblick in diverse Felder, in denen die Treffersortierung und Relevanzbeurteilung eine Rolle spielen. So ist beispielsweise in der Suchmaschinenoptimierung die Ausgestaltung der dargestellten Treffer inklusive ansprechender Links und Treffer-Kurzbeschreibungen von großer Bedeutung. Denn im Allgemeinen überfliegen die Suchmaschinennutzer die Ergebnisliste lediglich und sind nur dann zum Klick auf einen Treffer bereit, wenn sie in kürzester Zeit Anhaltspunkte für die Nützlichkeit einer bestimmten Seite ausmachen können (Lewandowski 2012, S. 106).

Die Beurteilung der Trefferrelevanz ist auch in meiner privaten Suchmaschinennutzung allgegenwärtig. Als Anwender steht man vor der ständigen Herausforderung, aus einer kaum überschaubaren Ergebnismenge die nützlichste Quelle zu ermitteln. Auch ich verlasse mich dabei in der Regel auf die Relevanz der vordersten Treffer, ohne den unteren Suchergebnissen Beachtung zu schenken. Dabei finde ich mich zum Teil in den Ergebnissen von Pan et al. (2007) wieder.

Aufgrund meiner fachlichen und privaten Anknüpfungspunkte an das Thema habe ich an der Erforschung des Status quo der Relevanzbeurteilung ein großes persönliches Interesse.

1.2 Aufbau der Arbeit

Die Arbeit gliedert sich in fünf Teilbereiche. Nach der Einleitung folgt zuerst die Darstellung des Forschungsstandes. Dieser ist in seiner Untergliederung an die Studie von Pan et al. (2007) angelehnt und bildet inhaltlich einen anderen Zeitraum ab. Der Forschungsfrage und den Hypothesen folgt der Teilbereich zur Methodik. Darin wird zum einen die Eye-Tracking-Methodik beschrieben und zum anderen detailliert auf die Bestandteile des Experiments eingegangen. Abgeschlossen wird die Arbeit durch die letzten beiden Bereiche. So werden zunächst die Ergebnisse des Experiments analysiert. Dies findet anhand der vier Analyseebenen der Hypothesen statt. Zum Schluss folgt die Diskussion, die neben einem Fazit auch die Grenzen der Studie aufzeigen sowie zukünftige Forschungsvorhaben skizzieren soll.

2 Stand der Forschung

In der folgenden Darstellung des Forschungsstandes soll zunächst auf relevante Eye-Tracking-Studien eingegangen werden, die einen Bezug zu Webseiten oder Suchmaschinen aufweisen. Dem schließt sich der Abschnitt zu Arbeiten über das Rechercheverhalten von Suchmaschinennutzern an, in denen kein Gebrauch von der Eye-Tracking-Methode gemacht worden ist.

2.1 Eye-Tracking-Studien mit Webseiten- oder Suchmaschinen-Bezug

Die älteste recherchierte Eye-Tracking-Studie, die einen Bezug zu Webseiten herstellt, stammt aus dem Jahr 1998 (Ellis, Candrea, Misner, Craig, Lankford, Huchinson). Untersucht wurde, welche Einflüsse vier verschiedene Versionen einer Webseite auf das Rechercheverhalten der Probanden haben. Die Eye-Tracking-Methode wurde lediglich unterstützend für die Interpretation der Ergebnisse eingesetzt. Einen bedeutenderen Stellenwert hatte die Methode in der Studie von Pan, Hembrooke, Gay, Granka, Feusner und Newman (2004). Die Autoren erbrachten grundlegende Erkenntnisse über das Blickverhalten von Internetnutzern auf Webseiten, unter anderem auf den Ergebnisseiten von Google und Yahoo. So konnte ermittelt werden, dass es unter anderem vom Geschlecht des Probanden sowie vom Webseiten-Typ abhängig ist, welche Teile der Webseite wie lange angesehen werden.

Betrachtet man die veröffentlichten Eye-Tracking-Studien, die das Rechercheverhalten von Suchmaschinennutzern untersuchen, ist zunächst die Arbeit von Hotchkiss, Alston und Edwards (2005) zu nennen. Auf sie geht der Begriff des “Goldenen Dreiecks” (*golden triangle*) zurück. Dieser beschreibt die nutzerseitige Eigenschaft, die ersten Ergebnisse einer SERP zu bevorzugen, wobei jeder weiter unten aufgeführte Treffer weniger Aufmerksamkeit erhält als der darüberstehende. Die folgende Abbildung zeigt das “Goldene Dreieck“ in Form einer Heatmap (s. *Abbildung 1*). Heatmaps stellen Ergebnisse aus Eye-Tracking-Untersuchungen grafisch dar (Lewandowski 2015, S. 295). Je wärmer (röter) die angezeigte Farbe, desto öfter oder länger wurde ein bestimmter Bereich per Blick fixiert (Tobii Technology 2010, S. 8).

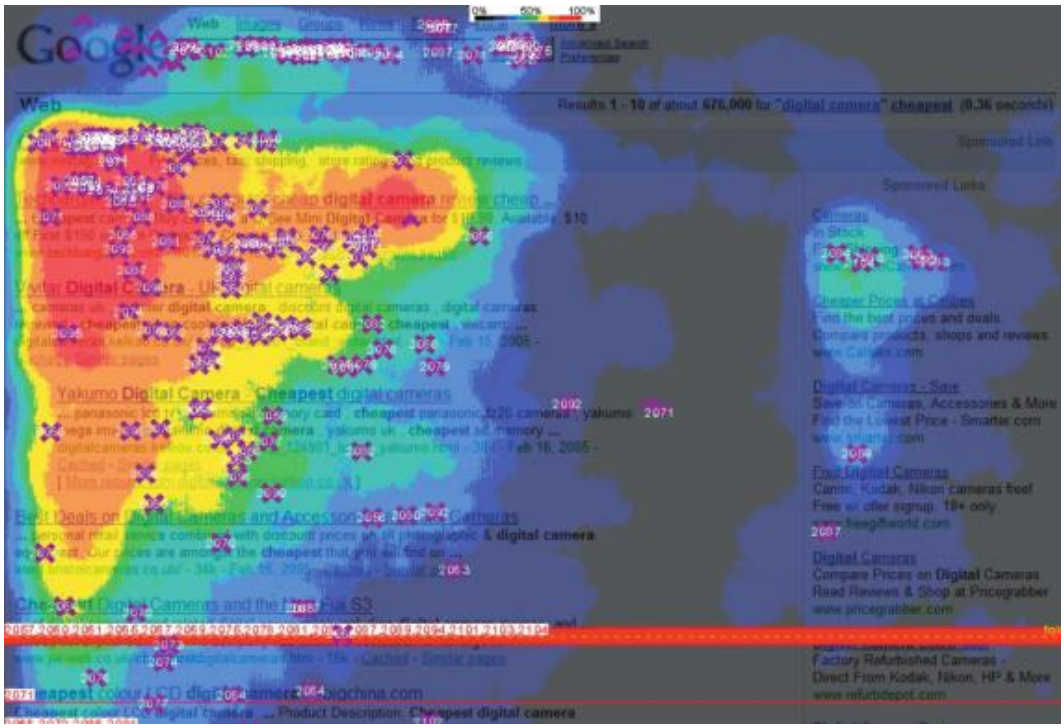

Abbildung 1: Goldenes Dreieck (Hotchkiss et al. 2005, S. 7)

Das Phänomen des “Goldenen Dreiecks“ gilt jedoch nur für nicht unterbrochene Listen (Lewandowski 2015, S. 131). Google bindet mittlerweile sogenannte Universal Search-Ergebnisse in die SERPs ein. Diese stammen aus Kollektionen wie Nachrichten, Bildern und Videos (Lewandowski 2015, S. 137). Dadurch verteilt sich die Aufmerksamkeit des Nutzers deutlich stärker auf der kompletten Ergebnisseite (Lewandowski 2015, S. 132; Usability.de 2009). Das “Goldene Dreieck” ist für die vorliegende Arbeit dennoch aus zweierlei Gründen relevant. Zum einen waren die SERPs in der zu wiederholenden Studie von Pan et al. (2007) auf die organischen Treffer reduziert. Dies ist im vorliegenden Experiment ebenso durchgeführt worden, sodass in beiden Fällen nicht unterbrochene Ergebnislisten vorliegen. Zum anderen ist das Vorkommen des “Goldenen Dreiecks” durch zahlreiche Studien bestätigt worden, welche in Auswahl nachfolgend aufgeführt werden.

Dem Autorenteam um Granka, Joachims und Gay zufolge war ihre Arbeit (2004) die erst zweite Studie, die sich per Eye-Tracking der Evaluation des Information Retrieval – in diesem Fall bei Google – annahm. Neben Nutzer-Verhaltensweisen, die schon bekannt waren, wie die erwähnte Bevorzugung der ersten Treffer, ermittelten die Autoren weitere nennenswerte Erkenntnisse.

So zeigten sie unter anderem Zusammenhänge zwischen der Betrachtungsdauer und der Auswahlhäufigkeit eines Ergebnisses auf. Die ersten beiden Treffer werden demnach zwar ähnlich lang betrachtet, erstgelegener Treffer jedoch deutlich häufiger ausgewählt.

Hier besteht ein direkter Zusammenhang zur vorliegenden Studie, in der die Kennzahlen “Betrachtungsdauer“ und “Anzahl der Klicks“ ebenfalls analysiert werden. Außerdem diente die Arbeit von Granka et al. (2004) als Grundlage der zu replizierenden Studie (Pan et al. 2007), in der die Praxis der Relevanzbeurteilung mittels manipulierter Ergebnislisten erforscht wurde.

Die Ergebnisse von Pan et al. (2007) offenbarten ein großes Vertrauen der Suchmaschinennutzer Google gegenüber. Dies konnte anhand des Blick- und Klickverhaltens der Teilnehmer ermittelt werden. Dazu führten die Autoren ein Eye-Tracking-Experiment durch, das in der vorliegenden Arbeit wiederholt und im weiteren Verlauf ausführlich beschrieben werden wird. Kernbestandteil der genannten Studie waren Rechercheaufgaben, die die Probanden zu lösen hatten. Dafür bekamen sie die Google-SERP entweder in normaler, in umgekehrter (*Treffer eins wird zu Treffer zehn usw.*) oder in vertauschter (*Treffer eins mit Treffer zwei*) Form präsentiert. Tendenziell bevorzugten die Teilnehmer dabei die erstplatzierten Ergebnisse. Das war auch bei denjenigen Probanden der Fall, die ausschließlich umgekehrte oder vertauschte Ergebnislisten präsentiert bekamen. Die Probanden hielten somit teilweise irrelevante Ergebnisse für relevant und wählten sie deshalb aus, weil diese einen der vordersten Ränge einnahmen.

Manipulierte Suchergebnisse setzten auch Cutrell und Guan (2007) in ihrem Experiment ein. Sie variierten die Länge der Beschreibungstexte und stellten fest, dass die Nutzer bei Recherchen nach Informationen durch längere Texte profitierten, nicht aber bei der Suche nach Orten. Dies drückt sich in der Geschwindigkeit aus, in der die Rechercheaufgaben absolviert worden sind.

Eine Gemeinsamkeit der aufgeführten Studien sind die durchgehend niedrigen Probandenzahlen. So akquirierten die Autoren zwischen 16 und 30 Teilnehmer für ihre Experimente. Außerdem wurden methodisch ähnlich vorgegangen. Die Eye-Tracking-Methode diente stets dazu, das Nutzerverhalten auf herkömmlichen oder manipulierten Suchergebnisseiten zu erforschen.

Die zu replizierende Studie von Pan et al. (2007) mit 16 studentischen Teilnehmern ist aus erstgenanntem Grund nicht repräsentativ und kann, wie auch die vorliegende Arbeit, keine Aussagen zur Gesamtheit der Suchmaschinennutzer treffen.

2.2 Studien zum Rechercheverhalten der Suchmaschinennutzer

Seit der Jahrtausendwende ist eine Vielzahl von Studien zum Rechercheverhalten von Suchmaschinennutzern veröffentlicht worden. Davon stützten sich einige auf sogenannte Transaktionsprotokolle bzw. “Clickthrough-Data“, die aus Datensätzen mit Suchanfragen, Ergebnis-Reihenfolgen sowie den ausgewählten Treffern zusammengesetzt sind. Diese Studien sind in ihrem Erkenntnisgewinn teilweise auf zeitliche und thematische Faktoren der Suche begrenzt und eher beschreibender Natur. Jansen, Spink und Saracevic (2000) analysierten Suchprotokolle der Suchmaschine Excite. Ozmutlu, Spink und Ozmutlu (2004) werteten vier Jahre später Excite- und Fast Web-Protokolle mit der Zielsetzung aus, Veränderungen des Rechercheverhaltens im Tagesablauf zu ermitteln. Bei Studien dieser Art bleiben wichtige Fragestellungen oftmals unbeantwortet. Es kann ermittelt werden, wann und wonach recherchiert worden ist. Darüber Hinausgehendes wie zum Kontext der Recherche oder zu Gründen, selbige abgebrochen zu haben, lässt sich nicht feststellen.

Aufschlussreicher ist die Studie von Keane, O’Brien und Smyth (2008), die zudem Parallelen zur vorliegenden Arbeit aufweist. Darin untersuchten die Autoren, inwieweit Suchmaschinennutzer von der Sortierung der Suchergebnisse beeinflusst werden. Dazu sollten die Teilnehmer 16 Rechercheaufgaben in zufälliger Reihenfolge durchführen. Die Suchergebnisse wurden entweder originalgetreu oder in umgekehrter Reihenfolge ausgegeben. Anhand des jeweils ersten Klicks eines Teilnehmers auf der SERP wurde festgestellt, wie die Probanden auf die beiden Szenarien reagieren. Feststellen ließ sich eine klare Bevorzugung der vordersten Treffer, auch wenn diese wie im Fall der manipulierten Ergebnislisten wenig relevant waren. Dahingegen wurden relevante, aber unterhalb aufgeführte Treffer erheblich seltener ausgewählt.

Die Autoren stellen neben der Theorie, dass ein solches Rechercheverhalten durch das nutzerseitige Vertrauen den Suchmaschinen gegenüber begründet sei, eine weitere mögliche Erklärung dar. Denn da nicht in allen Fällen ausschließlich die ersten Treffer ausgewählt wurden, streben die Suchmaschinennutzer den Ergebnissen der Studie zufolge eher nach zufriedenstellenden Ergebnissen.

Diese Annahme untermauern sie mit den Resultaten einer früheren Studie von O'Brien und Keane (2006). So findet ein letztplatzierter Treffer eher dann Beachtung, wenn ihm voran keine ähnlich relevanten Ergebnisse aufgelistet werden, die den Nutzer ebenso hätten zufriedenstellen können.

Da dennoch eine tendenzielle Bevorzugung der erstplatzierten Treffer vorliegt, sollten die Suchmaschinen den Verfassern zufolge dem sogenannten "rich-get-richer"-Effekt entgegenwirken. Denn sobald Suchergebnisse eine der vordersten Positionen einnehmen, erfahren sie durch das beschriebene Rechercheverhalten ein immer höheres Gewicht im Suchmaschinen-Ranking und untermauern dadurch mehr und mehr ihre populäre Position.

In einer weiteren Studie erforschten Bar-Ilan, Keenoy, Levene und Yaari (2009), ob Suchmaschinennutzer das eigentliche Ranking der Suchmaschine im Vergleich zu künstlich generierten Ergebnislisten bevorzugen. Dazu bekamen die Probanden Paare von Ergebnislisten zu zwölf Suchanfragen präsentiert, die aus zwei verschiedenen Versionen bestanden. Die erste Version stellte die originalen Reihenfolgen der Suchmaschinen Google, Yahoo oder Windows Live Search dar. Die andere Version bestand aus denselben Ergebnissen der jeweiligen Suchmaschine, jedoch in anderer Reihenfolge. Die Studie wies zwar nutzerseitig eine leichte Bevorzugung der originalen SERP nach, aber lediglich in sehr geringem Maße. Daraus resultiert die Annahme, dass der wichtigste Faktor für die Auswahl eines Suchergebnisses nicht dessen eigentliche Relevanz, sondern dessen Position auf der SERP ist.

Der Frage der Relevanz gingen auch Shani und Tractinski (2013) nach. Sie stellten dar, dass Suchmaschinennutzer eher gewillt seien, auf einen Treffer im unteren SERP-Bereich zu klicken, wenn an diesem eine Relevanzbewertung abgebildet wäre. Bei den Bewertungen handelt es sich um das suchmaschineneigene Relevanzurteil eines Treffers bezogen auf die getätigte Suchanfrage. Durch Zusatzinformationen wie diese werden Suchmaschinennutzer dazu angeregt, tendenziell eine größere Anzahl an Suchergebnissen zu betrachten. Der Gedanke, Suchergebnisse mit Relevanzbewertungen zu versehen und sie dadurch mehr in den Fokus des Anwenders zu rücken, wird im *Abschnitt 6.3 Zukünftige Forschungsvorhaben* nochmals aufgegriffen.

3 Forschungsfrage und Hypothesen

3.1 Forschungsfrage

Die Forschungsfrage lautet wie folgt:

„Lassen sich die Erkenntnisse aus der Studie von Pan et al. (2007) im zeitlichen und geografischen Kontext durch eine Replikation auf heutige, deutschsprachige Suchmaschinennutzer übertragen?“

Die Forschungsfrage ist so formuliert, dass neben der Frage der Relevanzbeurteilung auch die in der Einleitung genannten Überlegungen integriert wurden. Denn nicht nur die eigentliche Praxis der Relevanzbeurteilung stand im Fokus der Untersuchung. Auch sollte der Frage nachgegangen werden, ob es zeitlich und örtlich gesehen Unterschiede zu den Ergebnissen von Pan et al. (2007) geben würde.

3.2 Hypothesen

Die Hypothesen werden durch Analyseebenen gegliedert. H1 bis H3 wurden dabei aus der Studie von Pan et al. (2007, S. 811-812) übernommen. Inhaltlich wurden diese so belassen, weil es in der Literatur für eine Veränderung keine Anhaltspunkte gab. Hypothese H4 bestand so in der zu replizierenden Arbeit nicht, resultiert aber aus deren Ergebnissen bezüglich der Bevorzugung der Treffer-Position gegenüber dessen Relevanz. H5 dient dem Vergleich beider Untersuchungen.

- Analyseebene ‘SERP’

H1: Die Eye-Tracking-Daten werden in den drei Testszenarien (*normale, vertauschte und umgekehrte Trefferdarstellung*) voneinander abweichen. So werden die Probanden bei der umgekehrten Ergebnisdarstellung die SERP länger und öfter betrachten sowie deren Snippets häufiger anklicken, als dies bei der nicht manipulierten SERP der Fall sein wird. Als Snippets werden Trefferbeschreibungen auf SERPs bezeichnet, die jeweils aus anklickbarer Überschrift, URL sowie Beschreibungstext bestehen (Lewandowski 2015, S. 141).

- Analyseebene ‘Snippet‘

H2: In allen drei Testszenarien werden die Probanden den ersten beiden aufgelisteten Suchergebnissen nahezu gleichermaßen viel Aufmerksamkeit schenken, was sich durch die Analyse des Blickverhaltens belegen lassen wird. Dennoch werden die Probanden bei der umgekehrten Trefferdarstellung die letzten beiden Snippets (*die ersten beiden der normalen Darstellung*) länger betrachten, als in den anderen beiden Szenarien.

H3: Sowohl die Probanden der vertauschten (*Snippets eins und zwei miteinander vertauscht*) als auch die der umgekehrten Trefferdarstellung werden eher Snippets auswählen, die auf der normalen SERP niedriger gelistet werden und dort weniger Klicks erhalten.

- Analyseebene ‘Relevanz’

H4: Die Probanden werden in ihren Entscheidungen, Treffer auszuwählen, die Google-Positionen ihren eigenen Relevanzurteilen vorziehen. Dies wird sich statistisch dadurch ausdrücken, dass die Position einen stärkeren Einfluss auf das Blick- und Klickverhalten ausüben wird als die eigentliche Relevanz eines Treffers.

- Analyseebene ‘Vergleich‘

H5: Die Ergebnisse der vorliegenden Arbeit werden sich nicht grundlegend von denen der Studie von Pan et al. (2007) unterscheiden. Auch die Teilnehmer des vorliegenden Experiments werden ein großes Vertrauen Google gegenüber offenbaren.

4 Methodik

Nachfolgend soll zunächst die im Experiment angewendete Eye-Tracking-Methode erläutert werden. Dem wird sich die Darstellung des Studiendesigns der Untersuchung anschließen.

4.1 Eye-Tracking-Methodik

Beim Eye-Tracking handelt es sich um ein apparatives Verfahren zur Erfassung der Blickrichtung von Personen, beispielsweise von Untersuchungsteilnehmern. Die Ermittlung findet rezeptionsbegleitend statt, wobei als Rezeption im Falle der vorliegenden Arbeit die Betrachtung der Google-Ergebnisseiten bezeichnet werden kann (Blake 2013, S. 367).

Eye-Tracking findet in diversen Bereichen Anwendung, die in die Kategorien “diagnostisch” und “interaktiv” unterteilt werden können. Interaktive Eye-Tracking-Anwendungen verfolgen das Ziel, auf das Blickverhalten des Nutzers zu reagieren, beispielsweise als Zeigegerät am PC für körperlich beeinträchtigte Personen. Diagnostische Anwendungen dienen demgegenüber dem Zweck, anhand der Blickbewegungen Rückschlüsse auf die Aufmerksamkeit des Nutzers bei verschiedenen Stimuli feststellen zu können (Duchowski 2003, S. 131-132). So lassen sich Reihenfolge, Intensität und Schnelligkeit der Blickbewegungen erfassen, die ein Proband auf einem Element, beispielsweise auf einer Webseite (*Stimulus*) vollzieht (Quirmbach 2011, S. 247). Auch kommen diagnostische Verfahren häufig in den Neurowissenschaften wie der Psychologie, in der Informatik sowie in weiteren Disziplinen zum Einsatz (Duchowski 2003, S. 170).

Zur Erfassung der Blickdaten wird auf technischer Seite unter anderem die sogenannte Pupil Centre Corneal Reflection (*PCCR*)-Technik verwendet, die auch bei der vorliegenden Untersuchung zum Einsatz kam. Dafür gelangen zunächst auf beide Augen Infrarotstrahlen, deren Reflexionen durch eine im Eye-Tracker verbaute Kamera registriert werden. Im Anschluss werden die Reflexionen der Infrarotstrahlen auf Hornhaut (*Cornea*) und Pupillen herangezogen, um aus dem zwischen beiden Strahlen bestehenden Winkel die Blickrichtung berechnen zu können.

Die zwei Referenzwerte werden benötigt, da im Falle einer bloßen Auswertung der Pupillen-Reflexionen nicht festgestellt werden könnte, ob eine Positionsveränderung durch die Augen oder durch sonstige Körperbewegungen wie mit dem Kopf verursacht worden ist (Duchowski 2003, S. 60; Tobii Technology 2010, S. 6). Auf den im Experiment verwendeten Eye-Tracker sowie dessen Infrarottechnik wird im *Abschnitt 4.2.7 Technische Umsetzung* eingegangen.

Blickbewegungen werden in Sakkaden und Fixationen unterschieden. Sakkaden sind schnelle Augenbewegungen, die eine Dauer zwischen zehn und 100 Millisekunden haben können (Duchowski 2003, S. 44). Eine Sakkade wiederum kann ebenfalls in zwei Arten unterschieden werden. Sie kann einerseits reflexartig durch plötzlich auftretende Veränderungen innerhalb des betrachteten Bereichs auftreten. Andererseits können Sakkaden dadurch hervorgerufen werden, dass der Betrachter bewusst ein zunächst eher oberflächlich betrachtetes Objekt näher inspizieren möchte (Godijn, Theeuwes 2003, S. 3).

Um eine Informationsaufnahme zu ermöglichen, muss das Auge hingegen relativ bewegungslos sein. Diese Phasen des Stillstands im Blickverhalten werden als Fixationen bezeichnet. Dabei wird relativ stabil ein bestimmtes Objekt fixiert, was wiederum durch den nächsten Blickwechsel per Sakkade unterbrochen wird (Duchowski 2003, S. 48). Über die Dauer, wie lange sich die Augen im ruhigen Zustand befinden müssen, sodass von einer Fixation gesprochen werden kann, herrscht Uneinigkeit. Duchowski (2003, S. 49) nennt hierfür einen Mindestwert von 150 Millisekunden. Aus einer Übersicht von Hofer und Mayerhofer (2010, S. 153) geht hervor, dass 19 Studien zwischen 1975 und 2008 einen Mindestwert zwischen 50 und 240 Millisekunden für eine Fixation angesetzt haben.

Diese Uneinigkeit über die Fixationsdauer ist als sehr problematisch einzuschätzen. Die Fixationsdauer ist für Eye-Tracking-Untersuchungen ein Basismaß, das einen großen Einfluss auf die Ergebnisse ausübt. Studien, die auf den genannten unterschiedlichen Definitionen der Fixationsdauer fußen, sind nur sehr eingeschränkt miteinander vergleichbar (Blake 2013, S. 371). Im *Abschnitt 4.3 Auswertung der Eye-Tracking-Daten* wird die in der vorliegenden Arbeit verwendete Mindestdauer einer Fixation genannt und begründet.

Die Eye-Tracking-Methode verfügt über Schwächen bzw. Nachteile, die es bei derartigen Vorhaben zu beachten gilt. So ist der Zeitaufwand für die Rekrutierung und Betreuung der Probanden sowie für die Auswertung des umfangreich anfallenden Datenmaterials sehr hoch. Die benötigte Hard- und Software ist zudem sehr kostenintensiv (Quirmbach 2011, S. 250). Aufgrund dieser beiden Schwächen und der damit oftmals verbundenen geringen Probandenzahlen wurden diverse Studien veröffentlicht, die alternative Vorgehensweisen erproben. So bestehen direkte Zusammenhänge zwischen Mausbewegungen bzw. -klicks und dem Blickverhalten. Dadurch soll ermöglicht werden, Untersuchungen mit deutlich mehr Probanden als bei herkömmlichen Eye-Tracking-Studien, jedoch mit ähnlicher Aussagekraft durchzuführen (Guo, Agichtein 2010; Huang, White, Dumais 2011).

Als weiterer Nachteil der Eye-Tracking-Methode ist die Laborsituation bei den Untersuchungen zu nennen. Trotz angemessener Rahmenbedingungen wie der für die Probanden im Optimalfall kaum bemerkbaren Eye-Tracking-Technik verhalten sich die Untersuchungsteilnehmer in der Laborsituation nur selten genauso, wie sie es in der privaten Nutzung tun würden (Höchstötter 2007, S. 137). Dennoch trägt die Methode dazu bei, Fragestellungen zu erforschen, die ohne Eye-Tracking überhaupt nicht oder nur ungenauer behandelt werden könnten (Blake 2013, S. 384).

4.2 Studiendesign des Eye-Tracking-Experiments

Im Folgenden werden Ablauf und Details der Laboruntersuchung beschrieben. Begonnen wird mit der Zielsetzung, an die sich die Darstellung der Versuchsdurchführung anschließt. Die weiteren Abschnitte beleuchten die einzelnen Bestandteile der Untersuchung, welche mit der technischen Umsetzung abgeschlossen werden.

4.2.1 Zielsetzung

Wie in der Einleitung erwähnt verfolgt diese Arbeit die Zielsetzung, die Studie von Pan et al. (2007) zu replizieren. Die Ergebnisse beider Studien sollen miteinander verglichen werden, um Gemeinsamkeiten und Unterschiede im Rechercheverhalten der Probanden herauszustellen.

Die Wiederholung der Studie empfiehlt sich insbesondere aufgrund geografischer und zeitlicher Faktoren. So ist das ursprüngliche Experiment mit US-amerikanischen Studierenden durchgeführt worden. Diese bringen amerikanischen Konzernen wie Google vermutlich ein stärkeres Vertrauen entgegen als deutsche Studierende. Dieses Vertrauen könnte mitursächlich dafür gewesen sein, dass die Teilnehmer des Experiments von Pan et al. (2007) die durch Google zuerst platzierten Treffer präferiert haben.

Des Weiteren ist das Alter der zu replizierenden Studie hervorzuheben, welches zum Zeitpunkt der vorliegenden Arbeit knappe zehn Jahre betrug. Junge Menschen (*Digital Natives*) gelten heutzutage als sehr geschult im Umgang mit dem Internet. Der Begriff “Digital Native“ geht auf Prensky (2001) zurück und bezeichnet eine Person, die in das digitale Zeitalter hineingeboren wurde, wodurch sie mit digitalen Medien ähnlich selbstverständlich umgeht wie mit ihrer Muttersprache.

Die Teilnehmer beider Studien können als “Digital Natives“ bezeichnet werden. Beide Gruppen sind jedoch in unterschiedlich stark digitalisierten Welten aufgewachsen. Die Verheißung, dass daraus ein bewussterer Umgang mit Suchmaschinen resultiert, ist durch die Literatur jedoch nicht belegbar. Demzufolge soll die Studie auch Erkenntnisse darüber liefern, ob Studierende mittlerweile tatsächlich qualifizierter mit Suchmaschinen umgehen oder ob die Probanden wie in der Untersuchung von Pan et al. (2007) agieren.

4.2.2 Versuchsdurchführung

Der Ablauf des Versuchs entsprach in weiten Teilen dem aus der Arbeit von Pan et al. (2007). Auf die Unterschiedlichkeiten wird in den weiteren Abschnitten jeweils eingegangen. Zunächst soll die Durchführung des Versuchs in groben Zügen dargestellt werden.

Mit jedem der insgesamt 28 eingeladenen Probanden wurde ein individueller Termin für die Durchführung des Experiments vereinbart. Dafür war der Zeitraum vom 18.04.16 bis zum 29.04.16 vorgesehen, für den das Usability-Labor des Departments Information an der Hochschule für Angewandte Wissenschaften Hamburg zur Verfügung stand. Nach der Begrüßung folgte eine kurze Einführung, die den Ablauf der Untersuchung verdeutlichen sollte (*s. Anhang 3.1*). So wurde jeder Proband angewiesen, die Rechercheaufgaben genauso zu absolvieren, wie er es auch in einer privaten Nutzungssituation tun würde.

Getränke und Snacks sollten zu einer angenehmen Atmosphäre und zu einer Abmilderung der Versuchssituation beitragen. Den Teilnehmern wurde als voraussichtliche Dauer eines Durchlaufs ein Zeitraum zwischen 45 Minuten und einer Stunde genannt, wofür sie mit jeweils zehn Euro vergütet worden sind. Nach der Einführung wurden Unterschriften auf der Datenschutzvereinbarung (s. *Anhang 3.2*), der Einverständniserklärung (s. *Anhang 3.3*) sowie auf der Quittung für die Vergütung eingeholt.

Der dann folgenden Kalibrierung des Eye-Trackers schloss sich der Hauptteil des Experiments, die Absolvierung der in zufälliger Reihenfolge erschienenen Rechercheaufgaben, an. Jeder Teilnehmer bekam zehn Aufgaben präsentiert, die er anhand einer vorgegebenen SERP und den entsprechenden Webseiteninhalten der ersten zehn Treffer zu lösen hatte. Eigene Suchanfragen waren nicht möglich. Die SERPs lagen entweder in normaler, vertauschter oder umgekehrter Form vor. Auf die Manipulation der Suchergebnisseiten, welche den Probanden nicht bekannt war, wird später eingegangen.

Es bestand prinzipiell die Möglichkeit, alle Treffer einzusehen, wobei das Zeitlimit von drei Minuten pro Aufgabe nicht überschritten werden durfte. Die Zeitvorgabe ist ein wichtiger Bestandteil des Experiments, da sie eine Vergleichbarkeit der Teilnehmer untereinander herstellt. Erhielten einige Probanden mehr Zeit als andere, könnten diese unterschiedlichen Bedingungen ursächlich für ein abweichendes Verhalten, wie zum Beispiel für eine längere Betrachtungsdauer der Suchergebnisse, sein. Um solche Alternativerklärungen einzuschränken, sind mittels Kontrolltechniken möglichst identische Bedingungen bei Experimenten herzustellen (Sedlmeier, Renkewitz 2008, S. 130).

Wurde die korrekte Antwort gefunden oder das Zeitlimit erreicht, gelangte jeder Proband zur Relevanzbewertung. Diese erfolgte nach jeder einzelnen Rechercheaufgabe. Bewertet werden sollten die Relevanz der zuvor angezeigten Snippets sowie die der dazugehörigen Webseiteninhalte.

Während der gesamten Untersuchungsdauer saß der Testleiter links hinter dem Teilnehmer, sodass bei technischen Problemfällen oder Verständnisfragen zum Ablauf umgehend eingegriffen werden konnte.

Nachdem die Rechercheaufgaben inklusive Relevanzbeurteilungen aller Snippets und Webseiten absolviert worden waren, war der Durchlauf beendet.

4.2.3 Probanden

Über eine per Verteiler versendete E-Mail konnten insgesamt 28 Studierende akquiriert werden. Die E-Mail wurde zwei Mal mit dem Abstand von einer Woche versandt. 23 der Interessenten kamen aus den Bachelorstudiengängen Bibliotheks- und Informationsmanagement oder Medien und Information bzw. aus dem Masterstudiengang Information, Medien, Bibliothek. Fünf weitere Teilnehmer studierten entweder den Bachelorstudiengang Media Systems, Medientechnik oder Biologie bzw. den Masterstudiengang Sound/Vision. Die 28 Probanden wiesen ein Durchschnittsalter von 25,5 Jahren auf.

Den Probanden wurden durch das Tool, das im Abschnitt der technischen Umsetzung beschrieben wird, jeweils zufällig eine der drei Bedingungen (*normal, vertauscht, umgekehrt*) zugewiesen. Ein Teilnehmer wurde somit ausschließlich mit einer SERP-Variante konfrontiert. Die zufällige Verteilung erfolgte im Sinne der Ausbalancierung personengebundener Störvariablen, die nicht bzw. kaum kontrolliert werden können. So könnte die Intelligenz der Probanden das Rechercheverhalten beeinflussen. Diese jedoch vorab zu messen und die Teilnehmer dementsprechend zu verteilen, würde den Umfang der Arbeit übersteigen. Durch die unsystematische Einteilung per Zufall wird hingegen ohne entscheidenden Mehraufwand eine ausgewogene Aufteilung der Probanden auf die drei Versuchsbedingungen erreicht (Sedlmeier, Renkewitz 2008, S. 131-134).

Diese als Randomisierung bezeichnete Form der Verteilung sollte in der Regel bei einer großen Probandenanzahl eingesetzt werden. Die Teilnehmer der Untersuchung stellen jedoch eine homogene Population dar. So sind sie alle deutschsprachige Studierende ähnlichen Alters. Diese Homogenität klammert Störvariablen wie zu große Altersunterschiede aus und ermöglicht auch bei Probandenzahlen wie die im vorliegenden Experiment eine Randomisierung (Sedlmeier, Renkewitz 2008, S. 137).

Sinn und Zweck der Untersuchung wurden den Probanden weitestgehend vorenthalten. Dadurch sollte das Aufkommen sogenannter Demand Characteristics gering gehalten werden. Darunter werden Anforderungen des Experiments verstanden, die der Proband subjektiv empfindet. Das wiederum kann dazu führen, dass ein Proband sein Verhalten während des Experiments den empfundenen Anforderungen gemäß anpasst (Sedlmeier, Renkewitz 2008, S. 142-143).

Im Sinne der Kontrolle solcher Störvariablen wurden jedem Teilnehmer somit dieselben, oberflächlich gehaltenen Informationen vorab vermittelt (*s. Anhang 3.1*).

4.2.4 Rechercheaufgaben

In der Studie von Pan et al. (2007) wurden den Probanden zehn Rechercheaufgaben gestellt, jeweils zur Hälfte navigations- bzw. informationsorientiert.

Die Unterscheidung in informations-, navigations- und transaktionsorientierte Suchanfragen erfolgt nach Broder (2002) und zielt auf das Informationsbedürfnis hinter einer bestimmten Suchanfrage ab. Aufgrund ihres Einsatzes im Experiment sollen im Folgenden die ersten beiden Typen kurz beschrieben werden.

Navigationsorientierte Suchanfragen verfolgen die Zielsetzung, eine bestimmte Webseite zu finden, die dem Nutzer bekannt ist oder von der er annimmt, dass diese existiert. In der Regel wird der Nutzer bei solchen Anfragen mit einem bestimmten Ergebnis zufriedengestellt, anders als bei den informationsorientierten Suchanfragen. Bei diesen gibt es nicht nur einen richtigen Treffer, der gefunden werden soll. Vielmehr steht die Informationsrecherche über ein bestimmtes Thema im Vordergrund (Lewandowski 2015, S. 70).

In der Studie von Pan et al. (2007) wurden die Aufgaben laut vorgelesen, damit die Probanden ihren Blick nicht vom Bildschirm lösen und auf ein Aufgabenblatt richten mussten. Es wurde die originale Google-Oberfläche eingesetzt, auf der keine Fragen eingeblendet werden konnten. Da in der vorliegenden Studie jedoch ein Tool zur Verfügung stand, das die zu lösenden Aufgaben direkt auf dem Bildschirm präsentierte, konnte das Vorlesen derselbigen entfallen.

Die Fragestellungen aus der Studie von Pan et al. (2007) wurden nicht übernommen. Dies wird dadurch begründet, dass sich die Rechercheaufgaben teilweise stark auf regionale Themen bezogen. Neu erarbeitete Fragen mit Hamburg- und/oder Deutschland-Bezug sollten einen ähnlichen Kontext zu den Probanden herstellen. Es folgt eine Übersicht zur Verdeutlichung der für dieses Experiment gewählten Aufgabenstellungen (s. *Tabelle 1* und *Tabelle 2*). Auf die vorformulierten Suchanfragen (*Queries*) wird im weiteren Verlauf eingegangen.

Tabelle 1: Navigationsorientierte Suchanfragen des Eye-Tracking-Experiments

	I) Navigationsorientierte Suchanfragen			
	Aufgabenstellung der Studie von Pan et al. (2007, S. 808)	**Gegenstück mit deutschem oder Hamburger Kontext**	**Suchanfrage (Query)**	**Richtiger Treffer (wenn mehrere möglich, dann *u.a.*)**
01	Find the homepage of Michael Jordan, the statistician.	Finde die Startseite der Homepage vom Schulz von Thun Institut für Kommunikation.	friedemann schulz von thun institut	*http://www.schulz-von-thun.de/*
02	Find the page displaying the route map for Greyhound buses.	Finde eine Übersicht über verschiedene Anbieter von Hafenrundfahrten in Hamburg.	hafenrundfahrt hamburg übersicht	*http://www.hamburg.de/hafenrundfahrt/868760/hafenrundfahrten-hamburg/ (u.a.)*
03	Find the homepage of the 1000 Acres Dude Ranch.	Finde die Homepage des Tierparks Hagenbeck.	hagenbeck tierpark	*http://www.hagenbeck.de/startseite.html*
04	Find the homepage for graduate housing at Carnegie Mellon University.	Finde eine Webseite, die das Studentenwohnheim Hammerbrook überblicksartig darstellt, inklusive Angaben zu Miete und Kontakt.	wohnheim hamburg hammerbrook	*http://www.studierendenwerk-hamburg.de/studierendenwerk/de/wohnen/wohnanlagen/detail/?id=766*
05	Find the homepage of Emeril - the chef who has a television cooking program.	Finde die Webseite des bekannten Kochs Christian Rach.	christian rach	*http://www.christianrach.de/ (u.a.)*

Tabelle 2: Informationsorientierte Suchanfragen des Eye-Tracking-Experiments

	II) Informationsorientierte Suchanfragen			
	Rechercheaufgaben der Studie von Pan et al. (2007, S. 808)	**Gegenstück mit deutschem oder Hamburger Kontext**	**Suchanfrage (Query)**	**Richtiger Treffer & korrekte Antwort**
06	Where is the tallest mountain in New York located?	Welches Gründungsjahr und welchen Firmensitz hat die älteste Schuhfabrik Deutschlands?	älteste schuhfabrik deutschland gründungsjahr firmensitz	*https://de.wikipedia.org/wiki/Peter_Kaiser_%28Schuhfabrik%29 (u.a.)* *1838, Pirmasens (Peter Kaiser Schuhfabrik)*

	II) Informationsorientierte Suchanfragen			
	Rechercheaufgaben der Studie von Pan et al. (2007, S. 808)	**Gegenstück mit deutschem oder Hamburger Kontext**	**Suchanfrage (Query)**	**Richtiger Treffer & korrekte Antwort**
07	With the heavy coverage of the democratic presidential primaries, you are excited to cast your vote for a candidate. When are/were democratic presidential primaries in New York?	Wo arbeitet Norbert Hackbusch, Linken-Politiker und Mitglied der Hamburgischen Bürgerschaft?	norbert hackbusch linke hamburg	*https://de.wikipedia.org/wiki/Norbert_Hackbusch (u.a.)* *Gruner + Jahr*
08	Which actor starred as the main character in the original Time Machine movie?	Wer sind die Moderatoren des Fernsehsenders “Hamburg 1”?	hamburg 1 moderatoren	*https://de.wikipedia.org/wiki/Liste_der_Hamburg-1-Moderatoren* *Liste der Namen*
09	A friend told you that Mr. Cornell used to live close to campus - near University and Steward Ave. Does anybody live in his house now? If so, who?	Früher befand sich im Gebäude des Kunst- und Mediencampus die Frauenklinik Finkenau. Wann wurde diese gegründet?	frauenklinik finkenau gründung	*http://www.welt.de/print/die_welt/hamburg/article133115487/Wo-jedes-Jahr-6000-Babys-geboren-wurden.html (u.a.)* *1914*
10	What is the name of the researcher who discovered the first modern antibiotic?	Welcher Trainer ist in der Fußball-Bundesliga aktuell am längsten im Amt?	dienstältester trainer bundesliga aktuell	*http://www.rp-online.de/sport/fussball/bundesliga/bundesliga-diese-aktuellen-trainer-sind-am-laengsten-im-amt-bid-1.4871411* *Markus Weinzierl*

4.2.5 Manipulation der Suchmaschinen-Ergebnisseite (SERP)

Um zu ermitteln, welchen Einfluss die Position sowie die Relevanz eines Treffers auf die Auswahl des Nutzers haben, wurden die SERPs den Probanden in drei zum Teil manipulierten Formen präsentiert.

In der normalen Variante entsprach die präsentierte Reihenfolge der Suchergebnisse der originalen Sortierung von Google (*s. Anhang 2.1*).

Auf der vertauschten SERP wurden die ersten beiden Treffer (*Snippets*) miteinander vertauscht (*s. Anhang 2.2*), in der umgekehrten Variante die komplette Liste der Treffer eins bis zehn umgedreht. Treffer eins wurde zu Treffer zehn, Treffer zwei zu Treffer neun usw. (*s. Anhang 2.3*). Die normale Variante kann als Kontrollbedingung bezeichnet werden. Sie stellt sicher, dass eine Kovariation zwischen der Beschaffenheit der SERP und dem Verhalten des Probanden festgestellt werden kann. So ergibt sich aus dem Experiment mit einer manipulierten SERP nur dann eine Aussagekraft, wenn die Ergebnisse mit mindestens einer weiteren SERP-Variante bzw. Probandengruppe verglichen werden können (Sedlmeier, Renkewitz 2008, S. 128).

Suchmaschinennutzer werden stark von Inhalt und Qualität übriger Seitenbestandteile wie bezahlten Ergebnissen beeinflusst (Buscher, Dumais, Cutrell 2010, S. 49). Alle drei SERP-Varianten beschränkten sich daher auf die ersten zehn organischen Suchergebnisse. Dadurch sollte die Aufmerksamkeit der Probanden nicht auf Seitenbereiche gelenkt werden, die für das Experiment irrelevant waren. Sämtliche Bildbearbeitungen der SERPs wurden per Adobe Photoshop umgesetzt. Als Suchmaschine zur Erstellung der Screenshots wurde Google verwendet, wodurch die Google-Oberfläche gleichermaßen zum Bestandteil des Experiments wurde.

Wie in *Tabelle 1* und *Tabelle 2* zu sehen wurden zu den Fragestellungen entsprechende Suchanfragen (*Queries*) vorformuliert. Hier liegt ein weiterer Unterschied zu dem zu replizierenden Experiment vor. Dessen Forscher bedienten sich eines Proxy-Servers, der zwischen die Suchmaschine (*Google*) und dem Rechner des Probanden geschaltet wurde. Jede individuelle Suchanfrage ist somit durch den Server manipuliert worden, bevor die veränderte Ergebnisreihenfolge dem jeweiligen Teilnehmer präsentiert wurde (Pan et al. 2007, S. 808-809). Auf die technische Umsetzung der vorliegenden Studie wird im weiteren Verlauf eingegangen.

4.2.6 Explizite Relevanzbeurteilungen von Snippets und Webseiten

Nachdem die korrekte Antwort zu einer Rechercheaufgabe gefunden oder das Zeitlimit erreicht worden ist, gelangte der Proband zur Relevanzbewertung. Diese erfolgte nach jeder einzelnen Rechercheaufgabe und war in zwei Schritte unterteilt.

Zuerst wurden nacheinander und in zufälliger Reihenfolge Screenshots der zehn Snippets der zuvor betrachteten SERP angezeigt. Unterhalb jedes Snippets erschien jeweils ein Schieberegler, mit dem der Teilnehmer stufenlos auf einer Skala von 0-100 die Relevanz des jeweiligen Google-Ergebnisses bewerten sollte.

Der zweite Schritt erfolgte in identischer Art und Weise, jedoch mit Anzeige der zehn zu den Snippets gehörenden Webseiten. Dabei handelte es sich ebenfalls um Screenshots der Webseiten in kompletter Länge.

Als Resultat der Bewertungen entstand eine Excel-Tabelle mit 5000 Relevanzurteilen (*25 vollständige Datensätze mit jeweils zehn Aufgaben, je Aufgabe zehn Snippet- und zehn Webseiten-Bewertungen*).

In der Studie von Pan et al. (2007) fanden die Relevanzbewertungen in anderer Form statt. Dort wurden zusätzlich zu den Probanden sogenannte Relevanz-Juroren eingesetzt, die nicht an der eigentlichen Studie teilnahmen. Diese sollten sicherstellen, dass das dargestellte Google-Ranking die Treffer derart auflistet, wie dies auch von menschlichen Nutzern erfolgen würde. Verglichen mit der vorliegenden Studie ist ein wesentlicher Unterschied herauszustellen.

Hier wurden alle an der Eye-Tracking-Untersuchung teilnehmenden Studierenden auch zur Relevanzbewertung eingesetzt. Die Zielsetzung beider Studien ist in diesem Punkt hingegen dieselbe. Zusätzlich zur Trefferposition und weiterer Faktoren sollten die Relevanzurteile darüber Aufschluss geben, wovon sich die Probanden in ihrem Blick- und Auswahlverhalten am ehesten beeinflussen lassen.

4.2.7 Technische Umsetzung

Bezüglich der technischen Umsetzung des Experiments sind zwei Kernbestandteile zu unterscheiden. Zum einen war ein Eye-Tracker mit entsprechender Software erforderlich. Zum anderen wurde ein Tool benötigt, das innerhalb der Eye-Tracking-Software aufgerufen wurde und den Probanden durch die Rechercheaufgaben und Relevanzbeurteilungen leitete. Im Anschluss an beide Teile soll gesondert auf die während der Untersuchung aufgetretenen technischen Probleme eingegangen werden.

Als Eye-Tracker stand das Modell Tobii T60 zur Verfügung. Dabei handelt es sich um ein Gerät, welches Infrarotstrahlen auf die Hornhäute beider Augen des jeweiligen Probanden entsendet. Die dadurch entstehenden Reflexionen werden per Sensoren durch die Hardware erfasst. Mathematische Berechnungen ermitteln daraus das Blickverhalten des Probanden auf dem Bildschirm und bilden es ab. Die beschriebene Infrarottechnik ist in einen 17 Zoll TFT-Monitor integriert (Tobii AB 2016, S. 142).

Dadurch unterscheidet sich das Gerät äußerlich nicht entscheidend von einem herkömmlichen Monitor, was wiederum einer alltagsnahen Testsituation zuträglich ist.

Auf dem Rechner, mit dem der Eye-Tracker betrieben wurde, befand sich das Programm Tobii Studio in der Version 3.1.6. Innerhalb der Software ließ sich zunächst der Untersuchungsablauf definieren. Während sämtlicher Durchgänge zeichnete das Programm das Blick- und Klickverhalten der mit IDs versehenen Probanden auf. Die Auswertung der Daten, die ebenfalls in Tobii Studio geschah, wird im nächsten Abschnitt erläutert (Tobii AB 2016, S. 1).

Bevor ein Durchlauf starten konnte, erfolgte jeweils die Kalibrierung des Eye-Trackers. Als von der Software empfohlene Kalibrierungs-Methode wurde die Fünf-Punkt-Methode gewählt. Dafür nahm der Teilnehmer zunächst eine bequeme Sitzposition mit einem Abstand von 50-80 Zentimetern vor dem Monitor ein. Sobald sich Sitzposition und Abstand zum Monitor innerhalb des Toleranzbereichs bewegten, konnte die Fünf-Punkt-Kalibrierung gestartet werden. Dabei erschien ein roter Punkt auf dem ansonsten leeren Bildschirm. Dieser bewegte sich in alle vier Ecken des Monitors sowie in dessen Mitte. Folgte der Proband diesem Punkt in ausreichendem Maße, war die Kalibrierung erfolgreich (Tobii AB 2016, S. 34-36). Die Kalibrierung konnte bei allen Probanden erfolgreich abgeschlossen werden und sorgte für eine optimale Ausgangssituation. Dementsprechend stellten auch kleinere bis mittlere Bewegungen der Teilnehmer während der Untersuchung keine Hürde für das Programm dar. Die gelungenen Rahmenbedingungen äußerten sich auch in der Qualität der Eye-Tracking-Datensätze (*Samples*). Diese wiesen einen Durchschnittswert von 93,12 Prozent auf. Das bedeutet, dass der Eye-Tracker innerhalb aller produzierten Aufnahmen das Blickverhalten zu 93,12 Prozent erfassen konnte (Tobii AB 2016, S. 40).

Zur Absolvierung der zehn Rechercheaufgaben und zur Abgabe der Relevanzurteile wurde ein Tool eingesetzt. Dies wurde dankenswerterweise durch den Zweitprüfer der vorliegenden Arbeit programmiert, sodass sich die nachfolgenden Erläuterungen auf Konzeption und Funktionsweise beschränken werden.

Das Tool war online mittels URL verfügbar, sodass im Eye-Tracking-Programm lediglich die URL des Tools als Medien-Element (*Stimulus*) festgelegt werden musste (Tobii AB 2016, S. 8). Das Tool wurde jeweils nach der Kalibrierung gestartet und führte den Teilnehmer durch die Untersuchung.

Der Ablauf war für alle zehn Rechercheaufgaben derselbe und stellte sich folgendermaßen dar:

1. Anzeige der Rechercheaufgabe.
2. Anzeige der dazugehörigen SERP, die über alle Aufgaben hinweg entweder in normaler, umgekehrter oder vertauschter Ergebnis-Reihenfolge vorlag.
3. Anzeige der zehn Snippets der zuvor betrachteten SERP nacheinander inklusive Schieberegler zur Relevanzbewertung.
4. Anzeige der zehn zu den Snippets gehörenden Webseiten der zuvor betrachteten SERP nacheinander inklusive Schieberegler zur Relevanzbewertung.

Sowohl bei den SERPs als auch bei den Snippets und Webseiten für die Relevanzbewertung handelt es sich um Screenshots. Für jeden Screenshot wurde mittels des kostenlos zur Verfügung stehenden Programms “Online Image Map Editor“ ein HTML-Code für eine sogenannte Image-Map (*Verweissensitive Grafik*) erstellt (Maschek.hu 2016). Dabei wurden Koordinaten auf den Bereichen (*Snippets*) der Bilddateien (*SERPs*) definiert, die zu klickbaren Bereichen werden sollten (Münz, Gull 2013, S. 185-186). Dies ermöglichte, dass alle zehn Snippet-Titel auf den SERPs mit den dazugehörigen URLs der Webseiten ausgestattet wurden und ein möglichst originalgetreues Google-Abbild erzeugt werden konnte.

Sämtliche Relevanzurteile sowie Informationen zu den auf den SERPs angeklickten Treffern wurden in einer Excel-Tabelle gespeichert. Diese war wie auch das Tool selbst online verfügbar und zu jeder Zeit in aktueller Version abrufbar.

Während der Untersuchung kam es zu technischen Problemen, die verschiedene Ursachen hatten. So gab es insbesondere bei den ersten Probanden zahlreiche Verbindungsprobleme zum Server, auf dem das Tool bereitstand. Mit einer gewissen Wartezeit oder einer Wiederherstellung der Browsersitzung konnte aber jeder Durchgang fortgeführt werden, sodass kein Datenverlust entstand.

Größere Schwierigkeiten traten im Zusammenhang mit dem Tobii Studio-Programm auf. Zu Beginn der Untersuchung kam es zu drei Abstürzen. Die Fehlermeldungen wiesen auf Probleme im Arbeitsspeicher hin, bedingt durch die großen Datenmengen, die pro Teilnehmer erzeugt wurden. Die Problematik konnte dadurch behoben werden, für jeden Probanden ein separates Projekt innerhalb des Programms zu erstellen.

Die einzelnen Projekte mussten im weiteren Verlauf in das Hauptprojekt integriert werden, um die Auswertung vornehmen zu können. Aufgrund der Abstürze des Tobii Studio-Programms konnten die Datensätze dreier Probanden nicht verwendet werden. Deshalb mussten insgesamt 28 Probanden akquiriert werden, um auf die angestrebte Zahl von 25 vollständigen Datensätzen zu kommen.

4.3 Auswertung der Eye-Tracking-Daten

Der erste Schritt der Eye-Tracking-Auswertung war die Definition der Areas of Interest (*AOIs*). Eine AOI ist ein definierter Bereich wie beispielsweise ein Snippet auf einer SERP. Mittels mehrerer solcher AOIs kann festgelegt werden, welche Blickdaten auf welchen Teilen der angezeigten Darstellung zur statistischen Auswertung herangezogen werden sollen (Poole, Ball 2005, S. 10).

Die AOIs können in der Regel innerhalb des Tobii Studio-Programms auf den einzelnen Webseiten erstellt werden, die den Probanden präsentiert wurden. Aufgrund nicht bekannter technischer Probleme lieferte das Programm jedoch fehlerhaft bezeichnete und dargestellte Webseiten, wodurch diese nicht ausgewertet werden konnten.

Als Alternative wurden die Recordings, also die jeweils etwa einstündigen Aufnahmen der Untersuchungen zur Auswertung herangezogen. Das hatte den entscheidenden Nachteil, dass aufgrund der Dynamik der Aufnahmen jede Scroll-Bewegung aller Probanden berücksichtigt werden musste. Alle aktiven, also sichtbaren AOIs mussten somit manuell und Bild für Bild mitverschoben werden, da Tobii Studio die AOI-Positionen bei nutzerseitigen Bewegungen der dargestellten Seiten nicht automatisch anpasst (AOIs liegen starr auf dem Bildschirm). Ebenso mussten innerhalb der Aufnahmen alle SERPs als AOIs definiert werden. Dies war dafür notwendig, um in der Auswertung Aussagen wie zur Betrachtungsdauer auf den einzelnen SERPs treffen zu können (s. *Abschnitt 5.1 Analyseebene ‘SERP’*). Die folgende Abbildung zeigt einen Ausschnitt aus Tobii Studio. “SERP vertauscht” belegt die komplette SERP, die anderen AOIs jeweils die einzelnen Snippets (s. *Abbildung 2*).

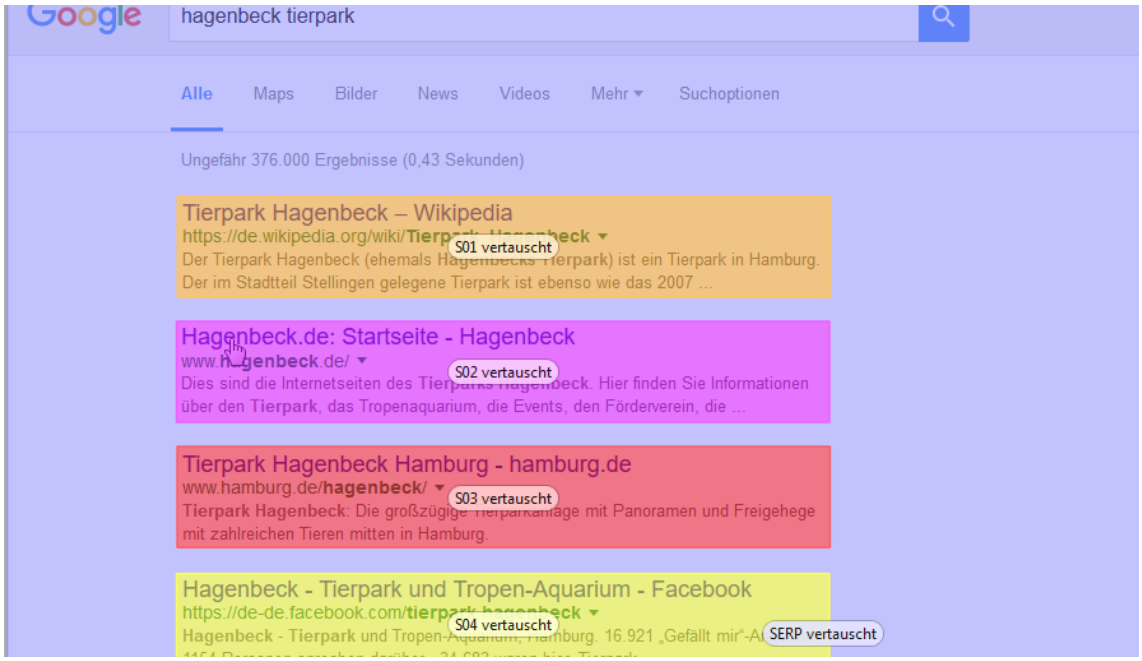

Abbildung 2: SERP mit AOIs in Tobii Studio

Tobii Studio bietet die Möglichkeit, AOIs zu aktivieren und deaktivieren. Aus zwei Gründen war dies wichtig. Einerseits konnten dadurch stets nur die für den Teilnehmer sichtbaren Snippets aktiv geschaltet und daher ausgewertet werden. Andererseits konnten sämtliche AOIs immer dann deaktiviert werden, sobald ein richtiger Treffer gefunden worden bzw. das Zeitlimit von drei Minuten abgelaufen war. Blicke auf den SERPs, die nur dem Auffinden des Buttons “zur Relevanzbewertung” dienten, konnten somit gezielt ausgeklammert werden. Dies wiederum ist als Vorteil der Auswertung per Recordings hervorzuheben. Denn das individuelle Deaktivieren der AOIs nach Beendigung der einzelnen Rechercheaufgaben wäre bei einer vollständigen Auswertung der Webseiten nicht möglich gewesen. Die Auswertung der einzelnen Aufnahmen ist somit zeitintensiver, die Präzision der erhobenen Daten aber höher.

Neben der beschriebenen Auswertung der kompletten Rechercheaufgaben von Beginn bis zur erfolgreichen Beendigung bzw. zum Ende des Zeitlimits wurde die Auswertung in einer zweiten Variante durchgeführt. In dieser wurden sämtliche AOIs immer bereits dann inaktiv geschaltet, sobald der Proband seinen ersten Mausklick auf einem Snippet getätigt hat. Beide Varianten (komplette Aufgabenlänge/Aufgabe bis zum ersten Klick) werden im Abschnitt der Ergebnisse des Experiments nochmals aufgegriffen.

Nachdem die AOIs definiert waren, wurden sie gruppiert. Demnach wurden alle elf AOIs eines Probanden (*SERP und zehn Snippets*) jeweils einer Gruppe hinzugefügt, damit die Daten bei der Auswertung aggregiert werden konnten. So bestand beispielsweise eine AOI-Gruppe “SERP umgekehrt”, in der die Blickdaten auf den SERPs aller Probanden zusammengefasst wurden, die sich in der umgekehrten Bedingung befanden.

Für die Auswertung wurden die Kennzahlen “Fixation Count”, “Fixation Duration” und “Mouse Click Count” herangezogen, auf die im weiteren Verlauf jeweils eingegangen werden wird.

Zuvor musste im Tobii Studio-Programm die Mindestdauer einer Fixation festgelegt werden. Wie unter *4.1 Eye-Tracking-Methodik* beschrieben hat sich die wissenschaftliche Fachliteratur hierfür bislang auf keinen empfehlenswerten Wert sondern lediglich auf eine Spanne einigen können. Da der Kernbestandteil der vorliegenden Arbeit die Replikation der Studie von Pan et al. (2007) darstellt und dort ein Mindestwert von 50 Millisekunden gewählt worden ist, wurde dieser ebenfalls verwendet. Somit sollte die Vergleichbarkeit beider Studien in diesem Aspekt sichergestellt werden.

5 Ergebnisse des Experiments

Für die Auswertung wurden die durch Tobii Studio gelieferten Werte zunächst in Excel-Tabellen gespeichert. Dabei wurde zwischen den Werten unterschieden, die die Blick- und Klickdaten bis zum jeweils ersten Klick der Probanden auf einer SERP darstellen sowie denjenigen, die diese Daten über die komplette Dauer der Aufgaben abbilden. Da wie beschrieben die Recordings, welche jeweils alle zehn Aufgaben enthielten und nicht die einzelnen Rechercheaufgaben ausgewertet werden konnten, lagen alle Daten in aggregierter Form vor.

Zur statistischen Auswertung wurde das Programm IBM SPSS Statistics in der Version 20 verwendet. Vorgehensweise sowie Unterschiede zur Studie von Pan et al. (2007) werden nachfolgend jeweils erläutert.

5.1 Analyseebene ‘SERP’

Die Datenanalyse bezüglich der einzelnen SERPs sollte Aufschlüsse darüber liefern, inwieweit sich Betrachtungsdauer, Anzahl der Fixationen sowie Anzahl der angeklickten Snippets gemäß der drei Bedingungen “normal”, “umgekehrt” und “vertauscht” unterscheiden. Hierfür wurden die Werte verwendet, die sämtliche Rechercheaufgaben von Beginn bis zum Ende abbilden.

5.1.1 Betrachtungsdauer auf den SERPs

Zunächst wurde die Betrachtungsdauer ermittelt. Dafür wurden zuvor in Tobii Studio die Werte für die Kennzahl “Visit Duration” erhoben. Diese Kennzahl misst die Dauer zwischen der ersten Fixation innerhalb einer AOI und der nächsten Fixation außerhalb der entsprechenden AOI, in diesem Fall der ganzen SERP (Tobii AB 2016, S. 109). Die SERP-AOIs wurden so definiert, dass sie beim Erscheinen jeder Google-Ergebnisseite den kompletten Bildschirm ausfüllten. Dadurch wurde sichergestellt, dass es keine Fixation außerhalb einer SERP-AOI geben konnte, die die Aufzeichnung der Visit Duration hätte unterbrechen können.

Mittels deskriptiver Statistiken wurden Mittelwerte zur Betrachtungsdauer einer SERP pro Rechercheaufgabe ermittelt.

Wie der *Tabelle 3* zu entnehmen ist, betrachteten diejenigen Probanden, die ausschließlich umgekehrte SERPs erhalten haben, diese mit durchschnittlich 37,6 Sekunden am längsten. Dahingegen verweilten die Teilnehmer der Bedingung "normal" pro Aufgabe 16,5 Sekunden lang auf der SERP und damit weniger als halb so lange. "N" gibt die Anzahl der Probanden in der jeweiligen Bedingung an.

Tabelle 3: Deskriptive Statistiken zur Betrachtungsdauer einer SERP pro Rechercheaufgabe

Vis_dur_mean_SERP

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
normal	8	16,5124	8,82719	3,12088	9,1327	23,8921	8,34	33,04
umgekehrt	9	37,6111	19,84428	6,61476	22,3574	52,8648	18,58	75,78
vertauscht	8	23,4331	9,31803	3,29442	15,6431	31,2232	14,56	37,19
Gesamt	25	26,3226	16,18535	3,23707	19,6416	33,0035	8,34	75,78

Um die Mittelwerte näher betrachten und Aussagen zur Signifikanz treffen zu können, wurde die Varianzanalyse gewählt. Diese ist in der Lage, die Mittelwerte von mehr als zwei Gruppen (*Bedingungen*) miteinander zu vergleichen, was sie vom t-Test unterscheidet. Letzterer ermöglicht nur den Vergleich zweier Gruppen miteinander (Rasch, Hofmann, Friese, Naumann 2010, S. 6). In SPSS kann die Varianzanalyse durch die einfaktorielle ANOVA (*Analysis of Variance*) durchgeführt werden. Einfaktoriell bedeutet, dass die Wirkung eines Faktors (*SERP-Bedingung*) auf eine Variable (*z.B. Betrachtungsdauer*) analysiert wird. Getestet wird bei der Varianzanalyse stets gegen die Nullhypothese. Erhält man ein signifikantes Ergebnis, kann die Nullhypothese verworfen und von einem signifikanten Unterschied der getesteten Bedingungen voneinander ausgegangen werden (Rasch et al. 2010, S. 35).

Welche der in diesem Falle drei Bedingungen signifikant voneinander abweichen und welche nicht, ist durch die Varianzanalyse allein nicht zu ermitteln. Dafür ist das Post-Hoc-Verfahren erforderlich, welches aufzeigt, zwischen welchen der getesteten Bedingungen signifikante Unterschiede vorliegen (Rasch et al. 2010, S. 27).

SPSS bietet in der verwendeten Version insgesamt 18 verschiedene Post-Hoc-Verfahren an. Diese werden unterschieden in Verfahren, für die eine Varianz-Gleichheit angenommen wird und in solche, für die keine Varianz-Gleichheit angenommen wird.

Betrachtet man die Voraussetzungen der Varianzanalyse, wird die Bedeutung der genannten Unterscheidung deutlich. So ist unter anderem neben der Normalverteilung des untersuchten Merkmals auch die Varianzhomogenität erforderlich. Diese besagt, dass die Varianzen der Populationen aller untersuchter Gruppen gleich sein müssen (Rasch et al. 2010, S. 49).

Da im Falle der Betrachtungsdauer auf den SERPs die Varianzhomogenität gegeben ist, wurde der Post-Hoc-Test nach Bonferroni gewählt, dessen Ergebnisse in nachstehender *Tabelle 4* zu sehen sind.

Tabelle 4: Post-Hoc-Test nach Bonferroni: Betrachtungsdauer SERP

Abhängige Variable: Vis_dur_mean_SERP

Bonferroni

(I) SERP_Art	(J) SERP_Art	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
normal	umgekehrt	-21,09874*	6,79613	,016	-38,7090	-3,4885
	vertauscht	-6,92075	6,99316	,999	-25,0415	11,2000
umgekehrt	normal	21,09874*	6,79613	,016	3,4885	38,7090
	vertauscht	14,17799	6,79613	,146	-3,4322	31,7882
vertauscht	normal	6,92075	6,99316	,999	-11,2000	25,0415
	umgekehrt	-14,17799	6,79613	,146	-31,7882	3,4322

*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Für die Interpretation der Ergebnisse des Post-Hoc-Tests wird die Spalte “Signifikanz” zurate gezogen, in der sich die *p*-Werte befinden. Der *p*-Wert gibt das empirische Signifikanzniveau an (Eckstein 2015, S. 368). Dies wird wie in der Sozialwissenschaft üblich als $\alpha = 0{,}05$ festgelegt.

Unterschreitet der dargestellte *p*-Wert dieses Niveau, kann die Nullhypothese verworfen werden, die besagt, dass kein Unterschied zwischen den jeweiligen Mittelwerten besteht. Damit gilt die Alternativhypothese, die auf einen signifikanten Unterschied deutet.

ANOVA gab an, dass ein signifikanter Unterschied zwischen einer der drei Gruppen besteht (F (2, 22) = 5.07, $p < .02$). Das Post-Hoc-Verfahren machte diesen nur zwischen den Bedingungen “normal” und “umgekehrt” aus ($p < .02$). In der Tabelle ist der entsprechende Eintrag mit einem *-Symbol markiert.

5.1.2 Anzahl der Fixationen auf den SERPs

Die Vorgehensweise zur Anzahl der Fixationen erfolgte weitestgehend in identischer Art und Weise wie bei der Betrachtungsdauer. Anstelle der “Visit Duration” wurden per Tobii Studio hierfür die Werte zur “Fixation Count” ausgelesen. Diese Kennzahl gibt die Anzahl der Fixationen innerhalb der hierfür gewählten SERP-AOI an (Tobii AB 2016, S.108).

Die deskriptiven Statistiken in *Tabelle 5* stellen dar, dass die Fixations-Anzahl auf der umgekehrten SERP pro Aufgabe mit durchschnittlich 140,4 Fixationen am höchsten ist. Auf der vertauschten SERP (*Snippet eins mit Snippet zwei*) finden durchschnittlich 79,6, auf der normalen SERP 56,9 Fixationen statt.

Tabelle 5: Deskriptive Statistiken zur Fixationsanzahl auf einer SERP pro Rechercheaufgabe

Fix_count_mean_SERP

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Mini-mum	Maxi-mum
					Untergrenze	Obergrenze		
normal	8	56,8625	28,47073	10,06592	33,0604	80,6646	27,20	110,50
umgekehrt	9	140,3778	72,99399	24,33133	84,2696	196,4859	59,30	262,50
vertauscht	8	79,5750	36,30532	12,83587	49,2230	109,9270	44,80	141,40
Gesamt	25	94,1960	61,09450	12,21890	68,9774	119,4146	27,20	262,50

Bezüglich des Post-Hoc-Tests bedarf es in diesem Fall einer veränderten Vorgehensweise, da die Homogenität der Varianzen nicht gegeben ist. Durch (p = .030) kann die Nullhypothese nicht verworfen werden. Die Varianzen unterscheiden sich demnach signifikant voneinander (s. *Tabelle 6*).

Tabelle 6: Test auf Homogenität der Varianzen

Fix_count_mean_SERP

Levene-Statistik	df1	df2	Signifikanz
4,113	2	22	,030

Für diesen Fall stehen in SPSS wie erwähnt Verfahren bereit, die diese Varianzinhomogenität tolerieren. Einer ist der sogenannte Games-Howell-Test, der als sehr robust gilt, weshalb er häufig bei fehlender Varianzhomogenität zum Einsatz kommt (Eckstein 2015, S. 148). Nachfolgend die Ergebnisse aus dem Post-Hoc-Test nach Games-Howell (s. *Tabelle 7*).

Tabelle 7: Post-Hoc-Test nach Games-Howell: Fixationsanzahl auf einer SERP pro Rechercheaufgabe

Abhängige Variable: Fix_count_mean_SERP

Games-Howell

(I) SERP_Art	(J) SERP_Art	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
normal	umgekehrt	-83,51528*	26,33128	,023	-155,0133	-12,0173
	vertauscht	-22,71250	16,31203	,373	-65,6843	20,2593
umgekehrt	normal	83,51528*	26,33128	,023	12,0173	155,0133
	vertauscht	60,80278	27,50951	,110	-12,5811	134,1866
vertauscht	normal	22,71250	16,31203	,373	-20,2593	65,6843
	umgekehrt	-60,80278	27,50951	,110	-134,1866	12,5811

*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

ANOVA wies auf Unterschiede zwischen mindestens einer der drei Bedingungen hin (F (2, 22) = 6.13, p < .01). Wie auch bei der Betrachtungsdauer der SERPs ist anhand des Post-Hoc-Verfahrens lediglich zwischen den Bedingungen “normal” und “umgekehrt” ein signifikanter Unterschied festzustellen, was die Anzahl der Fixationen auf den einzelnen SERPs angeht (p < .03).

5.1.3 Anzahl der angeklickten Snippets auf den SERPs

Um die Anzahl der angeklickten Snippets der drei SERP-Arten vergleichen zu können, wurde per Tobii Studio die Kennzahl “Mouse Click Count” erhoben. Sie gibt an, wie oft der bzw. die Teilnehmer die linke Maustaste drücken, während sich die Maus innerhalb einer bestimmten AOI befindet (Tobii AB 2016, S. 114). Hierfür wurde nicht wie in den vorausgehenden beiden Analysen die SERP-AOI verwendet. Dies wird dadurch begründet, dass dadurch auch Klickzahlen entstanden wären, wenn ein Proband auf eine leere Fläche der SERP, nicht aber auf ein einzelnes Snippet geklickt hat. Deshalb wurden die Klickzahlen der zehn Snippet-AOIs pro Teilnehmer addiert ausgewertet, um genannte Verzerrungen durch Klicks auf irrelevante Bereiche auszuklammern.

Mit durchschnittlich 2,14 Klicks auf Snippets pro Aufgabe liegt der Wert der umgekehrten SERP erneut an erster Stelle. Die normale SERP weist mit 1,40 Klicks auf Snippets einen ähnlichen Wert wie die vertauschte SERP (*1,43 Klicks*) auf, wie in *Tabelle 8* zu sehen ist.

Tabelle 8: Deskriptive Statistiken zur Klickhäufigkeit auf Snippets pro Rechercheaufgabe

Click_count_mean_SERP

	N	Mittelwert	Standard-abweichung	Standard-fehler	95%-Konfidenzintervall für den Mittelwert		Mini-mum	Maxi-mum
					Untergrenze	Obergrenze		
normal	8	1,4000	,33806	,11952	1,1174	1,6826	1,00	1,90
umgekehrt	9	2,1444	,92616	,30872	1,4325	2,8564	1,10	3,90
vertauscht	8	1,4250	,31510	,11140	1,1616	1,6884	1,10	1,90
Gesamt	25	1,6760	,69058	,13812	1,3909	1,9611	1,00	3,90

Die Ergebnisse aus ANOVA deuten knapp auf signifikante Unterschiede hin (F (2, 22) = 4.07, p < .04). Das Post-Hoc-Verfahren, das aufgrund Varianzinhomogenität erneut nach Games-Howell durchgeführt worden ist, kam zu anderen Ergebnissen. Demnach lassen sich in diesem Fall keine signifikanten Unterschiede zwischen den Gruppen feststellen (s. *Tabelle 9*)

Tabelle 9: Post-Hoc-Test nach Games-Howell: Klickhäufigkeit auf Snippets pro Rechercheaufgabe

Abhängige Variable: Click_count_mean_SERP

Games-Howell

(I) SERP_Art	(J) SERP_Art	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
normal	umgekehrt	-,74444	,33105	,109	-1,6474	,1586
	vertauscht	-,02500	,16339	,987	-,4529	,4029
umgekehrt	normal	,74444	,33105	,109	-,1586	1,6474
	vertauscht	,71944	,32821	,121	-,1799	1,6188
vertauscht	normal	,02500	,16339	,987	-,4029	,4529
	umgekehrt	-,71944	,32821	,121	-1,6188	,1799

5.1.4 Vergleich der Ergebnisse aus 5.1 mit der Hypothese H1

H1*: Die Eye-Tracking-Daten werden in den drei Testszenarien (normale, vertauschte und umgekehrte Trefferdarstellung) voneinander abweichen. So werden die Probanden bei der umgekehrten Ergebnisdarstellung die SERP länger und öfter betrachten sowie deren Snippets häufiger anklicken, als dies bei der nicht manipulierten SERP der Fall sein wird. Als Snippets werden Trefferbeschreibungen auf SERPs bezeichnet, die jeweils aus anklickbarer Überschrift, URL sowie Beschreibungstext bestehen (Lewandowski 2015, S. 141).*

Die Hypothese H1 kann in vollem Umfang bestätigt werden. Sowohl die Betrachtungsdauer als auch die Anzahl der Fixationen sind auf der umgekehrten SERP am höchsten. Auch klicken die Probanden in dieser Bedingung am häufigsten ein Snippet an.

Im Vergleich zur Studie von Pan et. al (2007) sind zwei veränderte Vorgehensweisen zu erwähnen, die technischer Natur sind.

Einerseits konnte in der vorliegenden Arbeit nicht analysiert werden, wie oft ein Teilnehmer in seinem Blick zu einem zuvor bereits betrachteten Snippet zurückgekehrt ist. Als Grundlage dieser Auswertung dienten Pan et al. (2007) sogenannte Scanpaths. Tobii Studio bietet eine ähnliche Funktion namens GazePlot an. Diese verfolgt allerdings lediglich Visualisierungszwecke. Ein Daten-Export zur Verwertung der GazePlot-Daten war demnach nicht möglich.

Der zweite Unterschied zum Untersuchungsdesign betrifft den *Abschnitt 5.1.3 Anzahl der angeklickten Snippets auf den SERPs*. Pan et al. (2007) analysierten in ihrer Arbeit anstelle der Klickzahlen die Anzahl der Snippets, die pro Rechercheaufgabe durchschnittlich fixiert worden sind (*z.B. 3,8 Snippets in der umgekehrten Bedingung*). Da es wie unter *4.3 Auswertung der Eye-Tracking-Daten* genannt jedoch nicht möglich war, die Rechercheaufgaben einzeln sondern nur die Aufnahmen auszuwerten, die jeweils zehn Aufgaben enthielten, war eine alternative Vorgehensweise erforderlich. So hätte in der vorliegenden Arbeit ermittelt werden können, wie viele Fixationen auf den einzelnen Snippets stattfanden. Damit wären aber keine Aussagen über Betrachtung oder nicht-Betrachtung innerhalb einzelner Rechercheaufgaben möglich gewesen, da jedes Snippet über alle Aufgaben hinweg vielfach fixiert worden ist.

Die Kennzahl “Mouse Click Count” wurde als Ersatz gewählt, da jedes Snippet pro Aufgabe nur einmal angeklickt wurde. Dadurch war es möglich, einen Durchschnittswert pro Rechercheaufgabe zu ermitteln.

5.2 Analyseebene ‘Snippet’

Nachdem die einzelnen SERPs analysiert worden sind, sollen nun die sich darauf befindlichen Snippets näher betrachtet werden. Begonnen wird mit einem Vergleich des Blick- und Klickverhaltens der Probanden aller drei Bedingungen in grafischer Form. Danach werden die Fixationen auf den einzelnen Snippets durch statistische Berechnungen analysiert.

Hierfür wurden die Werte verwendet, die sämtliche Rechercheaufgaben von Beginn der Aufgabe bis zum jeweils ersten Klick des Probanden auf einer SERP abbilden.

5.2.1 Blick- und Klickverhalten auf den Snippets

Die folgenden drei Abbildungen (*Abbildung 3, Abbildung 4, Abbildung 5*) vergleichen das Blick- und Klickverhalten der Probanden aller drei Bedingungen miteinander. Dargestellt wird zum einen die durchschnittliche Anzahl an Fixationen auf Snippets pro Rechercheaufgabe (*bis zum ersten Klick*). Zum anderen wird der prozentuale Anteil der Snippets an den ersten Klicks angegeben.

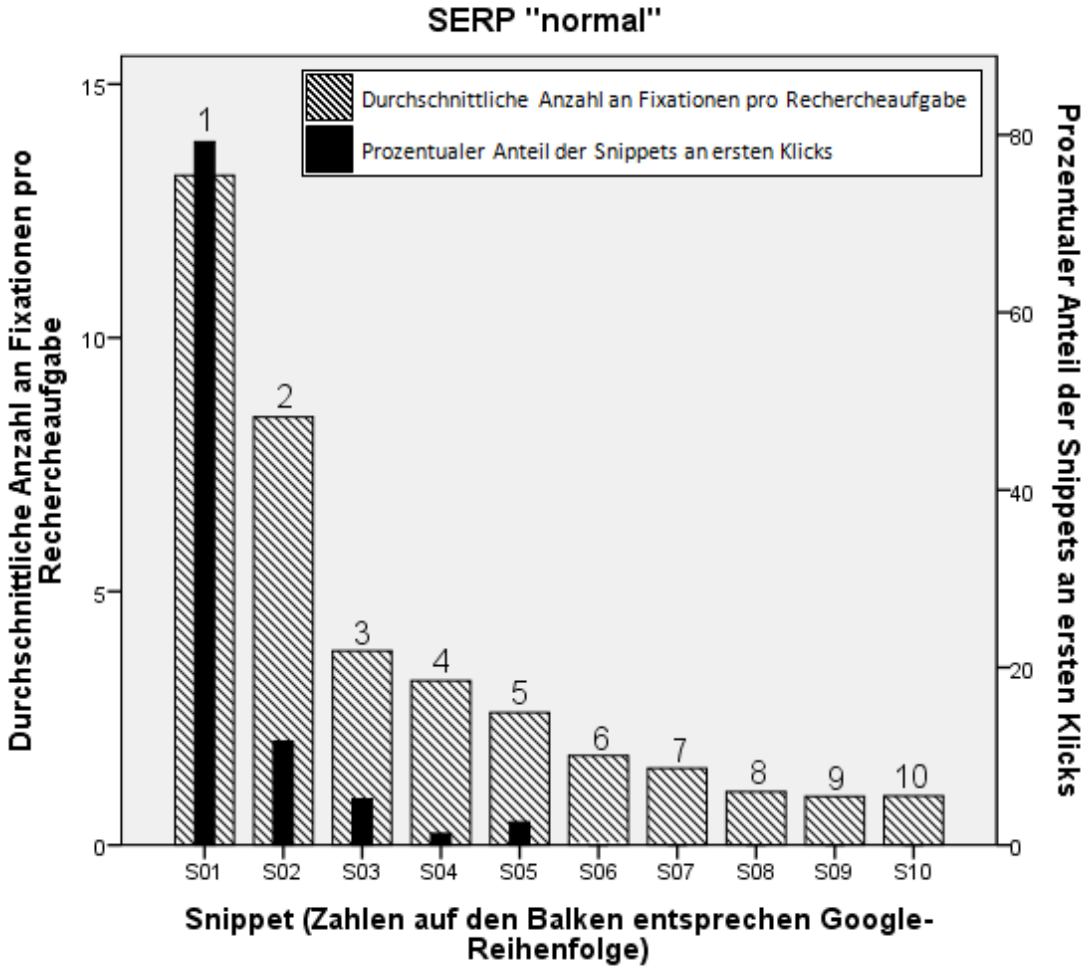

Abbildung 3: Blick- und Klickverhalten SERP "normal"

Die Teilnehmer, die die SERPs in normaler Form präsentiert bekamen, betrachteten das erste Snippet am häufigsten und wählten es auch mit annähernden 80 Prozent der Fälle zuerst aus. Auch das zweite Snippet kann einige Fixationen verzeichnen, wurde im Vergleich zum ersten aber deutlich seltener angeklickt. Die Snippets sechs bis zehn erhielten keine ersten Klicks.

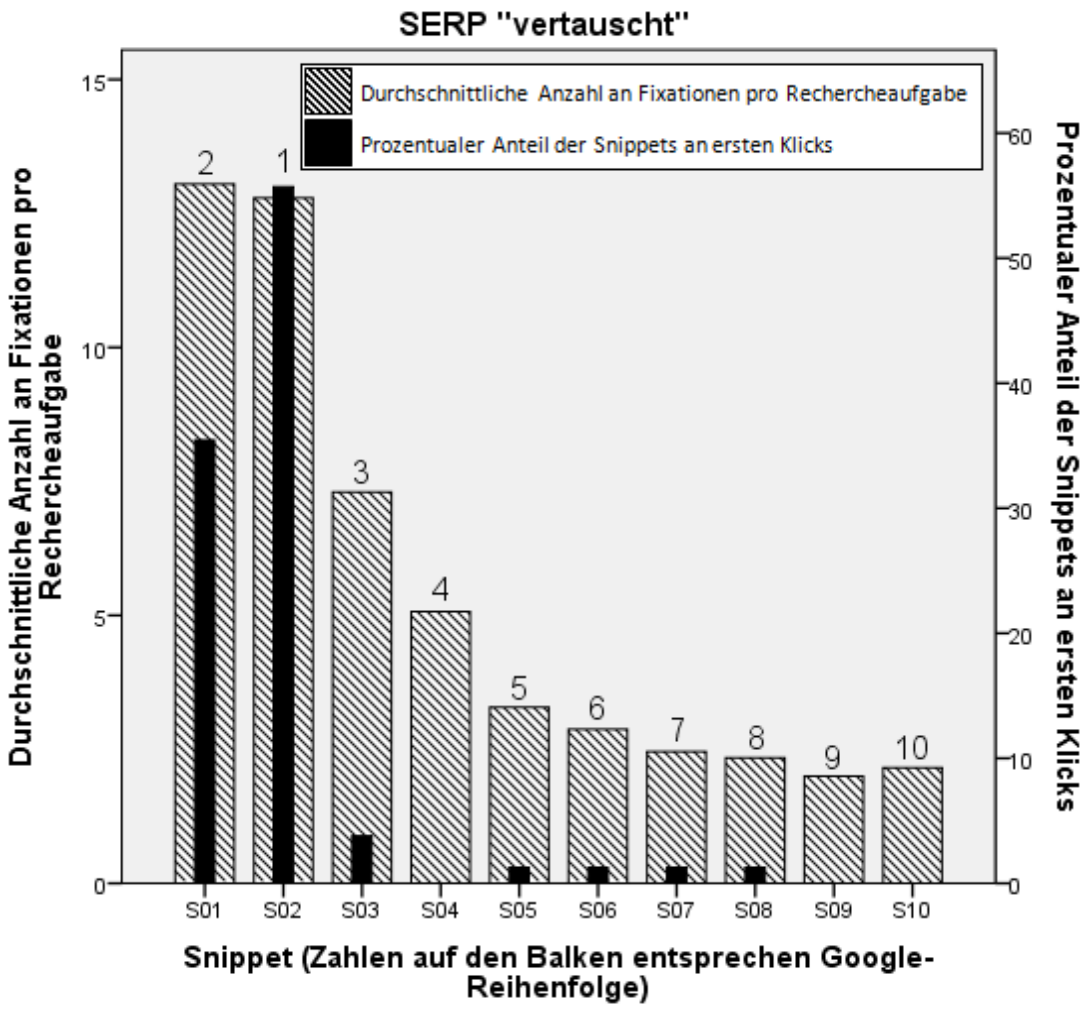

Abbildung 4: Blick- und Klickverhalten SERP "vertauscht"

In der vertauschten Bedingung ist eine klare Verschiebung des Blick- und Klickverhaltens erkennbar. So erhielten die ersten beiden Snippets fast gleichermaßen viele Fixationen, häufiger angeklickt wurde jedoch das Snippet auf der zweiten Position (*Snippet eins in der normalen Bedingung*). Vergleicht man die Klicks auf Snippet zwei der normalen mit denen auf Snippet eins der vertauschten SERP, wird jedoch eine Auffälligkeit deutlich. Beide genannten Snippets sind identisch und lediglich in der Reihenfolge vertauscht worden. Die Klickzahlen, die darauf entfallen, weichen jedoch erheblich voneinander ab.

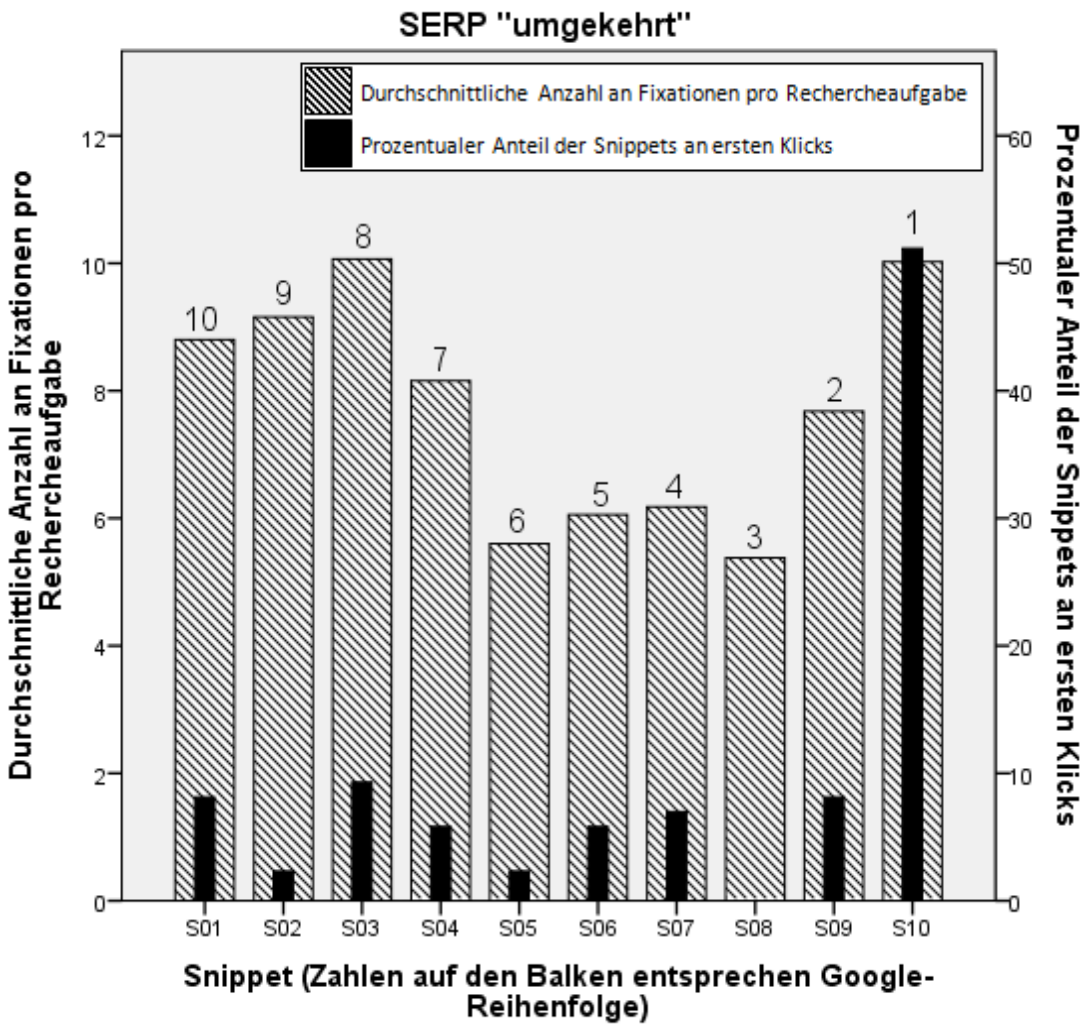

Abbildung 5: Blick- und Klickverhalten SERP "umgekehrt"

Bei der Betrachtung des Diagramms der umgekehrten SERP (*Abbildung 5*) wird deutlich, wie sich die bis zum ersten Klick gemachten Fixationen auf die zehn Snippets verteilen. Snippet zehn, normalerweise an erster Stelle, wird zu gut 50 Prozent der Fälle als relevant erkannt und als erstes angeklickt. Dennoch streuen die Klickzahlen mehr als bei den anderen beiden Bedingungen. So werden beispielsweise die Snippets auf den Plätzen eins bis fünf zu jeweils bis zu zehn Prozent der Fälle angeklickt, wohingegen dieselben Snippets in der normalen Bedingung (*dort auf den Positionen sechs bis zehn*) keine ersten Klicks erhalten haben.

Verglichen mit der Studie von Pan et al. (2007) sind die Diagramme leicht verändert erstellt worden. Deren Arbeit hat sowohl die Klickzahlen als auch die Blicke prozentual ausgedrückt. Es wurde also angegeben, welche Snippets bis zum ersten Klick betrachtet worden sind und welche nicht.

Aufgrund der aggregierten Werte und somit der fehlenden Möglichkeit, in der vorliegenden Arbeit nach Rechercheaufgaben zu unterscheiden, wurde als Alternative die durchschnittliche Fixations-Anzahl zurate gezogen.

5.2.2 Statistische Auswertungen zu den Fixationen auf den Snippets

Neben der Interpretation der vorherigen Balkendiagramme wurde das Blickverhalten der Teilnehmer auf den einzelnen Snippets statistisch ausgewertet. Hierfür kam in SPSS erneut einfaktorielle Varianzanalyse (*ANOVA*) zum Einsatz. Wurden signifikante Unterschiede zwischen den drei Bedingungen festgestellt, wurden diese per Post-Hoc-Verfahren verortet.

Zuerst wurden die durchschnittlichen Fixations-Zahlen auf den Snippets eins und zwei miteinander verglichen. Die Werte, die analysiert worden sind, gaben also für jeden Probanden an, wie oft er pro Rechercheaufgabe entweder das erste oder zweite Snippet betrachtet hat. Per ANOVA konnten keine signifikanten Unterschiede zwischen den drei Gruppen ermittelt werden. Das Post-Hoc-Verfahren ist an dieser Stelle nicht notwendig.

Probanden der normalen Bedingung unterschieden sich im Blickverhalten nicht signifikant von Teilnehmern der umgekehrten Bedingung, was die beiden zuerst aufgelisteten Treffer angeht. Dies ist insofern nennenswert, da die Probanden der umgekehrten SERP als erste Treffer die normalerweise letztplatzierten Ergebnisse präsentiert bekamen. Die erstplatzierten Ergebnisse wurden somit ähnlich oft betrachtet, unabhängig davon, welches Snippet sich dort wiedergefunden hat.

Als nächstes wurden die durchschnittlichen Fixations-Zahlen auf den Snippets neun und zehn miteinander verglichen. Das Verfahren ähnelte dem zuvor beschriebenen. Auch hier wurde zunächst eine ANOVA durchgeführt. Sie wies zwischen mindestens zwei Gruppen signifikante Unterschiede nach (F (2, 22) = 6.32, $p < .01$). Aufgrund fehlender Varianzhomogenität wurde das Post-Hoc-Verfahren nach Games-Howell gewählt, um zu erfahren, wo die signifikanten Unterschiede bestehen. Das Verfahren ergab, dass zwischen der normalen und der umgekehrten SERP ein signifikanter Unterschied besteht, was die Fixations-Anzahl auf Snippet neun bzw. zehn betrifft ($p < .05$).

Nun wurde ermittelt, wie sich das Blickverhalten der Probanden bezogen auf das Snippet unterscheidet, welches in der originalen Google-Reihenfolge den ersten Platz einnimmt. Verglichen wurden somit die Fixations-Daten zu Snippet eins der normalen, Snippet zwei der vertauschten sowie Snippet zehn der umgekehrten SERP. Es konnten keine signifikanten Unterschiede festgestellt werden. Aufgrund der Höhe der Balken der Diagramme aus *Abschnitt 5.2.1* ergibt sich eine ähnliche Einschätzung. Die Probanden ließen sich in dieser Hinsicht nicht durch die Trefferreihenfolge beeinflussen.

Aussagen zur Pupillenerweiterung auf den einzelnen Snippets, wie sie in der Studie von Pan et al. (2007) getroffen worden sind, waren in der vorliegenden Arbeit nicht möglich. Das Programm Tobii Studio bietet hierfür keine Auswertungsmöglichkeiten an.

5.2.3 Vergleich der Ergebnisse aus 5.2 mit den Hypothesen H2 und H3

***H2**: In allen drei Testszenarien werden die Probanden den ersten beiden aufgelisteten Suchergebnissen nahezu gleichermaßen viel Aufmerksamkeit schenken, was sich durch die Analyse des Blickverhaltens belegen lassen wird. Dennoch werden die Probanden bei der umgekehrten Trefferdarstellung die letzten beiden Snippets (die ersten beiden der normalen Darstellung) länger betrachten, als in den anderen beiden Szenarien.*

Diese Hypothese H2 kann bestätigt werden. Wie aus der statistischen Analyse hervorgeht, sind keine signifikanten Unterschiede in der Betrachtungshäufigkeit der ersten beiden Snippets über alle drei Bedingungen hinweg zu ermitteln. Bezogen auf die letzten beiden Snippets wird hingegen ein Unterschied deutlich. Dies erkennt man zunächst anhand der Diagramme aus *Abschnitt 5.2.1*. So weichen die Fixations-Zahlen auf den letzten beiden Snippets der umgekehrten SERP in signifikantem Ausmaß von denen der normalen SERP ab ($p < .05$).

***H3**: Sowohl die Probanden der vertauschten (Snippets eins und zwei miteinander vertauscht) als auch die der umgekehrten Trefferdarstellung werden eher Snippets auswählen, die auf der normalen SERP niedriger gelistet werden und dort weniger Klicks erhalten.*

Auch die Hypothese H3 wird bestätigt, was erneut durch die Diagramme verdeutlicht wird. Die Probanden der vertauschten Bedingung führten deutlich mehr Klicks auf dem erstplatzierten Snippet aus als die Probanden der normalen Bedingung auf demselben Snippet, welches dort als zweites aufgeführt worden ist. Werden die Bedingungen “normal” und “umgekehrt” miteinander verglichen, fällt auf, dass auf der normalen SERP keine ersten Klicks auf die Ergebnisse sechs bis zehn stattfinden. Dieselben Ergebnisse, die bei der umgekehrten SERP auf den Plätzen eins bis fünf stehen, erhalten demgegenüber jeweils bis zu zehn Prozent der ersten Klicks.

5.3 Analyseebene ‘Relevanz’

Zur Erweiterung der Ergebnisauswertung wurden die Relevanzurteile der Probanden herangezogen. Diese lagen sowohl für sämtliche präsentierten Snippets als auch für deren Webseiteninhalte vor. Demzufolge hat jeder Proband 200 Urteile abgegeben (*zehn Rechercheaufgaben mit jeweils zehn Snippet- und zehn Webseiten-Bewertungen*).

Zunächst sollte ermittelt werden, ob zwischen den Snippet- und den Webseiten-Bewertungen signifikante Unterschiede bestehen. Dazu wurden zunächst die Mittelwerte (*M*) beider Bewertungsformen ermittelt und anschließend ein t-Test für gepaarte Stichproben durchgeführt. Der Mittelwert aller Relevanzbewertungen von Snippets lag bei (M = 2,77, SD = 2,527), der Mittelwert der Webseiten-Bewertungen bei (M = 2,71, SD = 2,512). Die Ähnlichkeit beider Werte drückt sich auch in den Resultaten des t-Tests aus. Ein t-Test für gepaarte bzw. verbundene Stichproben berechnet die Differenz, die zwischen zwei Merkmalswerten (*Snippet- und Webseiten-Bewertungen*) innerhalb eines Merkmalsträgers (*Proband*) bestehen (Eckstein 2015, S. 136). Somit wurden die zwei Urteile, die ein bestimmter Proband pro Quelle abgegeben hat, direkt und dadurch paarweise miteinander verglichen. Die Unterschiede zwischen den Wertepaaren konnten nicht als signifikant nachgewiesen werden ($t(249) = 1{,}069$, $p = 0{,}286$). Es kann somit davon ausgegangen werden, dass es sich bei als relevant bewerteten Suchergebnissen auch tatsächlich um relevante Webseiteninhalte handelte.

5.3.1 Statistische Auswertung der Relevanzurteile per gemischter Modelle

Kernbestandteil der Analyseebene ‘Relevanz’ stellt die Wiederholung des in der Studie von Pan et al. (2007) als “mixed model analysis” bezeichneten Verfahrens dar.

In SPSS handelt es sich dabei um sogenannte “Lineare gemischte Modelle”. Diese ermöglichen es zu ermitteln, ob und wie stark sich gewisse Faktoren auf eine abhängige Variable auswirken. Als Faktoren wurden das Geschlecht des Probanden, die Bedingung (*normal, umgekehrt, vertauscht*), die Position des Snippets auf der manipulierten Ergebnisliste sowie die Relevanz des jeweiligen Snippets gewählt. Die Faktoren wurden als feste Effekte definiert, da von deren Einfluss auf die Mittelwerte beider abhängiger Variablen ausgegangen worden ist (SPSS Inc. 2002, S. 3).

Wie zuvor erwähnt wird in der Auswertung nicht nach Rechercheaufgaben unterschieden. Die aufgabenbezogenen Faktoren “Aufgabentyp” und “Aufgabenreihenfolge” der Arbeit von Pan et al. (2007) wurden somit nicht als Effekte übernommen.

Als abhängige Variable wurde zunächst die Anzahl an Fixationen und dann die Anzahl an Klicks bezogen auf die einzelnen Snippets festgelegt. Dadurch sollte zuerst festgestellt werden, welche Faktoren das Blickverhalten am stärksten beeinflussen und danach, wodurch die Auswahlentscheidung beeinträchtigt wird. Dafür wurden die Blick- und Klickdaten gewählt, welche die Rechercheaufgaben bis einschließlich zum ersten Klick je Aufgabe abbilden. Die Daten lagen als Durchschnittswerte vor. Der Datensatz, der als Grundlage für diese Auswertung diente, beinhaltete demnach für jeden Probanden zehn Fälle (*Snippets eins bis zehn*). Alle Fälle waren mit Durchschnittswerten zur Fixations- und Klickanzahl, zur Snippet-Relevanz sowie zur Angaben zur Bedingung, Snippet-Position und Geschlecht versehen.

Die folgende *Tabelle 10* stellt das gemischte Modell dar, welches den Einfluss auf die Anzahl der Fixationen auf den einzelnen Snippets pro Rechercheaufgabe beschreibt. Die Faktoren “Snippet-Position“, “Bedingung“ und “Snippet-Relevanz“ sind signifikant ($p < 0{,}01$), weshalb von deren Einfluss auf das Blickverhalten ausgegangen werden kann. Die *F*-Werte geben die Stärke dieser Einflüsse an. Demzufolge übt die Bedingung den mit Abstand größten Einfluss darauf aus, wie viele Fixationen auf den Snippets stattfinden. Der Bedingung folgt die Position, auf dem ein Snippet auf der präsentierten SERP steht. Den geringsten Einfluss auf die Häufigkeit der Betrachtungen übt die Relevanz des Snippets aus.

Tabelle 10: Lineares gemischtes Modell: Blickverhalten

Fester Effekt	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Snippet-Position	9	227	7,526	,000
Bedingung	2	227,000	13,559	,000
Geschlecht	1	227	1,130	,289
Snippet-Relevanz	10	227	3,576	,000

a. Abhängige Variable: Durchschnittliche Anzahl an Fixationen pro Snippet und Rechercheaufgabe.

Ein anderes Bild ergibt sich beim Blick auf die nun folgende *Tabelle 11*. Sie stellt die Resultate des zweiten gemischten Modells dar. Die abhängige Variable war in diesem Fall die durchschnittliche Anzahl an Klicks pro Rechercheaufgabe, die auf den zehn Snippets stattgefunden haben. Der mit Abstand stärkste Effekt auf die Klickhäufigkeit wird durch die Relevanz des jeweiligen Snippets hervorgerufen. Die zweite signifikante Einflussgröße ist die Snippet-Position. Die Auswirkung der SERP-Bedingung (*normal, umgekehrt oder vertauscht*) ist in diesem Fall nicht signifikant.

Tabelle 11: Lineares gemischtes Modell: Klickverhalten

Fester Effekt	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Snippet-Position	9	227	8,280	,000
Bedingung	2	227,000	1,187	,307
Geschlecht	1	227	,004	,947
Snippet-Relevanz	10	227	33,726	,000

a. Abhängige Variable: Durchschnittliche Anzahl an Klicks pro Rechercheaufgabe nach Snippets.

Zu erkennen ist, dass das Geschlecht des Probanden sowohl auf die Anzahl der Fixationen als auch auf die Anzahl der Klicks keinen Einfluss zu haben scheint. Beide Werte sind nicht signifikant.

5.3.2 Vergleich der Ergebnisse aus 5.3 mit der Hypothese H4

***H4**: Die Probanden werden in ihren Entscheidungen, Treffer auszuwählen, die Google-Positionen ihren eigenen Relevanzurteilen vorziehen. Dies wird sich statistisch dadurch ausdrücken, dass die Position einen stärkeren Einfluss auf das Blick- und Klickverhalten ausüben wird als die eigentliche Relevanz eines Treffers.*

Die Hypothese H4 muss verworfen werden. Das Blickverhalten wurde eher von der Position als von der Relevanz eines Treffers beeinflusst. Jedoch zogen die Teilnehmer der vorliegenden Untersuchung in der Entscheidung, einen Treffer anzuklicken, die Relevanz der Position eines Ergebnisses vor.

5.4 Analyseebene ‘Vergleich’

Nachfolgend sollen die Ergebnisse der vorliegenden Arbeit jeweils kurz rekapituliert sowie mit denen von Pan et al. (2007) verglichen werden.

Begonnen wird mit der Analyseebene ‘SERP’. Diejenigen Teilnehmer, die die Ergebnisseiten in umgekehrter Form präsentiert bekommen haben, unterschieden sich zum Teil signifikant von den Probanden der anderen beiden Bedingungen. Sie betrachteten die Ergebnisseiten eine längere Zeit, tätigten mehr Fixationen und klickten am häufigsten ein Snippet an. Trotz der erläuterten teilweise unterschiedlichen Vorgehensweise decken sich die Erkenntnisse in hohem Maße mit denen von Pan et al. (2007, S. 812).

Richtet man den Fokus auf die einzelnen Snippets, können konkretere Annahmen des Verhaltens auf der manipulierten Trefferliste getätigt werden. So klickten die Probanden der umgekehrten Bedingung insgesamt mehr Snippets an als die beiden anderen Gruppen. Zuerst ausgewählt wurde in etwa 50 Prozent der Fälle dennoch das zehnte Snippet, was in der normalen Bedingung an erster Stelle stand (*dort zu 80 Prozent zuerst ausgewählt*). Auch die Teilnehmer der vertauschten Bedingung ließen sich nur bedingt von der manipulierten Reihenfolge beeinflussen. Sie wählten zu etwa 55 Prozent den zweitplatzierten Treffer (*erstes Snippet der normalen SERP*) als erstes aus.

Diese Ergebnisse unterscheiden sich grundlegend von denen von Pan et al. (2007). Deren Probanden wählten in der umgekehrten Bedingung zu jeweils etwa 25 Prozent die beiden vordersten Snippets (*normale SERP: Snippets neun und zehn*) aus, welche von den Teilnehmern des vorliegenden Experiments nur acht bzw. zwei Prozent Anteil an den ersten Klicks erhielten. Hinzu kommt, dass das Letztplatzierte und dem Google-Ranking zufolge Relevanteste nur etwa fünf Prozent aller ersten Klicks verzeichnet hatte, wie in *Abbildung 6* zu sehen ist. Demgegenüber stehen die genannten 50 Prozent aller ersten Klicks auf derselben Snippet-Position der vorliegenden Arbeit.

Ein weiterer relevanter Unterschied ist im Blickverhalten der Probanden auf den Snippets der umgekehrten SERP auszumachen. Zur Veranschaulichung soll eine Gegenüberstellung der folgenden beiden Grafiken *Abbildung 4* und *Abbildung 6* dienen.

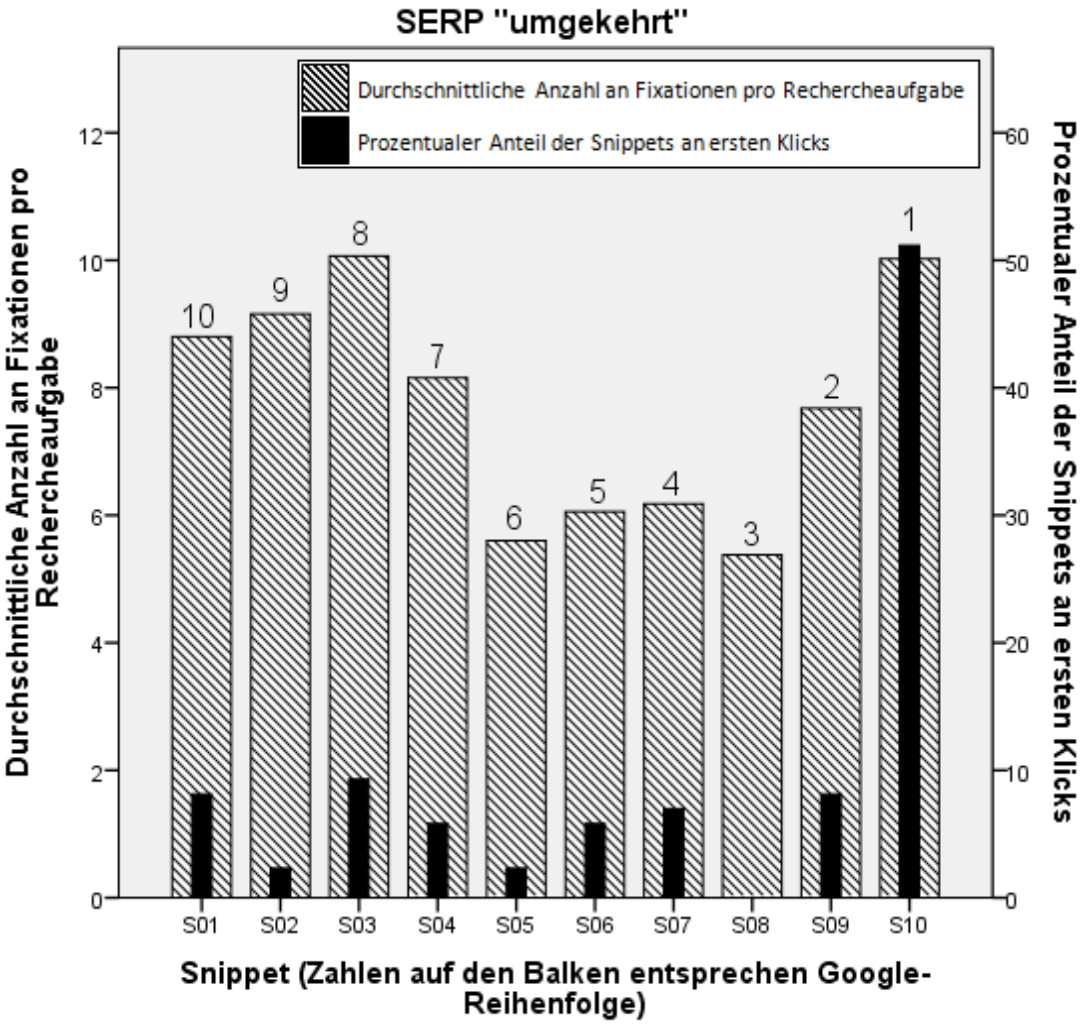

Abbildung 5: Blick- und Klickverhalten SERP "umgekehrt"

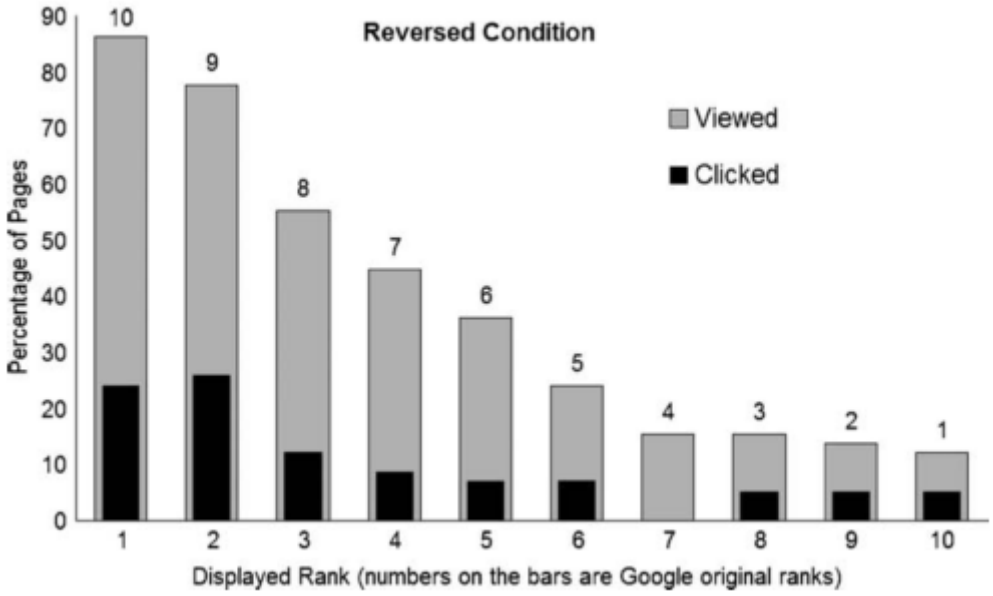

Abbildung 6: Blick- und Klickverhalten SERP "umgekehrt" (Pan et al. 2007, S. 814)

Wie bereits beschrieben lässt sich das Blickverhalten beider Diagramme nicht exakt miteinander vergleichen, da die Snippet-Betrachtung einmal in Prozent- und einmal in absoluten Werten angegeben wird. Ein Blick auf beide S10 (*Snippet zehn*) weist dennoch darauf hin, dass neben dem Klick- auch das Blickverhalten der Probanden beider Studien ein anderes ist. Wo in der Studie von Pan et al. (2007) das letztplatzierte Snippet in gut zehn Prozent der Fälle und damit am seltensten fixiert worden ist, ist die Anzahl an Fixationen in der vorliegenden Studie auf demselben Treffer mit am höchsten.

Zwischen den Fixationszahlen auf Snippet eins der normalen, Snippet zwei der vertauschten und Snippet zehn der umgekehrten SERP bestehen in der vorliegenden Studie keine signifikanten Unterschiede. Unabhängig von dessen Position bewegt sich die Aufmerksamkeit der Teilnehmer an diesem (*identischen*) Snippet somit auf einem ähnlichen Niveau. Auch davon weichen die Ergebnisse von Pan et al. (2007, S. 815) ab, wo signifikante Unterschiede zwischen den drei genannten Fällen bestanden.

Die Analyseebene ‘Relevanz’ untermauert die zuvor beschriebenen Ergebnisse. Die beiden gemischten Modelle der Studie von Pan et al. (2007, S. 816-817) weisen darauf hin, dass das Blick- und Klickverhalten der Probanden eher von der angezeigten Position als von der Relevanz eines Treffers bestimmt wurde.

Wie schon anhand der Analyseebene ‘Snippet’ zu erkennen ist, verhielt es sich bei den Probanden der vorliegenden Untersuchung anders. Die linearen gemischten Modelle belegen, dass die Trefferposition in dieser Arbeit nicht die stärkste Einflusskraft auf das Blick- und Klickverhalten darstellt. So war bezogen auf die Snippet-Fixationen die Bedingung gefolgt von Position und Relevanz der stärkste Effekt. Hinsichtlich der Klickzahlen auf den Snippets stellte die Relevanz vor der Trefferposition den einflussreichsten Faktor dar.

5.4.1 Vergleich der Ergebnisse aus 5.4 mit der Hypothese H5

***H5**: Die Ergebnisse der vorliegenden Arbeit werden sich nicht grundlegend von denen der Studie von Pan et al. (2007) unterscheiden. Auch die Teilnehmer des vorliegenden Experiments werden ein großes Vertrauen Google gegenüber offenbaren.*

Auch die Hypothese H5 muss verworfen werden. Da die Teilnehmer der vorliegenden Studie durch die Manipulation beeinträchtigt, in ihrem Auswahlverhalten aber nicht entscheidend beeinträchtigt worden sind, unterscheiden sie sich grundlegend von den Probanden von Pan et al. (2007).

6 Diskussion

Die Diskussion interpretiert die Ergebnisse der vorliegenden Arbeit. Dabei wird insbesondere auf die zum Teil unterschiedlichen Ergebnisse beider Studien eingegangen.

Die Probanden beider Arbeiten verweilten länger auf den umgekehrten Ergebnisseiten und tätigten dort mehr Fixationen und Klicks. Daraus kann geschlossen werden, dass auch die Teilnehmer des vorliegenden Experiments durch die umgekehrte Ergebnisreihenfolge einer gewissen Verunsicherung unterlagen. Da sich die relevanten Treffer am unteren Ende der SERP befanden, benötigten die Teilnehmer eine längere Zeit, diese zu erkennen und tätigten dadurch gleichermaßen mehr Fixationen. Die generelle höhere Klickzahl auf Snippets in der umgekehrten Bedingung weist darauf hin, dass je nach Aufgabenstellung zunächst irrelevante Treffer aufgrund ihrer prominenten Position angeklickt worden sind, bevor zum richtigen Ergebnis auf dem unteren Teil der SERP gelangt wurde.

Trotz dieser probandenseitigen Verunsicherung durch die manipulierten Ergebnislisten erkannten die Teilnehmer des vorliegenden Experiments etwa zehn Mal häufiger, dass sich das eigentlich zuerst platzierte Suchergebnis an letzter Position der umgekehrten SERP befand, als diejenigen der Untersuchung von Pan et al. (2007, S. 814). Diese gezielte Trefferauswahl geht mit einer hohen Fixationsanzahl auf dem letztplatzierten Snippet der umgekehrten SERP einher. Die Teilnehmer der vorliegenden Studie begutachteten die komplette Ergebnisliste offenbar bewusst genug, um den letztplatzierten Treffer zu erkennen, diesen als relevant einzuschätzen und schließlich auszuwählen.

Die linearen gemischten Modelle bestätigen diesen Eindruck. Im Gegensatz zur Studie von Pan et al. (2007) war in dieser Arbeit nicht die Snippet-Position der einflussreichste Faktor hinsichtlich des Blick- und Klickverhaltens. So wurde die Anzahl der Fixationen auf Snippets am stärksten durch die Bedingung (*normal, umgekehrt, vertauscht*) beeinflusst. Dies unterstreicht den geschilderten Eindruck der probandenseitigen Verunsicherung insbesondere in der umgekehrten Bedingung, die sich stark auf das Blickverhalten ausgewirkt hat.

Zusammengefasst liegt der Hauptunterschied beider Studien in der Evaluierung der Treffer. Denn in der Entscheidung, einen Treffer anzuklicken, zogen die Probanden der vorliegenden Arbeit die Treffer-Relevanz gegenüber der Position vor. Wo in der Arbeit von Pan et al. (2007) ein starkes Vertrauen der Probanden dem Google-Ranking gegenüber ausgemacht werden konnte, ließen sich die Teilnehmer des vorliegenden Experiments primär von ihrer eigenen Relevanzbeurteilung leiten. Sie erkannten die relevanten Treffer auch auf der umgekehrten SERP, was sich in einer gesteigerten Blick- und Klickhäufigkeit auf den entsprechenden Suchergebnissen äußerte.

Über die Ursachen dieses emanzipierten Rechercheverhaltens können lediglich Vermutungen angestellt werden. So könnte die unter *4.2.1* Zielsetzung genannte Annahme, die heutigen Suchmaschinennutzer seien bewusster im Umgang mit Suchmaschinen, mitursächlich für die Ergebnisse gewesen sein.

Des Weiteren studierten 20 der 25 Studentinnen und Studenten, die an der Untersuchung teilgenommen haben, einen informationswissenschaftlichen Bachelor- oder Masterstudiengang. Diese Dominanz hinsichtlich des Studienschwerpunktes lag in der Studie von Pan et al. (2007, S. 807) nicht vor. Es ist anzunehmen, dass die Ergebnisse unter anderem auf die studiumsbedingte Informations- und Recherchekompetenz zurückzuführen sind.

6.1 Fazit

Die Forschungsfrage lautet:

„Lassen sich die Erkenntnisse aus der Studie von Pan et al. (2007) im zeitlichen und geografischen Kontext durch eine Replikation auf heutige, deutschsprachige Suchmaschinennutzer übertragen?“

Die Forschungsfrage dieser Arbeit wird verneint. Die Erkenntnisse der Studie von Pan et al. (2007) lassen sich nicht auf die Suchmaschinennutzer der vorliegenden Untersuchung übertragen. Deren Hauptkriterium, ein Suchergebnis zu begutachten oder auszuwählen, war dessen Relevanz, nicht aber dessen Position. Begründet wird dies durch die Betrachtung der Hypothesen. H1 und H2 konnten bestätigt werden. Diese wiesen insbesondere auf eine probandenseitige Verunsicherung durch die manipulierten SERPs hin, was sich durch häufigere Klicks und Fixationen insbesondere auf der umgekehrten Ergebnisliste äußerte.

Auch H3 wurde bestätigt, da die Teilnehmer der umgekehrten Bedingung teilweise erste Klicks auf den vordersten fünf Treffern tätigten, die in der normalen Bedingung (*dort an Stelle sechs bis zehn*) keine ersten Klicks erhielten.

Hauptursächlich für die Verneinung der Forschungsfrage ist die verworfene Hypothese H4. Diese hat zusätzlich zum Blick- und Klickverhalten die Relevanzurteile miteinbezogen. Somit ergibt sich folgendes Gesamtbild:

Das Klick- und Blickverhalten der Probanden wurde durch die Manipulation der SERPs beeinträchtigt, auch wurden zum Teil irrelevante Treffer als erstes ausgewählt. In ihrer Entscheidung, einen Treffer auszuwählen, zogen die Teilnehmer dennoch das eigene Relevanzurteil der Trefferposition vor.

6.2 Grenzen der Studie

Zunächst sind probandenbezogene Aspekte zu nennen, die die vorliegende Arbeit in ihrer Aussagekraft einschränken. Alle Teilnehmer waren deutschsprachige Studierende einer ähnlichen Altersgruppe. Außerdem studierten wie beschrieben 80 Prozent der Teilnehmer einen informationswissenschaftlichen Bachelor- oder Masterstudiengang. Somit handelt es sich bei dem Probandenfeld um eine sehr homogene Population. Auch die Anzahl der Teilnehmer ist aufgrund der in Durchführung und Auswertung zeitintensiven Eye-Tracking-Methode begrenzt. Aufgrund dieser Ursachen können die Ergebnisse der Arbeit nicht als repräsentativ für durchschnittliche Studierende oder gar für die Durchschnittsbevölkerung angesehen werden.

Abseits des Faktors des Probandenfeldes sind technisch bedingte Grenzen der Studie aufzuführen. Von den Teilnehmern selbst eingegebene Suchanfragen, wie sie in der Arbeit von Pan et al. (2007) verwendet worden sind, waren in diesem Fall technisch nicht realisierbar. Das Tool, welches für die Experimente zur Verfügung stand, bot Google-Ergebnisseiten anhand entsprechend vorgefertigter Suchanfragen an.

Solch konstruierte Suchanfragen entsprechen nicht unbedingt den Informationsinteressen der Probanden, wodurch diese in ihrem natürlichen Rechercheverhalten zusätzlich beeinträchtigt werden. Hinzu kommt, dass sich durch die Anwesenheit des Versuchsleiters Probanden beobachtet fühlen könnten.

Dies wirkt sich verstärkend auf die von den Probanden ohnehin empfundene Laborsituation aus (Höchstötter 2007, S. 137).

6.3 Zukünftige Forschungsvorhaben

Sowohl die Arbeit von Pan et al. (2007) als auch die vorliegende bezogen sich lediglich auf die zehn ersten organischen Suchergebnisse der ersten SERP. Heutige Ergebnisseiten von Suchmaschinen sind jedoch stark von sogenannten Universal Search-Treffern (*Bilder, Nachrichten u.a.*) geprägt. Wie im Forschungsstand erwähnt ist das Konzept des "Goldenen Dreiecks" durch die Unterbrechung der organischen Trefferlisten nicht mehr gültig. Aufgrund dessen wäre es interessant zu untersuchen, wie die Relevanzbeurteilung durch Universal Search-Ergebnisse beeinflusst wird. Durch sie wird die Aufmerksamkeit des Anwenders gezielt durch die Suchmaschine beeinflusst (Lewandowski 2015, S. 9). Wie verhält es sich mit den Relevanzurteilen eines Nutzers, wenn beispielsweise ein Nachrichtenartikel entweder als organisches oder als Universal Search-Ergebnis präsentiert wird? Auf Fragen wie diese könnte eine ähnlich gelagerte Studie Antworten liefern und die Erkenntnisse zur Relevanzbeurteilung auf den Status quo der Suchmaschinen übertragen.

Das Experiment mit mehr Probanden und einer möglichst heterogen zusammengesetzten Population zu wiederholen wäre ebenfalls erforschenswert. Damit könnte ermittelt werden, ob die Erkenntnisse der vorliegenden Untersuchung auch für eine größere Grundgesamtheit gelten.

7 Literaturverzeichnis

Bar-Ilan, J., Keenoy, K., Levene, M., & Yaari, E. (2009). Presentation bias is significant in determining user preference for search results-A user study. *Journal of the American Society for Information Science and Technology*, *60*(1), 135–149. http://doi.org/10.1002/asi.20941

Blake, C. (2013). Eye-Tracking: Grundlagen und Anwendungsfelder. In W. Möhring & D. Schlütz (Eds.), *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft* (pp. 367–387). Wiesbaden: Springer Fachmedien Wiesbaden. http://doi.org/10.1007/978-3-531-18776-1_20

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

Buscher, G., Dumais, S. T., & Cutrell, E. (2010). The good, the bad, and the random. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10* (p. 42). New York, New York, USA: ACM Press. http://doi.org/10.1145/1835449.1835459

Cutrell, E., & Guan, Z. (2007). What are you looking for? In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 407). New York, New York, USA: ACM Press. http://doi.org/10.1145/1240624.1240690

Duchowski, A. T. (2003). *Eye Tracking Methodology : Theory and Practice*. London [u.a.]: Springer.

Eckstein, P. P. (2016). *Angewandte Statistik mit SPSS* (8. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. http://doi.org/10.1007/978-3-658-10918-9

Ellis, S., Candrea, R., Misner, J., Craig, C. S., Lankford, C. P., & Hutchinson, T. E. (1998). Using Eye Tracking Data to Help Build Better Web Pages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(23), 1613–1613. http://doi.org/10.1177/154193129804202328

Frees, B., & Koch, W. (2015). Internetnutzung: Frequenz und Vielfalt nehmen in allen Altersgruppen zu. *Media Perspektiven*, (9), 366–377. Retrieved from http://www.ard-zdf-onlinestudie.de/fileadmin/Onlinestudie_2015/0915_Frees_Koch.pdf

Godijn, R., & Theeuwes, J. (2003). The Relationship Between Exogenous and Endogenous Saccades and Attention. In *The Mind's Eye* (pp. 3–26). Elsevier. http://doi.org/10.1016/B978-044451020-4/50002-5

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04* (pp. 478–479). New York, New York, USA: ACM Press. http://doi.org/10.1145/1008992.1009079

Guo, Q., & Agichtein, E. (2010). Towards predicting web searcher gaze position from mouse movements. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10* (pp. 3601–3606). New York, New York, USA: ACM Press. http://doi.org/10.1145/1753846.1754025

Höchstötter, N. (2007). Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten. *Information, Wissenschaft & Praxis*, *58*(3), 135–140. Retrieved from http://www2.bui.haw-hamburg.de/pers/ulrike.spree/ws2007_2008/suchverhalten.pdf

Hofer, N., & Mayerhofer, W. (2010). Die Blickregistrierung in der Werbewirkungsforschung: Grundlagen und Ergebnisse. *Der Markt*, *49*(3-4), 143–169. http://doi.org/10.1007/s12642-010-0039-2

Hotchkiss, G., Alston, S., & Edwards, G. (2005). Google eye tracking report: How searchers see and click on Google search results. *Enquiro Search Solutions Inc.*, (June). Retrieved from http://searchengineland.com/figz/wp-content/seloads/2007/09/hotchkiss-eye-tracking-2005.pdf

Huang, J., White, R. W., & Dumais, S. (2011). No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (p. 1225). New York, New York, USA: ACM Press. http://doi.org/10.1145/1978942.1979125

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 & Management*, *36*(2), 207–227. http://doi.org/10.1016/S0306-4573(99)00056-4

Keane, M. T., O'Brien, M., & Smyth, B. (2008). Are people biased in their use of search engines? *Communications of the ACM*, *51*(2), 49–52. http://doi.org/10.1145/1314215.1314224

Lewandowski, D. (2015). *Suchmaschinen verstehen*. Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-662-44014-8

Lewandowski, D. (2012). Informationskompetenz und das Potenzial der Internetsuchmaschinen. In W. Sühl-Strohmenger (Ed.), *Handbuch Informationskompetenz* (pp. 101–109). Berlin: de Gruyter Saur. Retrieved from http://searchstudies.org/tl_files/Publikationen_PDFs/2012/Informationskompetenz und das Potenzial der Internetsuchmaschinen (2012).pdf

Maschek.hu. (2016). Maschek.hu - Online Image Map Editor. Retrieved May 9, 2016, from http://www.maschek.hu/imagemap/imgmap

Münz, S., & Gull, C. (2013). *HTML5 Handbuch* (9. Aufl.). Haar bei München: Franzis-Verl. Retrieved from http://m.ciando.com/img/books/extract/3645222073_lp.pdf

O'Brien, M., & Keane, M. T. (2006). Modeling result-list searching in the world wide web: The role of relevance topologies and trust bias. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Vol. 28, pp. 1881–1886). Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2006/docs/p1881.pdf

Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2004). A day in the life of Web searching: an exploratory study. *Information Processing & Management*, *40*(2), 319–345. http://doi.org/10.1016/S0306-4573(03)00044-X

Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). The determinants of web page viewing behavior. In *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004* (Vol. 1, pp. 147–154). New York, New York, USA: ACM Press. http://doi.org/10.1145/968363.968391

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. http://doi.org/10.1111/j.1083-6101.2007.00351.x

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. *Encyclopedia of Human-Computer Interaction*, 211–219. Retrieved from http://www.csd.uoc.gr/~hy564/files/material/panel_08/Eye Tracking in Human-Computer Interaction and Usability Research.pdf

Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1. *On the Horizon*, *9*(5), 1–6. http://doi.org/10.1108/10748120110424816

Quirmbach, S. (2011). Usability und User Experience in Suchmaschinen. In D. Lewandowski (Ed.), *Handbuch Internet-Suchmaschinen 2* (pp. 229–263). Heidelberg: Akademische Verlagsgesellschaft AKA GmbH.

Rasch, B., Hofmann, W., Friese, M., & Naumann, E. (2010). *Quantitative Methoden Band 2: Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (3. Aufl.). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-05270-5

Sedlmeier, P., & Renkewitz, F. (2008). Experimentelle Designs. In *Forschungsmethoden und Statistik in der Psychologie* (pp. 123–180). München: Pearson Studium.

Shani, G., & Tractinsky, N. (2013). Displaying relevance scores for search results. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13* (p. 901). New York, New York, USA: ACM Press. http://doi.org/10.1145/2484028.2484112

SPSS Inc. (2002). *Linear mixed-effects modeling in SPSS : an introduction to the MIXED procedure*. Retrieved from http://www.spss.ch/upload/1107355943_LinearMixedEffectsModelling.pdf

Tobii AB. (2016). Tobii Studio User's Manual. *Tobii*, 1–161. Retrieved from http://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf

Tobii Technology. (2010). Tobii Eye Tracking - An introduction to eye tracking and Tobii Eye Trackers, 14. Retrieved from http://www.tobii.com/Global/Analysis/Training/WhitePapers/Tobii_EyeTracking_Introduction_WhitePaper.pdf?epslanguage=en

Usability.de. (2009). Goodbye Golden Triangle. Retrieved from http://www.usability.de/usability-user-experience/publikationen/download/start.html

Anhang 1: Beigabe (CD)

Inhalt der CD:

1. Bachelorarbeit als PDF-Version
2. Testleitfaden
3. Datenschutzvereinbarung
4. Einverständniserklärung
5. *Screenshots*\Screenshots aller SERPs, Snippets und Webseiten
6. *Rohdaten*\Kennzahlen aus Tobii Studio und Relevanzurteile
7. *SPSS*\Daten und Outputs

Anhang 2: Suchergebnisseiten zu einer Rechercheaufgabe

Nachfolgend Screenshots der drei SERP-Varianten (*normal, vertauscht, umgekehrt*) zur Rechercheaufgabe „Finde die Homepage des Tierparks Hagenbeck“. Die Screenshots aller Rechercheaufgaben in vollständiger Größe befinden sich auf der beigelegten CD (s. *Anhang 1*).

Anhang 2.1 SERP “normal“

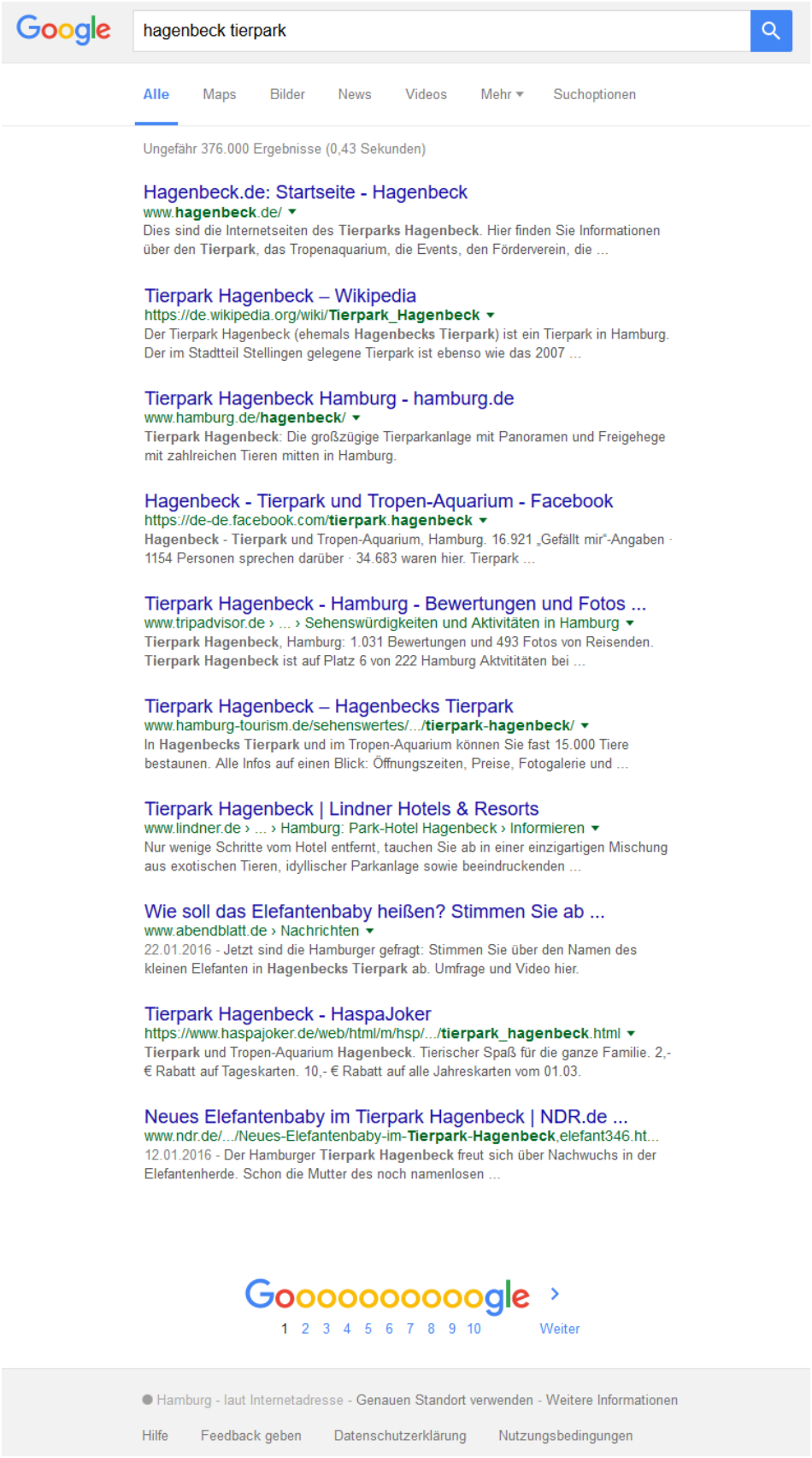

Abbildung 7: Beispiel SERP "normal"

Anhang 2.2 SERP “vertauscht“

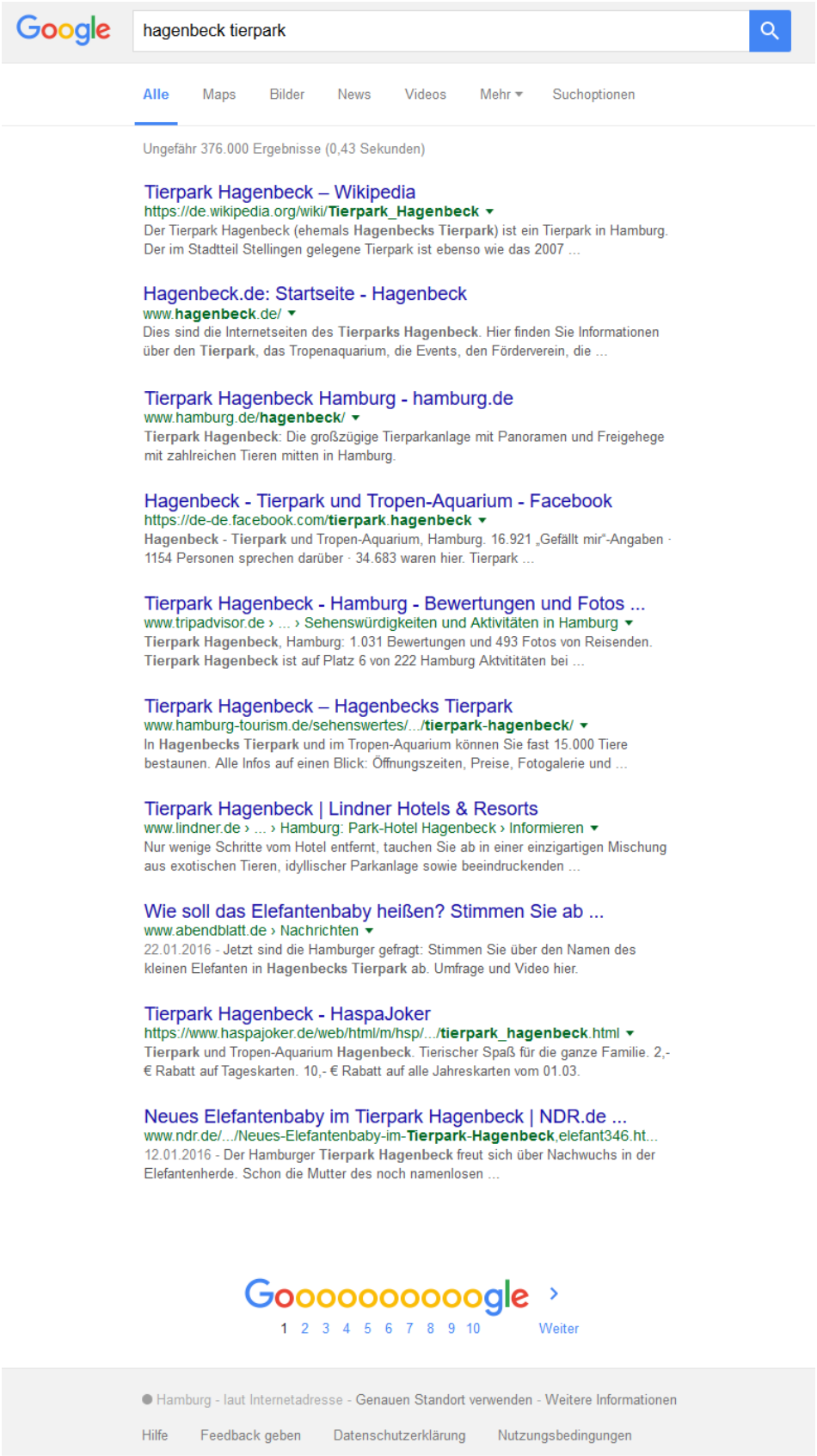

Abbildung 8: Beispiel SERP "vertauscht"

Anhang 2.3 SERP “umgekehrt“

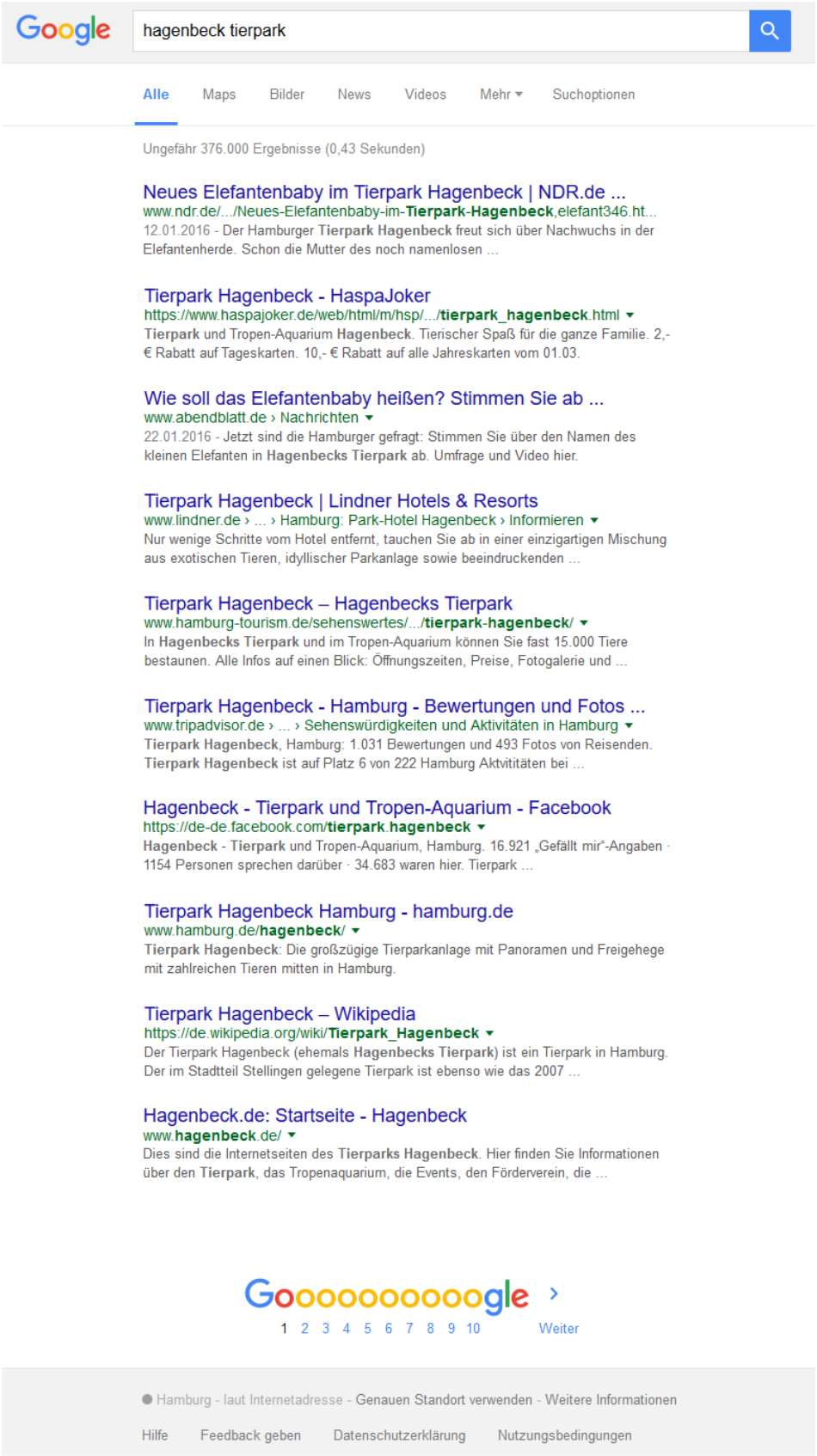

Abbildung 9: Beispiel SERP "umgekehrt"

Anhang 3: Dokumente zur Durchführung der Experimente

Anhang 3.1 Testleitfaden

Eye-Tracking-Untersuchung im Rahmen der Bachelorarbeit von Herrn Sebastian Schultheiß

Testleitfaden

---

1. Vorbereitungen

- **1.1 Rechner starten**
- **1.2 Raum lüften**
- **1.3 Festplatte für Datensicherung bereitlegen**
- **1.4 Getränke, sauberes Glas und Kekse bereitlegen**
- **1.5 Probandenvergütung und Quittung bereitlegen**
- **1.6 Einverständniserklärung und Datenschutzvereinbarung bereitlegen**

2. Begrüßung und Einführung Testperson

- **2.1 Ggf. Testperson abholen**
- **2.2 Begrüßung**

Hallo Frau/Herr xy/Vorname,

- *damit Sie sich nicht wundern, warum ich diesen Text ablese: Das mache ich zum einen, damit nichts vergessen wird, und zum anderen, um sicherzustellen, dass alle Teilnehmer/innen die Studie unter denselben Bedingungen durchführen.*
- *Möchten Sie geduzt oder gesiezt werden?*
- *Bedienen Sie sich auch gerne jederzeit an den Getränken und Keksen.*

herzlich willkommen am Department Information der HAW. Es freut mich sehr, dass Sie sich zur Teilnahme an meiner Untersuchung bereit erklärt haben.

Mein Name ist Sebastian Schultheiß, ich studiere im 6. Semester Bibliotheks- und Informationsmanagement an der HAW Hamburg und schreibe derzeit meine Bachelorarbeit.

Kurz gefasst geht es darin um die Relevanzbeurteilung von Google-Nutzern. Viel mehr möchte ich vorab nicht erzählen, da das Ihr Verhalten im Experiment beeinflussen könnte.

Kernbestandteil der Bachelorarbeit ist ein Eye-Tracking-Experiment, wofür wir heute hier sind und was etwa 45 bis maximal 60 Minuten in Anspruch nehmen wird.

Wenn Sie Fragen haben, stellen Sie diese gerne zu jedem Zeitpunkt. Je nach Fragestellung werde ich jedoch eventuell keine Antwort geben dürfen, ebenfalls, um die Untersuchung nicht zu verfälschen. Sämtliche anfallenden Daten werden selbstverständlich in nicht personenbezogener Form ausgewertet (mittels IDs). Dadurch werden durch die Testergebnisse keinerlei Rückschlüsse auf Ihre Person möglich sein. *(Verweis auf Datenschutzerklärung)*

3. Instruktionen

Wie erwähnt handelt es sich bei der Untersuchung um ein Eye-Tracking-Experiment. Ein Eye-Tracker ist ein Gerät, das mittels Infrarot die Positionen und Bewegungen Ihrer Pupillen erkennt und aufzeichnet. Dieses Verfahren ist gänzlich unschädlich und während des Experiments kaum wahrnehmbar. Zusätzlich zu Ihrem Blick- wird auch Ihr Klickverhalten auf sämtlichen dargestellten Webseiten aufgezeichnet. Das Experiment wird wie folgt ablaufen:

3.1 Rechercheaufgaben

3.1.1 Sie erhalten 10 Rechercheaufgaben.
3.1.2 Bitte führen Sie die Recherche genauso durch, wie Sie es auch privat tun würden.
3.1.3 Die Suchanfragen können Sie nicht frei eintippen. Sie erhalten vorgefertigte Google-Ergebnislisten mit jeweils 10 Ergebnissen.
3.1.4 Sie können alle Ergebnisse aufrufen.
3.1.5 Ich bitte Sie jedoch, innerhalb der einzelnen Webseiten nicht mehr weiter zu klicken. Dies ist für die Beantwortung der Rechercheaufgaben auch nicht erforderlich.
3.1.6 Das Zeitlimit pro Aufgabe beträgt 3 Minuten.

3.2 Relevanzbeurteilungen

3.2.1 Nach jeder Rechercheaufgabe bitte ich Sie, Relevanzbeurteilungen vorzunehmen.
3.2.2 Diese erfolgen per Schieberegler von links (Wert 0) nach rechts (Wert 100).
3.2.3 Alle 10 Treffer der Google-Liste (Snippets).
3.2.4 Alle 10 Webseiten-Inhalte (per Screenshots).

4. Abschluss der Instruktionen / Unterschriften einholen

4.1 Bevor wir mit der Studie beginnen, benötige ich noch Ihre Unterschriften auf der Einverständniserklärung, auf der Datenschutzvereinbarung sowie auf der Quittung Ihrer Vergütung:

- Einverständniserklärung unterschreiben lassen
- Datenschutzvereinbarung unterschreiben lassen
- Vergütung aushändigen und Quittung unterschreiben lassen

4.2 Hinweis: Bitte drücken Sie nicht „Esc“ bzw. „F10“, da dadurch die Aufzeichnung des Eye-Trackers gestoppt wird.
4.3 Haben Sie noch ungeklärte Fragen?
4.4 Gerne können wir nun mit dem Experiment beginnen.

5. Kalibrierung des Eye-Trackers

5.1 Bitte finden Sie eine bequeme Sitzposition.
5.2 Selbstverständlich haben Sie einen gewissen Bewegungsspielraum. Die Kalibrierung stellt lediglich eine optimale Anfangssituation her, sodass das Gerät auch bei Bewegungen Ihrerseits nicht den Kontakt zu Ihren Pupillen verliert.
5.3 Gleich wird ein roter Punkt auf dem Bildschirm erscheinen. Bitte folgen Sie diesem einfach mit den Augen.

6. Start der Untersuchung: Absolvieren der Rechercheaufgaben

	I) Navigationsorientierte Suchanfragen			
	Aufgabenstellung der Studie von Pan et al. (2007, S. 808)	**Gegenstück mit deutschem oder Hamburger Kontext**	**Suchanfrage (Query)**	**Richtiger Treffer (wenn mehrere möglich, dann *u.a.*)**
01	Find the homepage of Michael Jordan, the statistician.	Finde die Startseite der Homepage vom Schulz von Thun Institut für Kommunikation.	friedemann schulz von thun institut	*http://www.schulz-von-thun.de/*
02	Find the page displaying the route map for Greyhound buses.	Finde eine Übersicht über verschiedene Anbieter von Hafenrund-fahrten in Hamburg.	hafenrundfahrt hamburg übersicht	*http://www.hamburg.de/hafenrundfahrt/868760/hafenrundfahrten-hamburg/ (u.a.)*
03	Find the homepage of the 1000 Acres Dude Ranch.	Finde die Homepage des Tierparks Hagenbeck.	hagenbeck tierpark	*http://www.hagenbeck.de/startseite.html*
04	Find the homepage for graduate housing at Carnegie Mellon University.	Finde eine Webseite, die das Studenten-wohnheim Hammerbrook überblicksartig darstellt, inklusive Angaben zu Miete und Kontakt.	wohnheim hamburg hammerbrook	*http://www.studierendenwerk-hamburg.de/studierendenwerk/de/wohnen/wohnanlagen/detail/?id=766*
05	Find the homepage of Emeril - the chef who has a television cooking program.	Finde die Webseite des bekannten Kochs Christian Rach.	christian rach	*http://www.christianrach.de/ (u.a.)*

	II) Informationsorientierte Suchanfragen			
	Rechercheaufgaben der Studie von Pan et al. (2007, S. 808)	**Gegenstück mit deutschem oder Hamburger Kontext**	**Suchanfrage (Query)**	**Richtiger Treffer & korrekte Antwort**
06	Where is the tallest mountain in New York located?	Welches Gründungsjahr und welchen Firmensitz hat die älteste Schuhfabrik Deutschlands?	älteste schuhfabrik deutschland gründungsjahr firmensitz	*https://de.wikipedia.org/wiki/Peter_Kaiser_%28Schuhfabrik%29 (u.a.)* *1838, Pirmasens (Peter Kaiser Schuhfabrik)*
07	With the heavy coverage of the democratic presidential primaries, you are excited to cast your vote for a candidate. When are/were democratic presidential primaries in New York?	Wo arbeitet Norbert Hackbusch, Linken-Politiker und Mitglied der Hamburgischen Bürgerschaft?	norbert hackbusch linke hamburg	*https://de.wikipedia.org/wiki/Norbert_Hackbusch (u.a.)* *Gruner + Jahr*
08	Which actor starred as the main character in the original Time Machine movie?	Wer sind die Moderatoren des Fernseh-senders “Hamburg 1”?	hamburg 1 moderatoren	*https://de.wikipedia.org/wiki/Liste_der_Hamburg-1-Moderatoren* *Liste der Namen*
09	A friend told you that Mr. Cornell used to live close to campus—near University and Steward Ave. Does anybody live in his house now? If so, who?	Früher befand sich im Gebäude des Kunst- und Mediencampus die Frauenklinik Finkenau. Wann wurde diese gegründet?	frauenklinik finkenau gründung	*http://www.welt.de/print/die_welt/hamburg/article133115487/Wo-jedes-Jahr-6000-Babys-geboren-wurden.html (u.a.)* *1914*
10	What is the name of the researcher who discovered the first modern antibiotic?	Welcher Trainer ist in der Fußball-Bundesliga aktuell am längsten im Amt?	dienstältester trainer bundesliga aktuell	*http://www.rp-online.de/sport/fussball/bundesliga/bundesliga-diese-aktuellen-trainer-sind-am-laengsten-im-amt-bid-1.4871411* *Markus Weinzierl*

7. Abschluss der Untersuchung und Verabschiedung

7.1 Nochmals vielen Dank für Ihre Teilnahme an meiner Untersuchung. Dadurch haben Sie mir für meine Bachelorarbeit sehr weitergeholfen.

7.2 *Ggf. Testperson zum Ausgang begleiten.*

8. Nachbereitung

8.1 Prüfen, ob alle Dokumente mit ID der Testperson versehen sind.

8.2 Prüfen, ob alle Unterschriften vorhanden sind. (Einverständniserklärung, Datenschutzvereinbarung, Quittung)

8.3 Usability-Labor aufräumen und ggf. direkt für den nächsten Probanden vorbereiten.

8.4 Datensicherung mit externer Festplatte:

8.4.1 Nach jedem Probanden: Excel-Tabelle der Relevanzurteile (exportieren)

8.4.2 Am Ende des Tages: kompletten Projekt-Ordner

Anhang 3.2 Datenschutzvereinbarung

Datenschutzvereinbarung

zur Eye-Tracking-Untersuchung im Rahmen der Bachelorarbeit von
Herrn Sebastian Schultheiß an der Hochschule für Angewandte Wissenschaften Hamburg

Hiermit erkläre ich mich einverstanden, dass während des Experiments mein Blick- und Klickverhalten aufgezeichnet bzw. dokumentiert werden wird. Ein Eye-Tracker wird für die Aufzeichnung und Analyse meines Blickverhaltens zum Einsatz kommen.

Ich stimme zu, dass die erhobenen Daten gegebenenfalls zu Forschungszwecken vor einem geschlossenen Kreis präsentiert werden.

Die Aufzeichnung, Auswertung und Präsentation sämtlicher Daten und anderer Angaben wird in nicht personenbezogener Form erfolgen. Eine Weitergabe an Dritte erfolgt nicht.

Vollständiger Name in Druckbuchstaben _____

Unterschrift _____

Ort, Datum _____

Anhang 3.3 Einverständniserklärung

Einverständniserklärung

zur Teilnahme an einer Eye-Tracking-Untersuchung im Rahmen der Bachelorarbeit von Herrn Sebastian Schultheiß an der Hochschule für Angewandte Wissenschaften Hamburg

Ziel der Studie

Im Rahmen meiner Bachelorarbeit innerhalb des Studienganges Bibliotheks- und Informationsmanagement an der Hochschule für Angewandte Wissenschaften (HAW) Hamburg untersuche ich die Relevanzbewertung von Google-Nutzern in der Praxis. Mit der Teilnahme an dieser Studie helfen Sie mir sehr, meine Bachelorarbeit voranzubringen und Erkenntnisse über meine Fragestellung zu erlangen.

Ablauf

Kernbestandteil der Untersuchung sind 10 Rechercheaufgaben. Die Eye-Tracking-Methode wird eingesetzt, um zusätzlich zu Ihrem Klick- auch ihr Blickverhalten aufzuzeichnen. Nach jeder Rechercheaufgabe werde ich Sie darum bitten, Relevanzbewertungen abzugeben. Bewertet werden soll die Relevanz der dargestellten Google-Ergebnisse (Snippets) sowie der Webseiteninhalte im Kontext der jeweiligen Fragestellung. Insgesamt wird Ihre Teilnahme am Experiment etwa 45-60 Minuten dauern.

Aufwandsentschädigung

Für die vollständige Teilnahme an dieser Studie erhalten Sie 10 Euro.

Abbruch

Ihre Teilnahme an dieser Studie ist komplett freiwillig. Sie können jederzeit abbrechen, ohne dass Ihnen daraus Nachteile entstehen.

Fragen

Wenn Sie noch Fragen haben, können Sie diese jetzt oder jederzeit während der Studie stellen. Wenn Sie nach der Studie Fragen haben, können Sie mich per E-Mail an sebastian.schultheiss@haw-hamburg.de kontaktieren. Mit Ihrer Unterschrift bestätigen Sie, dass Sie mit den hier genannten Bedingungen einverstanden sind und dass Sie mir die Erlaubnis geben, Ihre Ergebnisse aus dieser Studie gemäß der beiliegenden Datenschutzvereinbarung zu verwenden.

Vollständiger Name in Druckbuchstaben _____

Unterschrift _____

Ort, Datum _____

Eidesstattliche Erklärung

Ich versichere, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen oder dem Sinn nach entlehnten Passagen sind durch Quellenangaben kenntlich gemacht.

Ort, Datum

Unterschrift
(Sebastian Schultheiß)