

Sprachgesteuerte Websuche im Test – Ein Erfahrungsbericht aus der Perspektive von Auftraggeber, Auftragnehmer und Durchführendem

Sebastian Sünkler¹, Friederike Kerkmann¹, Sonja Quirnbach² und Sebastian Schultheiß¹

Abstract: Beschrieben wird ein studentisches Projekt, das im Auftrag eines Telekommunikationsunternehmens die sprachgesteuerte Websuche vier verschiedener Systeme evaluierte und aus den Erkenntnissen Empfehlungen für die Gestaltung einer optimalen Voice Web Search ableitete. Der Beitrag versucht, die unterschiedlichen Perspektiven und Erwartungen von Auftraggeber, Projektleitung und durchführenden Studierenden zu beleuchten und miteinander in Beziehung zu bringen. Mit diesem Erfahrungsbericht sollen zum einen die konkreten Ergebnisse in ihren Grundzügen vorgestellt, zum anderen die unterschiedlichen Blickwinkel und Anforderungen der Beteiligten reflektiert sowie die wissenschaftlichen und forschungspraktischen Herausforderungen des innovativen Themas dargestellt werden.

Keywords: Websuche, Sprachsteuerung, Smartphone, Google Now, Apple Siri, Microsoft Cortana, Amazon Fire OS, Usability, Evaluation, Suchmaschinen

1 Einleitung

Die Popularität der Sprachsteuerung hat durch den rasanten Anstieg der Absatzzahlen von Smartphones stark zugenommen und wird genutzt, um Anrufe aufzubauen, Textnachrichten zu verfassen, Apps zu öffnen oder Websuchen durchzuführen. Aktuelle Studien zeigen, dass jeder zweite Deutsche die Sprachsteuerung anwendet [Bi16] und in den USA knapp 40% der Besitzer auf die Eingabe von Befehlen per Sprache zurückgreifen [PA16]. Wie effektiv, effizient und zufriedenstellend aber arbeiten derzeitige Systeme, speziell im Bereich der Websuche? Wie gut funktioniert „googeln“, wenn der Nutzer seine Suchanfrage nicht schreibt, sondern spricht? Diesen Fragen ist ein studentisches Projekt nachgegangen, das im Auftrag eines Telekommunikationsunternehmens die sprachgesteuerte Websuche von vier verschiedenen Systemen evaluierte und aus den gewonnenen Erkenntnissen Handlungsempfehlungen zur Gestaltung einer eigenen Voice Search Anwendung für den Auftraggeber ableitete.

In diesem Beitrag wird versucht, die drei unterschiedlichen Perspektiven (1) Auftraggeber, (2) Auftragnehmer bzw. Projektleitung und (3) Durchführender auf das

¹ Hochschule für Angewandte Wissenschaften, Fakultät Design, Medien und Information, Department Information, Finkenau 35, 22081 Hamburg, vorname.nachname@haw-hamburg.de

² Deutsche Telekom AG, Products & Innovation, T-Online-Allee 1, 64295 Darmstadt, s.quirnbach@telekom.de

Thema, das Projekt sowie die Ergebnisse zu beleuchten. Mit diesem Erfahrungsbericht sollen zum einen die konkreten Ergebnisse in die Fachcommunity transportiert, zum anderen die unterschiedlichen Blickwinkel und Anforderungen der Beteiligten reflektiert werden. Gleichzeitig sollen auch die wissenschaftlichen und forschungspraktischen Herausforderungen eines innovativen Themas wie die Websuche per Sprachsteuerung, für das es kaum etablierte Maße und Methoden gibt, diskutiert werden.

Der Beitrag gliedert sich in drei Abschnitte: Zunächst werden als Rahmen das Projekt sowie die Grundlagen der sprachgesteuerten Websuche vorgestellt. Im zweiten Abschnitt beschreiben nacheinander der Auftraggeber, die Auftragnehmer bzw. Projektleiter und ein Student, der mit der Durchführung beauftragt war, ihre jeweiligen Erfahrungen und Einschätzungen im Hinblick auf die Zielsetzung, die Durchführung und die Ergebnisse. Der Erfahrungsbericht endet mit einem gemeinsamen Fazit, das diese unterschiedlichen Perspektiven bündelt, die Chancen und Herausforderungen der Auseinandersetzung mit der sprachgesteuerten Websuche zusammenfasst und weitere Forschungsbedarfe konkretisiert.

2 Das Projekt „Ok Google... The End of Search as we know it“

Das Projekt „Ok Google... The End of Search as we know it. Sprachgesteuerte Suche im Test“ fand als studiengangübergreifendes Projekt der beiden Studiengänge *Medien und Information* und *Bibliotheks- und Informationsmanagement* am Department Information der Hochschule für Angewandte Wissenschaften (HAW) Hamburg im Wintersemester 2015/16 statt. Auftraggeber war die Deutsche Telekom AG; teilgenommen haben 15 Studierende beider Studiengänge unter der Leitung von zwei Projektleitern.

Die Zielsetzung bestand darin, die drei marktführenden Systeme zur sprachgesteuerten Websuche *Google Now*, *Apple Siri* und *Microsoft Cortana* sowie das inzwischen nicht mehr erhältliche *Amazon Fire OS* auf ihre Stärken und Schwächen hin zu evaluieren. Ausgehend davon sollten Handlungsempfehlungen für die Entwicklung einer sprachgesteuerten Websuche abgeleitet und eine optimierte Oberfläche konzipiert werden. Die Ergebnisse wurden in einem Abschlussbericht aufgearbeitet und dem Auftraggeber in einem Workshop vorgestellt und dort gemeinsam mit diesem weiterentwickelt.

3 Sprachgesteuerte Websuche

Die Geschichte der modernen sprachgesteuerten Websuche ist jung. Sowohl *Google* als auch *Apple* veröffentlichten im Jahr 2010 ihre Dienste *Google Voice Search* und *Apple Siri*. *Microsoft Cortana* wurde 2014 als Bestandteil von Windows 8.1 auf den Markt gebracht. Die drei genannten Systeme sind momentan die Marktführer in ihrem Segment, da sie integrierter Bestandteil der Betriebssysteme der Anbieter sind. Sie

nutzen dabei vergleichbare Technologien zur Verarbeitung von Spracheingaben und sind selbstlernende Systeme, die auf Cloud-Server-Architekturen zugreifen, um die Eingaben per Stimme durch ihre Nutzer auszuwerten und Websuchen durchzuführen. In der praktischen Anwendung bedeutet dies, dass ein Nutzer eine Suchanfrage einspricht, die durch Websuchmaschinen verarbeitet wird und die Ergebnisse dem Nutzer anschließend präsentiert werden. Die Darstellung und Verarbeitung hängen dabei von der Integration der Websuchmaschinen in das Betriebssystem des Gerätes ab. So wird die akustische Suchanfrage entweder direkt in Text umgewandelt und an eine Suchmaschine weitergeleitet oder es findet eine Aufbereitung der Suchergebnisse bezogen auf das System statt.

Alle diese Anbieter setzen dabei auf Suchmaschinentechnologien, die sie entweder selbst entwickelt haben (*Google, Microsoft Bing*) oder verwenden eine Kombination vorhandener Suchtechnologien. *Apple* nutzt *Google, Bing, Yandex* und weitere Suchmaschinen. Die Ergebnispräsentation unterscheidet sich bei den Anbietern dahingehend, dass *Google* auch bei Suchanfragen per Sprache eine Darstellung analog zur traditionellen Search Engine Result Page (SERP) bei textuellen Suchanfragen nutzt, während *Microsoft* und *Apple* in diesem Fall jeweils ein eigenes Layout der Ergebnisse präsentieren. Vereinzelt unterscheidet sich die Ergebnispräsentation auch dadurch, dass bestimmte Suchanfragen durch Spracheingabe auch akustisch beantwortet werden. Zusätzlich hebt sich die sprachgesteuerte Websuche von der traditionellen Suche dadurch ab, dass die Systeme zum Teil in der Lage sind, mit dem Anwender „zu sprechen“, indem sie in den Dialog mit ihm treten und sich so auf vorangegangene Suchen beziehen können (z. B. „Wie heißt der Präsident der Vereinigten Staaten?“ → Antwort → „Wie groß ist er?“ → Antwort).

Das Fire Phone von *Amazon*, dessen Sprachsteuerung keinen eigenen Namen trägt und von anderen sprachbasierten Amazon-Diensten wie bspw. *Alexa* abzugrenzen ist, ist inzwischen nicht mehr käuflich erhältlich. Das Gerät, das durch Netlock ausschließlich an das Netz der Telekom gebunden war, dennoch in den Test einzubeziehen, war Wunsch des Auftraggebers.

Eine detaillierte Darstellung des derzeitigen Entwicklungsstandes und Funktionsumfangs zu allen Geräten findet sich im Projektbericht, der unter <http://searchstudies.org/ok-google/> online zur Verfügung steht.

4 Perspektive des Auftraggebers

4.1 Zielsetzung und Erwartungen an das Projekt

Innerhalb meiner Arbeit in der Produktentwicklung von innovativen Informationstechnologien bei der Deutsche Telekom AG werde ich zunehmend mit technisches Endgeräten konfrontiert, die

- nur ein minimalgroßes Interface haben, wie Smartwatches,
- eine intelligente Schnittstelle zum Kunden bekommen, deren Benutzerschnittstellen noch nicht standardisiert sind bzw. sich erst entwickeln, wie Geräte rund um Smarthome oder Connected Car,
- mit neuen Steuerungsmöglichkeiten zur Mensch-Maschine-Kommunikation ausgestattet werden, wie Geste oder Sprache.

Diese Endgeräte sind für die Nutzer noch neu, weil sie zum einen noch nicht in deren Lebensalltag angekommen sind und zum anderen andere bzw. bisher ungelernete Benutzerschnittstellen aufweisen. Es gibt also eine große Veränderung der Benutzerschnittstellen. Das heißt, sie werden vielfältiger werden und nicht jede wird in der gewohnten Art erscheinen und ein klassisches grafisches interaktives Frontend haben und über Tastatur, Maus oder Berührung steuerbar sein. Festzuhalten ist aber, dass die neuen Endgeräte eine Suchfunktion besitzen bzw. über diese auch Suchen über verschiedene Indexe möglich sind; sei es eine Websuche oder eine Suche über eine Applikation in einem Katalog.

Eine schnelle Entwicklung vollzieht sich zur Zeit insbesondere im Bereich der Sprache als Interaktions- und Steuerungselement zwischen Mensch und Maschine, die zunehmend ausgereifter und intelligenter werden, z. B. *Siri* oder *Amazon Echo*.

Als Folge darauf war es für mich naheliegend zu forschen, inwieweit es unter Nutzern schon verbreitet ist, auch die Websuche über die Sprache zu bedienen. Aus diversen selbst durchgeführten Nutzerstudien zur Sprachsteuerung weiß ich, dass es noch nicht selbstverständlich ist, die Sprache als Steuerungselement zu nutzen. Die meisten Befragten gaben dabei an, Sprache bei Navigationssystemen im Auto, bei Telefonhotlines oder maximal bei Sprachnachrichten über *WhatsApp* zu nutzen.

Dies gab mir den Anlass aufzuspüren, wie Nutzer aktuell die Spracheingabe verwenden, wie sie damit suchen und wie ihre Erwartungshaltung ist. Da ich davon ausgegangen bin, dass es Qualitätsunterschiede zwischen den Anbietern der sprachgesteuerten Websuchen von *Google Now*, *Apple Siri*, *Microsoft Cortana* und *Amazon Fire OS* gibt, war auch ein Benchmark dieser Systeme als Grundlage nahe. Die marktführenden Systeme sollten auf ihre Stärken und Schwächen evaluiert werden. Ein weiterer Fokus war für mich zu erfahren, wie Nutzer nach einer Sucheingabe mit den Ergebnissen interagieren.

Ein besonderer Schwerpunkt war dabei auch die Präsentation der Suchergebnisse und wie Nutzer mit diesen umgehen. Hier besteht die grundlegende Frage darin, den Medienbruch Sprache und visuelle Ausgabe der Ergebnisse gleichzeitig in einem nutzerfreundlichen Kontext zusammenzubringen. Eine der Stärken von Smartphones ist, dass sie in einem mobilen Kontext von unterwegs verwendet werden können und der Nutzer jederzeit online ist. Es wurde in der Studie deshalb davon ausgegangen, dass sich die Nutzer geographisch bewegen und ihre Suchmotivation innerhalb eines bestimmten Kontexts heraus entsteht (bspw. abgebildet über die sog. Micro-Moments, von Google definierte Suchintentionen im mobilen Kontext, s. Tab. 2).

4.2 Methodendiskussion

Schon innerhalb meiner eigenen durchgeführten Recherche zum Thema fiel mir auf, dass es kaum öffentlich verfügbare Studien gibt, die sich mit dem Nutzerverhalten im Zusammenhang mit der Sprachsteuerung befassen. Das hatte sich dann auch bei der begleitenden Recherche der Studierenden bestätigt. Umso mehr wurden wir alle gleich schon zu Beginn des Projektes darin bestätigt, zu diesem Thema zu forschen.

Mehr denn je sind die Einschätzungen zur User Experience zur Spracheingabe in Bezug auf Websuchen oder Suchen allgemein noch ungenau und subjektiv. Daraus eine allgemeingültige Ableitung zu treffen, ist dann die Herausforderung für weitere Studien. Sprache ist ein sehr individuelles Kommunikationsmedium, das grundlegend von den soziokulturellen Bedingungen der Nutzer abhängt. Hierzu gehört beispielsweise auch das verwendete Vokabular zu Suche, Satzbau und vor allem das Wissen um die Sprachbefehle und die Auswirkungen auf den Erfolg der Suche.

Aufbauend auf meinen Studien rund um Nutzerevaluierung in Suchmaschinen [Qu12] ist hier eine Erweiterung der anzuwendenden Evaluationsmethoden und Einschätzung der User Experience notwendig. Es sollten Kriterien identifiziert werden, um ein optimales System zu konzipieren und darauf aufbauend auch um dieses Systeme von Beginn an in der Gebrauchstauglichkeit und letztlich seinem Erfolg beurteilen zu können.

Im Folgenden zwei Fragestellungen, welche die Komplexität der Nutzung über die Sprache verdeutlichen:

1. Welche akustischen Ankerpunkte bekommt der Nutzer zur Sprachausgabe mitgeliefert, wenn zu einer Suchanfrage viele Ergebnisse gefunden werden und die ersten fünf Suchergebnisse von der Liste vorgelesen werden? Wie viele Ergebnisse kann der Nutzer für sich im Kopf behalten und sich erinnern, um dann das für ihn passende auszusuchen?
2. Wie kann der Nutzer effektiv mit den Ergebnissen arbeiten, um komplexe Aktionen miteinander zu verknüpfen (bspw. in Form von Formulierungen wie „Suche mit den neuen Trailer von Ice Age 4 auf Deutsch, ohne Werbung und sende ihn per Whats App an meine Schwester.“).

Hier gibt es in Sachen Nutzerverhalten und User Experience noch großes Entwicklungs- und Forschungspotential, insbesondere da bis dato noch keine Standards bezüglich Sprachsteuerung vorhanden sind.

Aus der Studie kam heraus, dass sich die Probanden eine intelligente Nutzerführung seitens des Systems wünschen. Deswegen wurden *Google* und *Siri* etwas besser bewertet, da sie den Probanden intelligenter erschienen. Zudem wird ein stärkere Emotionalisierung der Sprache bevorzugt, denn es gefällt wenn das System lustig und natürlich reagiert, wie das bei *Siri* der Fall ist. Ein natürliches Kommunikationsverhalten wird also als besser empfunden, als eine Stimme, die mechanisch und künstlich klingt.

4.3 Einschätzung der Ergebnisse

Zusammenfassend zeigt die Studie, dass die Bedienung der Geräte über Sprache vielen Nutzern noch sehr fremd ist. Teilweise war die Nutzung der Spracheingabe in öffentlichen Raum für die Probanden sogar unangenehm, da es für sie schon in ihre Privatsphäre eindringt, wenn sie in der Öffentlichkeit ihre Suchanfragen preisgeben. Hinter jeder Sucheingabe befinden sich persönliche Vorlieben und Motivationen, die man eben nicht gern öffentlich bekannt geben möchte.

Eine andere Frage ist, wie die Nutzerführung aussieht. Einerseits finden Probanden es hilfreich und erwarten es, wenn mehrere Ergebnisse vorgelesen werden. Andererseits stellt sich die Frage danach, inwieweit man sich auch mehrere Ergebnisse merken kann, um diese zu reflektieren und das passende auszusuchen. Die intelligente Nutzerführung wird bei der Suche über die Sprachsteuerung somit einen großen Fokus bekommen.

Es gibt also noch viele offene Fragestellungen, die bisher weder im Kontext der Nutzung noch aus der Perspektive der Nutzer detaillierter untersucht worden sind. Interessant in diesem Zusammenhang sind auch die Fachgebiete Suchmaschinenoptimierung (Search Engine Optimization, SEO) und Online Marketing (OM). Beispielsweise wäre interessant herauszufinden, wie spezielle Topergebnisse oder Anzeigen über die Sprache ausgegeben werden können. Es ist aktuell noch kein Wissen vorhanden, um über die Möglichkeiten und vor allem Chancen von SEO oder OM in Bezug zur Sprache Aussagen zu machen.

Die Nutzung der Suche über die Sprache ist somit dato noch nicht im Fokus der UX-Betrachtung. Deshalb kann noch keine Handlungsempfehlung bezüglich User Experience getroffen werden. Jedoch können erste Tendenzen abgeleitet werden, insbesondere die Erwartungshaltung der Nutzer gegenüber dem System. Diese Studie stellt somit Weichen und kann als Grundlage für weitere Evaluationen herangezogen werden. Ich hoffe mir natürlich aus meiner Sicht, dass weitere Projekte diesbezüglich stattfinden.

5 Perspektive der Projektleiter

5.1 Zielsetzung und Erwartungen an das Projekt

Der allgemein formulierte Projektauftrag, Stärken und Schwächen der Systeme zu erheben, musste für die praktische Durchführung auf konkret zu beurteilende Qualitätskriterien heruntergebrochen werden. Analog zu den Qualitätsdimensionen traditioneller Desktop-Suchmaschinen nach [LH07] sollten daher auch die Voice Web Search Systeme anhand ihrer Treffer-Qualität, Recherche-Qualität sowie Usability beurteilt werden. Die Index-Qualität, die üblicherweise den vierten Qualitätsaspekt von Suchmaschinen definiert, wurde nicht gesondert betrachtet, da zu solchen Evaluierungen

Aspekte der Vollständigkeit, Abdeckung des Webs und der Aktualität untersucht werden müssten, die zum einen methodisch durch die Dynamik des Webs schwer zu messen und zum anderen für den genannten Anwendungsfall nicht relevant waren. Die Treffer-Qualität wurde dabei durch die Retrievaleffektivität, den Einsatz von Personalisierungsmaßnahmen sowie die Präsentation der Ergebnisse abgebildet, die Recherche-Qualität über die Verarbeitung der Suchanfragen und die Dialogfähigkeit der Systeme. Die Beurteilung der Usability erfolgte anhand des Zugangs, der Bedienung, der Barrierefreiheit sowie der Bereitstellung von Hilfen.

Zur Evaluation wurde den Studierenden ein Multimethoden-Ansatz bestehend aus verschiedenen experten- und nutzerbasierten Verfahren vorgegeben. Diese Methodenvielfalt diente sowohl dem konkreten Erkenntnisgewinn als auch dem Kennenlernen der Geräte und Systeme. Gleichzeitig bestand der Anspruch darin, den Studierenden ein breites Methodenrepertoire an die Hand zu geben. Tabelle 1 zeigt die eingesetzten Methoden sowie ihre jeweilige Zielsetzung in der Übersicht:

	Methode	Ziel inhaltlich	Ziel methodisch
		Qualitätsdimension	
expertenbasiert unter Laborbedingungen	Literaturrecherche	State of the Art	Kennenlernen des Systems/Geräts, Einüben von wissenschaftlichem Arbeiten
	Heuristik	Identifikation von ‚bedeutenden‘ / ‚katastrophalen‘ Usability-Problemen, Vergleich der Geräte	Kennenlernen des Systems/Geräts, Einüben der Methode
		Usability	
	Szenarienbasierter Walkthrough	Identifikation von Usability-Problemen aus Nutzersicht, Hineinversetzen in verschiedene Nutzertypen und Nutzungsszenarien	Kennenlernen der jeweils anderen Systeme/Geräte, Einüben der Methode + Persona-Entwicklung
		Usability	
	Retrievaltest	quantitative Analyse der Retrievaleffektivität unter Laborbedingungen, Vergleich der Geräte	Einüben der Methode
Treffer-Qualität			
nutzerbasiert unter Realbedingungen	Vorabbefragung	Erfassung demographische Daten und Vorerfahrung	Einüben der Methode, „Angst vorm Nutzer verlieren“
	Teilnehmende Beobachtung	Identifikation von Stärken und Schwächen unter realen Bedingungen	Einüben der Methode
		Recherche-Qualität, Treffer-Qualität, Usability	
	Retrospective Thinking Aloud	Identifikation von Stärken und Schwächen unter realen Bedingungen	Einüben der Methode
		Recherche-Qualität, Treffer-Qualität, Usability	
SUS-Fragebogen	quantitative Analyse der	Einüben der Methode	

		Gebrauchstauglichkeit, Vergleich der Geräte	Einüben der Methode
		Usability	
	Nachbefragung	Identifikation von Stärken und Schwächen in der Rückschau	
		Recherche-Qualität, Treffer-Qualität, Usability	

Tab. 1: Eingesetzte Methoden und Ziele

Dem Walkthrough, dem Retrievaltest sowie der Nutzerstudie zugrunde lagen Suchanfragen, die sich jeweils an den von *Google* formulierten Micro-Moments³ [Go15] orientierten, um so die verschiedenen Suchintentionen, die im mobilen Kontext auftreten können, gleichmäßig zu berücksichtigen. Tabelle 2 zeigt die vier genutzten Micro-Moments in der Übersicht:

Micro-Moment	Suchintention	Beispiel
WANT-TO-KNOW	informationsorientierte Suchanfragen	Ein Nutzer will sich über ein Produkt informieren, das er in diesem Moment in einer TV-Werbung sieht.
WANT-TO-GO	lokale Suche	Ein Nutzer sucht nach einem italienischen Restaurant in seiner Nähe.
WANT-TO-DO	Howto-Suche	Ein Nutzer sucht nach einem Rezept für einen Apfelkuchen.
WANT-TO-BUY	transaktionsorientierte Suche	Ein Nutzer befindet sich einem Geschäft und nutzt die Suche als Entscheidungshilfe beim Produktkauf.

Tab. 2: Micro-Moments

5.2 Methodendiskussion

Anwendungen der Sprachsteuerung lassen sich bei der praktischen Nutzung nicht immer trennscharf voneinander unterscheiden; auch die Abgrenzung der im Rahmen des Projektes relevanten Websuche von allgemeiner Suchfunktion gestaltete sich nicht einfach. Häufig bildet die Suche den Zugang zu der eigentlichen Anwendung. Dies trifft insbesondere auf die Nutzung von Personal Assistants zu, die Anwender bei der Erfüllung von Aufgaben unterstützen sollen. Z. B. wird vor dem Aufbau eines Anrufs eine Suche in der Kontaktliste auf dem Smartphone durchgeführt. Dieser Prozess zählt zu den Voice-Search-Anwendungen der Sprachsteuerung ohne Bezug auf eine Websuche. Voice-Search-Anwendungen lassen sich von Voice Web Search dadurch abgrenzen, dass Voice-Search sich sowohl auf Offline- und Online-Inhalte bezieht, dabei aber nicht zwangsweise als Websuche agiert. Voice Web Search hingegen bezieht sich ausschließlich auf die Inhalte aus Websuchmaschinen und ist damit unabhängig von den

³ Mobile Suchen sind im Gegensatz zur Suche an stationären Geräten praktisch von überall möglich und damit stärker von Spontaneität und Ungeduld der Nutzer getrieben. Dieses Verhalten bilden die Micro-Moments ab, indem sie die verschiedenen Momente, in denen ein Nutzer mit seinem Smartphone suchen kann, und die sich daraus ergebenden Suchbedürfnisse beschreiben.

gespeicherten Inhalten auf dem Gerät. Diese inhaltliche und sprachliche Abgrenzung war im Projekt ebenso wie die Differenzierung von Gerät vs. System nicht immer einfach und blieb auch bis zum Projektabschluss z.T. unklar – Ein Zustand, der bisweilen zu Unsicherheit und Unschärfe führte.

Die methodische Herausforderung des Vorhabens bestand im Fehlen etablierter Standardmethoden für die Evaluation von Voice Web Search bzw. in deren (Nicht-)Übertragbarkeit. So existiert bislang keine Heuristik für diesen Bereich. Ansätze, wie die Guidelines zur Navigation und Orientierung auf Smartphones aus der Grundlagenstudie von [eR14] bilden zwar eine gute Ausgangsbasis für mobile Websites, sind jedoch nicht auf Sprachsteuerung und Websuche übertragbar. Mangels etablierter Heuristiken wurde eine eigene Heuristik entwickelt, die folgende potentielle Problemkategorien abprüfte: Zugang/Einstieg, Bedienung, Suchanfrageverarbeitung, Personalisierung, Barrierefreiheit, Ergebnispräsentation und Hilfestellung [SK16]. Die Überprüfung der Tragfähigkeit dieser selbstentwickelten Heuristik steht noch aus. Auch bei der Durchführung des Retrievaltests mussten gegenüber der klassischen Variante Einschränkungen in Kauf genommen werden: So waren den Juroren die Systeme bei der Bewertung bekannt. Außerdem variierte die Darstellung bzw. Ausgabe der Ergebnisse (akustisch vs. herkömmliche Trefferdarstellung mit Ergebnisliste) so stark zwischen den Systemen, dass eine Aussage über die Retrievaleffektivität auf Dokumentenebene nicht möglich war.

Im Rahmen der Nutzertests blieb der Grundsatz, Probanden die Anwendung ausschließlich mit dem Endgerät testen zu lassen, dass sie auch selbst besitzen (bzw. bestenfalls ihr eigenes Gerät zur Studie mitzubringen) [eR13], unberücksichtigt. Hintergrund war zum einen die geringe Verbreitung des Microsoft- und Fire-Phones, zum anderen hätten Verzerrungen durch Personalisierungsmaßnahmen auf dem eigenen Gerät nicht systematisch ausgeschlossen werden können.

5.3 Einschätzung der Ergebnisse

Der genutzte Methodenmix war ein Versuch, etablierte Methoden der Usability- und Suchmaschinen-Forschung für ein neues Anwendungsfeld zu adaptieren und miteinander zu kombinieren. Dabei kam es zu den beschriebenen Methodenschwächen und zum Teil nur bedingt aussagekräftigen Ergebnissen. Stärke und Schwäche von Usability-Evaluierungen ist die überwiegende Nutzung „weicher“, subjektiver qualitativer Usability-Kriterien. Das zeigt sich auch in den Projektergebnissen, die nach der Durchführung der Tests keinen eindeutigen Sieger festmachen können, wobei Google von allen getesteten Systemen noch die besten Ergebnisse erzielte. Gründe dafür liegen zum einen in den „weichen“, wenig greifbaren Maßen, zum anderen in der Anzahl der Evaluatoren, die dieselben Sachverhalte möglicherweise unterschiedlich beurteilten. Schwierigkeiten ergaben sich auch durch eine Affinität der Studierenden und Probanden zu bestimmten Systemen, was eine objektive Beurteilung teilweise einschränkte.

Dennoch konnten die Ergebnisse der verschiedenen Untersuchungsschritte genutzt werden, um daraus Erfolgskriterien abzuleiten, die eine nutzerfreundliche Voice Web Search kennzeichnen. Diese Kriterien können herangezogen werden, um bestehende Systeme in ihrer Gebrauchstauglichkeit zu beurteilen oder um ein optimiertes System zu konzipieren. Tabelle 3 zeigt die gesammelten Erfolgskriterien in der Übersicht:

Zugang/ Einstieg	Bedienung	Suchanfrage- verarbeitung	Personalisierung	Barrierefreiheit	Ergebnispräsentation	Hilfestellung
Eindeutig als Sprachsteuerung erkennbar	Lange Sprechpausen möglich	Akzente und Dialekte werden korrekt erkannt	Standort wird bei der Suchanfrage einbezogen	Grafiken in den Ergebnissen werden in ausreichender Größe dargestellt	Sprachausgabe bei Faktentrecherche	Fehlermeldungen sind hilfreich
	Brauchbare Ergebnisse bei Stichwortsuchen	Englische Anfragen werde korrekt erkannt	Ausgabesprache wird der Eingabesprache angepasst	Die Schriftgröße bei den Ergebnissen ist ausreichend	Akustisches Feedback zu den Ergebnissen nach der Suchanfrageverarbeitung	Aktive Unterstützung bei der Bedienung
	Interaktion mit Suchergebnissen per Sprache möglich	Unbekannte Abkürzungen können interpretiert werden		Kontrast bei der Ergebnisdarstellung ist angemessen	Sprachausgabe direkt zu den Suchergebnissen wird unterstützt	Links zur Hilfe sind vorhanden und gut sichtbar
	Suchverarbeitung per Sprache beenden	Korrekte Erkennung der Anfrage		Akustische Wiedergabe aller Suchergebnisse	Passende Darstellungen zu den Micro-Moments	Hilfestellung in Textform wird angeboten
	Unterstützung von Dialogen	Autokorrektur wird unterstützt		Schriftvergrößerung wird unterstützt	Die Quellen der Suchergebnisse sind immer erkennbar	Hilfe bei Null-Treffer-Seiten und nicht verstandenen Anfragen
	Echtzeitfeedback bei der Verarbeitung				KWIC wird unterstützt	
	Zuverlässige Erkennung bei lauten Umgebungsgläuschen				Die Trefferbeschreibungen sind verständlich	
					Die Ergebnisse sind auf dem aktuellen Stand	

Tab. 3: Erfolgskriterien von Voice Web Search Anwendungen [SK16]

Für die praktische Arbeit mit den Ergebnissen ergeben sich aber andere Anforderungen als an die wissenschaftliche Aussagekraft. Für eine Sicherstellung der wissenschaftlichen Aussagekraft ist die Entwicklung eines Forschungsdesigns mit Berücksichtigung der Reliabilität, Validität und Objektivität der Studien notwendig, was aber nicht im Fokus der Projektarbeit stand. Vielmehr ging es darum festzulegen mit dem Einsatz wissenschaftlichen Methoden praxistaugliche Ergebnisse generiert werden. Dabei fand eine ständige Abwägung zwischen Wissenschaftlichkeit und Pragmatismus statt, die zu den teilweise unscharfen Bewertungs- und Erfolgskriterien von Voice Search Anwendungen führten. Die Projektergebnisse bieten in jedem Fall einen nutzbaren ersten Einblick zu den Stärken und Schwächen bestehender Anwendungen der sprachgesteuerten Websuche. Die Handlungsempfehlungen sind ein brauchbarer Einstieg bei der Gestaltung und Entwicklung eines eigenen Dienstes.

6 Perspektive des Durchführenden

6.1 Zielsetzung und Erwartungen an das Projekt

Als Studierender stieg ich mit einer relativ klaren Erwartungshaltung in das Projekt ein. Diese wurde insbesondere durch den hohen Innovationsgehalt des Themas der sprachgesteuerten Websuche geprägt. Denn insbesondere aufgrund fehlender etablierter Evaluationsmethoden war von Beginn an deutlich, dass ein gewisser Handlungsspielraum gegeben war.

Zur privaten Vorbereitung machte ich mich zunächst mit wesentlichen Funktionen der sprachgesteuerten Websuche vertraut, die ich als Privatanwender bislang nicht verwendet hatte. Somit war die Zielsetzung an das Projekt zweigeteilt. So wollte ich die auf dem Markt verfügbaren Systeme einerseits grundlegend kennenlernen. Insbesondere von Interesse war allerdings die Anwendung diverser Methoden zur Evaluation der Systeme, worauf im Folgenden eingegangen werden wird.

6.2 Methodendiskussion

Der unter 5.1 beschriebene Multimethoden-Ansatz erwies sich für uns Studierende als Gewinn. Vor deren Durchführung konnten wir die verschiedenen Methoden zunächst kennenlernen, indem wir diese auf ihre Einsetzbarkeit hin überprüft haben. Ausschlaggebend für den Lernerfolg war schließlich die unmittelbare Anwendung, die in experten- und nutzerbasierte Evaluation unterteilt war.

Die expertenbasierte Evaluation bestand aus Literaturrecherche, Heuristik, szenariobasiertem Walkthrough sowie einem Retrievaltest. Diese war aus zweierlei Gründen wichtig. Zum einen waren die Voice Web Search Systeme den meisten Studierenden in ihren Funktionsweisen unbekannt. Insbesondere die Durchführung von Recherche und Heuristik haben demnach sichergestellt, dass sich jeder Projektteilnehmer umfangreich mit den jeweiligen Geräten beschäftigt hat, bevor die Nutzertests begannen. Des Weiteren diente die expertenbasierte Evaluation als wichtige Grundlage für die Nutzertests. Es konnten bereits vorab bestimmte Schwächen der Systeme ausgemacht werden, denen in den Nutzertests besondere Beachtung geschenkt worden ist. So wurde während der Durchführung der Heuristik ermittelt, dass die Systeme oftmals nicht abwarten, bis der Nutzer die Suchanfrage zu Ende gesprochen hat. Anhand der Nutzertests konnte dieser Eindruck bestätigt werden.

Die Nutzertests können als Kernbestandteil der gesamten Evaluation bezeichnet werden. Umrahmt von einer Vor- und Nachbefragung wurden die Probanden während der Tests gebeten, ihre Eindrücke zu schildern (Retrospective Thinking Aloud) und zum Schluss einen Fragebogen zur Gebrauchstauglichkeit (System Usability Scale, SUS [Br86]) auszufüllen.

Für die Tests konnten Probanden akquiriert werden, die ein sehr heterogenes Probandenfeld abbildeten. So war der jüngste Proband 8, die älteste 46 Jahre alt. Diese Vielfalt stellte sich schon während der einzelnen Durchläufe als vorteilhaft heraus, weil sich die unterschiedlichen Altersgruppen in der Benutzung der sprachgesteuerten Websuche teilweise stark unterscheiden haben. Da reale Anwender solcher Systeme ebenso divers sind, ist dieser Aspekt der Evaluation als besonders positiv hervorzuheben.

Als Resultat der Evaluation wurden Empfehlungen für die Nutzerführung einer optimalen sprachgesteuerten Websuche ermittelt. Dafür wurde per Flussdiagramm ein Suchprozess erstellt, angefangen von der Sucheingabe bis hin zur Ergebnisauswahl. Abbildung 1 zeigt beispielhaft den ersten Schritt des Sucheinstiegs innerhalb einer optimalen Voice Web Search Anwendung.

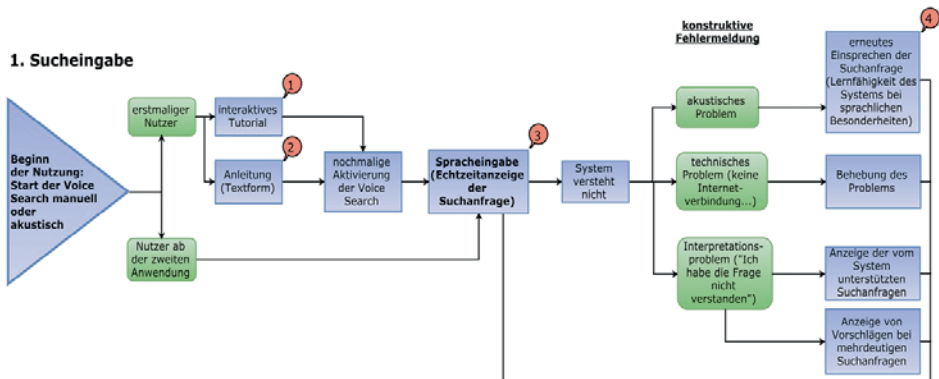


Abb. 1: Sucheinstieg innerhalb einer optimalen Voice Search [SK16]

Zusätzlich wurden mögliche Nutzerinterfaces skizziert, die einzelne Schritte innerhalb des Prozesses visualisieren sollten. Sämtliche dieser Empfehlungen wurden gruppenübergreifend erarbeitet. Dadurch vereinigten sich die Evaluierungsergebnisse aller inhaltlichen Arbeitsgruppen und es wurde deutlich, wie ähnlich die Erfahrungen aus Experten- sowie Nutzersicht waren. Abbildung 2 illustriert die Oberflächen für ein interaktives Tutorial sowie eine schriftliche Anleitung für den Nutzer, der die sprachgesteuerte Websuche das erste Mal nutzt.



Abb. 2: Sucheinstieg innerhalb einer optimalen Voice Search [SK16]

6.3 Einschätzung der Ergebnisse

Die schon unter 5.3 erwähnten „weichen“ Usability-Kriterien erschwerten auch für uns Studierende die Durchführung und Auswertung der einzelnen Methoden.

Dies äußerte sich bereits in den Heuristiken, die Usability-Probleme identifizieren und gewichten sollten. Hat beispielsweise eine Gruppe fehlende Funktionen zur Barrierefreiheit als „Usability-Katastrophe“ gewichtet, könnte eine andere Gruppe diese Gewichtung sehr unterschiedlich vorgenommen haben, obwohl deren System identische Mängel aufweist. So kann ein Vergleich der einzelnen Geräte auch lediglich als grobe Einschätzung betrachtet werden.

Bei der Konzipierung der optimalen Voice Web Search hingegen waren sich alle Gruppen zügig einig, obwohl als Grundlage dieselben, subjektiven Ergebnisse verwendet worden sind. Dies kann dadurch begründet werden, dass die grundsätzlichen Stärken und Schwächen der Systeme lediglich in ihren Ausprägungen, nicht aber in ihrem Vorhandensein zur Diskussion standen.

Eine interessante Fragestellung für die Zukunft wäre es zu ermitteln, wie Probanden auf einen Prototyp der konzipierten Voice Web Search reagieren würden. Dies wurde selbstverständlich nicht durch die Zielsetzung des Projekts abgedeckt, könnte aber die vorliegenden Erkenntnisse aufgreifen und diese in funktionierende Prototypen umsetzen.

7 Fazit

Das Projekt kombinierte Methoden der Usability- und Suchmaschinenevaluierung ebenso wie Grundlagen des prototypischen Designs und passte diese jeweils auf mobile Endgeräte und die Steuerung durch Sprache an. Dabei basierten diese Adaptionen überwiegend auf eigenen, praxisgetriebenen Überlegungen und weniger auf wissenschaftlich bestätigten Ansätzen. Dies zeigt die Chancen und Herausforderungen, die die Auseinandersetzung mit einem innovativen Thema, für das es keine Erfahrungen gibt, mit sich bringt: Zum einen beinhaltet ein solches Projekt eine große Freiheit, in dem die Methoden, Vorgehensweisen und Anwendungsbezüge mangels Standards frei gewählt werden können. Zum anderen jedoch existiert kein in der Wissenschaft oder Praxis geprüfter Handlungsrahmen, an dem sich das Vorhaben orientieren kann. Gründe für eine fehlende wissenschaftliche Grundlage zu der systematischen Evaluierung der sprachgesteuerten Suche liegen in der jungen, innovativen Technologie bzw. der erst jetzt relevanten Verbreitung vorangetrieben durch die Verfügbarkeit und den hohen Absatz von smarten Devices. Wie auch in anderen Bereichen der Forschung und Suchmaschinenforschung lassen sich solche Daten erst erheben, wenn sich solche Technologien etablieren. Bisherige Studien zur Nutzung werden momentan durch die Marktführer selber durchgeführt. Die Methoden und die Rohdaten stehen der Wissenschaft aber nicht zur Verfügung.

Damit zeigt sich erheblicher weiterer Forschungsbedarf: Auf Methodeneben werden bspw. eine auf ihre Tragfähigkeit hin überprüfte Heuristik benötigt, Ansätze, wie akustische und herkömmliche Treffer gleichermaßen in ihrer Effektivität beurteilt werden können oder Erkenntnisse über die Auswirkungen des retrospektiven Lauten Denkens oder den Einfluss eines Fremdgerätes im Rahmen einer Nutzerstudie. Auch der Frage, wie Akustik beschrieben und dargestellt werden kann, stellt sich.

Auf inhaltlicher Ebene lässt sich zunächst der Bedarf an einem allgemein anerkannten Begriffsverständnis festhalten. Die in diesem Zusammenhang relevanten Elemente (Anbieter / System / Gerät, Suche / Web-Suche / Sprachsuche, Persönlicher Assistent / Sprachsteuerung etc.) müssen eindeutig voneinander differenziert und beschrieben werden. Weiterer Forschungsbedarf ergibt sich im Hinblick auf Fragen der Personalisierung, des Datenschutzes oder der tatsächlichen Barrierefreiheit für Menschen mit Behinderungen.

Trotz der beschriebenen Schwächen und weiterhin offenen Fragen kann das Projekt als erfolgreich bezeichnet werden. Die identifizierten Erfolgskriterien bilden einen soliden Ausgangspunkt für die weitergehende Auseinandersetzung mit der sprachgesteuerten

Websuche und die Weiterentwicklung entsprechender Anwendungen. Die Studierenden, die bis zum Projekt weder über Erfahrung in der sprachgesteuerten Websuche noch in der Durchführung von Usability- oder Suchmaschinenevaluationen verfügten, konnten von der vermittelten Methodenvielfalt profitieren und für das Thema der sprachgesteuerten Websuche sensibilisiert werden.

8 Literaturverzeichnis

- [Bi16] Bitkom: Sprachsteuerung setzt sich bei Smartphones durch, <https://www.bitkom.org/Presse/Presseinformation/Sprachsteuerung-setzt-sich-bei-Smartphones-durch.html>, 2016.
- [Br86] Brook, J.: SUS – A quick and dirty usability scale, http://cui.unige.ch/isi/icle-wiki/_media/ipm:test-suschart.pdf, 2016.
- [eR14] eResult, <http://www.eresult.de/ux-wissen/forschungsbeitraege/einzelansicht/news/mobile-web-navigation-navigieren-und-orientieren-auf-dem-smartphone/>, Stand: 11.04.2016
- [eR13] eResult, <http://www.eresult.de/ux-wissen/forschungsbeitraege/einzelansicht/news/usability-testing-von-responsive-webdesigns/>, Stand: 11.04.2016
- [Go15] Google, <https://think.storage.googleapis.com/images/micromoments-guide-to-winning-shift-to-mobile-download.pdf>, Stand: 01.04.2016
- [LH07] Lewandowski, D.; Höchstötter, N.: Qualitätsmessung bei Suchmaschinen. System- und nutzerbezogene Evaluationsmaße, Informatik Spektrum 30(2007)3.
- [PA16] Parks Associates: Almost 40% of U.S. smartphone owners use voice recognition software, <http://www.parksassociates.com/blog/article/pr-01272016>, 2016.
- [Qu12] Quirnbach, S.: Suchmaschinen User Experience, Usability und nutzerzentrierte Website-Gestaltung, Springer-Vieweg, 2012
- [SK16] Sünkler, S.; Kerkmann, F.: Ok Google... The End of Search as we know it. Projektbericht zum Studienprojekt „Ok Google... The End of Search as we know it. Sprachgesteuerte Suche im Test“, <http://searchstudies.org/ok-google/>, Stand: 04.05.2016