

Gero Wenderholm

Alexa, wozu braucht man eine Computer-Maus?

Sprachassistenten verbreiten sich deutlich schneller als Handys, Tablets oder jedes andere technische Gerät zuvor, und große Konzerne wie Microsoft, Apple oder Samsung, allen voran aber Google und Amazon, liefern sich einen erbitterten Kampf um die Vorherrschaft als Gatekeeper für das Internet der Zukunft. Die Vorteile für Nutzer und die Chancen für Firmen sind gleichermaßen groß; ein Milliardenmarkt entsteht. Gleichzeitig ist das Angebot an sinnvollen Anwendungen für die digitalen Helfer noch vergleichsweise gering. Doch statt Aufbruchsstimmung herrscht vielerorts Zurückhaltung und Ratlosigkeit. Dabei ist der Einstieg einfacher als gemeinhin angenommen ...

Platon bezeichnete das geschriebene Wort in seinem Phaidros-Dialog einst als „leblose Kopie der beseelten Sprache“, die sich ihren „Adressaten nicht aussuchen“ könne und zudem „Klarheit und Vollständigkeit“ vermissen lasse.

Überträgt man diese frühe Medienkritik in die heutige Zeit und kehrt sie um, werden einige Vorzüge einer Sprachschnittstelle gegenüber der bislang überwiegend textuellen Bedienung des Internets deutlich:

- » Ein verbaler Dialog ist emotionaler und natürlicher als das Bedienen einer Tastatur.
- » Bedürfnisse können mit Sprache exakt artikuliert und ggf. korrigiert werden.
- » Sprachassistenten lernen Nutzer kennen und ermöglichen ein individuelles Erlebnis.

Hinzu kommt, dass Menschen im Schnitt dreimal so schnell sprechen wie sie schreiben und – laut einer Untersuchung von Neuro Insights (www.jwtintelligence.com/trend-reports/speak-easy/) – gesprochene Suchen wesentlich entspannter durchgeführt werden als jene mit einer Tastatur. Wer mit der Stimme sucht, formuliert oft ganze Sätze und Fragen, wie sie einem realen Gegenüber gestellt würden. Wäre es möglich gewesen, dann hätten wir – so scheint es – das Internet schon immer mithilfe unserer Stimme bedient. Yariv Adan, Chefentwickler des Google Assistants, vermutet gar, dass uns eine viele größere Revolution bevorsteht als jene des mobilen Webs durch Smartphone und Tablets: die Ubiquität des Internets.

Internet überall für alle

Durch die rasante Verbreitung von Smartphones hat inzwischen mehr als die Hälfte der Weltbevölkerung Zugang zum Internet und auch die vergleichsweise einfache Bedienung eines Tablets hat dazu geführt, sehr junge und ältere Menschen erstmalig online zu bringen. Die Sprachsuche ebnet nun den Weg für einen Personenkreis, dem die klassische Internetnutzung schwerfällt oder bislang unmöglich war; sei es

„This time, the connected device is not just changing‘ – it’s disappearing all together. We are no longer interacting with a physical device, but rather with an ambient AI entity that can be accessed from ... everywhere!“ –

Y. Adan, Product Lead Google Assistant

DER AUTOR



Gero Wenderholm seit ist über 15 Jahren im Online-Marketing aktiv. Der frühere Tchibo Head of SEO/SEA und BVDW-Experte unterstützt große Firmen als freier Berater bei der Entwicklung zukunftsweisender Online-Strategien.

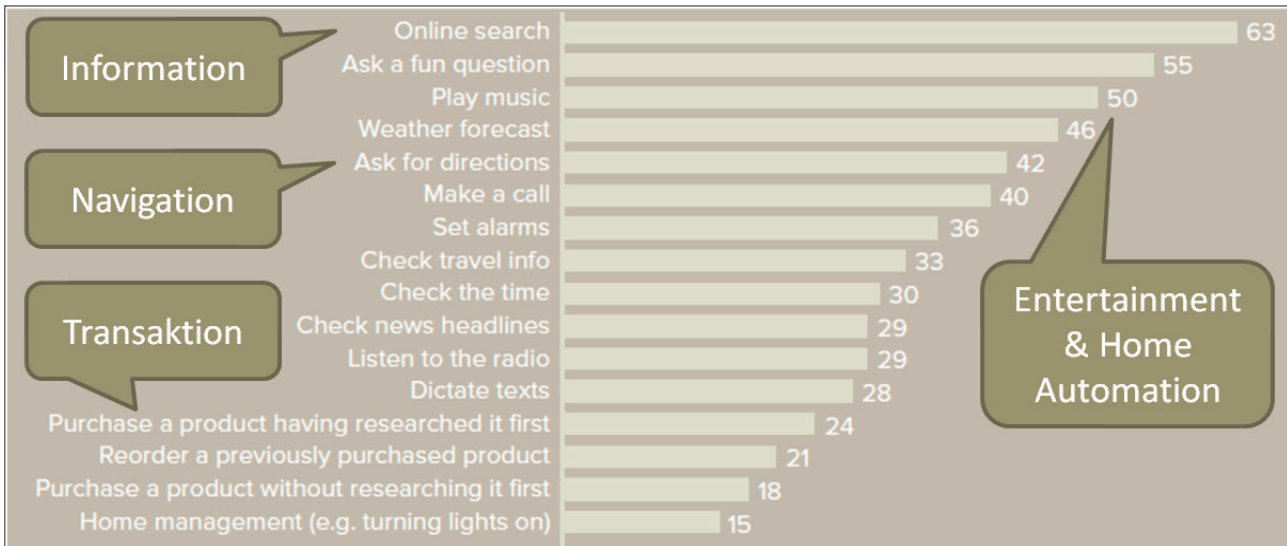


Abb. 1: Wofür werden Sprachassistenten genutzt – *Speak-Easy-Studie 2017*

durch eingeschränkte Sehkraft oder sonstige körperliche Beeinträchtigungen. Auch die 20 % Analphabeten, die laut einer aktuellen Studie der EU-Kommission in Europa leben, oder solche Nutzer, die den komplizierten Umgang mit dem Internet bewusst gemieden haben, könnten als neue Zielgruppen interessant werden. Das Wall Street Journal prognostiziert, dass die nächste Milliarde Internetnutzer vor allem ihre Stimme benutzen werden.

Da viele Sprachassistenten mangels Display keine visuelle Aufmerksamkeit benötigen und eine freihändige Bedienung ermöglichen, ergeben sich vollkommen neue Anwendungsszenarien, z. B. im Auto, im Dunkeln oder beim Kochen. Und auch Werbenden bieten sich neue Chancen: Statt QR-Codes auf Plakate zu drucken oder Webadressen im Fernsehen einzublenden, könnte ein bei Amazon gelisteter Händler potenziellen Kunden eine direkte Bestellung soufflieren, die dieser dann postwendend an sein Echo-Gerät weitergibt, ohne den Umweg über eine Webseite.

Bei der Voice-Optimierung zählt nicht nur Google

Laut Google soll zum Ende des Jahrzehnts jede zweite Websuche mittels Sprache erfolgen, viele davon über Sprachassistenten. Die Suchintentionen bei diesen Geräten unterscheiden sich dabei erst mal nicht wesentlich von

denen bei der normalen Websuche: Auch hier geht es um informations-, navigations- & transaktionsgetriebene Anfragen. Neben der Steuerung von Geräten und zur Unterhaltung ist Suche die meistgenutzte Funktion (vgl. Abbildung 1).

Bei der Antwortoptimierung für Assistenzsysteme gelten hingegen ganz andere, zum Teil gerätespezifische Regeln als beim klassischen SEO. Statt einer reinen Website-Optimierung für Google muss beim Voice-SEO beachtet werden, dass Amazons Alexa und Microsofts Cortana beispielsweise Bing als Quelle für ihre Antworten nutzen. Bei der Suche nach einem guten Restaurant werden zudem externe Plattformen wie Yelp, TripAdvisor oder Google MyBusiness herangezogen. Wer dort nicht gut gelistet und bewertet ist, bleibt für die Sprachsuche möglicherweise unauffindbar, zumal im Gegensatz zu den zehn Ergebnissen der normalen Websuche oft nur ein einziges Ergebnis präsentiert bzw. vorgelesen wird.

Aufbereitung von Webinhalten für Voice Search

Für informationsgetriebene Sprachsuchen spielen die sog. Featured Snippets – Antworten, die direkt in den Suchergebnissen eingeblendet werden – eine große Rolle. Diese tauchen nicht nur bei Google und Bing in zunehmendem Maße auf, sondern werden häufig auch

von Sprachassistenten vorgelesen (wie man Featured Snippets bekommt, kann man z. B. hier nachlesen: de.ryte.com/magazine/erstellt-du-erfolgreiche-featured-snippets-eine-case-study-von-sixt). Auch die Auszeichnung der eigenen Webinhalte mit strukturierten Daten für Produkte, Events oder Fragen- & Antwortbereiche bietet Vorteile; eine gesonderte Auszeichnung für explizit vorlesbare Inhalte ist bei schema.org gerade in der Prüfung (pending.schema.org/speakable).

Darüber hinaus hat Google Anfang des Jahres unter dem Titel „Evaluation of Search Speech – Guidelines“ eine Ergänzung zu den Quality Rater Guidelines veröffentlicht (storage.googleapis.com/guidelines-eyesfree/evaluation_of_search_speech_guidelines_v1.0.pdf), aus denen sich ableiten lässt, worauf es bei der Optimierung von Texten für eine optimale Sprachausgabe grundsätzlich ankommt:

1. Wird die Frage des Nutzers komplett beantwortet?
2. Ist die Länge der Antwort der Frage angemessen?
3. Sind Formulierung und Grammatik korrekt und natürlich?
4. Sind Aussprache und Rhetorik einwandfrei?

Punkt 4 meint dabei die flüssige Vorlesbarkeit eines Textes durch eine

„Let Users Design your Chatbot, not Developers!“ –

W. King – CEO Hi9

Maschine, die z. B. durch Abkürzungen, Sonderzeichen oder Zahlwerte beeinträchtigt werden kann. Bei einem Satz wie: „Hamburg hat 1.810.438 Einwohner (Männer: 735.068/Frauen: 793.238 – Stand: 2016)“ würden die Ziffern einzeln vorgelesen und die Klammern/Striche ignoriert. Für die Sprachausgabe sollte man eher schreiben: „Hamburg hat circa 1,8 Millionen Einwohner, davon sind etwas mehr als die Hälfte Frauen.“ Diese Antwort ist zwar nicht ganz so exakt, aber für die meisten Menschen wohl ausreichend genau.

Amazon Skills & Google Actions – die neuen Apps?

Neben einer sprachgerechten Aufbereitung der eigenen Inhalte zur Verbesserung der Auffindbarkeit bei Sprachsuchen sowie der Nutzung externer Portale gibt es noch die Möglichkeit, Software direkt für Sprachassistenten zu entwickeln: Für Google-Home-Geräte heißen die Anwendungen „Actions“ und bei der Amazon-Echo-Serie heißen sie „Skills“. Diese können u. a. über spezielle Verzeichnisse oder gezielte Abfragen an den Assistenten gefunden werden (vgl. Abbildung 2.)

Diese Programme sind vergleichbar mit den Apps für Handys und für viele Einsatzzwecke gibt es in den jeweiligen Entwicklungsumgebungen fertige Vorlagen, die auf die eigenen Bedürfnisse angepasst werden können, sowie Schnittstellen für Log-ins, Transaktionen und den Datenaustausch mit dem eigenen Server. So ist ein Einstieg auch ohne große Programmierkenntnisse möglich.

Allerdings ist es alles andere als trivial, dem Nutzer eine natürlich empfun-

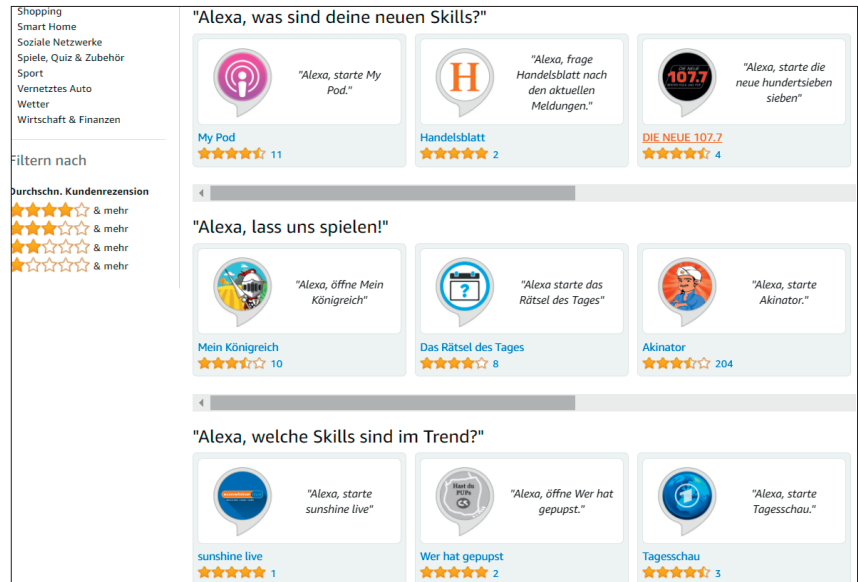


Abb. 2: Skill-Übersicht auf amazon.de

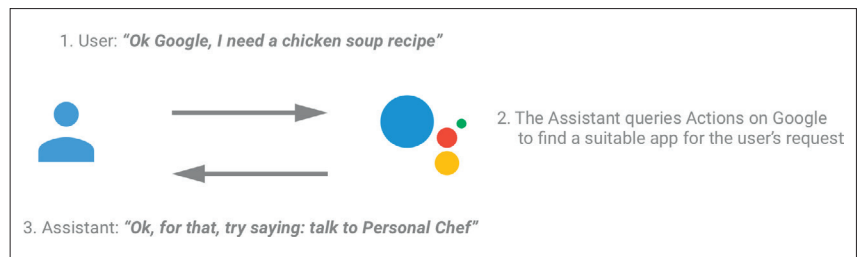


Abb. 3: Der Google Assistant schlägt selbstständig eine App vor – Quelle: Google

dene Konversation mit einem digitalen Assistenten zu ermöglichen, und es zeigt sich für viele Entwickler, wie komplex die menschliche Sprache ist: Aufeinander aufbauende Fragen, unterschiedlichste Ausdrucksmöglichkeiten derselben Intention, Verständnisprobleme, persönliche Vorlieben und Einbeziehung der eigenen Umgebung sowie ein gewisses Maß an Empathie lassen sich nicht so einfach abbilden.

Wo King, CEO von Hi9, einer Agentur für AI und Bot Design, sieht die Verwendung von Personas als einen möglichen Schlüssel und empfiehlt, seine Zielgruppe bei der Produktentwicklung eng einzubinden. Andere Firmen setzen auf Rollenspiele oder nutzen Callcenter-Aufzeichnungen, um typische Gesprächsverläufe zu analysieren.

Der Name ist Programm

Nutzer können die Skills bzw. Actions durch direkten Aufruf (= Explicit Invocation) des Feature-Namens

(„Alexa, öffne Schachspiel“) starten – wobei beim Amazon Echo im Gegensatz zum Google Home im Gegenseitigen die Aktivierung der Software nötig ist – oder direkt eine bestimmte Funktion innerhalb einer Anwendung auslösen („O. k. Google, frag das goldene Orakel, ob ich die Website Boosting abonnieren soll!“).

Da diese Namen nur einmal je Plattform vergeben werden und es in Deutschland noch nicht einmal 5.000 verfügbare Alexa Skills und noch deutlich weniger Google Actions gibt (zum Vergleich: Allein im Google Play Store gibt es knapp vier Millionen Apps), sind noch viele attraktive Namen verfügbar und Anwendungen unbesetzt. Bedenken Sie bei der Namenswahl, dass sich Ihre Nutzer den Namen gut merken und in einem bestimmten grammatikalischen Zusammenhang aussprechen müssen. Bei kreativen Fantasienamen können die Maschinen wiederum Verständnisschwierigkeiten haben.

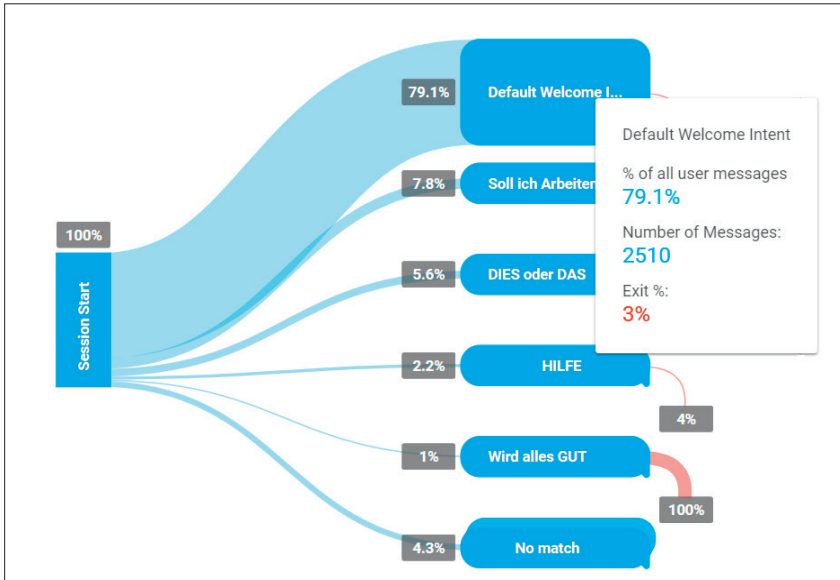


Abb. 4: Analysebericht einer Google Action – Quelle: Dialogflow

Achten Sie zudem bei der Einreichung Ihres Skills/Action auf einen möglichst attraktiven Verzeichniseintrag: Beschreiben Sie die Funktion, den Nutzen und die Möglichkeit des Aufrufs sehr exakt.

Implicit Invocation – das neue SEO

Im Gegensatz zum direkten Aufruf einer Action kommt es vor, dass der Google Assistant Nutzern mit einem bestimmten Problem selbstständig Actions empfiehlt, die eine Lösung oder Antwort bereithalten. Man spricht hier von Implicit Invocation (indirektem Aufruf – vgl. Abbildung 3), eine Funktion, die auch Amazons Echo beherrscht. Auf eine Frage wie: „O. k. Google, wer hilft mir bei meiner Steuererklärung?“, könnte eine Antwort lauten: „Rede doch mal mit dem Steuerberatungsservice.“ So bekommt man neue Nutzer, unabhängig vom Verzeichnislisting oder Branding.

Wer via Implicit Invocation vorgeschlagen werden will, sollte folgende Ratschläge beherzigen:

- » Informiere den Nutzer stets über seine Handlungsoptionen.
- » Vermeide Fehler und logische Ungeheimheiten.
- » Zwingen den Nutzer nicht früh zu einem Log-in oder der Eingabe sensibler Daten.

- » Vermeide offene Gesprächsenden und ermögliche Nachfragen und Rücksprünge.
- » Verwende aussagekräftige Aktivierungsphrasen.

Sprache liefert mehr Signale als nur Klicks!

Wer eine Lösung für Sprachassistenten anbietet, sollte diese sehr sorgfältig konzipieren, testen und ihren Erfolg messen – nicht nur, um positive Bewertungen zu bekommen oder die Chancen für eine indirekte Aktivierung zu erhöhen. Publisher sollten sich bewusst sein, dass jede Voice Session eine Vielzahl an z. T. subtilen Signalen an den Provider (z. B. Amazon) liefert, die es bei der Webanalyse und als SEO-Rankingfaktoren nicht gibt bzw. die nicht eindeutig messbar sind.

Klickt ein genervter Nutzer frustriert auf einer Webseite umher oder lässt ein anderer sein Browserfenster offen, während er fernsieht, freut sich der Analyst im Zweifel zu Unrecht über eine hohe Anzahl an Seitenaufrufen und eine lange Aufenthaltsdauer. Sprachassistenten bekommen sehr genau mit, ob eine interessierte oder verzweifelte Nutzung vorliegt oder ob der Bediener abgelenkt ist. So werden Flüche und Begeisterungsausprägungen, genau wie Hintergrundgeräusche, Pausen oder Ver-

ständnisfragen übertragen; an der Wortwahl und Ausdrucksweise lassen sich Stimmungen ablesen. In Zukunft werden auch emotionale Zustände am Tonfall erkennbar sein, sodass ein Servicebot seine Antworten anpassen kann oder ein digitaler Verkäufer nur entspannten Kunden Upselling-Empfehlungen gibt.

Metriken & Rankingsignale für Sprachassistenten

- » Anzahl der Aktivierungen (+ ggf. Installationen)
- » Nutzungsintensität: Frequenz, Dauer und wiederholte Nutzung
- » Erfolgreiche Nutzungen sowie Abbrüche, Rücksprünge, Schleifen
- » Performancemetriken für einzelne Intents (Teilfunktionen) der Anwendung
- » Nutzerzufriedenheit – messbar über subtile Signale

„Das nächste Interface für die digitale Welt wird die Stimme ...

[...] und die Folgen werden so tiefgehend sein wie beim Smartphone“, so kommentiert Sascha Lobo bei Spiegel Online die rasante Verbreitung von Sprachassistenten. Tatsächlich sind Alexa & Co. die Speerspitze des grundsätzlichen Wandels unseres alltäglichen Umgangs mit dem Internet, dessen sich Firmen und Marketer bewusst sein sollten. Einfach, intuitiv und emotional ansprechend – diesem Ideal jagen Webdesigner zumeist vergeblich hinterher und jeder Verkäufer in der realen Welt verlöre seinen Job, würde er so schlecht überzeugen, wie es die meisten Online-Shops tun. Auch wenn bei vielen sprachgesteuerten Systemen die angebotenen Lösungen noch nicht perfekt sind, so wird sich ihre bequeme und als natürlich empfundene Handhabung in vielen Bereichen schnell durchsetzen. Website Relaunch oder Voice Assistant – über diese Budget-Entscheidung sollte man reden. ¶