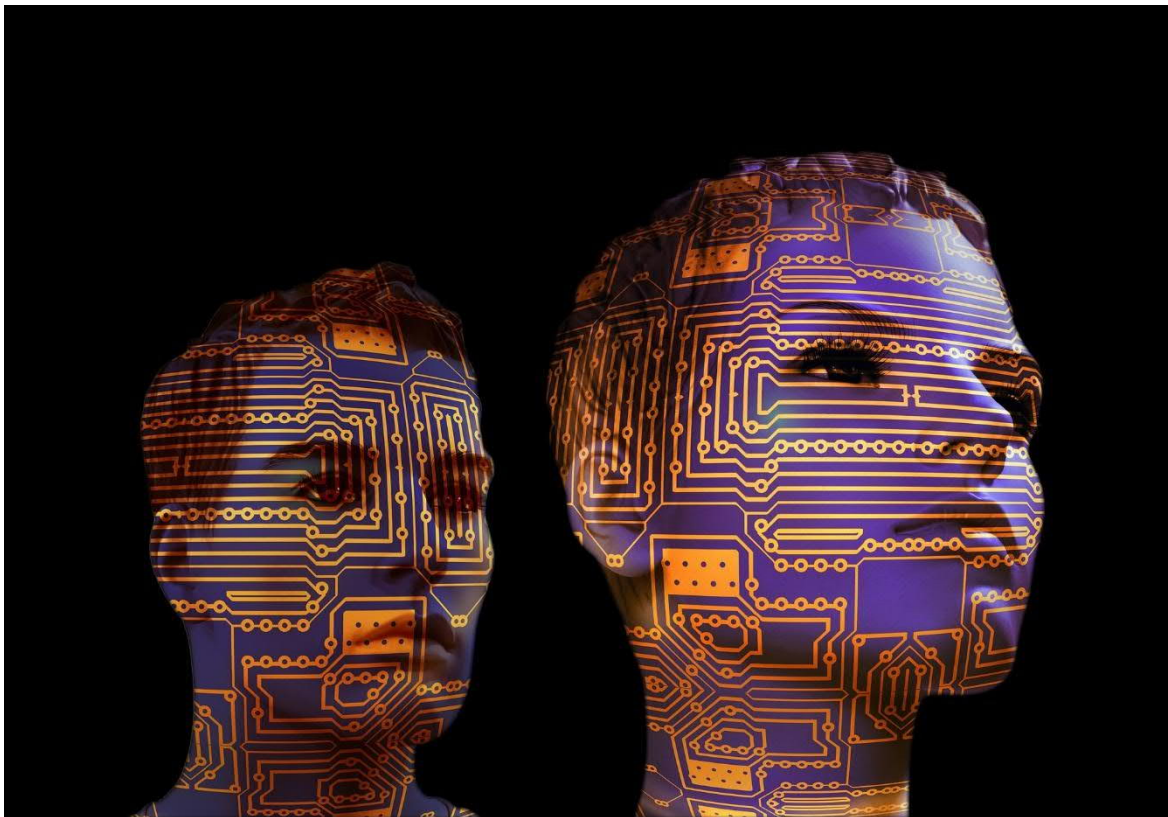


SPRACHASSISTENTEN IM FOKUS

Kurzübersicht



INHALTSVERZEICHNIS

| | | |
|----------|---|----------|
| 1 | TESTKONZEPT FÜR DEN SMARTPHONE-PRAXISTEST VON SPRACHASSISTENTEN..... | 1 |
| 2 | EXPLORATIVER VS. DETERMINISTISCHER TESTANSATZ | 2 |
| 3 | TESTMETHODIK IM PRAXISTEST SPRACHASSISTENTEN | 2 |
| | Testanalyse Praxistest Sprachassistenten | 2 |
| 3.1 | Testvorbereitung Praxistest Sprachassistenten | 3 |
| 3.2 | Testparameter im Praxistest Sprachassistenten | 3 |
| 4 | TESTDURCHFÜHRUNG PRAXISTEST SPRACHASSISTENTEN | 4 |
| 5 | TESTERGEBNISSE PRAXISTEST SPRACHASSISTENTEN..... | 5 |
| 5.1 | Zusammenfassung der Testergebnisse | 5 |
| | Nutzbare sprachliche Grundfunktionen bei Sprachassistenten | 5 |
| 5.2 | Unausgeprägte Sprachkompetenz | 5 |
| 5.3 | Fehlende Dialogfähigkeit | 6 |
| 5.4 | Vergleich bestandener und gescheiterter Testfälle | 6 |
| 5.5 | Abweichungen pro Funktion | 6 |
| 5.6 | Fehlerklassen der Abweichungen der Ergebnisse des Praxistests Sprachassistenten | 7 |
| 5.7 | Anteil der Fehlerklassen im Praxistest Sprachassistenten | 7 |
| 6 | ZUSAMMENFASSUNG..... | 8 |
| | Ausführliche Darstellung der Ergebnisse des Praxistests Sprachassistenten | 8 |

Der Praxistest Sprachassistenten und seine Ergebnisse werden hier kurz vorgestellt. Vier marktgängige Sprachassistenten wurden mit identischen Testäußerungen geprüft und das Ergebnis wird dargestellt. Sprachassistenten haben Stärken in alltäglichen Frage-Antwort-Situationen. Die Leistungsfähigkeit lässt nach, wenn komplexe Anliegen geäußert werden.

1 TESTKONZEPT FÜR DEN SMARTPHONE-PRAXISTEST VON SPRACHASSISTENTEN

Im Softwaretest gibt es die Grundregel "Kein Test ohne Anforderung". Fünf goldene Regeln für erfolgreiches Testmanagement - heise.de Diese Grundregel ist uns auf den Fuß gefallen. Denn welche Anforderungen den Sprachassistenten zugrunde liegen, lässt sich nur erraten.

Vorliegende Tests von Sprachassistenten beziehen sich im Allgemeinen auf Heimplautsprecher. Darüber hinaus fokussieren sie auf die Nützlichkeit und basieren auf Benutzerwertungen bei Verkaufsplattformen. Es ist nicht erkennbar, in welchem Maße die vorhandenen Testansätze vergleichbar sind.

Der Fokus der Speech & Phone GmbH liegt in der Spracherkennung und was liegt näher als Sprachassistenten auf Spracherkennung und auf das Voice User Interface hin zu untersuchen. So wurde dieser Satz unser Credo für dieses Projekt.

Mit dem Praxistest Sprachassistenten wollten wir die Testkandidaten genauer unter die Lupe nehmen und den explorativen Ansatz verlassen. -Diethelm Dahms, Speech & Phone GmbH

Klar ist, dass Sprachassistenten im alltäglichen Einsatz gute Dienste leisten. Dieser Nachweis muss nicht erbracht werden. Aus diesem Grund haben wir den Fokus auf die sprachliche Kompetenz der Sprachassistenten gelegt. Dazu gehören diese Aspekte

- Spracheingabe im Sinne der Verwandlung von akustischen Signalen in Text
- Sprachausgabe als Wandlung von geschriebener Sprache in akustische Signale
- Dialogfähigkeit das Wechselspiel einer Kommunikation zwischen Mensch und Maschine

2 EXPLORATIVER VS. DETERMINISTISCHER TESTANSATZ

Explorative Ansätze des Tests von Sprachassistenten gehen von der Subjektivität der nutzenden Personen aus. Dieses Vorgehen ist eine gute Möglichkeit die Fähigkeiten eines Systems zu ermitteln. Zusammen mit dem Wespennestprinzip ergibt sich eine kaum auflösbare Wahrnehmungsverzerrung, nämlich der Bestätigungsfehler oder confirmation bias. Wer ein System für unschlagbar hält, hält die gefundenen Abweichungen für erklärbar. Wer andererseits Abweichungen findet, wird eventuell diese als Grundlage seiner Ablehnung sehen. Einerseits wird die Softwarequalität überbewertet, andererseits wird sie unterbewertet.

3 TESTMETHODIK IM PRAXISTEST SPRACHASSISTENTEN

Testanalyse Praxistest Sprachassistenten

Im Rahmen deterministischen Testvorgehens erfassen wir zunächst die Funktionen des Sprachassistenten. Dabei kamen wir auf die folgenden Funktionen

- Alltagsorganisation
- Musik und Medien
- Reise
- Wetter
- Wissensfragen

Der Blick auf Sprachassistenten ist durch die Identifikation von Sprachassistenten und Smartspeaker verzerrt. Deren Hauptfunktion ist die Wiedergabe von Medien. Daher werden andere Funktionen als Bonus empfunden. Da der Einsatzbereich der Sprachassistenten im Praxistests bei Smartphones und Desktops liegt, wird der Fokus auch darauf geschärft, dass Sprachassistenten nicht nur Sprache verstehen, sondern sich auch verständlich äußern können.

In der Analysephase des Testprozesses wurden deshalb ausgehend von den oben genannten Funktionen Testparameter ermittelt. Als signifikante Merkmale sprachlicher Äußerungen wurden in das Testfalldesign folgende Parameter aufgenommen.

- Formulierungsart

- Nutzung adverbialer Bestimmungen
- Einsatz von Redewendungen
- Verwendung sprachlicher Bilder
- Äußerung von Bedürfnissen
- Mehrdeutige Begriffe

3.1 Testvorbereitung Praxistest Sprachassistenten

Aus diesen abstrakten Testfällen wurden konkrete Testfälle gemäß der definierten Testparameter erstellt. Aus dem abstrakten Testfall im Funktionsbereich Wissensfragen entstehen so mehrere konkrete Testfälle. Die ersten Beispiele beziehen sich auf die Uhrzeit. Für mehrdeutige Begriffe wurde das Wort rock gewählt. Es ist übrigens verblüffend, was dabei entsteht. Dazu wurde der Artikel über die fehlende Erkennung von Sprachen geschrieben.

3.2 Testparameter im Praxistest Sprachassistenten

| Parameter | Testäußerung |
|---------------------------|---|
| Formulierungsart | Wie spät ist es? Sag mir bitte, wie spät es ist. |
| | Kannst du mir sagen, wie spät es ist? |
| adverbiale Bestimmungen | Wie spät ist es in Tokio? |
| Einsatz von Redewendungen | Was hat die Uhr geschlagen? |
| sprachliche Bilder | Was zeigt die Uhr? |
| Bedürfnisse | Ich habe die Zeit vergessen. |
| Mehrdeutigkeit | Zeige Rock Bilder |

4 TESTDURCHFÜHRUNG PRAXISTEST SPRACHASSISTENTEN

Auf diese Art der Testfallerstellung wurden für die fünf Funktionsbereiche 83 Testfälle notiert und für die folgenden vier Sprachassistenten durchgeführt. Das Testfeld umfasste folgende Sprachassistenten

| Sprachassistent | Plattform | Betriebssystem |
|------------------|------------|----------------|
| Alexa | Smartphone | Android 10 |
| Cortana | Desktop | Windows 10 |
| Google-Assistant | Smartphone | Android 10 |
| Siri | Smartphone | iOS X |

Die Verteilung der Testfälle auf die Funktionen verteilt sich gemäß dieser Prozentangaben.

| Funktion | Anteil |
|----------|--------|
| Alltag | 24 |
| Medien | 14 |
| Reise | 18 |
| Wetter | 13 |
| Wissen | 31 |

5 TESTERGEBNISSE PRAXISTEST SPRACHASSISTENTEN

5.1 Zusammenfassung der Testergebnisse

Zusammenfassend gesagt: Sprachassistenten sind als Spracheingabe für eine Websuche nutzbar. Das ist für die meisten Anwendungsfälle sehr gut geeignet und ausreichend. Es macht sich gleichzeitig hinderlich bemerkbar, wenn die gesprochene Antwort in dem Hinweis auf die Websuche besteht. Es ist für multimodale Umgebungen geeignet, aber nicht für barrierefreie Anwendungen. Dadurch werden grafische Schnittstellen jedoch keinesfalls abgelöst und sind durch die aktuell implementierten Sprachassistenten nicht durchgängig gedacht. Das ist insofern bemerkenswert, weil Sprachschnittstellen in diesem Moment besser bei der Eingabe, als bei der Sprachausgabe funktionieren. Dieses Design hält die Hände zwar frei, aber nicht die Augen.

Für komplexe Anliegen sind die Assistenten nichts weiter als das Spracheingabefeld einer Suchmaschine. *-Amos Dahms, Analyst bei Speech & Phone GmbH*

Nutzbare sprachliche Grundfunktionen bei Sprachassistenten

Gleichzeitig wird beobachtet, dass Sprachassistenten Grundfunktionen beherrschen, und dass sie für spezielle Anwendungen auf ein kostenpflichtiges Angebot verweisen. Dann verweisen sie auf fehlende Funktionen bei der Antwort und sie scheinen als Vertriebsbeauftragte für kostenpflichtige Updates der Anbieter engagiert zu sein. Sprachassistenten sind so ein Freemium-Angebot. Besonders erinnerlich sind hier Alexa und Siri.

5.2 Unausgeprägte Sprachkompetenz

Die Sprachkompetenz und die Sprachleistung der Sprachassistenten ist bei der Spracheingabe und beim Antwortverhalten schwach ausgeprägt. Das zeigt sich daran, dass Objekte, Adverbialbestimmungen oder unterschiedliche Präpositionen ignoriert werden.

Beispiel

Die Antwort auf die Frage “Wie ist die Temperatur im See?” beinhaltet Auskünfte über das Wetter des aktuellen Ortes, das Wetter

eines Ortes namens See, die Anomalie des Wassers in Binnenseen oder zur Biologie von Seen.

5.3 Fehlende Dialogfähigkeit

Gerade in Situationen wie diesen macht sich die sehr eingeschränkte Dialogfähigkeit der Sprachassistenten unrühmlich bemerkbar. Im oben genannten Beispiel wäre es zielführender gewesen, zurückzufragen, welches Wetter oder welcher Ort gemeint sind. Durch den Verzicht auf diese Rückfrage vergeben sich die Sprachassistenten die Möglichkeit als Assistent wahrgenommen zu werden. Aus meiner Sicht wirken sie in dieser Art nur besserwisserisch.

5.4 Vergleich bestandener und gescheiterter Testfälle

Die folgende Tabelle zeigt den Anteil der bestandenen und gescheiterten Testfälle pro Sprachassistent.

| Sprachassistent | Bestanden | Gescheitert |
|------------------|-----------|-------------|
| Alexa | 60 % | 40 % |
| Google-Assistant | 63 % | 37 % |
| Cortana | 38 % | 62 % |
| Siri | 66 % | 34 % |

5.5 Abweichungen pro Funktion

Bei den Abweichungen, also den gescheiterten Testfällen, zeigen die verschiedenen Funktionen Alltag, Medien, Reise, Wetter und Wissen unterschiedliche Schwankungsbreite.

| Funktion | Schwankungsbreite |
|----------|-------------------|
| Alltag | 18 % - 24 % |
| Medien | 13 % - 27 % |

| | |
|-------|-------------|
| Reise | 11 % - 27 % |
|-------|-------------|

| | |
|--------|-------------|
| Wetter | 18 % - 27 % |
|--------|-------------|

| | |
|--------|-------------|
| Wissen | 12 % - 27 % |
|--------|-------------|

5.6 Fehlerklassen der Abweichungen der Ergebnisse des Praxistests Sprachassistenten

Wenn wir genauer auf die Ursache der Abweichungen schauen, finden wir folgende Einteilung

- Eingabe: Die sprachliche Äußerung wird nicht oder fehlerhaft interpretiert.
- Ausgabe: Bei der Ausgabe werden Abweichungen von der Normsprache festgestellt.
- Dialog: Es fehlen Rückfragen, wenn die Äußerung bei der Eingabe mehrdeutig oder unklar ist.
- Backend: Das beste Dialogsystem kann nicht besser sein als die Daten, die das Backend liefert.

5.7 Anteil der Fehlerklassen im Praxistest Sprachassistenten

Diese Tabelle zeigt die Verteilung der Fehlerklassen für alle Sprachassistenten insgesamt.

| Fehlerklasse | Anteil an Abweichungen |
|--------------|------------------------|
|--------------|------------------------|

| | |
|---------|------|
| Eingabe | 71 % |
|---------|------|

| | |
|---------|-----|
| Ausgabe | 1 % |
|---------|-----|

| | |
|--------|------|
| Dialog | 13 % |
|--------|------|

| | |
|---------|------|
| Backend | 15 % |
|---------|------|

6 ZUSAMMENFASSUNG

Schließlich bleibt als Tester nur die unrühmliche Aufgabe, den Finger in die Wunde zu legen. Das Jagdfieber ist geweckt, sobald die erste Abweichung, das erste Stolpern bemerkt wird. Gleichzeitig wissen alle Menschen im Test:

*Ein Diamant leuchtet nur, wenn er geschliffen ist.
Zakhar Bron*

Wir sehen Sprachassistenten die am Markt und in der breiten Masse etabliert sind. Gleichzeitig werden mit der Marktbreite zusätzliche Ansprüche geweckt. Solange Sprachassistenten für bestimmte Funktionen wie Banking oder Telefonvermittlung gedacht waren, konnte über mangelnde Fähigkeiten bei Sprachleistung oder Sprachkompetenz hinweg gesehen werden. Dies geht mit breiter Marktpenetranz nicht mehr. Neben diese Parameter treten Dialogfähigkeit und auch die Fähigkeiten Äußerungen für die Antworten so aufzubereiten, dass sie normsprachlich korrekt sind und sie in der Lage sind, Webseiten sprachlich vorzulesen.

Ausführliche Darstellung der Ergebnisse des Praxistests Sprachassistenten

Die Foliensammlung mit griffiger Zusammenfassung können Sie auf Nachfrage per Email an contact2019@speech-and-phone.de erhalten.