Maschinen, die den Menschen verstehen - IT-Methoden und IT-Standards bei der Gestaltung mobiler Sprachdialogsysteme

Dirk Nordwig

dawin gmbh
Belgische Allee 50
D-53842 Troisdorf
nordwig@dawin.de

Abstract: Sprachgestützte Datenerfassungstechnologien auf Basis verschiedener moderner IT-Methoden sind praxistauglich und bieten eine Reihe von Vorteilen. Eine benutzergerecht gestaltete Dialogschnittstelle erhöht die Akzeptanz und die Datenqualität, Sprachtechnologie reduziert den Erfassungsaufwand und macht die Konzentration auf den Arbeitsinhalt möglich.

1 Sprachtechnologie für mobile Anwendungen

Mobiles Arbeiten gehört für immer mehr Menschen zum Alltag. Sprache kann dabei einen wichtigen Beitrag leisten, um mobile Datenerfassung sicherer, bequemer und effizienter zu machen.

Vor dreißig Jahren war für uns der sprechende Bordcomputer aus dem Film Raumschiff Enterprise sensationelle Science Fiction. Heute betrachten wir mobiles Telefonieren und die Sprachkommunikation mit Maschinen als ebenso natürlich wie den Plausch mit dem netten Kollegen. Sprachanwendungen, wie z. B. Siri, bieten die Möglichkeit, schnell bestimmte Funktionen auf unserem iPhone aufzurufen.

Mehr und mehr Geräte und Softwareanwendungen lassen sich per Sprache bedienen. Dennoch scheint es, dass die meisten Einsatzmöglichkeiten von Sprache (in der Industrie und Wirtschaft jenseits von Consumer-Anwendungen) noch brach liegen. Sprachgesteuerte Software ermöglicht die bequeme Bedienung eines Gerätes und gibt Augen und Hände für andere Aufgaben frei.

Die Vorteile einer sprachgestützten Maschinen-Mensch-Kommunikation sind:

- Sprechen ist für die meisten Menschen einfacher als tippen und lesen
- Sprachsteuerung kann gegebenenfalls auf Anzeige- und Tastaturmedien verzichten. Die notwendige Hardware kann also handlicher, leichter und billiger werden

- Mobile Datenerfassungsaufgaben jenseits des Büroarbeitsplatzes lassen sich berührungsfrei per Sprachsteuerung effizienter erledigen, Hände und Augen sind für die Arbeitsaufgabe frei. In bestimmten Situationen - besonderere hygienische Anforderungen oder schmutzige oder kontaminierte Umgebung – ist der Sprachdialog der Nutzung von Tastatur oder Maus deutlich überlegen.

2 Anwendungsbeispiele

Doch wie funktioniert das Sprechen mit Maschinen? Die dawin gmbh aus Troisdorf hat sich auf die Entwicklung von sprachgestützten Softwarelösungen spezialisiert. Im ersten Anwendungsbeispiel wird die effiziente Erfassung von Fachinformationen via Telefonie durch Umwandlung von frei gesprochenen Berichten in digitale und strukturierte Daten gezeigt. Die wesentlichen Anwendungskennzeichen sind:

- Überall verfügbares Endgerät (Telefon)
- Unabhängig von Internetanbindungen
- Einfach in der Anwendung sowie überall und jederzeit nutzbar

Hier handelt es sich um eine servergestützte Lösung, welche die komfortable Datenerfassung mit herkömmlichen Telefonen (Festnetz, Mobiltelefon) ermöglicht. Auf den ausschließlichen Sprachdialog abgestimmte Erfassungsroutinen ermöglichen hier den Verzicht auf eine grafische Benutzeroberfläche bei der Arbeit. Damit steht eine interessante Alternative zur mobilen Datenerfassung ohne zusätzliche Hardwareinvestition zur Verfügung. Im zweiten Beispiel wird eine Lösung für Bonituren (Gewächshaus oder Freiland) per Sprachsteuerung vorgestellt. Gerade in Laborumgebungen oder bei der Arbeit mit Pflanzenschutzmitteln und Giftstoffen kann die sprachgesteuerte Datenerfassung die Datenqualität und die Sicherheit der Mitarbeiter verbessern. Der Pflanzenprüfer arbeitet mit einem Headset und dokumentiert seine Befunde berührungsfrei per Sprachbefehl in einer Datenbank.Die Vorteile dieser Lösung: Hände und Augen bleiben frei für die Pflanzenprüfung. Die Gefahr einer Kontaminierung der Mitarbeiter wird minimiert und die Datenqualität und Geschwindigkeit der Bonitur erhöht.

3 Spezielle Anforderungen an Sprachtechnologieanwendungen

Die Softwarelösungen lassen sich je nach Bedarf auf verschiedenen Geräten (Laptops, Tablets, Smartphones, PDAs und konventionelle Telefone) nutzen. Sie ermöglichen die Erfassung vielfältigster Daten, wie z. B. alpha-numerischen Informationen, Auswahlfelder und Textblöcke. Selbst die Aufnahme von Bilddaten, GPS-Koordinaten oder die Erfassung von Barcodes und RFID können per Sprachbefehl gesteuert werden. Selbst die Erfassung und Erkennung von frei diktiertem Text zur Dokumentation von Informationen, Arbeitsschritten oder Inspektionsergebnissen ist möglich.

Abhängig von der Aufgabenstellung können komplexe Lösungen unter Verwendung verschiedenster IT-Methoden und IT-Standards entwickelt werden. In dem gezeigten Beispiel wurde das Zusammenspiel der folgenden Technologien realisiert:

- Komplexe strukturierte Abfragedialoge auf der Basis von Voice-XML
- Voip-Telefonie und Cloudtechnologie zur Zwischenspeicherung der analogen Telefoniedaten (IVR-Server - Interactive Voice Response) und zur Verarbeitung (Transkription) in digitale und strukturierbare Informationen
- LV ASR (Large Vocabulary speech recognition / dictation) zur fachgerechten Transkription von gesprochenem frei formuliertem Text

Mittels moderner Sprachtechnologie ist es möglich, Dialogtext in einen Kontext zur Eingabesituation bzw. zum Anwendungszusammenhang zu setzen. Dadurch wird der relevante Wortschatz für die Spracherkennung drastisch reduziert (kontextsensitives Keyword-spotting) und die Erkennungsrate sehr hoch. Dennoch betont auch dawin, dass es für sprachgesteuerte Programme keine 100 prozentige Trefferquote gibt. Die gibt es allerdings auch beim Menschen nicht. Auch in zwischenmenschlichen Gesprächen bitten wir unseren Gesprächsteilnehmer gelegentlich um Wiederholung, weil wir es akustisch oder inhaltlich nicht verstanden haben. Diese Art der Interaktion und des Nachfragens ist auch mit guter Software durch eine entspechend sorgfältige Gestaltung der Dialogschnittstelle (Vocal User Interface vs. Graphic User Interface) möglich. Der Sprachdiaog Mensch – Maschine hat die folgenden Eigenschaften:

- Eine nahezu unbegrenzte Anzahl an Sprachbefehlen (auch in verschiedenen Nationalsprachen) ist möglich, die Bedienung ist nutzerunabhängig (kein individuelles Sprachtraining erforderlich) und robust gegenüber Dialekten und Mundarten
- Die Sprachdialoge sind frei gestaltbar und nach dem Erfahrungsgrad der Anwender skalierbar (Anfängermodus / Profimodus / interaktive Hilfe-Funktionen)
- Das aktuelle Vokabular wird zur Laufzeit dynamisch generiert
- Barge-in (dazwischen sprechen) Funktionen ermöglichen die Verkürzung der Dialoge durch erfahrene Nutzer, die Detailiertheit des Dialogs ist für die konkreten Arbeitsbedingungen einstellbar
- Eine mehrstufige Hintergrundgeräuschunterdrückung gestattet die Arbeit auch lauten Produktionsumgebungen oder im Freiland (Windgeräusche, Verkehrslärm)

4 Gestaltung von konkreten Sprachtechnologielösungen

Die maßgeblichen Anforderungen an die Technologiewahl und die Gestaltung der Software ist immer der konkrete Arbeitsprozess und die Arbeitsbedingungen. Die Lösung wird dem Prozess angepasst, so dass der Anwender in seiner Kernarbeitsaufgabe optimal unterstützt werden kann.

Die Gestaltung bzw. Auswahl der jeweiligen Softwaretechnologie, der technischen Infrastruktur und der Endgeräte ist abhängis vom Einsatzort, den Einsatzbedingungen und dem Zielsystem für die erfassten Daten. So ist die sprachgestützte Erfassung von strukturierten Informationen (z.B. Checklisten) in einer sogenannten „embedded-Lösung“ auf dem mobilen Gerät (Tablet-PC) lokal und ohne Netzwerkanbindung möglich. Komplexe Aufgaben, welche die Erkennung und Erfassung von frei diktierten Texten und Informa-

tionen erfordern, werden mit Hilfe von Serverarchitekturen mit diversen Endgeräten gestaltet.

Ein wichtiges Kriterium für die Akzeptanz und damit den Einsatz von Sprachsoftware ist der Nutzerdialog. Während uns grafische Benutzeroberflächen sowie deren Bedienung lange vertraut sind, ist die Bedienung per Sprache (VUI) eine weitaus komplexere Interaktion, als eine begrenzte Anzahl Tasten zu drücken. Unter den bekannten Kommunikationskanälen ist die Sprache die natürlichste und menschlichste Art, zu kommunizieren.

Sprache ist für uns Menschen spontan und unkompliziert – sie enthält jedoch auch sehr komplexe innere Strukturen, die die menschlichen kognitiven Muster wiedergeben. Diese komplexen Informationen lassen sich in der Regel nicht mit Hilfe der hierarchischen Baumstrukturen der bekannten grafischen Benutzeroberflächen abbilden. Eine gute Dialoggestaltung berücksichtigt also in jedem Dialogschritt den Gesprächszusammenhang (Kontext) sowie Annahmen bezüglich der Absichten und Erwartungen (Semantik) des Nutzers.

Dem Anwender muss zu jedem Zeitpunkt des Dialoges geläufig sein, welche Befehle oder Abfragen im aktuellen Kontext möglich sind. Gegebenenfalls werden situationsbedingte akustische Hilfetexte angeboten oder eine Reihe von Synonymen bei den Nutzereingaben akzeptiert. (z.B. positive Antwort: „Ja“, „OK“, „in Ordnung“)

Darüberhinaus werden die VUI je nach Erfahrungsgrad (Neuling, erfahrener Nutzer, Experte) skaliert gestaltet, um die Länge und Detailiertheit des Bedien- und Eingabedialoges dem aktuellen Arbeitsprozess anpassen zu können.

5 Weitere Beispiele und Ausblick der künftigen Entwicklungen

Ein drittes Beispiel zeigt die sprachgestützte Erfassung von Daten aus der Landwirtschaft. Diese Softwarelösung wurde auf Basis der generischen Plattform „dawin checkMaster“ realisiert. Hier kann der Anwender selbst die von ihm benötigten Dialoge für die Felddatenerfasung gestalten und ist so flexibel und unabhängig.

Neben dieser Plattform zur Erfassung von Checklisten steht die mobile Sprachtechnologie auch als Software-Development-Kit (SDK) zur Verfügung und kann für die Sprachaktivierung vorhandener Software sowie bei der Entwicklung neuer sprachgesteuerter Produkte von Softwareanbietern verwendet werden.

Die dawin gmbh beobachtet sehr aufmerksam die aktuellen Entwicklungen der mobilen IT-Technologie und ist bereits auf eine Erweiterung des Einsatzspektrums ihrer Sprachtechnologie auf weiteren Plattformen (Android, ggf. IOS) vorbereitet. Ein wesentlicher Schwerpunkt ist aktuell die Gestaltung der Dialogschnittstellen Mensch – Maschine (VUI). Die Entwicklung von intuitiven und zunehmend natürlichsprachlichen Dialogen wird weitere Anwendungen ermöglichen und damit unserer täglichen Arbeit vereinfachen und erleichtern. Unsere Vision ist es, dem Menschen im Dialog mit der Maschine wieder seine natürliche Art und Weise der Kommunikation zurück zu geben.