Voice Extensible Markup Language

Referent:
Roland Ramthun
Universität Trier, FB II
Dozent: Dr. S. Naumann

31.05.2010

Einordnung VoiceXML

- VoiceXML ist eine XML-Anwendung
- VoiceXML definiert Abläufe in Sprachdialogsystemen
- VoiceXML ist eine W3C Empfehlung und damit standardisiert

Was war nochmal XML?

- XML (Extensible Markup Language) ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten
- Ein XML-Dokument besteht nur aus Text, Binärdaten sind nicht enhalten

```
1 <?xml version="1.0" encoding='UTF-8'?>
2 <painting>
3   <img src="madonna.jpg" />
4   <caption>This is Raphael's "Foligno" Madonna, painted in
        <date>1511</date>–<date>1512</date>.</caption>
5 </painting>
```

xmlbsp.xml

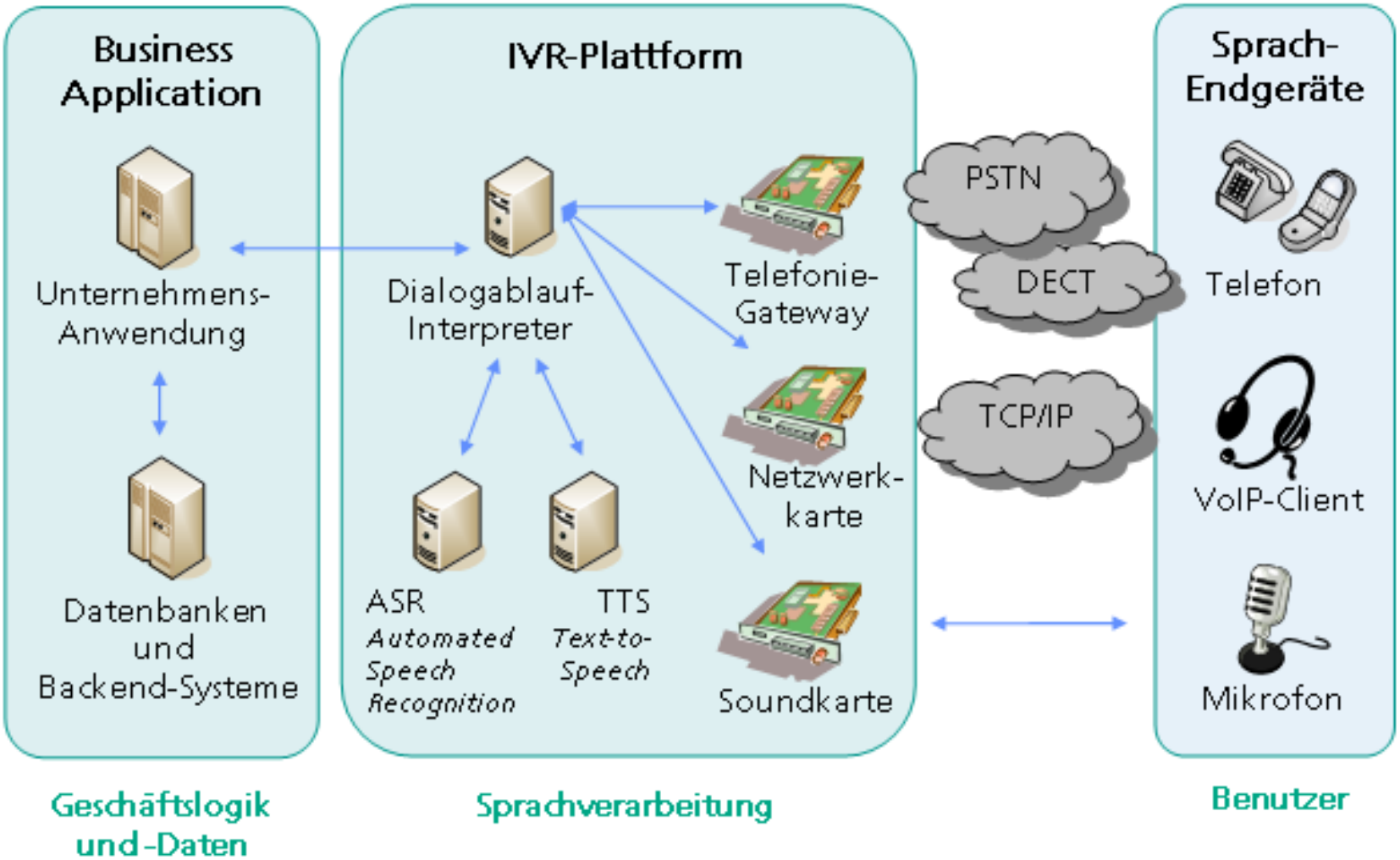
Was war nochmal XML?

- Korrekte XML-Dokumente müssen zwei Bedingungen genügen
  - Wohlgeformtheit, d.h. dass das Dokument die allgemeinen XML-Regeln einhält (ein Wurzelelement, öffnende/schließende Tags, ...)
  - Gültigkeit, d.h. dass das Dokument einer Grammatik (DTD, XML-Schema) genügt. Auf diese wird im XML-Dokument verwiesen, weshalb sie anwendungsspezifisch sein kann.

Sprachdialogsysteme

- mit Sprachdialogsysteme können Benutzer automatisierte Dialoge führen, z.B. per Telefon
- werden daher auch IVR-System (Interactive Voice Response) genannt
- Benutzereingaben entweder über DTMF oder natürliche Sprache

Sprachdialogsysteme



Sprachdialogsysteme - Einsätze

- B2C
  - z.B. Auskünfte, Buchungen, Vermittlung, intelligente Warteschleifen, Voting, Gewinnspiele, Störungsansagen-Management
- B2E
  - z.B. Warenanname, Inventur, Inspektion, Fern-/Vor-Ort-Diagnostik
- Geräteintegriert
  - z.B. Freisprecheinrichtungen, Navigationssysteme, Computerspiele

Entwicklung der Beschreibungssprachen

- Früher: keine Trennung von Applikation und Plattform, Dialogverläufe waren fest programmiert
  - Vorteil: schnell gemacht und zuverlässig
  - Nachteil: änderungsunfreundlich, Anwendungsentwickler für Änderungen nötig
- Heute: Trennung von Applikation und Plattform
  - Vorteil: änderungsfreundlicher
  - Nachteil: per se immer noch proprietär, d.h. kein einfacher Wechsel der Infrastruktur möglich
- An dieser Stelle kommt VoiceXML ins Spiel

Entwicklung der Beschreibungssprachen

- Die Ursprünge von VoiceXML lagen 1995 in einem AT&T Projekt namens Phone Markup Language (PML), das die Entwicklung von Spracherkennungsanwendungen erleichtern sollte. Nach internen Reorganisationen bei AT&T, arbeiteten AT&T, Lucent und Motorola mit eigenen PML-Varianten weiter.
- 1998 gab das W3C eine Konferenz zum Thema Sprachbrowser, bei der klar wurde, wie viele Sprachen es mittlerweile gab, um Ähnliches zu erledigen (PML (AT&T), VoxML (Motorola), SpeechML (IBM), TalkML (HP), VoiceHTML (PipeBeach))
- AT&T, IBM, Lucent und Motorola gründeten daraufhin das VoiceXML-Forum, um ihre Bemühungen zu bündeln

Entwicklung der Beschreibungssprachen

- 2000 gab das VoiceXML-Forum VoiceXML 1.0 frei
- Die Arbeit des W3C bzw. der darin vertretenen Firmen und der Öffentlichkeit mündeten 2004 in VoiceXML 2.0, was als W3C Recommendation verabschiedet wurde
- 2007 kam das abwärtskompatible VoiceXML 2.1 heraus, das 2.0 verfeinert
- Aktuell wird Version 3.0 entwickelt (Working Draft 4 March 2010)

VoiceXML Aufgaben

VoiceXML-Dokumente beschreiben:

- Gesprochene Ausgaben mithilfe synthetischer Sprache
- Ausgaben von vorgefertigten Audiodaten
- Erkennung von gesprochenen Wörtern und Sätzen
- Erkennung von Tastentönen (DTMF)
- Aufnahme gesprochener Eingaben
- Kontrolle des Dialogflusses
- Telefoniekontrolle (Anruftransfer und Auflegen)

Wie man bereits hier sieht, erfordert das Design sprachgestützter User-Interaktion andere Überlegungen als das Design von GUIs.

Dialog Design

- Gegenüber grafischen Benutzeroberflächen (GUI) haben sprachbasierte Benutzeroberflächen (VUI) einige Besonderheiten
  - Akustische Information kann nur linear und mit der vorgegebenen Geschwindigkeit aufgenommen werden
  - Kommunikationsfehler sind schwerer zu beheben
  - Informationsvermittlung dauert länger
  - Schallsignal ist schnell vergänglich, die Information existiert danach nur noch im begrenzten Kurzzeitgedächtnis des Hörers
- Eine VUI wird schneller unbedienbar als eine GUI

Dialog Design - Dialogführung

- Geführter Dialog
  - Das System übernimmt die Initiative

```
1 System: "Von wo aus moechten sie fliegen?"
2 Nutzer: "Von Frankfurt"
3 System: "Wohin moechten Sie von Frankfurt aus
      fliegen?"
4 Nutzer: "Nach Berlin"
5 System: "Und wann moechten Sie von Frankfurt nach
      Berlin fliegen?"
```

- Gemischt-initiativer Dialog
  - Sowohl System, als auch Nutzer können den Dialog vorantreiben

```
1 Nutzer: "Ich moechte von Frankfurt nach Berlin
      fliegen"
2 System: "Und wann moechten Sie von Frankfurt nach
      Berlin fliegen?"
```

VoiceXML Grundkonzepte

- Die *Session* beginnt, sobald ein Benutzer anfängt mit der Plattform zu interagieren und bleibt bestehen, bis sie durch Nutzer, Plattform oder VXML beendet wird.
- Eine *Anwendung* ist eine Menge von *Dialogzuständen*, die dasselbe Anwendungshauptdokument teilen. Von jedem Dialogzustand aus wird auf den nächsten mit Hilfe von *URLs* verwiesen.

> *W3C Recommendation für VXML 2.0: A VoiceXML document (or a set of related documents called an application) forms a conversational finite state machine*

VoiceXML Grundkonzepte

- Jeder *Dialog* besteht aus *Menüs* und *Formularen*.
  - Ein Menü präsentiert dem Benutzer eine Reihe von Optionen sowie die Übergänge zu anderen Dialogzuständen, die auf der Auswahl des Benutzers basieren.
  - Ein Formular definiert eine Interaktion, die Werte für alle *Felder* in dem Formular sammelt. Jedes Feld kann eine Eingabeaufforderung, den erwarteten Input und Evaluierungsregeln spezifizieren.

VoiceXML Grundkonzepte

- Dialogzustände besitzen assoziierte Grammatiken, die gültige Nutzereingaben beschreiben
- Grammatiken kommen aus
  - dem Dialog selbst
  - externen Grammatik, die durch Links referenziert werden
  - dem Dokument, das den Dialog enthält (sofern sie als global aktiv markiert sind)
  - aus dem Anwendungshauptdokument

VoiceXML Grundkonzepte

- Grammatiken sind bevorzugt in grXML (Speech Recognition Grammar Specification Version 1.0, W3C Recommendation 16 March 2004) geschrieben, teilweise auch in GSL (Nuance proprietär) oder ABNF (angereicherte Backus-Naur-Form)

```
1 <?xml version= "1.0"?>
2 <grammar xmlns="http://www.w3.org/2001/06/grammar"
      xml:lang="de" mode="voice" root="main">
3   <rule id="myrule" scope="public">
4     <one-of>
5       <item>Hund</item>
6       <item>Katze</item>
7     </one-of>
8   </rule>
9 </grammar>
```

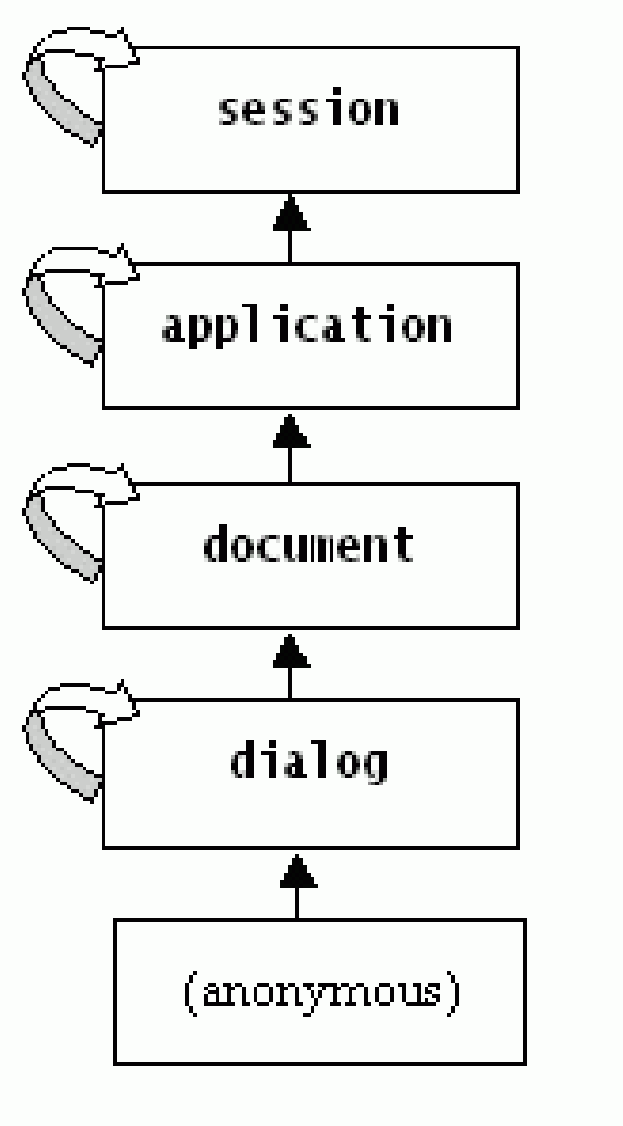
grammarbsp.xml

VoiceXML Grundkonzepte

- *Subdialoge* erlauben es, zu einem anderen Dialog herauszurufen und danach zum Originaldialog zurückzukehren, ohne dass lokale Variablen verloren gehen. Sinnvoll z.B. für Bestätigungen.
- VoiceXML kennt *Variablen*. Diese folgen einem Vererbungsmodell und können benutzt werden, um Eingaben zu speichern, konditionalen Code zu aktivieren oder in Grammatiken verwendet werden.

VoiceXML Grundkonzepte

- *Ereignisse* entstehen, wenn der Benutzer etwas tut (Auflegt, eine unverständliche Antwort liefert, gar nicht antwortet). Für diese Ereignisse können Handler implementiert werden - da die Ereignisse auch „nach oben“ vererbt werden auch in übergeordneten Dokumenten



VoiceXML Grundkonzepte

- VoiceXML benutzt ECMAScript zur Ablaufsteuerung

```
1  <form id="ecmabsp">
2    <block>
3      <script>
4        <![CDATA[
5          function sayasDigits(number)
6          {
7            var digitNumber = number.charAt(0);
8            for(var i = 1; i < number.length; i++)
9            {
10             digitNumber += ' ' + number.charAt(i);
11           }
12           return digitNumber;
13         }
14       ]]>
15     </script>
16   <prompt>
17     <value expr="sayasDigits(100)"/>
18   </prompt>
19   </block>
20 </form>
```

ecmascriptbsp.xml

Quellen und Bildnachweis

- Dokumente aus Richtung des W3C
  - http://www.w3.org/TR/voicexml20/
  - http://www.w3.org/Voice/Guide/
  - http://meiert.com/de/w3/Voice/Guide/
- Dokumente der Voxeo Corporation
  - http://www.vxml.org/
- Wikimedia und Wikipedia
  - http://commons.wikimedia.org/wiki/File:IVR-Systemarchitektur.png
  - http://de.wikipedia.org/w/index.php?title=Sprachdialogsystem&oldid=73793699
  - http://de.wikipedia.org/w/index.php?title=Dialog_Design&oldid=60414041

VoiceXML

Das war der Vortragsteil, haben Sie noch Fragen?