

VoiceXML in a Multimodal Web

Ajit Paul Singh
Guangwu Liu

Introduction

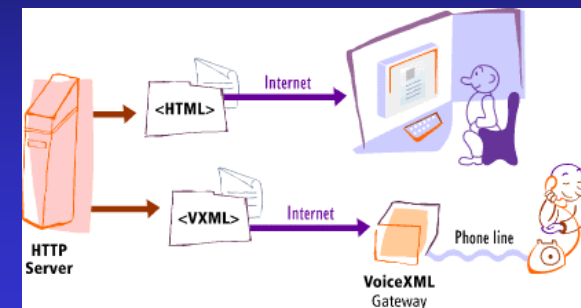
- Multimodal Web Browsing
- What is VoiceXML ?
- The VoiceXML Architecture
- Dialogs
- Structure of a VoiceXML Document
- Control Structures & Grammars
- Example fragment

Multimodal Web Browsing

- Text-based browsing is difficult:
 - If your using a cell phone
 - Driving / Need hand-free access
 - Are visually-impaired
- Telephones are:
 - Widely available
 - Easier to use than computers

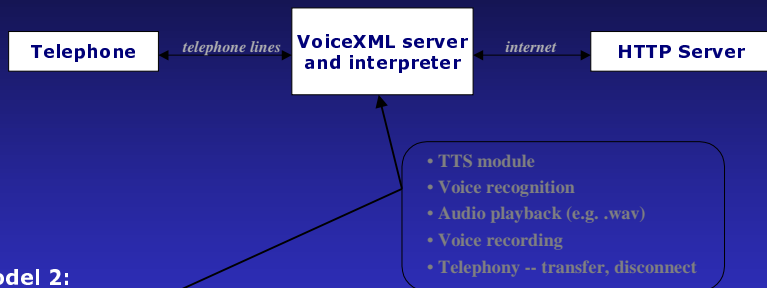
What is VoiceXML ?

- VoiceXML is an XML schema.
- VoiceXML is a markup language for representing human-computer dialogs



The VoiceXML Architecture

Model 1:



Model 2:



Dialogs



Where will you be departing from ?

Edmonton. And I'm going to Jamaica.

And when will you be leaving ?

Tomorrow. What is the cheapest flight ?

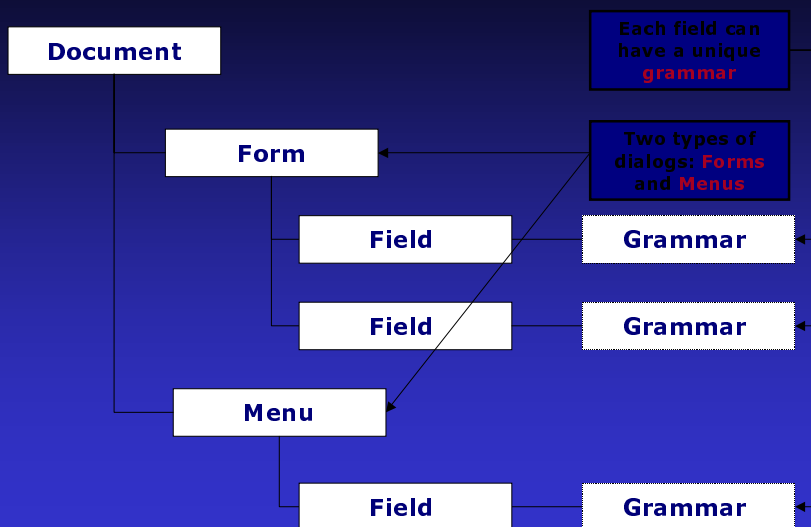
Dialogs (cont)

- HTML forms are asymmetric, user-initiative.
- Conversations are more symmetric, can be mixed-initiative.
- Can view form filling as a series of dialogs.
- Dialogs have context.

Why we need dialogs

- General voice recognition is difficult and prone to error.
- A computer will accept only a limited range of inputs to a question.
- Define the expected answers in a grammar.
 - Limiting the hypothesis space
 - Context-free, stochastic grammars

Structure of a VoiceXML Document



Control Structures & Grammars

- Goto statement controls the execution sequence of a voiceXML document


```
<goto next="#say_goodbye">
```

Where say_goodbye is the label to a dialog in the file

```
<goto "aDocument.vxml">
```

where aDocument.vxml is another VoiceXML document
- Grammars
 - What grammar for?
 - specify the set of answers to a dialog
 - Provides a corresponding string value or set of attribute-value pairs to describe the information or action.
 - Type of grammars
 - Java Speech Grammar Format is commonly used
 - DTMF(Telephone keypad tones)
 - e.g. DTMF_YES_NO[dtmf-1{return("yes")} dtmf-2{return("no")}]

Example of a VoiceXML Fragment

```

<form id="get_card_info">
  <field name="card_number">
    <prompt>
      what kind of credit card do you
      have ?
    </prompt>
    <grammar>
      vis          {visa}
      | master     {mastercard}
    </grammar>
    </help>
    Please say visa, or mastercard
  </help>
</field>
<menu>
  <prompt> Thank you. Please select
  an option </prompt>
  <choice><next="cancel.vxml">
  Cancel order</choice>
  <choice><next="checkout.vxml">
  Checkout</choice>
  <noinput>please say one of
  <enumerate><noinput >
</menu>
</form >

```

Example conversation:

C: what kind of credit card do you have?
H: Discover
C: I did not understand what you said (default)
C: Please say visa, or mastercard
H: Visa
C: Thank you. Please select an option. Cancel order, or checkout?
H: <long pause>
C: Please say one of cancel or checkout
H: checkout.
<go to the checkout.vxml>

Links

VoiceXML Tutorials / Introduction

<http://www.voicexml.org/tutorials/intro1.html>
<http://www.arsdigita.com/asj/vxml/>
<http://www-106.ibm.com/developerworks/library/ibm-voicex/?dwzone=ibm>
<http://www.xml.com/pub/a/2000/08/23/didier/>

VoiceXML Specification

<http://www.w3.org/TR/voicexml>

Related W3C Specifications

<http://www.w3.org/voice>

Voice Portals

Tellme.com: <http://www.tellme.com>
myAudioPoint.com: <http://www.myaudiopoint.com>
VoiceGenie: <http://www.voicegenie.com>