Can Web understand Croatian language?

Albert.Novak@CARNet.hr

Croatian Academic and Research Network

Introduction

- Some people only use telephone for distance communication

- In some situations speech is only appropriate communication channel (e.g. Car driving)

- Can we offer our web application to these users?

- Yes, we can! We have standards such as VoiceXML or SALT!
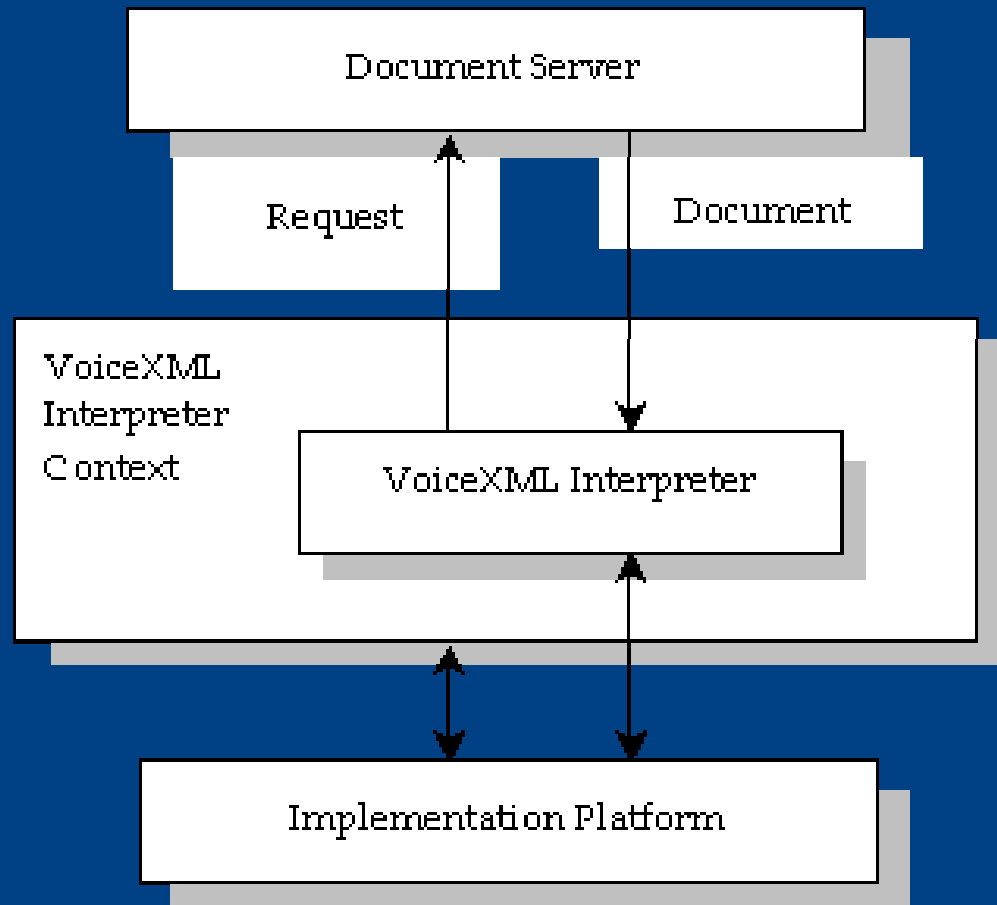
VoiceXML

Can Web understand Croatian language?

- is an XML language for writing Web pages you interact with by listening to spoken prompts and jingles, and control by means of spoken input
- is not a HTML
- is designed to control spoken dialog between the user and the application

VoiceXML Architecture

VoiceXML – Key Concepts

- Session – starts with phone call and lasts until the comunication is disconnected
- VoiceXML documents define an application
- Basic element is dialog
  - Forms: collect data
  - Menus: offer several options
- Subdialogs – similar behaviour as a function call
- Events – signal unexpected situation
- JavaScript – allows additional controll

VoiceXML example

```
<?xml version="1.0" encoding="iso-8859-2"?>
<vxml version="2.0" lang="hr">
<menu>
  <prompt>Odaberite područje vijesti: <enumerate/></prompt>
  <choice next="http://www.carnet.hr/users.vxml">
    CARNet korisnici</choice>
  <choice next="http://www.CARNet.hr/network.vxml">
    mreža</choice>
  <choice next="http://www.carnet.hr/educa.vxml">
    obrazovanje</choice>
  <choice next="http://www.carnet.hr/admin.vxml">
    sistemci</choice>
</menu>
</vxml>
```

VoiceXML example

Computer: Odaberite područje vijesti: CARNet
korisnici; mreža; obrazovanje; sistemci.

User: Sigurnost

Computer: Nisam razumio što ste rekli. (a platform-
specific default message)

Computer: Odaberite područje vijesti: CARNet korisnici;
mreža; obrazovanje; sistemci.
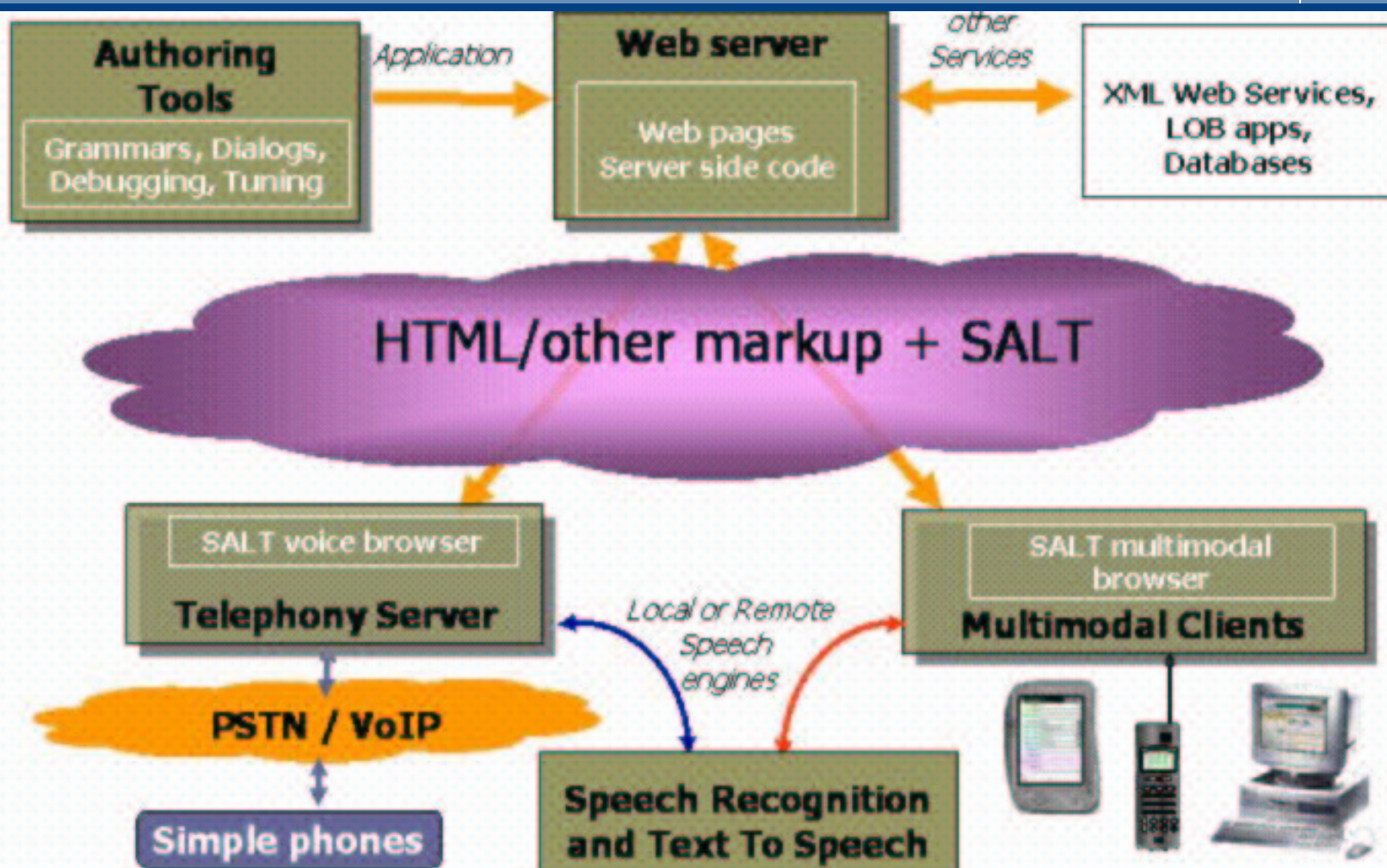
User: CARNet korisnici

Computer: (proceeds to http://www.CARNet.hr/users.vxml)

SALT

- Speech Application Langugage Tags
- SALT is not independent language
- SALT and VoiceXML have different technical goals
  - VoiceXML is designed for telephony-based application and uses a document-based approach
  - SALT focuses on adding multimodal capability to web-based application and takes programming approach

SALT Architecture

SALT elements

- Very small set of tags

<listen ...> configures the speech recognition engine, executes recognition process and handles speech input events

– <grammar>, <bind>, <record>

<prompt ...> configures the text-to-speech engine and plays out defined prompts

<dtmf ...> configures and control DTMF in telephony applications

– <grammar>, <bind>

SALT example

```
<html xmlns:salt=http://www.saltforum.org/2002/SALT>
<body onload="Run()">
  <form id="UserForm" method="post" action="login.php">
  <input name="Name" type="text"></form>
  <salt:prompt id="hello">Hello, please enter Your name</salt:prompt>
  <salt:listen id="recoName" onreco="processName()">
  <salt:grammar src="name.xml"/></salt:listen>
</body>
<script>
 function Run()
{   hello.Start();
    recoName.Start();  }
 function processName()
{   UserForm.Name = recoName.text;
    UserForm.submit();  }
</script>
</html>
```

Voice and Multimodal browsers solutions

Can Web understand Croatian language?

- Commercial solutions:
  - Very expensive
  - Difficult to find speech implementation with Croatian support
  - BeVocal Café, Conita, HeyAnita, IBM VoiceServer, Microsoft .Net Speech, SpeechWorks
- Open Source solutions:
  - Incomplete
  - Open source speech implementation (we can build Croatian language support)
  - OpenVXI, OpenSALT

Open Source solutions OpenVXI

- **SpeechWorks open source release**
  - Full VoiceXML 1.0 implementation
  - Covers the most of VoiceXML 2.0
- **Included in release:**
  - VoiceXML interpreter, XML parser (Apache Xerecs), JavaScript (Mozzila SpiderMonky interpreter), Internet and cache librarry, and logging interface
- **Missing in release:**
  - Speech recognition engine, text-to-speech engine, and call-controll

Open Source solutions OpenSALT

Can Web understand Croatian language?

- Carnegie Mellon University open source project
- First release at the end of 2002
- System base is:
  - Mozilla web browser
  - Sphinx speech recognition engine
  - Festival text-to-speech engine
- Support for Windows and Linux platform

Speech recognition - Sphinx

Can Web understand Croatian language?

- Group of products for building speech recognition engine from Carnegie Mellon University
  - Sphinx 2 and sphinx 3 for speech recognition
  - SphinxTrain for acoustic model training
  - Web-based language modeling tool and CMU-Cambridge Statistical Language Modeling Toolkit for building language models

Speech synthesis - Festival

Can Web understand Croatian language?

- Complete open source solution for building text-to-speech system
- With Festival we can build new voices and play existing voices
- Festival can use text and linguistic analysis from MBROLA project to generate waveform
  – Croatian is one of MBROLA languages

Something about Speech Recognition Theory

Given a sequence o_1^T, a word sequence $\hat{W}$ is generated by the speech recognition system with a search process based on the rule:

$$\hat{W} = \arg\max_{W} p(o_1^T \mid W)\, p(W)$$

- Probability $p(o_1^T \mid W)$ is computed by Acoustic Models (AM)

- Probability $p(W)$ is computed by Language Models (LM)

Acoustic models

- Describe basic unit of acoustic signals
  - Phone (monophone, allophone, biphone, triphone)
  - Word
- Modeled by HMM (Hidden Markov Model)
- Appropriate training database (recorded speech)
  - All expected word, different speaker (different age, diferent gender, etc.), transcription of recorded speech, large amount of data
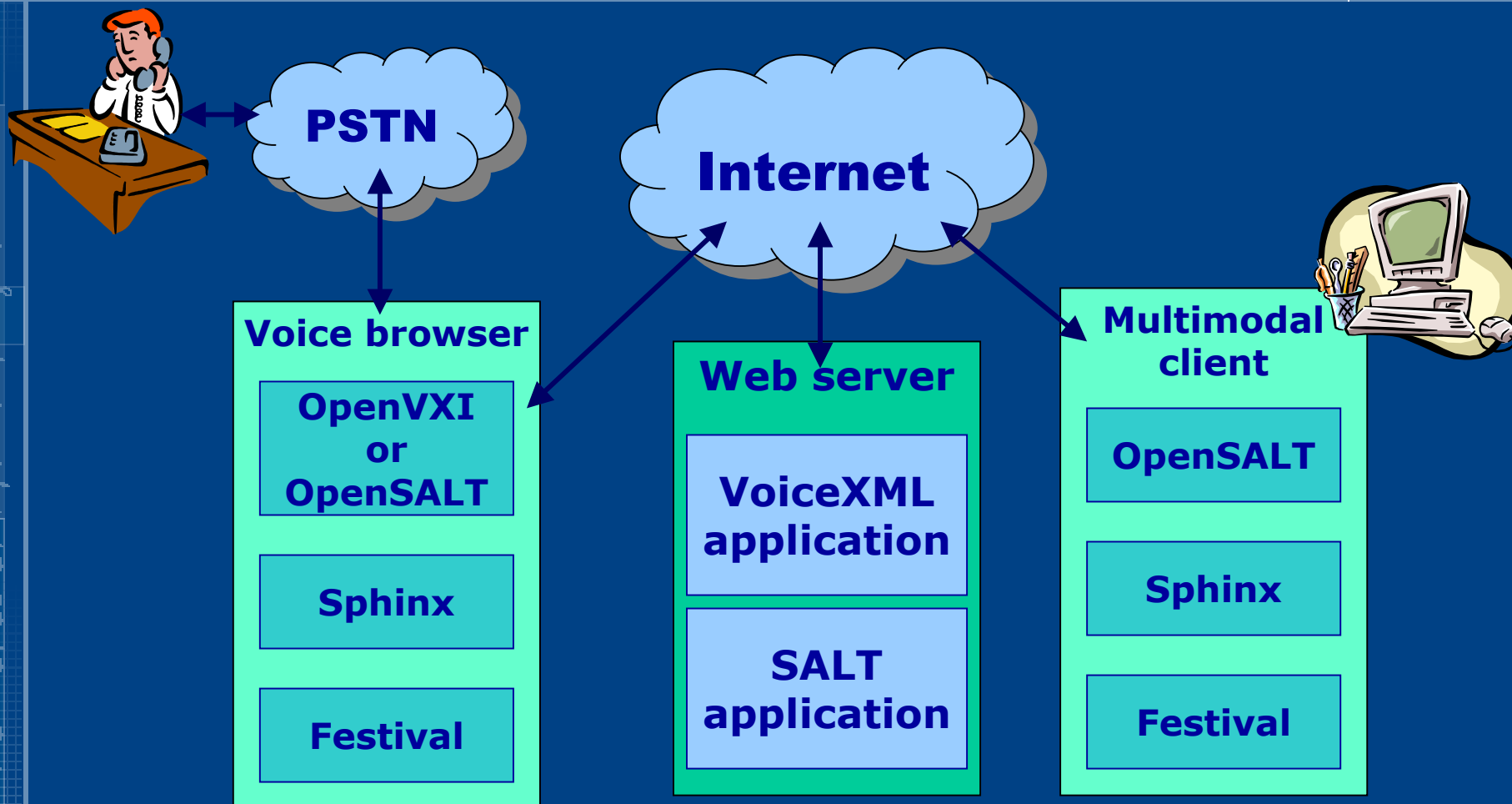
Language models

- Give context or recognition enviroment for speech recognition
- N-gram model for language modelling
  - Represent statistic context of N-word
  - Can be directly estimated from text data
- Training database for Language modelling is a corpus of text data adequate for our application

Open Source System

Can Web understand Croatian language?

PSTN

Internet

Voice browser

OpenVXI or OpenSALT
Sphinx
Festival

Web server

VoiceXML application
SALT application

Multimodal client

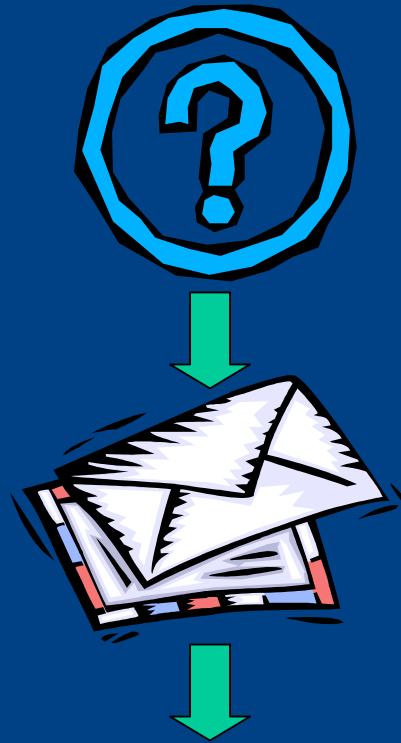| OpenSALT |
| --- |
| Sphinx |
| Festival |

Conclusion

Can the web understand Croatian language?

Yes, it can!

We have the technology and if we use it in the right way we can achieve the aim, "build a useful Croatian voice portal", with more or less efforts

Question?

Can Web understand Croatian language?

Albert.Novak@CARNet.hr