

SpeechOO: Reconhecimento de Voz para o BrOffice.org

Pedro Batista¹ William Colen²

Grupo FalaBrasil - LaPS
Universidade Federal do Pará
<http://www.laps.ufpa.br/falabrasil>¹
BrOffice - CoGrOO²

LatinoWare - 2010



Sumário

Introdução

- Definição

- Motivação

- Uma Aplicação: SpeechOO

A LaPSAPI

- O que é a LaPSAPI?

- O Uso da LaPSAPI

Como o Reconhecimento é Possível?

- Definição

- Os Modelos

- O Sistema

Desempenho

- Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

- Arquitetura

- Instalação

Considerações Finais

- Trabalhos Futuros

- Outros Trabalhos do Grupo FalaBrasil

Sumário

Introdução

- Definição

- Motivação

- Uma Aplicação: SpeechOO

A LaPSAPI

- O que é a LaPSAPI?

- O Uso da LaPSAPI

Como o Reconhecimento é Possível?

- Definição

- Os Modelos

- O Sistema

Desempenho

- Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

- Arquitetura

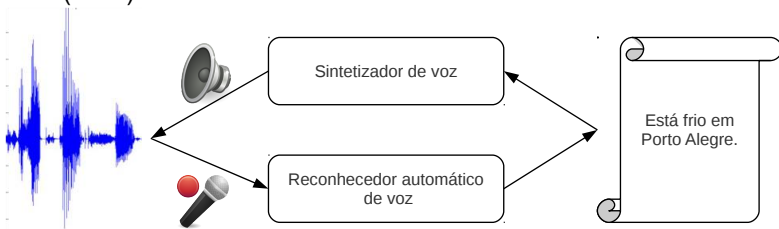
- Instalação

Considerações Finais

- Trabalhos Futuros

- Outros Trabalhos do Grupo FalaBrasil

- Reconhecimento Automático de Voz (RAV) e Síntese de Voz (TTS).



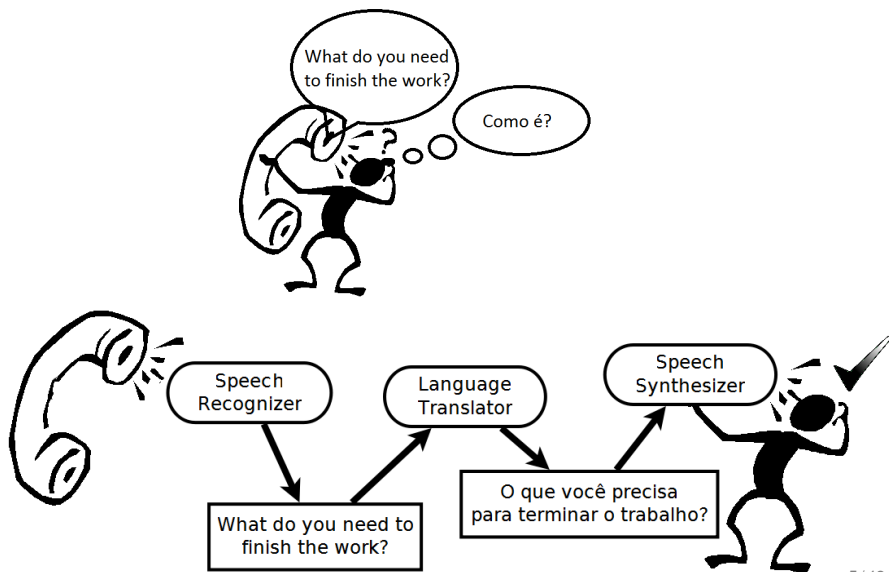


Por que Reconhecimento Automático de Voz?





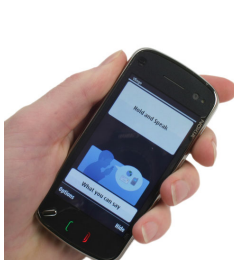
Por que Reconhecimento Automático de Voz?



Por que Reconhecimento Automático de Voz?



Speech Controlled Robotic Arm Edge



SpeechOO o que é?

- Faz uso do reconhecimento de voz para possibilitar ditado para o BrOffice.org Writer.
- Comandos por voz para controlar o BrOffice.org.

Acessibilidade

- Atalhos de teclado.
- Gestos de mouse.
- Joysticks.
- Reconhecimento de Voz - SpeechOO.

Sumário

Introdução

Definição

Motivação

Uma Aplicação: SpeechOO

A LaPSAPI

O que é a LaPSAPI?

O Uso da LaPSAPI

Como o Reconhecimento é Possível?

Definição

Os Modelos

O Sistema

Desempenho

Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

Arquitetura

Instalação

Considerações Finais

Trabalhos Futuros

Outros Trabalhos do Grupo FalaBrasil



O Decodificador Julius e a LaPSAPI

- Um importante integrante do RAV é o decodificador. Ele utiliza os modelos acústicos e de linguagem para realizar a conversão dos sinais de fala para texto.
 - Modos de operação:
 - Comando e controle.
 - Ditado (ou fala espontânea).
 - Dependência de locutor.
 - Principais métricas de avaliação: precisão e velocidade.
- É quem faz o reconhecimento propriamente dito.
- Para funcionar necessita de modelos acústico e de linguagem.
- LaPSAPI:
 - API de fácil uso.
 - Voltada para o desenvolvimento de aplicações.

Como eu Posso Usar Essa API?

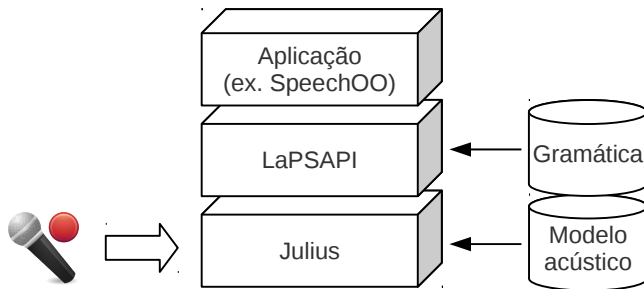
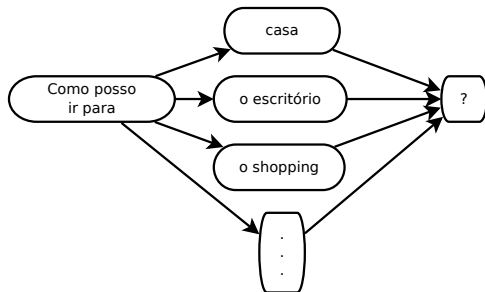


Figure: Interação entre a aplicação do usuário, a LaPSAPI e o Julius

Informando para a Engine o que Reconhecer

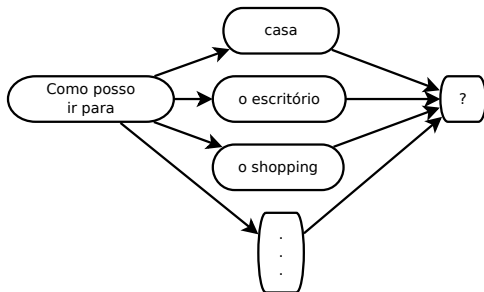
- Você deve definir uma gramática para informar a engine o que ela deve esperar do usuário.



- Você pode também usar modelo de língua para que "tudo" possa ser reconhecido.

Informando para a Engine o que Reconhecer

- Você deve definir uma gramática para informar a engine o que ela deve esperar do usuário.



- Você pode também usar modelo de língua para que "tudo" possa ser reconhecido.

Sumário

Introdução

Definição

Motivação

Uma Aplicação: SpeechOO

A LaPSAPI

O que é a LaPSAPI?

O Uso da LaPSAPI

Como o Reconhecimento é Possível?

Definição

Os Modelos

O Sistema

Desempenho

Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

Arquitetura

Instalação

Considerações Finais

Trabalhos Futuros

Outros Trabalhos do Grupo FalaBrasil

Como Funciona Reconhecimento Automático de Voz?

- A fala é uma sequência de palavras.
- Cada palavra consiste em uma série de sons (fonemas).
- Dicionário fonético: conversão de uma sequência de caracteres em sequência de fonemas.
- Modelos estatísticos baseados em probabilidades:
 - Acústica: cadeias escondidas de Markov (HMM).
 - Da língua: modelos n-gramas.
- Modelos não-probabilísticos: gramáticas livre de contexto.

O Conversor Grafema Fonema (G2P)

- O dicionário fonético nos dá a informação da correspondência entre a forma ortográfica e a pronúncia de uma palavra.

notícia -> n o t s i s i a
(grafema) (fonema)

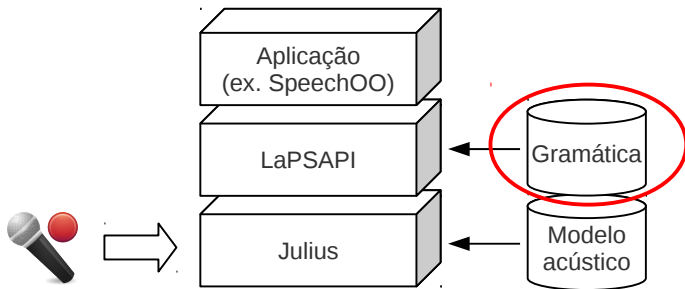
- O LaPS disponibiliza um conversor G2P baseado em regras. Usamos este para gerar o dicionário fonético com 65 mil palavras.

O Conversor Grafema Fonema (G2P)

- O dicionário fonético nos dá a informação da correspondência entre a forma ortográfica e a pronúncia de uma palavra.

notícia -> n o t s i s i a
 (grafema) (fonema)

- O LaPS disponibiliza um conversor G2P baseado em regras. Usamos este para gerar o dicionário fonético com 65 mil palavras.



O Corpora de Texto LaPSNews

- O corpus de texto é simplesmente um conjunto de sentenças.

<s> a cidade não vai se livrar do minhocão </s>

<s> segundo a escola os dois casos não têm relação </s>

<s> meu filho estava a poucos metros de casa </s>

...

- LaPSNews:
 - Baseado na coleta diária de textos da internet (crawling).
 - Foram coletadas mais de 1 milhão de sentenças.

O Corpora de Texto LaPSNews

- O corpus de texto é simplesmente um conjunto de sentenças.

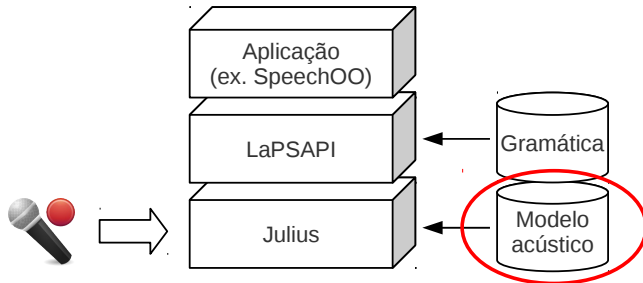
<s> a cidade não vai se livrar do minhocão </s>

<s> segundo a escola os dois casos não têm relação </s>

<s> meu filho estava a poucos metros de casa </s>

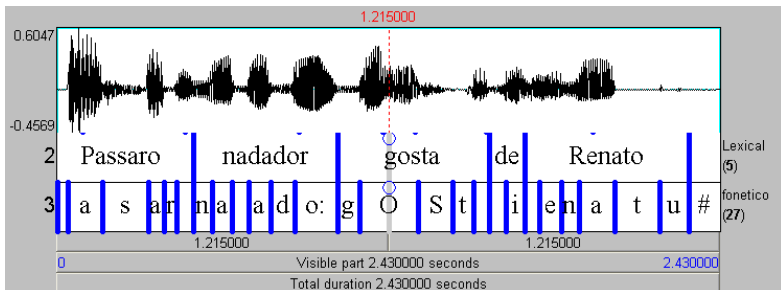
...

- LaPSNews:
 - Baseado na coleta diária de textos da internet (crawling).
 - Foram coletadas mais de 1 milhão de sentenças.



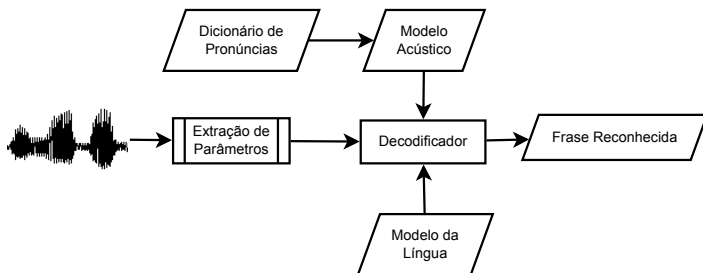
Corpora de Áudio

- Para obter um AM robusto necessita-se de uma grande quantidade de áudio transcrito.
- Para Português brasileiro nos temos certa de 20 horas de áudio de boa qualidade transcrito.
- Outras línguas como o Inglês tem corpus com mais de 240 horas.





Processo de Reconhecimento





É Difícil Construir um Sistema RAV?

- Dados para treino dos modelos acústico e de linguagem.
 - Dificuldade na segmentação da fala.
- Variação no ritmo, timbre e intensidade da fala.
- Diferenças de sotaque.
- Ruído ambiente.

Sumário

Introdução

- Definição

- Motivação

- Uma Aplicação: SpeechOO

A LaPSAPI

- O que é a LaPSAPI?

- O Uso da LaPSAPI

Como o Reconhecimento é Possível?

- Definição

- Os Modelos

- O Sistema

Desempenho

- Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

- Arquitetura

- Instalação

Considerações Finais

- Trabalhos Futuros

- Outros Trabalhos do Grupo FalaBrasil



Quão Bom é o Reconhecedor?

- Três software (modelos) foram avaliados.
 - HDecode.
 - Julius.
 - Software comercial (IBM ViaVoice).
- Medidas de performance.
 - WER (word error rate).
 - xRT (real-time factor).
- A avaliação foi dividida em duas etapas:
 - Modelo dependente de locutor.
 - Modelo independente de locutor.



Comparação usando Modelos Independentes de Locutor

- Os modelos LaPSAM e LaPSLM foram usados com Julius e HDecode.

Decoder	WER(%)	xRT
Julius	29	0.99
HDecode	20.13	1.2
IBM ViaVoice	29.30	-

Table: Comparação dos sistemas usando modelos independentes de locutor.



Comparação usando Modelos Dependentes de Locutor

- Dois modelos dependentes de locutor foram comparados.
- Para cada adaptação foram usados 10 minutos de áudio.

Decoder	CWR(%)	xRT
Julius	13.30	0.99
HDecode	6.81	0.91
IBM ViaVoice	17.30	-

Table: Comparação dos sistemas usando modelos dependentes de locutor.

Sumário

Introdução

Definição

Motivação

Uma Aplicação: SpeechOO

A LaPSAPI

O que é a LaPSAPI?

O Uso da LaPSAPI

Como o Reconhecimento é Possível?

Definição

Os Modelos

O Sistema

Desempenho

Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

Arquitetura

Instalação

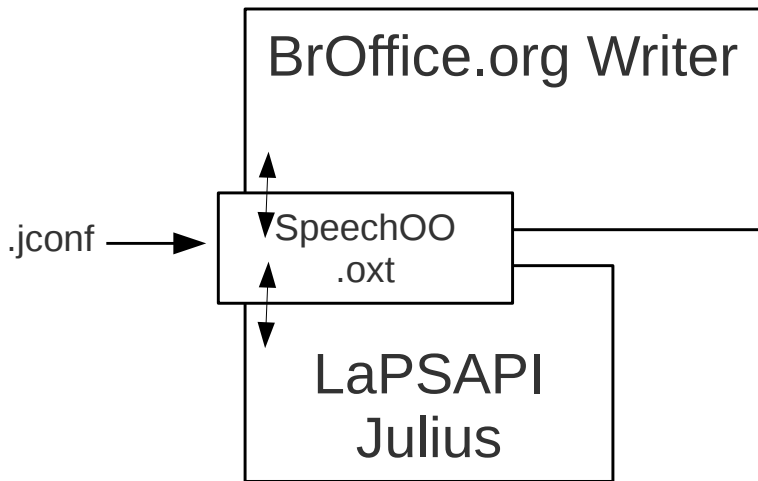
Considerações Finais

Trabalhos Futuros

Outros Trabalhos do Grupo FalaBrasil



Arquitetura do SpeechOO





Página do Projeto


speechoo - Project Hosting on Google Code - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://code.google.com/p/speechoo/

speechoo - Project Hosting on ...

william.colen@gmail.com | My favorites | Profile | Sign out

 **speechoo**
A dictation pad for OpenOffice.org

Project Home Downloads Wiki Issues Source Administer

Summary | Updates | People

Tip: Discuss and then document [each teammate's project duties](#). [hide](#)

SpeechOO is a voice recognition extension for OpenOffice.org. It is still alpha stage. Fell free to download the source code and build it to try. Check our milestones for new alpha releases.

An initial prototype is available, but not connecting to the voice recognition libraries yet. It proves the concept of running mixed extension: Java plus C++ wrapped with JNI.

Are you able to develop OOo extensions or would like to learn how to do it? Join this group and become member of the development team!

Alpha version available

A first alpha release (0.0.1-apha001) of SpeechOO is vallable in downloads section. This version works only under GNU/Linux systems. This can get utterances from Julius and append it to the current document. To use the extension, install the .oxt using Extension Manager and configure SpeechOO using Tools > Options > OpenOffice.org Writer > SpeechOO Options. Set the path to a Julius configuration file (.jconf) which gives the models for the recognition process. We suggest the free models below.

VoxForge English Model <http://www.voxforge.org/home/downloads>


FalaBrasil Brazilian Portuguese Model (dictation) <http://www.laps.ufpa.br/falabrasil/downloads.php>

warning: the extension is experimental and can damage your OOo profile folder.

Requirements: Java 1.6 (or similar). Ubuntu requires the installation of openoffice.org-java-common package.

The project is managed by members of [Fala Brasil/LaPS](#) and [CCSL-JME/USP](#)

Star this project

Activity:  [High](#)

Code license:
[GNU General Public License v3](#)

Labels:
[ASR](#), [coruja](#), [broffice](#)

Feeds:
[Project feeds](#)

Groups:
[Dev group](#)

Owners:
[edokdz](#), [william.colen](#), [nelsonneto@ufpa.br](#)

[People details »](#)



Os Modelos Para Reconhecimento

FalaBrasil - Laboratório de Processamento de Sinais - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.laps.ufpa.br/falabrasil/downloads.php

FalaBrasil - Laboratório de Proc...

FalaBrasil

Reconhecimento de Voz para o Português Brasileiro

Última atualização: 13/06/10

- Descrição
- Integrantes
- Status
- Downloads
- Referências

Softwares:

- **Coruja 1.0** - Software para reconhecimento de voz em Português Brasileiro. Composto por uma API em C++ com suporte a CLR (Common Language Runtime). O engine do reconhecedor é o decodificador **Julius**, que permite uma grande gama de opções além de suportar modelos (acústicos e de linguagem) criados com a ferramenta HTK.
 - **Coruja 1.0** (Windows) - Composto do LaPSAM v1.3 + LapsAPI.
 - **Coruja 1.0** (Linux)
 - **Coruja para SpeechOO** (Ditado no OpenOffice). Download (Linux): [x686](#) [x64](#)
 - **Documentação** Documentation:
 - Português: [PDF - HTML](#)
 - English: [PDF - HTML](#)
 - [Lista de Discussão](#)

Download da Extensão


Downloads - speechoo - Project Hosting on Google Code - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://code.google.com/p/speechoo/downloads/list



Downloads - speechoo - Project...

william.coten@gmail.com | My favorites | Profile | Sign out

 **speechoo**
A dictation pad for OpenOffice.org

Project Home Downloads Wiki Issues Source Administer

New download | Search Current downloads for

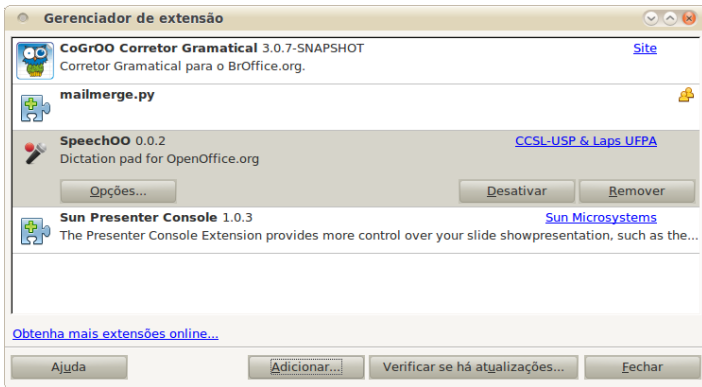
Filename	Summary + Labels	Uploaded	Size	DownloadCount	...
  SpeechOO-0.0.2.oxt	Better terminate handling	5 days ago	1.3 MB	4	

©2010 Google - [Terms](#) - [Privacy](#) - [Project Hosting Help](#)
Powered by [Google Project Hosting](#)

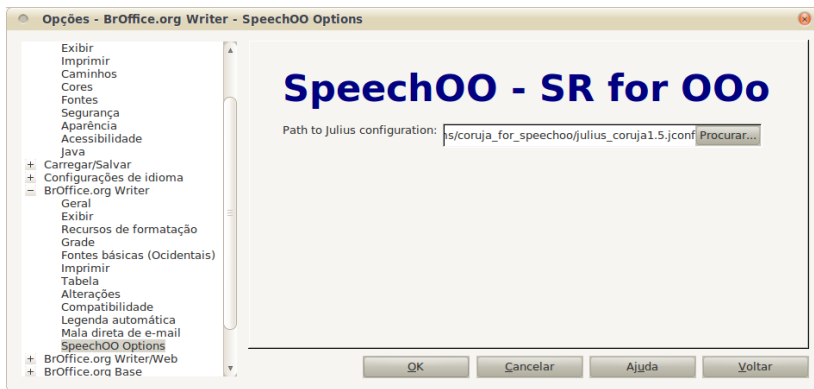
30 / 42



Instalação da Extensão

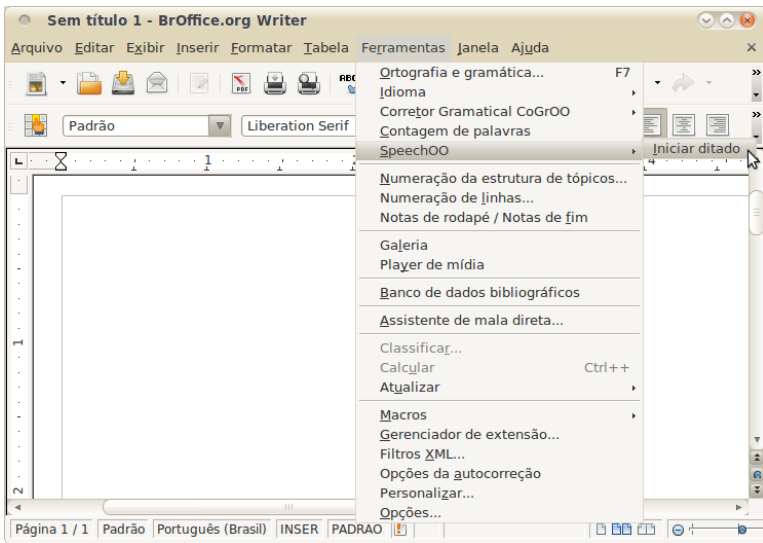


Configurando o SpeechOO



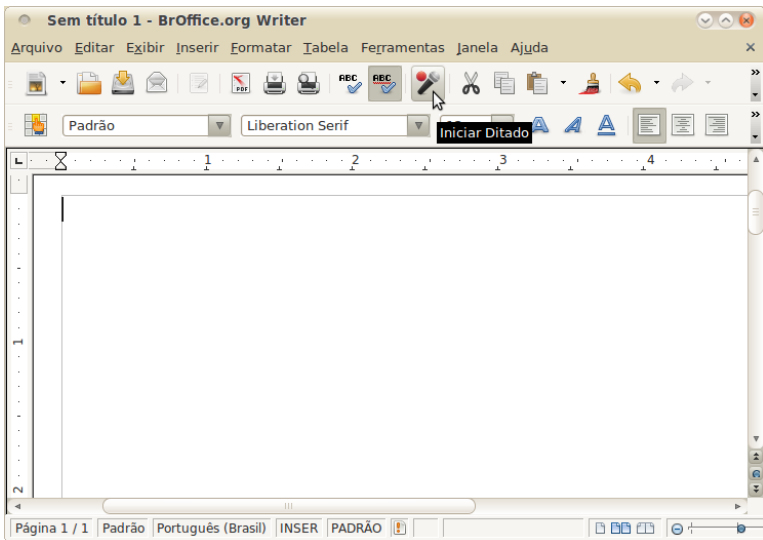


Iniciar Ditado



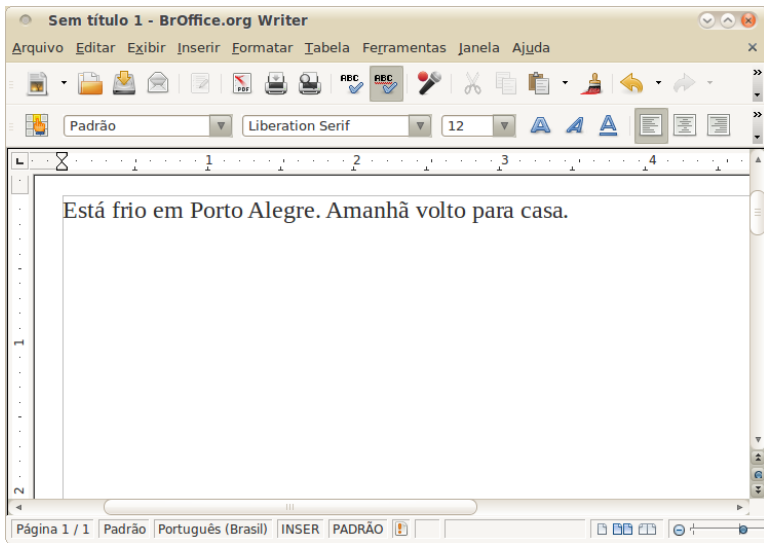


Iniciar Ditado





Ditando



Sumário

Introdução

- Definição

- Motivação

- Uma Aplicação: SpeechOO

A LaPSAPI

- O que é a LaPSAPI?

- O Uso da LaPSAPI

Como o Reconhecimento é Possível?

- Definição

- Os Modelos

- O Sistema

Desempenho

- Comparação entre HDecode, IBM ViaVoice e Julius

SpeechOO

- Arquitetura

- Instalação

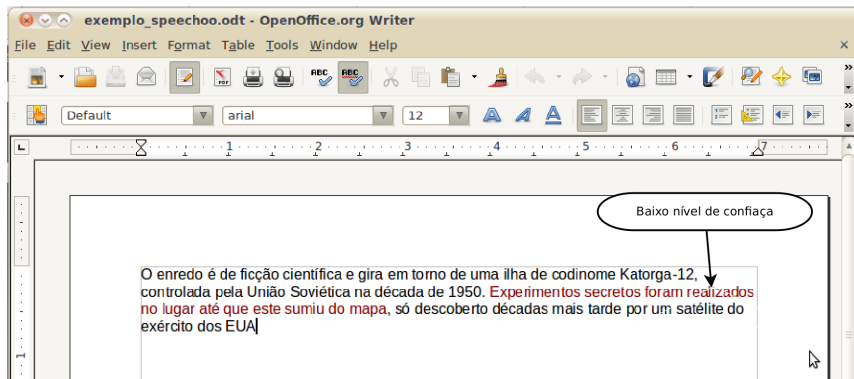
Considerações Finais

- Trabalhos Futuros

- Outros Trabalhos do Grupo FalaBrasil

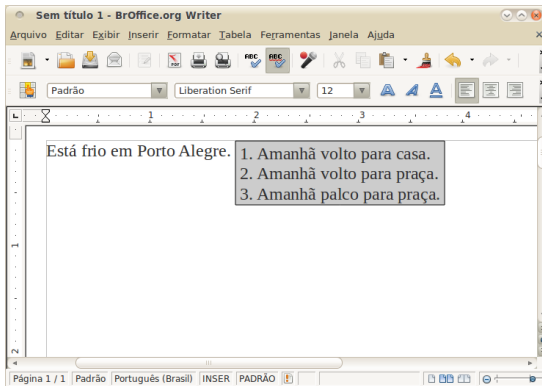


Trabalhos Futuros - SpeechOO





Trabalhos Futuros - SpeechOO





Trabalhos Futuros - SpeechOO

- Possibilitar comandos, como: Negrito, tamanho da fonte, etc.
- Passar a sentença pelo corretor gramatical para corrigir possíveis erros.
- Módulo para fazer recase da sentença reconhecida.
- Permitir a correção de uma palavra que foi reconhecida errado.
- Acrescentar módulo de adaptação de locutor (aguardando módulo do Coruja).



Versão para Português Brasileiro do Simon

- Simon é um controlador do KDE que dispensa o uso do teclado e mouse.
- Permite digitar textos, simular sequências de atalhos, iniciar programas, controle do mouse, etc.
- Possibilita a criação de comandos personalizados.
- Baseado no decodificador Julius e modelos HMM (HTK).
- SimonBr:
 - Uso dos modelos LaPS como base.
 - Criação de comandos (ambientes).
 - Tradução da interface.

Versão para Português Brasileiro do Simon

- Simon é um controlador do KDE que dispensa o uso do teclado e mouse.
- Permite digitar textos, simular sequências de atalhos, iniciar programas, controle do mouse, etc.
- Possibilita a criação de comandos personalizados.
- Baseado no decodificador Julius e modelos HMM (HTK).
- SimonBr:
 - Uso dos modelos LaPS como base.
 - Criação de comandos (ambientes).
 - Tradução da interface.



CorujaNavigator: Um Navegador Web não visual e hands-free

- Navegação web exclusivamente através de texto.
 - O usuário diz da página atual que quer acessar.
 - O CorujaNavigator acessa e sintetiza o texto da nova página.
- Atualmente necessita de métodos para cada domínio a ser acessado.
 - Estamos trabalhando para que o acesso seja livre de domínio.



Obrigado! Perguntas?



CCSL CENTRO DE
COMPETÊNCIA EM
SOFTWARE LIVRE
FLOSS Competence Center



<http://ccsl.ime.usp.br>

