

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO

CONVERSÃO TEXTO-FALA DE PÁGINAS HTML

MAYCON JOHNATHAN FRANCISCO

BLUMENAU
2010

MAYCON JOHNATHAN FRANCISCO

CONVERSÃO TEXTO-FALA DE PÁGINAS HTML

Proposta de Trabalho de Conclusão de Curso
submetida à Universidade Regional de
Blumenau para a obtenção dos créditos na
disciplina Trabalho de Conclusão de Curso I
do curso de Ciência da Computação —
Bacharelado.

Profa. Joyce Martins - Orientadora

**BLUMENAU
2010**

1 INTRODUÇÃO

Vive-se hoje na sociedade da informação e do conhecimento, tanto que chega ser um dos fatores cruciais para o sucesso, estar a par das últimas notícias, sociais ou profissionais. Um dos principais acessos a essas notícias se dá através das tecnologias de informação e comunicação. Assim, a internet desempenha um papel fundamental dentre essas tecnologias, por facilmente romper barreiras físicas e espaciais. Portanto, ela deveria estar disponível para o maior número possível de cidadãos, evitando-se exclusão social. Sendo assim, como pode ser possível o uso da internet pelos portadores de deficiência visual? Uma das soluções é o uso dos sistemas de conversão texto-fala (*Text-to-speech*) ou síntese de fala. Segundo Oechsler (2009, p.13), “Este tipo de software, difundido principalmente na língua inglesa, pode auxiliar pessoas com deficiência visual, e até mesmo analfabetos, a usarem um computador”.

Diante do exposto, propõe-se disponibilizar uma ferramenta para sintetizar páginas *HyperText Markup Language* (HTML) através da conversão texto-fala. Para tanto será transcrito para a linguagem de programação Java, bem como aprimorado, o protótipo desenvolvido por Oechsler (2009), o FurbTTS, que é um sistema que efetua a conversão texto-fala a partir do processamento de texto com vocabulário irrestrito escrito em língua portuguesa. A ferramenta aqui proposta será uma extensão do FurbTTS e terá como entrada uma página HTML e como saída a síntese de voz do conteúdo da mesma. A problemática está na obtenção da página HTML, sendo necessário desenvolvimento de um navegador. Faz-se necessário também efetuar o tratamento das páginas HTML, já que as mesmas não estão em formato de texto padrão, para que então seja sintetizada, através da ferramenta proposta.

1.1 OBJETIVOS DO TRABALHO

O objetivo deste trabalho é desenvolver um navegador para conversão texto-fala de páginas HTML.

Os objetivos específicos do trabalho são:

- a) migrar a ferramenta FurbTTS desenvolvida em Delphi para Java;
- b) processar páginas HTML, extraindo o conteúdo propriamente dito, ou seja,

- separando o conteúdo do código fonte, incluindo texto, *links* e título da página;
- c) sintetizar a página HTML processada.

1.2 RELEVÂNCIA DO TRABALHO

A internet possibilita a descentralização da informação, da cultura e da educação, bem como favorece a exploração de atividades profissionais além de outras tantas possibilidades. Pretende-se desenvolver uma ferramenta que permita que pessoas portadoras de deficiência visual possam fazer uso da internet, sem ajuda de outros. Portanto, considera-se o desenvolvimento da ferramenta proposta relevante tanto do ponto de vista social quanto técnico, já que será necessário desenvolver um sistema navegador que obtenha as páginas HTML, conforme requisição do usuário. Pretende-se construir um navegador, totalmente implementado em Java, capaz de sintetizar em voz o conteúdo das páginas visitadas.

1.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- a) levantamento bibliográfico:
- realizar levantamento de bibliografia sobre sistemas de conversão texto-fala,
 - estudar o protótipo FurbTTS, incluindo as etapas de processamento de texto implementadas (pré-processador, analisador lingüístico, identificador fonético e processador prosódico),
 - estudar o funcionamento do sintetizador MBROLA, bem como o formato de entrada e de saída,
 - buscar e estudar bibliotecas para auxiliar no desenvolvimento da ferramenta proposta, tal como a *Application Program Interfaces (API) Java Speech Markup Language (JSML)*, verificando como poderá ser integrada ao sintetizador MBROLA;
- b) migração do FurbTTS para Java: transcrever o código do protótipo FurbTTS para a linguagem Java, junto ao ambiente NetBeans e o sintetizador MBROLA;

- c) elicitação dos requisitos: detalhar os requisitos, observando as necessidades e peculiaridades levantadas nas etapas anteriores;
- d) especificação: especificar a ferramenta com análise orientada a objeto utilizando a *Unified Modeling Language* (UML). Será utilizado o Enterprise Architect para a elaboração dos diagramas de caso de uso, de classe, sequência e de estado;
- e) implementação: implementar a ferramenta especificada na etapa anterior para tratamento e conversão em texto padrão de páginas HTML e posterior sintetização do mesmo. Será utilizada a linguagem Java junto ao ambiente NetBeans, bem como o sintetizador MBROLA e as bibliotecas definidas na etapa (a);
- f) testes: efetuar testes para validação da conversão em texto padrão, bem como da sintetização de páginas HTML.

As etapas serão realizadas nos períodos relacionados no Quadro 1.

etapas / quinzenas	2011									
	fev.		mar.		abr.		maio		jun.	
	1	2	1	2	1	2	1	2	1	2
levantamento bibliográfico										
migração do FurbTTS para Java										
elicitação dos requisitos										
especificação										
implementação										
testes										

Quadro 1 - Cronograma

2 REVISÃO BIBLIOGRÁFICA

Esse capítulo está organizado em 4 seções: a primeira trata de sistemas sintetizadores texto-fala, a segunda descreve o protótipo FurbTTS, a terceira e a quarta apresentam, respectivamente, o sintetizador MBROLA e a API JSML, e a última seção traz trabalhos correlatos à ferramenta proposta.

2.1 SISTEMA DE CONVERSÃO TEXTO-FALA

Os primeiros sistemas sintetizadores de voz começaram a surgir a partir de 1922, data do primeiro sintetizador que se tem registro (GOMES, 1998, p. 5). Um sistema sintetizador texto-fala tem como principal objetivo efetuar a síntese de um texto em voz. Em outras palavras, transforma texto em áudio. Para tanto o sistema faz uso dos seguintes procedimentos: pré-processamento, analisador lingüístico, analisador fonético, processador prosódico,

O pré-processamento do texto “analisa o texto de entrada para construções especiais da língua [...] tratamentos especiais são requeridos para as abreviaturas, acrônimos, datas, épocas, números, moeda corrente [...]” (MARANGONI; PRECIPITO, 2006). Oechsler (2009, p. 20) afirma que “um sistema de conversão texto-fala deve primeiramente eliminar sentenças estranhas e formatar o texto de acordo com o vocabulário suportado”. Além da verificação das palavras válidas/existentes dentro do vocabulário suportado, deve-se haver um cuidado com símbolos especiais, siglas e números devem ser transcritos para sua forma extensa, bem como fazer o tratamento de palavras herdadas do latim, ex: exame para ezame.

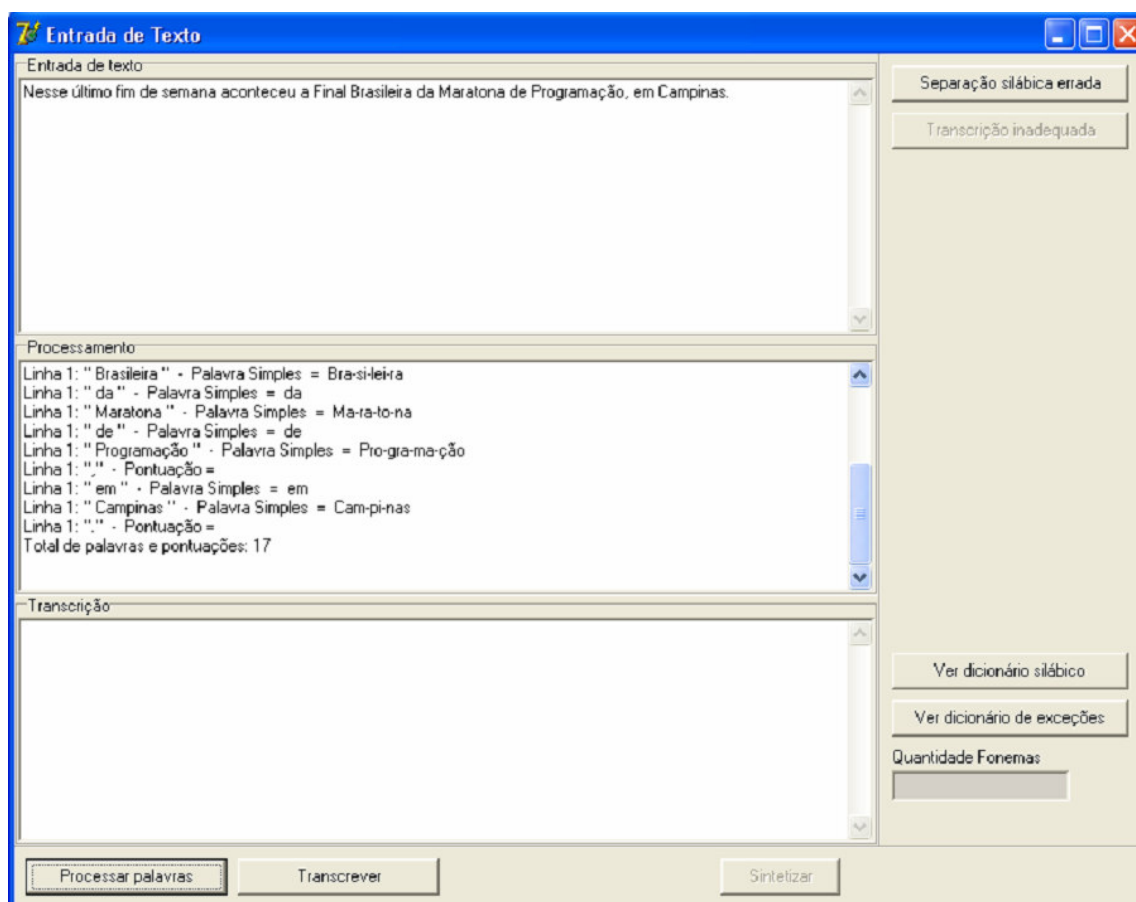
O analisador lingüístico é responsável por definir as classes gramaticais das palavras encontradas no texto de entrada, depois de pré-processado. É com o resultado dessa classificação que o programa irá definir a pronúncia de cada palavra. A classificação pode se dar por duas técnicas: por comparação e por regras posicionais (OECHSLER, 2009, p. 21).

Já o analisador fonético transforma uma sequência ortográfica em um conjunto de símbolos que represente a sequência de sons que compõe cada uma das palavras do texto. (SIMÕES, 1999, p. 52 apud OECHSLER, 2009, p. 22). É ele também quem vai analisar a sílaba tônica da palavra, muito importante para a naturalidade do resultado final.

Marangoni e Precipito (2006) afirmam que “Frasear corretamente uma sentença, para produzir a melodia correta para uma sentença e para enfatizar corretamente palavras” é função do processador prosódico. O processador prosódico é então, responsável por tratar da entonação, duração do segmento e intensidade sonora.

2.2 FURBTTS

O FurbTTS é um protótipo sintetizador texto-fala para língua portuguesa com suporte as regras do novo acordo ortográfico gramatical. A partir de uma entrada de texto em português, o protótipo faz pré-processamento do texto, fazendo o isolamento e tratamento inicial das palavras, números, siglas e símbolos especiais; separa as palavras silabicamente, transcreve-as em fonemas com seu tempo e informação prosódica; para que, por fim, seja sintetizada com ajuda do sintetizador MBROLA e um banco de fonemas em português, denominado BR3 (COSTA, 2005). A figura 1 mostra o protótipo de Oechsler (2009).



Fonte: Oechsler (2009, p. 56).

Figura 1 – FurbTTS

No campo `Entrada de texto` (figura 1) é escrito o texto que será sintetizado. No campo `Processamento` são exibidas informações do processamento do texto original, gerados através do botão `Processar palavras`. Pressionando o botão `Transcrever` é apresentado no campo `Transcrição` os fonemas e suas informações prosódicas. Para síntese do texto é utilizado o botão `Sintetizar`.

A operacionalidade do protótipo é bastante simplificada. É importante salientar que o mesmo não está preparado para ser operado por pessoas com deficiência visual severa (que não enxergam). Isso se deve ao fato de que o objetivo principal da interface é testar a qualidade do processamento textual que é realizado inteiramente pelo pacote (biblioteca) desenvolvido com nome FurbTTS. (OECHSLER, 2009, p. 54).

Para o pré-processamento do texto de entrada, foram definidas duas gramáticas, a fim de evitar erros léxicos e sintáticos. Junto ao analisador léxico foi implementado o analisador linguístico, cujo objetivo é verificar a classificação das palavras, verificando se não é uma abreviação por exemplo. Além disso, o analisador linguístico possui a responsabilidade de fazer a separação silábica, que por sua vez, é de extrema importância para o identificador fonético, que consiste em identificar os fonemas presentes em cada palavra. O identificador fonético possui um método de transcrição para cada classe de palavra, sendo o método `TranscreverClassePalavra` o mais importante. Neste método encontram-se as regras para identificação dos fonemas do alfabeto completo, incluindo os encontros consonantais, a acentuação etimológica e as regras do novo acordo ortográfico gramatical.

O quadro 2 apresenta o resultado da transcrição da frase “bom dia” pelo FurbTTS, para posterior envio ao sintetizador MBROLA

b	105	50	90
om	90		
d	100	50	90
i	80		
a	130	10	86 90 82
_	15	10	90

Quadro 2 – Transcrição da frase “bom dia” pelo FurbTTS

2.3 MBROLA

Sintetizador de voz que teve início em 1995 pelo *TCTS Lab* da *Faculté Polytechnique de Mons* (Bélgica), cujo objetivo do projeto de mesmo nome é disponibilizar sintetizadores de alta qualidade para tantas línguas quanto possível, livre para uso em aplicações não

comerciais (DUTOIT et al., 2005). O MBROLA está baseado na concatenação de difones¹. Tem como entrada uma lista de fonemas, juntamente com a informação prosódica (duração do fonema, descrição linear de entonação), e como saída produz amostras em 16 bits (linear), na frequência do banco de difones utilizado, o que não o torna um sistema texto-fala, já que não aceita texto bruto como entrada.

Segundo Przygoda (2010, p. 25), “O algoritmo MBROLA é derivado de outro algoritmo chamado *Pitch Synchronous OverLap Add* (PSOLA), que foi originalmente desenvolvido pela *France Telecom*”. O PSOLA é basicamente dividido em três passos (MBROLA TEAM, 2000 apud PRZYGODA, 2010, p. 25):

- a) análise da fonte de som;
- b) separação de sinais através da filtragem subtrativa;
- c) reconstrução dos sinais e sintetização da voz.

Segundo Oechsler (2009, p. 41), “a instalação do MBROLA requer apenas o uso de uma biblioteca (`mbrplay.dll`) que deve acompanhar o sistema que utilizar o sintetizador.

2.4 API JSML

A API JSML, especificada pela Sun Microsystems (2009), tem como objetivo fornecer aos sintetizadores de voz, através de anotações textuais, informações sobre qualidade, inteligibilidade e naturalidade do texto a ser sintetizado. Além de estabelecer estilos estruturais, enfatiza a pronuncia de palavras e a natureza das declarações, determina pausas, controla o nível sonoro de forma dinâmica e o ritmo da pronuncia.

A JSML é derivada da *eXtensible Markup Language* (XML). “Documentos JSML podem prover ao sintetizador várias fontes de dados, como livros, documentos técnicos, páginas web, e-mails, autoatendimento e informações em terminais aéreos para que se produza a voz sintetizada como saída de áudio.” (PRZYGODA, 2010, p. 18). O quadro 3 apresenta o formato de um arquivo JSML com frase “bom dia”.

¹ Difones “são pequenas sequências de áudio que mostram a transcrição da metade de um fonema para a metade do outro” (MACHADO, 2007).

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<jaml lang="pt-br">
  <voice gender="male" age="adult" variant="6" name="br3"></voice>
  <sayas>bom dia</sayas>
</jaml>
```

Quadro 3 – Frase “bom dia” em formato JSML

2.5 TRABALHOS CORRELATOS

A seguir são apresentadas ferramentas que utilizam sintetizadores de voz, sintetizando por meios próprios ou através de sintetizadores externos, com o objetivo de prover subsídio a invisuais no uso do computador, as quais são: WebVox (PORTO, 2001) e WebbIE (KING, 2009).

2.5.1 WebVox

O WebVox é o navegador da família DOSVOX², que na exibição de uma página HTML, traduz a informação gráfica contida na mesma para informação sonora, através de síntese de voz para transformação dos textos. Utiliza também sons gravados para *tags* HTML, como por exemplo, na existência de uma referência externa `<href>` durante o texto, é reproduzido um som, para que o usuário saiba da existência do mesmo. Desta forma, se ganha em acessibilidade.

O WebVox permite quatro opções de leitura (PORTO, 2001, p. 77):

- a) textual: não é sintetizada nenhuma *tag*, consistindo em uma forma para leitura rápida;
- b) resumida: as *tags* recebem apenas “bips” e as figuras com nomes alternativos (*alternate name*) são sintetizadas;
- c) normal: todas as *tags* são sonorizadas;
- d) detalhada: todas as *tags* são sintetizadas, como por exemplo “aqui uma linha

² DOSVOX é um sistema, cuja interação homem-máquina é obtida através da síntese de voz, objetivando o uso de computadores por deficientes visuais (BORGES et al, 2002). O projeto foi iniciado em 1993 e é desenvolvido no Núcleo de Computação Eletrônica (NCE), da Universidade Federal do Rio de Janeiro (UFRJ).

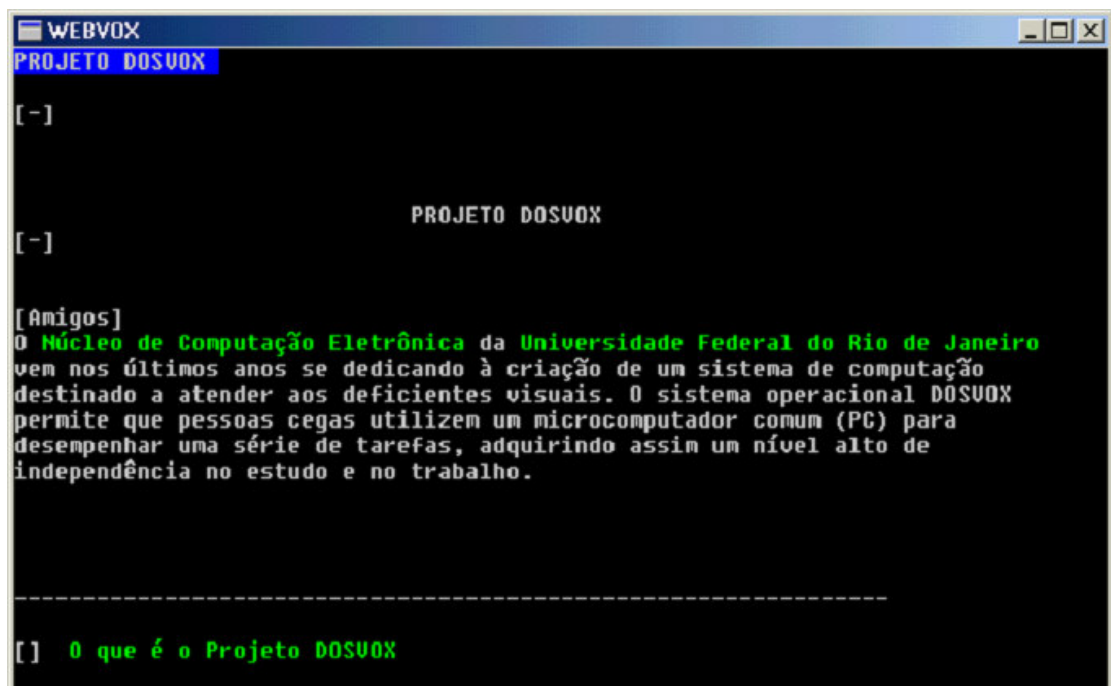
horizontal”.

As figuras 2 e 3 mostram, respectivamente, uma mesma página HTML em um navegador convencional e no WebVox.



Fonte: Porto (2001, p. 74).

Figura 2 – Página HTML em um navegador convencional (Microsoft Internet Explorer)



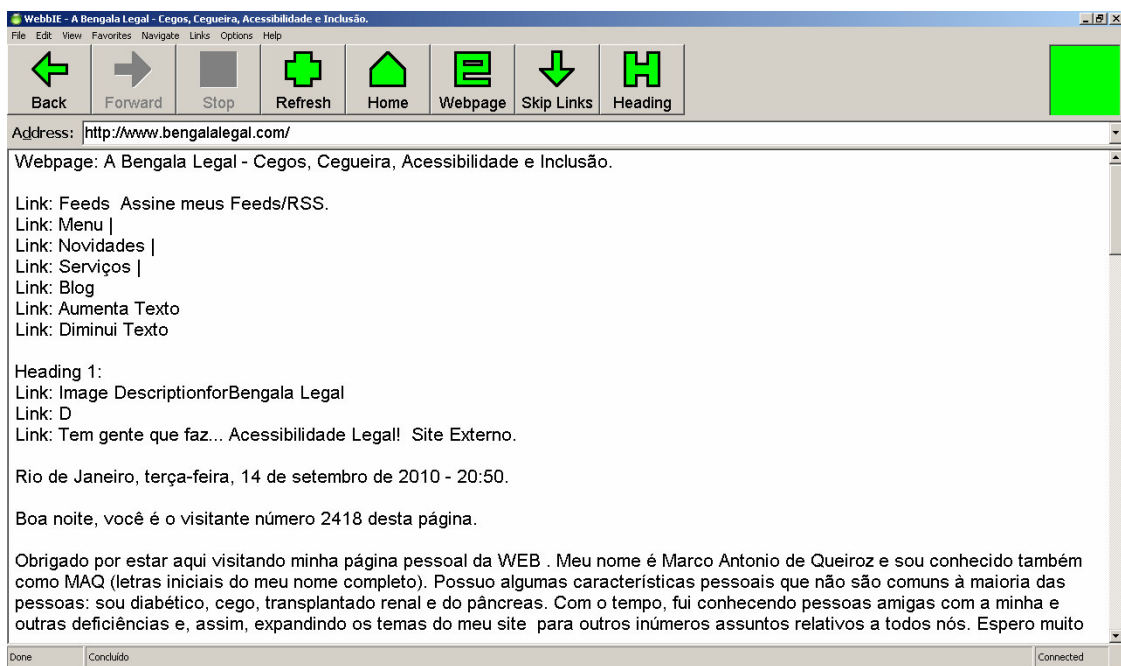
Fonte: Porto (2001, p. 75).

Figura 3 – A mesma página no WebVox

2.5.2 WebbIE

O WebbIE, segundo King (2009), é um navegador baseado no Microsoft Internet Explorer, que faz uso do componente Microsoft WebBrowser ActiveX. O WebbIE é um projeto desenvolvido por estudantes do departamento de Ciência da Computação da *University of Manchester Institute of Science and Technology* (UMIST) desde de 2002, ano de seu lançamento. Foi projetado para ser utilizado em conjunto com softwares sintetizadores leitores de tela, ou seja, ele próprio não faz síntese de voz, porém, possui um modo de exibição. O WebbIE transcreve uma página HTML em texto, separando *links* e textos, para que então sejam sintetizados por qualquer leitor de tela.

As figuras 4 e 5 ilustram o modo de exibição textual e o convencional, respectivamente.



Fonte: Queiroz (2010).

Figura 4 – WebbIE em modo textual



Fonte: Queiroz (2010).

Figura 5 – WebbIE em modo convencional

3 REQUISITOS DO SISTEMA A SER DESENVOLVIDO

O protótipo de sistema de conversão texto-fala deverá:

- a) permitir ao usuário informar o endereço de uma página na web (Requisito Funcional - RF);
- b) carregar a página solicitada pelo usuário (RF);
- c) efetuar o tratamento de páginas HTML, extraíndo de seu código fonte apenas o conteúdo propriamente dito, incluindo textos, *links* e título da página (RF);
- d) sintetizar os textos extraídos das páginas HTML (RF);
- e) ser implementado em Java (Requisito Não-Funcional - RNF);
- f) utilizar uma base de dados de fonemas em português compatível com o MBROLA (RNF);
- g) executar no sistema Windows XP (RNF).

4 CONSIDERAÇÕES FINAIS

Sabe-se que invisuais sofrem com diversas dificuldades e o quão importante é para eles serem os mais autônomos possíveis. Sistemas de conversão texto-fala são uma das formas para proporcionar mais conforto e autonomia para os mesmos. Observa-se que a evolução desses sistemas é constante.

Este projeto terá como ponto de partida a migração do FurbTTS para linguagem de programação Java, bem como implementação de um navegador web que seja capaz de extrair o conteúdo propriamente dito das páginas HTML, tal como textos, título da página e textos dos *links*, para que então, com a ajuda da API JSML, seja possível gerar, a partir do conteúdo extraído das páginas, seus textos com suas respectivas informações prosódicas. Por fim, os textos extraídos serão sintetizados com apoio do sintetizador MBROLA. Desta forma, ao final do projeto, pretende-se disponibilizar um navegador web para que portadores de deficiência visual possam fazer uso, embora limitado, da internet.

Quanto aos trabalhos correlatos, tem-se o WebVox que converte páginas HTML para texto padrão, para então sintetizá-las, tal como a ferramenta proposta. O WebVox reproduz sons gravados, identificado rotinas, como o fim de carregamento de página por exemplo. Além disso, possui 4 diferentes tipos de síntese, textual, resumida, normal e detalhada, conforme descrito na seção 2.5.1. O WebbIE também converte páginas HTML para texto padrão, porém não faz a síntese por si próprio. Seu objetivo é apenas tratar as páginas para dar suporte a leitores de tela, diferente da ferramenta proposta que fará a síntese de páginas HTML. Com o WebbIE tem-se a visualização da página em formato padrão, como em um navegador tradicional, bem como a visualização da página tratada, sendo exibidos apenas textos em formato padrão.

REFERÊNCIAS BIBLIOGRÁFICAS

BORGES, José. A. S. et al. **Projeto DOSVOX**. [Rio de Janeiro], [2002]. Disponível em: <<http://intervox.nce.ufrj.br/dosvox/>>. Acesso em: 12 set. 2010.

COSTA, Denis R. **BR3 database: brazilian portuguese male**. [S.l.], 2005. Disponível em: <<http://tcts.fpms.ac.be/synthesis>>. Acesso em: 13 set. 2010.

DUTOIT, Thierry D. et al. **The MBROLA project: towards a freely available multilingual speech synthesizer**. [Bélgica], 2005. Disponível em: <http://tcts.fpms.ac.be/synthesis/mbrola/mbrola_entrypage.html>. Acesso em: 19 set. 2010.

GOMES, Leandro C. T. **Sistema de conversão texto-fala para a língua portuguesa utilizando a abordagem de síntese por regras**. 1998. 107 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas. Disponível em: <<http://cutter.unicamp.br/document/?code=vtls000132102&fd=y>>. Acesso em: 09 set. 2010.

KING, Alasdair. **Webbie**: the free web browser for blind people with little or no sight. [S.l.], [2009?]. Disponível em: <<http://www.webbie.org.uk/>>. Acesso em: 10 set. 2010.

MACHADO, Lucas. Síntese de voz para produção de livros falados e inclusão social para deficientes visuais. In: FEIRA DE INICIAÇÃO CIENTÍFICA, 16., 2007, Porto Alegre. **Anais eletrônicos...** Porto Alegre: UFRGS, 2007. Disponível em: <<http://ufrgsweb.ufrgs.br/node/247?page=1>>. Acesso em: 19 set. 2010.

MARANGONI, Josemar B.; PRECIPITO, Waldemar B. Reconhecimento e sintetização de voz usando Java Speech. **Revista Científica Eletrônica de Sistemas de Informação**, Garça, v. 2, n. 4, não paginado, fev. 2006. Disponível em: <<http://www.revista.inf.br/sistemas04/notas/notatecnica01.pdf>>. Acesso em: 09 set. 2010.

OECHSLER, Thiago M. **Processamento de texto escrito em linguagem natural para um sistema conversor texto-fala**. 2009. 73 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) – Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau. Disponível em: <<http://www.inf.furb.br/tcc/index.php?cd=11&tcc=1218>>. Acesso em: 01 set. 2010

PORTO, Bernard C. **WEBVOX: um navegador para a world wide web destinado a deficientes visuais**. 2001. 115 f. Dissertação (Mestrado em Informática) - Instituto de Matemática e Núcleo de Computação Eletrônica, Universidade Federal do Rio De Janeiro, rio de Janeiro. Disponível em: <<http://intervox.nce.ufrj.br/dosvox/textos.htm>>. Acesso em: 19 set. 2010.

PRZYGODA, Wendel, D. **Implementação do suporte à programação e interpretação da fala no ambiente LTD**. 2010. 67 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) – Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau. Disponível em: <<http://www.inf.furb.br/tcc/index.php?cd=11&tcc=1258>>. Acesso em: 18 set. 2010.

QUEIROZ, Marco A. **A bengala legal: cegos, cegueira, acessibilidade e inclusão**. Rio de Janeiro, 2010. Disponível em: <<http://www.bengalalegal.com>>. Acesso em: 22 set. 2010.

SUN MICROSYSTEMS. **API markup language specification**. [S.l.], [2009?]. Disponível em: <<http://java.sun.com/products/java-media/speech/forDevelopers/JSML/Specification.html#19137>>. Acesso em: 13 set. 2010.