

# SISTEMA COMPUTACIONAL DE LEITURA AUTOMÁTICA DE TEXTO IMPRESSO

Camilo de Lelis Tosta Paula, Keiji Yamanaka

Universidade Federal de Uberlândia, Faculdade de Engenharia Elétrica/Computação, Uberlândia – Minas Gerais

**Resumo** - O objetivo deste documento é apresentar um sistema computacional voltado para acessibilidade visual, acoplado em um sintetizador de voz, que proporciona ao usuário leitura em tempo real de mídia impressa, podendo ser utilizado por pessoas com deficiência visual total ou parcial. Esse sistema poderá ser utilizado também para os que não conseguem ou têm dificuldades para aprendizagem ao utilizar os métodos convencionais de inclusão na sociedade.

**Palavras-Chave** – acessibilidade, artificiais, educação, neurais, redes, sintetizador.

## COMPUTER SYSTEM FOR AUTOMATIC READING OF PRINTED TEXT

**Abstract** – This paper aims at presenting a computer system aimed at visual accessibility, coupled to a voice synthesizer, which provides the user with real time reading print media. It can be used by people with visual impairments in whole or in part. This system can also be used for those who can not or have trouble learning to use conventional methods of inclusion in society.

**Keywords** - accessibility, artificial, education, neural networks, synthesizer.

### I. INTRODUÇÃO

Leonhard Euler, matemático do século XVIII disse - “Realmente, seria uma invenção considerável, a de uma máquina capaz de imitar nossa fala, com seus sons e articulações, eu acho que isso não é possível”. Com o passar dos anos a afirmação de não ser possível se tornou persuasão falsa. O que para a época não era alcançável, hoje é praticamente uma realidade. Tanto voz como articulações já são realizadas por máquinas, o desafio está em aplicar este conhecimento para que a sociedade tire o melhor proveito.

A distribuição da informação como elemento fundamental da sociedade pode ser realizada através da cultura e educação, o que a torna um veículo de disseminação dos conhecimentos mais eficazes para uma gradual formação de uma população cidadã. O sistema proposto neste trabalho é utilizado para realizar leitura de textos a partir da imagem da página, e irá abrir novos horizontes para o deficiente visual, sendo que este poderá ler (por assim dizer, ouvir) livros, revistas, jornais baseados na mídia impressa. Dessa forma, será aplicada a Tecnologia Assistiva (TA) que, de acordo com [11], “a TA deve ser então entendida como um auxílio que promoverá a ampliação de uma habilidade funcional deficitária ou possibilitará a realização da função desejada e que se encontra impedida por circunstância de deficiência ou

pelo envelhecimento.” O projeto “Pequenos Olhares”, do Conselho Brasileiro de Oftalmologia apontou para números semelhantes aos da OMS (Organização Mundial de Saúde). Segundo a OMS, existem 45 milhões de cegos e 135 milhões de portadores de baixa visão, caracterizando assim uma visão subnormal (baixa) [4]. Imaginemos esse universo de pessoas tendo acesso à leitura em formato de áudio, crianças ouvindo leituras, adolescentes podendo discutir livros de literatura sem a necessidade de outra pessoa realizar a leitura, otimizando, assim, o processo de aprendizagem e, principalmente, realizando a inclusão desse cidadão na sociedade sem a necessidade da realização de um retrabalho, que seria a criação de mídias em estúdio para atender ao deficiente visual, e, assim qualquer livro em qualquer lugar poderia ser uma obra acessível para ele. Sendo assim, o sistema promove a acessibilidade do deficiente visual no mundo da leitura. O artigo está dividido nas seguintes seções: a seção A contém os fundamentos teóricos nos quais o sistema será desenvolvido, as características que envolvem redes neurais, seu treinamento e aprendizado; na seção B, os trabalhos relacionados ao projeto aqui proposto, as principais semelhanças e diferenças; na seção C, será apresentada a arquitetura do sistema, seus módulos de execução, métodos, classes e bibliotecas utilizadas para trabalhar com redes neurais; na seção D, é apresentada a conclusão sobre o tema estudado.

#### A. Fundamentos teóricos

##### a) Java Application;

De acordo com [13] é uma tecnologia constituída de uma linguagem de programação e um programa para execução chamado *virtual machine*. Quando se propõe a usar essa tecnologia, usa-se a linguagem de programação Java e um ambiente de desenvolvimento Java para gerar um *software* que será executado em um ambiente de distribuição Java. Baseada em Orientação a Objetos é hoje uma das mais conceituadas linguagens de programação comercial do mundo.

##### b) Processamento digital de imagens;

De acordo com [12] processar uma imagem consiste em transformá-la sucessivamente com o objetivo de extrair mais facilmente a INFORMAÇÃO nela presente. Existem duas linhas que abordam essas técnicas complexas, que seriam a parametrização da imagem, usada para descrever diferentes informações de parâmetros, passados através de medidas, e a *image Enhancement*, que permite ter uma melhoria de qualidade, utiliza níveis de cinza para aumentar o contraste ou para evidenciar uma determinada área da imagem. O sistema proposto utiliza a segunda linha, usando a filtragem digital para definição de áreas específicas.

c) Principais conceitos e características de uma rede neural Artificial;

Possuem características muito apropriadas à ação proposta. Será dividido de acordo com os fatores de treinamento definido pelo aprendizado supervisionado, e será proposto um padrão a ser reconhecido juntamente com a resposta que a rede deve fornecer ao encontrar o mesmo padrão. Caso a resposta seja diferente é gerado um sinal de erro, que é usado para ajustar os pesos sinápticos da rede e assim minimizar os erros. Esse ajuste vai acontecendo até encontrar o padrão desejado. O algoritmo usado por esse tipo de aprendizado é o adaptativo por correção de erros que, de acordo com [1], é aplicado no decorrer do processo de treinamento de redes neurais artificiais multicamadas denominado algoritmo de propagação de erros (backpropagation).

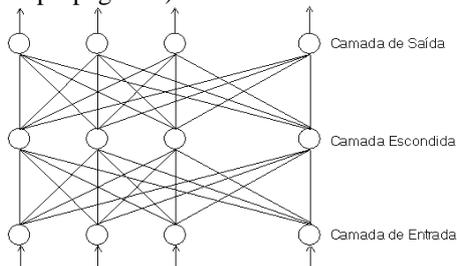


Figura 3 – Estrutura do método Backpropagation

Como mostra a figura 3, a estrutura de uma rede neural artificial multicamadas tem alimentação progressiva para diante ou *feedforward*, não tendo conexões sobre si mesma, seja diretamente ou indiretamente.

d) Treinamento e aprendizado

De acordo com [10] basicamente a aprendizagem por retropropagação de erro consiste em encontrar os pesos que resolvem o problema com *performance* aceitável. Assim, o desempenho da rede se torna rápido e confiável.

e) Banco de dados PostgreSQL;

De acordo com [8] “é um projeto de Sistema Gerenciador de Banco de Dados *open-source* que foi iniciado em 1986, na Universidade de Berkeley, na Califórnia”. Com esse tempo de mercado nos garante robustez, eficiência e rapidez nas consultas.

f) Java Speech;

É uma *Application programming interface* (API); utilizada nos pacotes Java, e de acordo com [9] foi criada com o intuito de prover síntese de reconhecimento de voz nas aplicações feitas em Java.

### B. Trabalhos relacionados

DOSVOX - de acordo com [3] o aplicativo contém vários pacotes de interação com o usuário. Entre eles destacam-se o Sistema de síntese de fala SAPI e o Editor, leitor e impressor/formatador de textos; também não tem incompatibilidade com outros pacotes para deficientes visuais, ou seja, é uma excelente ferramenta para apoio ao deficiente visual. Sua principal diferença do sistema apresentado neste trabalho é o fato de não capturar imagens de textos e convertê-las em sons.

OPENBOOK - também de acordo com o *site* [6] é uma ferramenta de apoio ao deficiente visual, com ótima interação com o usuário, mas não contém arquivo de leitura e voz em português e não é livre; oferece uma versão *demo free*, que mostra apenas algumas ferramentas do aplicativo; depende de um *scanner* para realizar a varredura e captura da imagem e de outra pessoa para manipular o editor de textos nativo para executar a separação dos textos.

NATURAL READER - de acordo com [7] esse software é de fácil utilização e pode converter qualquer texto escrito como MS Word, páginas Web, ficheiros PDF e *e-mails* em palavras faladas; sua deficiência é não ter arquivos de voz em português; também depende da inserção do texto ser realizada pelo usuário.

### C. Arquitetura do Sistema

A Figura 1 mostra os módulos que compõem o sistema de leitura automática de textos impressos.

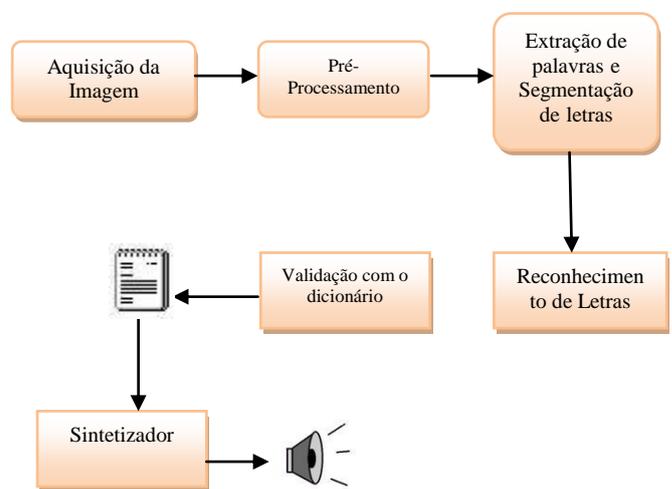


Figura 1 – Módulos do projeto

#### C.1. Módulo de aquisição da imagem.

A aquisição da imagem é feita através de uma *webcam* com resolução mínima de 2 *megapixels*, fixada a uma estrutura articulada mostrada na Figura 2, para realizar a captura da imagem do texto. Foi construído um módulo em Java que realiza o reconhecimento da câmera e, através de um toque com o botão do lado direito do *mouse*, é realizada a captura da imagem e enviada para o módulo seguinte.



Figura 2 – suporte para captura de imagem

### C.1.1. Testes realizados para o módulo de aquisição da imagem.

Para realizar uma varredura da imagem com qualidade superior, foram realizados alguns testes com equipamentos voltados para captura digital de imagens. A figura 3 mostra a imagem capturada através de uma câmera de 2 Mega Pixels (MP) denominada microscópio digital USB 2.0 com iluminação independente e uma câmera digital Mirage Visio de 14 MP. A primeira não conseguiu capturar toda a página, pois sua resolução é de 1600x1200 pixels, focando apenas em 50% do conteúdo pretendido e, assim, não atendeu ao intuito do projeto.

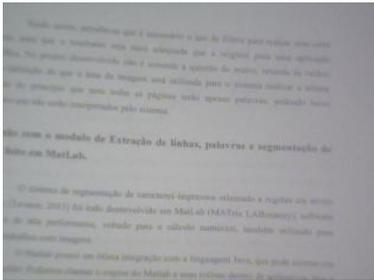


Figura 3 – imagem capturada pelo microscópio digital.

Na figura 4, a imagem capturada é mais nítida e completa, devido à resolução de 2532X1944 pixels utilizada pela câmera. Com a aplicação dos filtros para detecção de bordas e para dar maior nitidez à imagem é realizada a sua preparação para ser analisada pelo módulo de segmentação de letras.

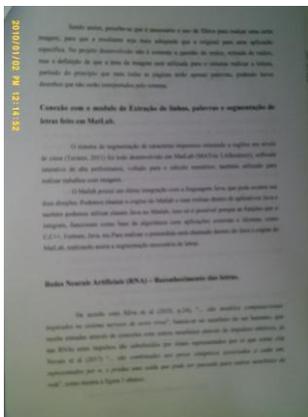


Figura 4 – imagem capturada pela câmera digital.

### C.2. Módulo de Pré-Processamento.

O pré-processamento caracteriza-se pela separação das formas do texto impresso. A princípio, o trabalho será realizado apenas para definição e separação das bordas do texto, definindo os limites da página onde encontram-se os textos, páginas que contenham gravuras ainda não serão trabalhadas.

### C.2. Módulo de Pré-Processamento.

O pré-processamento caracteriza-se pela separação das formas do texto impresso. A princípio, o trabalho será realizado apenas para definição e separação das bordas do

texto, definindo os limites da página onde encontram-se os textos; páginas que contenham gravuras ainda não serão trabalhadas.

### C.3. Módulo de Extração de palavras e Segmentação de letras.

Foi realizada a segmentação de texto e palavras utilizando o método Segmentação de Linhas e Palavras baseado no Operador de Energia de Teager (SLP-TEO) conforme descrito em [2].

### C.4 Módulo de Reconhecimento de letras

Após o recebimento das imagens contendo as letras, é realizado o reconhecimento dos padrões das letras através de uma RNA (Rede Neural Artificial), multicamada com treinamento utilizando a retropropagação do erro.

### C.5. Pós-Processamento, validação com o dicionário de palavras válidas.

Após realização do reconhecimento de padrões, cada letra é gravada em um vetor dinâmico de caracteres, respeitando o tamanho de cada palavra; após definida a palavra é realizada uma comparação com um dicionário de português acomodado em um servidor local de banco de dados PostgreSQL 9.1. Ele foi escolhido por ser confiável de fácil manuseio; no momento em que a palavra não for reconhecida é sugerida uma palavra em substituição à anterior, o que seria algo próximo do que estava definido.

### C.6. Módulo do sintetizador de voz.

Todas as palavras serão gravadas em um arquivo texto, seguindo os parágrafos do texto e após cada parágrafo montado, há acoplado um sintetizador de voz que realizará a leitura do arquivo texto. Para essa função foi definido o Java *Spenck* e o IBM Via Voice 5.0. Este não traz uma voz menos robotizada, mas para o nosso intuito, irá nos atender muito bem.

## D. CONCLUSÕES

A Tecnologia Assistiva é um dos grandes movimentos para inclusão de deficientes na sociedade. O *software* desenvolvido irá auxiliar o deficiente visual, seja ele com deficiência total ou parcial na interação com o meio. Será de potencial ajuda na educação, no dia-a-dia de todas as pessoas que tenham dificuldades em realizar leituras de textos e será mais atrativo que as leituras realizadas pelo método Braille, o qual nem todas as pessoas manipulam com facilidade e rapidez, tendo assim um ganho extremamente alto de qualidade de vida.

As redes neurais artificiais multicamadas estão cada vez mais próximas de nossa realidade e, se ontem parecia ficção, hoje é quase que atual no dia-a-dia de todos nós. A rapidez e a confiabilidade demonstrada pela rede nos mostra que *software* desenvolvidos usando essa tecnologia estarão cada vez mais presentes.

## REFERÊNCIAS

[1] SILVA, Ivan Nunes da, SPATTI, Danilo Hernanes, FLAUZINO, Rogério Andrade, *Redes Neurais Artificiais*, Editora Artiliber, 1ª Edição, São Paulo, 2010.

- [2] TAVARES, Josimeire do Amaral, *Sistema de segmentação de caracteres impressos orientado a regiões em níveis de cinza*, UFU, 2010.
- [3] Acesso Brasil (2009). *DOSVOX - Uma nova realidade educacional para Deficientes Visuais*. Acedido em 02 de Maio de 2012, em: [tecnologia-assistiva.org.br/produtos.php?det=s&id\\_produto=7603](http://tecnologia-assistiva.org.br/produtos.php?det=s&id_produto=7603).
- [4] Portal da Oftalmologia(2005), *Visão subnormal*. Acedido em 30 de abril de 2012, em: <http://www.portaldaretina.com.br/home/saibamais.asp?cod=118>
- [5] Assistiva – Tecnologia e Educação (2012). *O que é Tecnologia Assistiva*. Acedido em 02 de Maio de 2012, em: **Erro! A referência de hiperlink não é válida.**
- [6] Freedom Scientific (2012), *OpenBook Scanning and Reading Software*, Acedido em 02 de Maio de 2012, em: <http://www.freedomscientific.com/products/fs/openbook-product-page.asp>
- [7] Natural Reader 10 (2012). *The most Popular Personal Reader*, acedido em 02 de maio de 2012. em: <http://www.naturalreaders.com>
- [8] Comunidade PostgreSQL Brasil (2011). *O banco de dados de código aberto mais avançado do mundo*, acedido em 02 de maio de 2012, em: <http://www.postgresql.org.br/>
- [9] Wiki Pública do Departamento Acadêmico de Informática – Universidade Tecnológica Federal do Paraná (2010). *A API Java Speech*. Acedido em 02 de maio de 2012, em: <http://www.dainf.ct.utfpr.edu.br/wiki/index.php/2010bEquipe11>.
- [10] TIERRA, Alfonso, FREITAS, Silvio Correia de, *Predição de anomalias gravimétricas de ar-livre e bouguer para dados esparsos*, UFPE, 200?.
- [11] BERSCH, Rita, *Introdução à Tecnologia Assistiva*, CEDI, Porto Alegre, 2008.
- [12] ALBUQUERQUE, Márcio Portes de, ALBUQUERQUE, Marcelo Portes de, *Processamento de imagens: métodos e análises*. CBPF/MCT, Rio de Janeiro. Acedido em 02 de Maio de 2012, em: <http://www.cbpf.br/cat/download/publicacoes/pdf/ProcessamentoImagens.PDF>
- [13] JavaFree, *O que é Java?*, Acedido em: 01 de maio de 2012, em: <http://javafree.uol.com.br/artigo/871498/>.