



T0934

SISTEMA PARA CONVERSÃO ORTOGRÁFICO-FONÉTICA DO PORTUGUÊS BRASILEIRO: IMPLEMENTAÇÕES EMPREGANDO REGRAS LINGÜÍSTICAS E ÁRVORES DE DECISÃO

Tatiana Silva (Bolsista PIBIC/CNPq), Edmilson S. Morais (Co-orientador) e Prof. Dr. Fábio Violaro (Orientador), Faculdade de Engenharia Elétrica e de Computação - FEEC, UNICAMP

A conversão ortográfico-fonética é uma dentre as diversas etapas da síntese de fala. O presente trabalho consistiu no desenvolvimento de dois sistemas para conversão ortográfico-fonética do português brasileiro. O primeiro sistema foi baseado em regras lingüísticas e listas de exceções, enquanto o segundo foi baseado em algoritmos estatísticos e treinado a partir de transcrições fonéticas fornecidas pelo primeiro sistema. A estrutura básica do primeiro sistema havia sido desenvolvida pelo doutorando da FEEC/UNICAMP: Edmilson Morais. As contribuições do projeto de iniciação científica com relação a este sistema foram: documentação, otimização/teste de todas as funções e projeto/construção de uma interface gráfica, escrita em linguagem Python. Esta interface permite a inserção, a remoção ou a alteração das regras de conversão ortográfico-fonética e dos alfabetos fonéticos, sem a necessidade de conhecimentos sobre linguagens de programação. O segundo sistema foi desenvolvido pela aluna de iniciação científica. Ao final, analisaram-se os dois sistemas e a abordagem estatística se mostrou superior conforme o esperado.

Conversão ortográfico-fonética - Síntese de fala - Árvores de decisão