

Processador de Thesaurus

José Carlos Lima Martins A78821 and Miguel Miranda Quaresma A77049

University of Minho, Department of Informatics, Braga, Portugal
e-mail: {a78821,a77049}@alunos.uminho.pt

Resumo Este documento serve de apoio ao desenvolvimento de um processador Thesaurus explicando todas as decisões tomadas de modo a obter o mesmo. Inicialmente será explicado a estrutura dos ficheiros de entrada (extensão .mdic) bem como será mostrado as decisões tomadas de modo a chegar ao produto final. Por fim, será apresentado exemplos de utilização do mesmo.

1 Introdução

O presente projeto (Processador de Thesaurus) visa ser um sistema de processamento de dicionários Thesaurus que recorre a expressões regulares (ERs) para filtrar e transformar os mesmos, extraindo e tratando a informação mais relevante de forma eficiente. Para atingir este objetivo é usada a linguagem de filtragem e tratamento de dados **GAWK** visto ser uma **DSL**(Domain Specific Language) com foco em dados semi-estruturados.

2 Desenvolvimento

3 Conclusão