

Paradigmas da Programação

MiECom (2º ano)

Trabalho Prático nº 4

Ano lectivo 2009/2010

1 Objectivos e Organização

Este trabalho prático tem como principais **objectivos**:

- praticar o uso de expressões regulares para especificar filtros de texto, recorrendo a autómatos finitos deterministas para fazer o reconhecimento das frases que seguem os padrões;
- usar Geradores de Programas baseados em especificações – no caso concreto incentivando o uso do gerador FLex baseado em sistemas de regras ”*padrão(ER)-acção*”
- aumentar a experiência de programação na linguagem lógica Prolog para resolução geral de problemas, aplicada à implementação de sistemas periciais com explicação;

Para o efeito, esta folha contém 1 único enunciado que consiste num classificador automático de textos jornalísticos. Cada grupo deve escolher duas classes/categorias diferentes para classificar os seus textos, dando assim lugar a 9 implementações diferentes da mesma proposta geral. O programa desenvolvido—**a realizar até ao fim do semestre, dia 13 de Junho**—deve ser enviado por email aos docentes por cada grupo, mal esteja concluído, acompanhado de um relatório de desenvolvimento.

O **relatório** a elaborar deve ser claro e, além do respectivo enunciado, da descrição do problema, e das escolhas/decisões que determinaram a sua implementação final, deverá conter exemplos de utilização e o código do programa. Como é de tradição, o relatório será escrito em L^AT_EX.

2 Classificador de Notícias de Jornais

Neste trabalho, pretende-se que desenvolva em C e Prolog um programa para classificar artigos de jornal de forma simples, isto é, dizendo se um dado texto pertence a uma de duas classes pré-definidas (à escolha de cada grupo). O artigo analisado deverá ser associado à classe que tem maior probabilidade, se for acima de 75% pelo menos em uma delas; ou indicado como não pertencente a nenhuma delas, se ficar abaixo dos 75% em ambas.

Para efeito deste trabalho, considere que uma notícia tem os seguintes blocos (marca-se com ’?’ os campos facultativos:

Titulo, Subtítulo?, Destaque?, Corpo, Roda-pé?, Fotos*

O classificador em Prolog deve implementar um conjunto de regras de classificação (a definir por cada grupo) baseadas nas medidas de frequência de ocorrência das palavras discriminantes de cada classe; além disso, esse programa deve explicar o raciocínio que seguiu para a sua decisão final

A frequência de ocorrência dos discriminantes das classes, dentro de cada bloco do documento em análise, (a dita notícia de jornal) deve ser contada por um *filtro de texto* em C construído automaticamente com recurso ao *Gerador de Analisadores Léxicos a partir de Expressões Regulares Flex*.