

# Processamento de Linguagens I

## LESI + LMCC (3º ano)

1º Ficha Prática

Ano Lectivo de 01/02

### 1 Objectivos

Esta ficha prática contém exercícios para serem resolvidos nas aulas teórico-práticas com vista a sedimentar os conhecimentos relativos a:

- uso de Expressões Regulares para definir (gerar) Linguagens Regulares;
- uso de Expressões Regulares para desenvolver programas eficientes, baseados em algoritmos standard guiados por Autómatos Finitos Deterministas, para reconhecer Linguagens Regulares;
- uso de Autómatos Deterministas Reactivos, para processar Linguagens Regulares, isto é para desencadear Acções específicas ao reconhecer frases que derivam de Padrões (definidos com base em ERs) —princípio da Programação baseada em regras *Condição-Reacção*;
- geração automática de programas a partir de especificações formais;
- uso da ferramenta Lex, disponível em ambiente Linux, para geração automática de processadores de Linguagens, nomeadamente criação de *Filtros de Texto* e de *Analísadores Léxicos*, definidas por ERs.

### 2 Enunciados

#### Filtro de Questionários

Suponha que ao fim de cada entrevista um Repórter produz um texto com as perguntas e respostas, distinguindo umas das outras porque as perguntas começam com "EU:" no início da linha e as respostas começam com "ELE:", também no início da linha.

Nesse contexto, pretende-se desenvolver um filtro de texto para processar os questionários que:

- a) simplesmente retire do texto original as tais marcas "EU:" e "ELE:", devolvendo todo o resto da entrevista sem qualquer alteração.  
Melhore o filtro, de modo a tratar as marcas, quer estejam escritas em maiúsculas, quer em minúsculas;
- b) substitua a marca "EU" pelo nome do entrevistador e retire do texto todas as respostas, incluindo a marca "ELE:"

## Expansor de Abreviaturas

Quando se retiram apontamentos, ou de uma forma geral, se tem de escrever muito depressa é hábito usar abreviaturas que correspondam a uma ou mais palavras vulgares e longas. Suponha que criou esse costume e resolveu inserir nos seus textos as ditas abreviaturas (2 ou mais letras) precedidas pelo carácter "\-por exemplo "\qq"(qualquer), ou "\mb"(muito bom), ou ainda "\sse"(se e só se).

Desenvolva, então, um processador que lhe devolva o texto original mas com todas as abreviaturas (definidas à partida) devidamente expandidas.

Melhore o seu filtro de moda a contemplar ainda o tratamento do carácter "/"no fim de uma palavra, representando o sufixo "mente", e o carácter "~"no início de uma palavra, representando o prefixo "in".

## Filtro para XML

Como sabe um Documento XML é u texto vulgar semeado de anotações, ou marcas, que são identificadores especiais intercalados entre os caracteres "<"e ">".

Desenvolva então um filtro que receba um texto XML e o devolva após ter retirado todas as marcas.

Evolua o seu filtro de modo a:

- a) contar o número de marcas de abertura e o número de marcas de fecho (distinguem-se porque as marcas de fecho começam por "</"), indicando Erro sempre que se verifique um desequilíbrio entre ambas;
- b) verificar a concordância entre as marcas de abertura e as marcas de fecho, isto é, garantir que as marcas se fecham por ordem inversa que se abrem.