



UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS E TECNOLOGIA

Departamento de Engenharia Informática

Trabalho nº 3 – “Language Annotator” Algoritmos e Estruturas de Dados

2015/2016 – 2º Semestre

Meta A: 14/Março/2016, 23h00

Meta B: 28/Março/2016, 23h00

**Meta C: não submetido no mooshak
listagem anexa ao relatório,**

Relatório: 11/Abril/2016, 17h00

Objectivos: Árvores de Pesquisa e outras estruturas semelhantes.

Notas Importantes:

- O relatório do trabalho devem ser realizados, MANUSCRITOS, nos formulários que serão oportunamente fornecidos.
- Só os alunos que entreguem o relatório do trabalho dentro do prazo-limite estabelecido têm acesso à avaliação neste trabalho.

Descrição do Problema

A empresa DigiWorld Lda pretende melhorar a sua aplicação “Language Annotator” para telemóveis. Esta aplicação é usada por estudantes de inglês para guardar novas palavras e sua tradução para português. As palavras guardadas podem ser consultadas de diversas formas conforme for mais útil ao processo de aprendizagem.

A aplicação compreende um dicionário off-line em que cada entrada contém:

- Palavra de entrada no dicionário (cadeia de carateres, max. 40 carateres)
- Tradução da palavra (cadeia de carateres, max. 100 carateres)
- Número de consultas (inteiro)
- Marcador (booleano, a *verdadeiro* se a palavra está marcada)

Nota: todas as palavras e respetiva tradução são inseridas em **minúsculas** e é **retirada toda a acentuação**¹

Considerar que o número de entradas no dicionário off-line é na ordem das 200 000 e que o número de palavras acrescentadas ao dicionário pelo utilizador é de ordem bastante inferior bem como o número de palavras distintas consultadas no dicionário.

Tomar as decisões sobre estruturas de dados que maximizem a eficiência temporal do programa, mesmo que a custo de espaço em memória.

¹ A razão destas limitações prende-se com a submissão no Mooshak e com a forma como este lida com ficheiros com palavras acentuadas originados em diferentes sistemas operativos, bem como com as possíveis ambiguidades na utilização de maiúsculas nas palavras.

As operações (comandos) possíveis sobre o “Language Annotator” são:

1. CARREGA

Carrega o dicionário inicial linha a linha no formato “<palavra> < tradução>” até encontrar a linha com o termo “fim\$dicionario”. Depois de carregar o dicionário devolve “DICCIONARIO CARREGADO”

2. ACRESCENTA <palavra> <tradução>

Acrescenta uma nova entrada no dicionário “<palavra> < tradução>”

Se palavra já existe devolve “PALAVRA JA EXISTENTE” e não faz mais nada. Se a palavra e respetiva tradução foi inserida devolve “PALAVRA ACRESCENTADA”

3. PESQUISA <palavra>

Pesquisa por palavra

Se palavra não existe no dicionário devolve “PALAVRA NAO EXISTENTE” e não faz mais nada. Se existe devolve “<palavra> < tradução>”.

4. MARCA <palavra>

Procura palavra e marca-a

Se palavra não existe no dicionário devolve “PALAVRA NAO EXISTENTE” e não faz mais nada. Se existe devolve “<palavra> MARCADA”

5. LISTA_ALFANUM

Lista por ordem alfabética todas as palavras no dicionário

Mostra por linha “palavra”. Termina com uma linha com “FIM LISTA”

6. LISTA_MARCADAS

Lista por ordem alfabética as palavras marcadas

mostra por linha “palavra”. Apresenta as palavras por ordem alfanumérica. Termina com uma linha com “FIM MARCADAS”

7. LISTA_MAIS_ACESSOS

Lista as palavras por ordem decrescente de número de consultas + número de marcações (para palavras com o mesmo número de consultas+marcações apresenta por ordem alfabética), só lista palavras com 1 ou mais consultas+marcações (implementar, não integrado no Mooshak)

mostra por linha uma palavra, termina com uma linha com “FIM LISTA ACESSOS”

Exemplos de *input* e *output*

O input corresponde aos comandos já indicados (um por linha). As operações deverão terminar quando for lida uma linha vazia.

O output será constituído pelas respostas da aplicação aos comandos tal como descrito acima.

Para os exemplos abaixo considerar que o dicionário tem inicialmente as seguintes entradas

a um about sobre after depois again de novo all todo almost quase also tambem always sempre and e any qualquer beautiful bonito because porque before antes big grande body corpo but mas by por city cidade day dia down para baixo every todo everybody todos everyday todo dia everything tudo everywhere em todos os lugares far longe first primeiro for por forever para sempre friend amigo from de good bom happy feliz	he ele her ela here aqui him ele his dele how como i eu if se in em last ultimo many muitos me mim more mais most o mais much muito my meu near perto never nunca new novo no nao now agora of de on no only apenas or ou other outro people povo please por favor price preco same o mesmo she ela so tao something algo	sometimes as vezes our nosso still ainda than do que that que the o their deles them eles there la they eles thing coisa this este time tempo to para today hoje under sob us nos usually geralmente very muito way jeito we nos what o que when quando where onde which o qual who que why por que with com without sem word palavra yes sim you voce your seu
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Exemplo de *Input*:

```
CARREGA[\n] <insere o dicionario acima seguido de linha com "fim$dicionario">
PESQUISA city[\n]
PESQUISA acknowledge[\n]
ACRESCENTA acknowledge reconhecer[\n]
PESQUISA acknowledge[\n]
PESQUISA also[\n]
MARCA also[\n]
MARCA about[\n]
PESQUISA acknowledge[\n]
PESQUISA city[\n]
PESQUISA acknowledge[\n]
PESQUISA a[\n]
LISTA_MARCADAS [\n]
LISTA_MAIS_ACESSOS [\n]
[\n]
```

Exemplo de Output:

```
DICIONARIO CARREGADO[\n]
city cidade[\n]
PALAVRA NAO EXISTENTE[\n]
PALAVRA ACRESCENTADA[\n]
acknowledge reconhecer[\n]
also tambem[\n]
also MARCADA[\n]
about MARCADA[\n]
acknowledge reconhecer[\n]
city cidade[\n]
acknowledge reconhecer[\n]
a um[\n]
about[\n]
also[\n]
FIM MARCADAS[\n]
acknowledge[\n]
also[\n]
city[\n]
a[\n]
about[\n]
FIM LISTA ACESSOS [\n]
```

Nota: as linhas no dicionário podem aparecer no final com um espaço em branco que deve ser eliminado nas saídas da operação PESQUISA para ser aceite no mooshak.

Meta A 14/Março/2016, 23h00 código anexo ao relatório

Desenvolver as operações 1 a 6 usando Árvores Binárias de Pesquisa.

Meta B 28/Março/2016, 23h00 código anexo ao relatório

Desenvolver as operações 1 a 6 usando Árvores Binárias de Pesquisa Equilibrada. Ter como objetivo a eficiência temporal.

Meta C código anexo ao relatório

Desenvolver a operação 7 incluindo eventualmente novas estruturas de dados auxiliares que otimizem temporalmente esta operação.

Relatório Final – 11/Abril/2016, 17h00 (no cacifo Prof. Carlos Bento)

Deverá entregar o relatório final a 11/Abril/2016 até às 17:00 no cacifo do Docente da Aula Teórica. A forma de entrega e o formato padrão do relatório, que irá incluir medições de desempenho e análise de complexidade, será oportunamente divulgado.