



Universidade Do Minho

Mestrado Integrado Em Engenharia Biomédica

Bibliotecas Digitais

MATLAB

Exercício 2

Braga, Março 2013



Universidade Do Minho

Mestrado Integrado Em Engenharia Biomédica

Bibliotecas Digitais

MATLAB

Exercício 2

Docente: Joaquim Macedo

Discentes: Hugo Gomes (58536)
Marta Moreira (54459)
Telma Veloso (58521)

1 Contagem de ocorrência da palavra

De forma a contar o número de ocorrências da palavra **text** — enquanto *string* e não apenas como sequência de caracteres —, começou-se por declarar um contador (**counter**), inicializado a zero. De seguida, abriu-se o ficheiro **.txt** através do comando **fopen**, operação cuja execução retorna um inteiro (**fid**), utilizado nas rotinas seguintes como identificador do ficheiro. Por intermédio de um ciclo *while*, implementou-se um conjunto de acções a executar até que se atingisse o final do ficheiro — altura em que **feof(fid) == 1** —, iniciado com a leitura do ficheiro (linha a linha). Numa primeira instância, utilizou-se o comando **fgetl** para o efeito, o que resultou na emissão da seguinte mensagem de erro na consola:

```
Error using textscan
First input can not be empty.
```

Assim, concluiu-se que a função mais adequada, neste caso específico, seria a **fgets**, uma vez que, contrariamente à anterior, esta não remove os caracteres de nova linha, colmatando a questão de *inputs* vazios explicitada na mensagem de erro. Posteriormente, tornou-se necessário proceder à leitura dos dados de texto para um *array*, utilizando **%s** como especificador de conversão e assegurando, assim, a segregação de cada uma das *substrings* em células distintas. De forma a assegurar que as palavras associadas a caracteres especiais eram igualmente contabilizadas, construiu-se uma função com o intuito de eliminá-los. Devido ao facto de o código ter sido testado em versões do **MATLAB** e sistemas operativos distintos, este ponto divergiu obrigatoriamente entre duas construções distintas (com uma ligeira variação na parte inicial do código), uma vez que não foi possível validar ambas em todos os sistemas, simultaneamente. Assim, considerem-se:

```
1. fline = textscan(fid,'%','delimiter',{' ','.',',',';','"',',','/','\b','\t','\t'},'MultipleDelimsAsOne',1);

2. fline(~isstrprop(fline,'alphanum')) = ' ';
```

O passo seguinte passou por veicular a comparação, utilizando o comando **strcmpi**, das *strings* contidas no *array* com a *string* cuja ocorrência se pretende contabilizar, sendo que se especificou a variante do comando que permite ignorar a *case*, no sentido de possibilitar a identificação das *strings* não obstante o seu tipo de escrita. Os valores lógicos de ocorrência foram somados através da função **sum** e incrementados de forma iterativa ao contador, de forma a obter a contagem pretendida, quando é atingido o final do ficheiro. Por último, computou-se o fecho do ficheiro, através do comando **fclose**, com a respectiva identificação (**fid**) passada como atributo.

A primeira parte do problema proposto foi então resolvida recorrendo às sequências de código abaixo indicadas, ambas retornando o número esperado de ocorrências da palavra pretendida — 43.

1. MATLAB R2012a (Mac OS X)

```
counter = 0;

fid = fopen('BD2.txt' , 'rt');

while feof(fid) == 0
```

```

    fline = fgets(fid);

    fline(~isstrprop(fline,'alphanum')) = ' ';

    splitstring = textscan(fline, '%s');

    splitstring = splitstring{1};

    matches = sum(strcmpi('text', splitstring));

    if matches > 0

        counter = counter + matches;

    end;

end;

fclose(fid);

counter

counter =

    43

```

2. MATLAB R2012b (Windows)

```

counter = 0;

fid = fopen('BD2.txt','rt');

while feof(fid) == 0

    scanned = textscan(fid,'%s','delimiter',{ ' ','.',',',';','"', '/', '\b', '\t', '\t' }, 'MultipleDelimsAsOne',1);

    scanned = scanned{1};

    matches = sum(strcmpi('text',scanned));

    if matches > 0

```

```
        counter = counter + matches;
    end;

end;

fclose(fid);

counter

counter =

    43
```

2 Contagem de ocorrência do padrão

A segunda parte do enunciado pressupunha a identificação de todas as ocorrências da sequência de caracteres `'text'` o que, contrariamente ao caso anterior, inclui também *strings* que contenham a referida sequência como *substring*. Atendendo a esta especificidade, começou-se igualmente por declarar um contador e abrir o ficheiro, iniciando-se um ciclo a executar até que surgisse o final do mesmo. Porém, a leitura foi efectuada recorrendo ao comando `fgetl`, contando que não se mostrou relevante eliminar os caracteres de nova linha, e o método utilizado para encontrar as ocorrências da sequência socorreu-se do comando `strfind` — que permite a procura com base num padrão, neste caso `'text'` —, aplicado a um *array* modificado de forma a que o seu conteúdo não apresentasse variações de *case*. A incrementação do contador foi feita com base no comprimento dos *arrays* assim obtidos, sendo que, no final, a contagem ascendeu às esperadas 45 ocorrências (`43 + textfile + texting`). Apresenta-se, de seguida, o trecho de código que levou à resolução da segunda parte do problema proposto.

```
counter = 0;

fid = fopen('BD2.txt', 'rt');

while feof(fid) == 0

    tline = fgetl(fid);

    matches = strfind(lower(tline), 'text');

    num = length(matches);

    if num > 0

        counter = counter + num;

    end;

end;

fclose(fid);

counter

counter =

    45
```