

**Relatório**

**Algoritmos e Estruturas de Dados II**

**Alunos: Luís Lamela & Artur Mendes**

**Número: 14822 & 14870**

**Professor: Mário Vale**

**Licenciatura em Engenharia de Sistemas Informáticos**

Barcelos, 3 de Abril, 2020

Índice de Figuras

[Figura 1 7](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539552)

[Figura 2 8](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539553)

[Figura 3 8](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539554)

[Figura 4 9](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539555)

[Figura 5 10](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539556)

[Figura 6 10](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539557)

[Figura 7 11](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539558)

[Figura 8 11](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539559)

[Figura 9 12](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539560)

[Figura 10 12](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539561)

[Figura 11 12](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539562)

[Figura 12 13](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539563)

[Figura 13 14](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539564)

[Figura 14 15](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539565)

[Figura 15 16](file:///C:\Users\filas\Desktop\Trabalho%20Artur%20Github\Estrutura%20Relatório%20AED%20-%208Abril.odt#_Toc37539566)

Índice

[1. Introdução 4](#_Toc37539567)

[1.1. Contextualização 4](#_Toc37539568)

[1.2. Motivação e Objetivos 6](#_Toc37539569)

[1.3. Interpretação e Desenvolvimento da Aplicação 7](#_Toc37539570)

[1. Conclusão 17](#_Toc37539571)

[2. Bibliografia 18](#_Toc37539572)

# Introdução

## Contextualização

O presente trabalho foi desenvolvido no âmbito da UC de Algoritmos e Estruturas de Dados II.

O Processamento de Linguagem Natural, uma sub-área da Inteligência Artificial, dedica-se ao estudo de métodos computacionais para o estudo das línguas humanas, ditas naturais. Um dos processos típicos na preparação de um texto para o seu processamento é a Etiquetação Morfossintática em que, a cada palavra, se associa uma categoria morfológica (ex. verbo, substantivo, adjetivo) e também de informação morfológica adicional (ex. número ou género nos substantivos, ou tempo e pessoa, nos verbos).

Neste trabalho prático pretende-se criar uma ferramenta que seja capaz de analisar um qualquer ficheiro já etiquetado com informação morfossintática.

Cada ficheiro é composto por várias linhas, uma para cada palavra ou sinal de pontuação. Cada linha é composta por várias colunas, sendo que a primeira corresponde à palavra da frase original, a segunda define a **raíz da palavra** e a terceira à sua **análise morfossintática**.

A última coluna inclui a certeza da ferramenta em relação à análise realizada.

**As colunas estão separadas por um caracter de espaço.**

**Propostas para o Desenvolvimento**

**1)** Inicialmente, ler todas as linhas e guardá-las numa ou mais estruturas de dados dinâmicas. As linhas em branco, e as linhas correspondentes a sinais ortográficos podem ser ignorados (ou seja, só deverão ser consideradas as linhas referentes a palavras). A estrutura de dados a usar deverá ser escolhida de modo a ser possível responder às seguintes questões.

**2)** Construa uma tabela de frequências absolutas, relativas e acumuladas, referente à categoria gramatical usada (terceira coluna). Apresente a tabela ordenada por ordem crescente de frequência absoluta.

**3)** Construa uma tabela de frequências absolutas, relativas e acumuladas, referente ao tamanho das palavras existentes no texto.

**4)** Para cada tipo de análise morfológica (terceira coluna), calcule com base na medida de certeza de etiquetação (quarta coluna) a sua média aritmética e desvio padrão

**5)** Calcule as seguintes medidas de localização e dispersão, relativas ao tamanho das palavras: média aritmética, mediana, moda e desvio padrão.

**6)** Com base nas frequências das palavras (o seu número de ocorrências), calcule os quartis e implemente uma função que, dada uma palavra indicada pelo utilizador, permita obter, o quartil a que pertence.

**7)** Obtenha os dados necessários para a construção de um histograma das probabilidades (quarta-coluna). Defina o número de classes e as respetivas frequências a considerar

## Motivação e Objetivos

Este trabalho tem como objetivo salientar os conhecimentos adquiridos durante as aulas. Pretendem promover a autoaprendizagem através da investigação individual, e da aplicação das matérias lecionadas no decurso do semestre na unidade curricular de Algoritmos e Estruturas de Dados II.

## Interpretação e Desenvolvimento da Aplicação

Face ao Exercício 1 do Trabalho Prático, foram desenvolvidas algumas estruturas de dados para suportarem os dados da melhor maneira na aplicação:

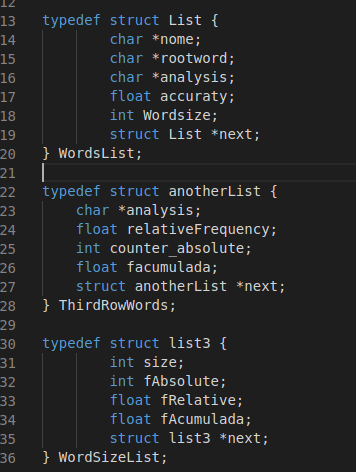


Figura 1

A Lista ligada chamada WordsList irá suportar as palavras que estão no documento de texto a analisar, e guarda as palavras contidas no ficheiro, e o tamanho da palavra original (para simplificar cálculos noutros exercicios).

Quando o utilizador abre o programa pode escrever o nome do ficheiro que pretende analisar (o programa assume que está no mesmo diretório que o executável), e posteriormente é feito o acesso ao ficheiro:

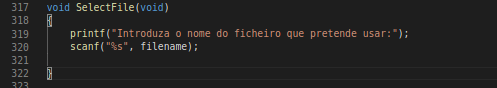


Figura 2

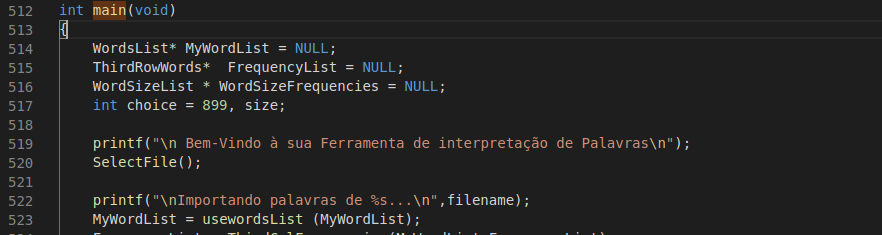


Figura 3

Após guardar o nome do ficheiro numa variável global (filename), é feito o acesso ao ficheiro, onde as palavras são respetivamente adicionadas à Lista Ligada de Palavras, como mostra a imagem abaixo.

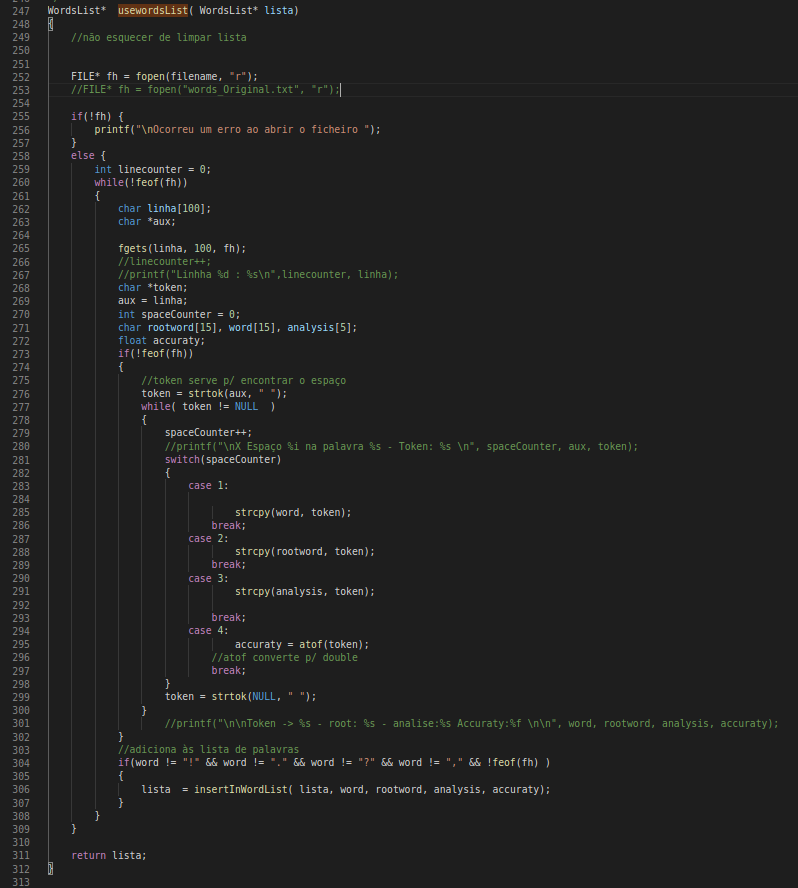


Figura 4

Para o Exercício 2 foi desenvolvida uma função que acede à lista de palavras e guarda a categoria gramatical numa nova lista e calcula as frequências da mesma.

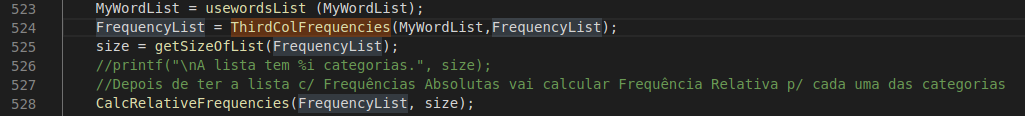


Figura 5

A função ThirdColFrequencies é responsável por guardar as palavras na FrequencyList, enquanto que a função CalcRelativeFrequencies é responsável por calcular as frequências absolutas, relativas e acumuladas das categorias.

A função GetSizeofList calcula o nº de palavras existentes na Lista.

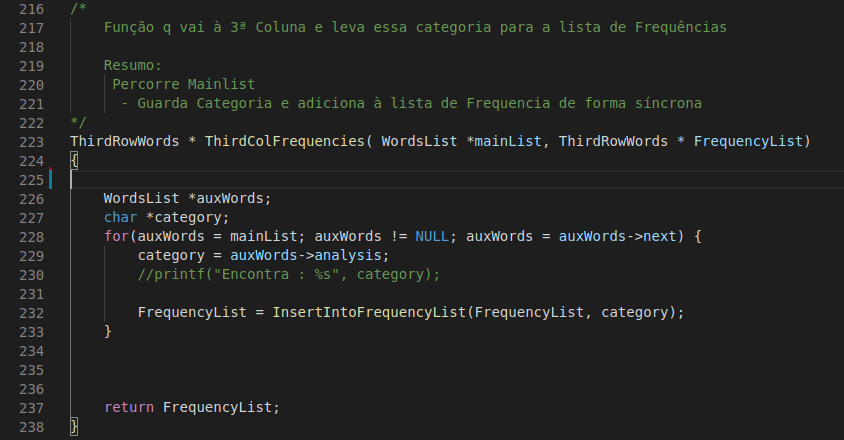


Figura 6

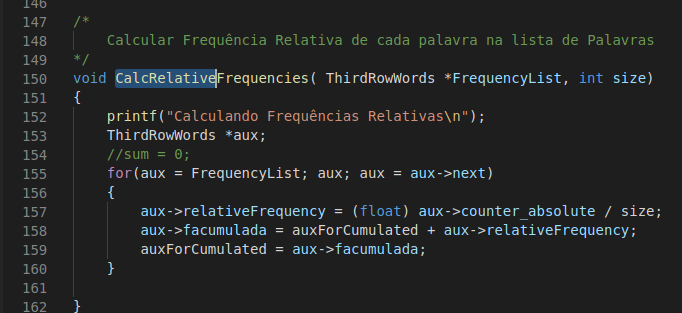


Figura 7

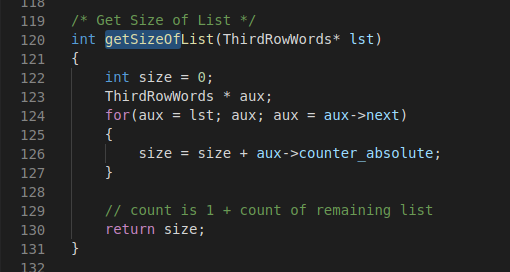


Figura 8

A proposta do Exercício 3 tem uma resolução muito semelhante à do Exercício 2, alterando a lista que é trabalhada.

Criou-se uma lista WordSizeList para poder trabalhar sem conflitos com os dados.

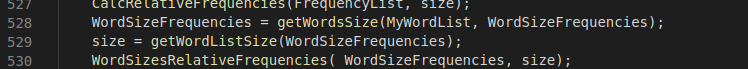
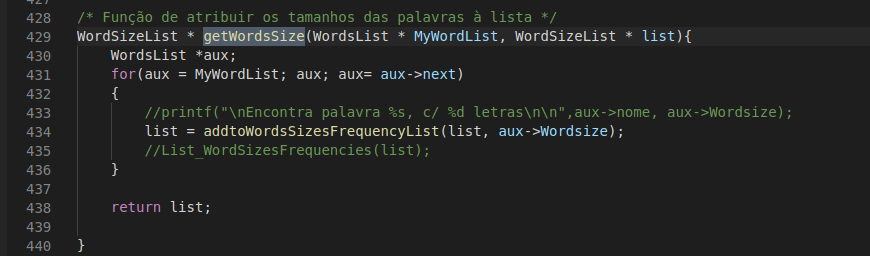


Figura 9

Figura 10

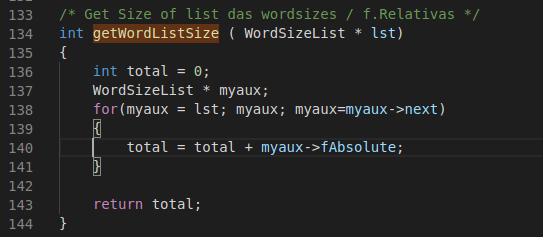


Figura 11

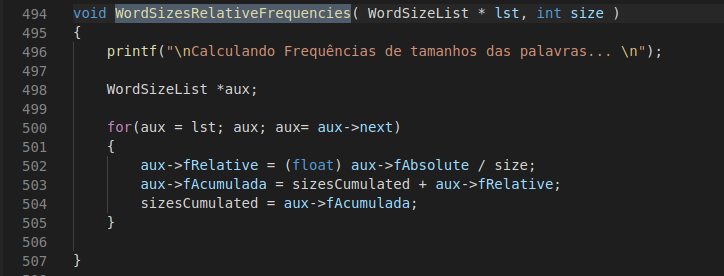


Figura 12

Quando chegamos à resolução do exercício 3 começamos a explorar formas de desenvolver o *main*, então demos ao utilizador um menu com várias opções.

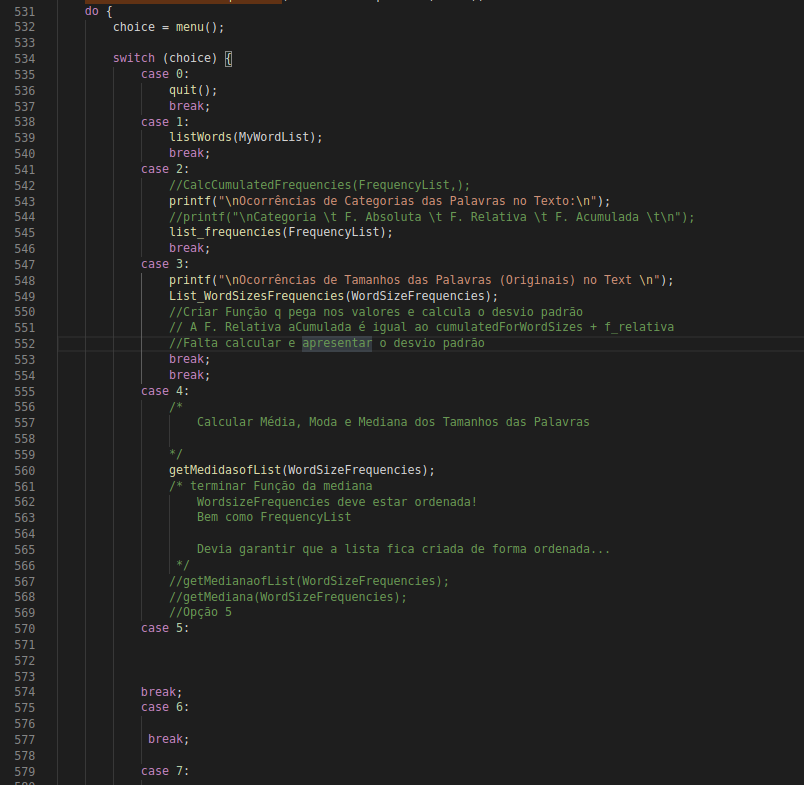


Figura 13

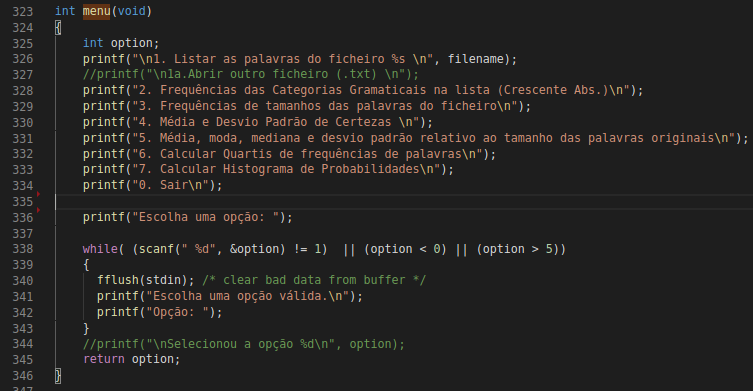
****

Figura 14

Quando o utilizador seleciona a opção 5 são calculadas as medidas relativas aos tamanhos das palavras:



Figura 15

Para o a proposta de desenvolvimento 6, será desenvolvida uma função que indique os quartis da lista de palavras Originais, e que recolha uma palavra escrita pelo utilizador e indique a que quartil pertence.

Para o Exercício 7 é necessário definir o número de classes e frequências a considerar

# Conclusão

O trabalho prático ajudou-nos a desenvolver capacidades de organização que antes não tínhamos. Tivemos que usar o GitHub para trabalharmos remotamente, e tivemos várias dificuldades ao interpretar inicialmente o enunciado, mas com tempo conseguimos desmistificar os objetivos.

Desenvolvemos as nossas capacidades com a linguagem C, e, apesar de não estarmos a trabalhar presencialmente, conseguimos resolver o enunciado e chegar a um resultado final.

Esperamos no próximo trabalho ter tempo e capacidade para chegar ao fim do enunciado.

# Bibliografia

**Não existem origens no documento atual.**