

**Relatório do Projeto**

**Licenciatura em Engenharia Informática**

**Sistemas Operativos**

** Carlos Filipe de Moura Braz (2203845)**

** Inês Alexandra Ribeiro Machado (2200723)**

**Carlos Filipe de Moura Braz (2203845) e Inês Alexandra Ribeiro Machado (2200723) declaram sob compromisso de honra que o presente trabalho (código, relatórios e afins) foi integralmente realizado por nós, sendo que as contribuições externas se encontram claramente e inequivocamente identificadas no próprio código. Mais se declara que os estudantes acima identificados não disponibilizaram o código ou partes dele a terceiros.**

**Leiria, abril de 2021**

1. **Descrição e estado das estruturas**

|  |  |
| --- | --- |
| **Funcionalidade** | **Estado** |
| Modo C  Modo P  Modo T | Totalmente operacional  Totalmente operacional  Totalmente operacional |

1. **Apresentação e justificação das opções de implementação tomadas**

O relatório deve ainda explicar a metodologia de contagem, indicando quais as opções tomadas (se números foram consideradas palavras, qual a abordagem em relação à pontuação, etc.).

Cat – para aceder ao ficheiro e guardar o conteúdo numa variável = CONTENT

Pdftotext – para ler o conteúdo de ficheiro pdf para uma string = CONTENT

Gnuplot – para gerar o gráfico conferir se o utilitário existe na máquina do utilizador

Flag nopgbrk: exclui da string a informação de quebra de página que não é relevante para a contagem de palavras

Verificação de requisitos:

* Conferir se a aplicação pdftotext existe na máquina do utilizador para correr o programa com ficheiros pdf
* Limitação dos idiomas suportados pelo script para apenas inglês ou português

Metodologia de contagem:

Através do comando “tr” foram excluídos números com a opção [:digit:] e a pontuação com a opção [:punct:]

Palavras compostas foram separadas e consideradas palavras independentes, por exemplo, sistemas-digitais, contou-se como (1) sistemas e (1) digitais.

Optou-se por converter todos os caracteres para lowercase para uniformizar a contagem das palavras

A contagem das palavras realizou-se por meio de um pipe em que através da flag sort se ordenaram as palavras alfabeticamente, seguido da flag uniq -c para contar as ocorrências de palavras únicas e finalmente um sort invertido, flag reverse -r, para gerar uma lista ordenada decrescente do número de ocorrências, flag -n.

Regex para validar o valor das word-stats-top, o valor “05”, por exemplo não foi considerado válido, seguindo a mesma lógica, zero também não é aceite.