

INF01124 - Classificação e Pesquisa de Dados - Exercício 3

Professor João Comba

1 Radix Sort

Implemente o Radix Sort como visto em aula e ordene os arquivos de texto *frankenstein_clean.txt* e *war_and_peace_clean.txt* disponíveis no moodle da disciplina e no site do Projeto Gutenberg. Para ambos os arquivos, considere apenas as **palavras contendo pelo menos 4 caracteres e apenas letras do alfabeto inglês, isto é, ignore numerais e símbolos**. Após essa filtragem inicial dos arquivos, o programa deve gerar como saída dois arquivos texto (*frankenstein_ordenado.txt* e *war_and_peace_ordenado.txt*) listando em ordem lexicográfica as palavras e o número de ocorrências de cada palavra por linha.

A saída deve seguir o formato:

```
palavra1 #ocorrencias_1
palavra2 #ocorrencias_2
palavra3 #ocorrencias_3
...
```

Por exemplo, para War and Peace seria assim:

```
ABACK 3
ABACUS 1
ABANDON 25
ABANDONED 54
ABANDONING 26
ABANDONMENT 14
ABANDONS 1
ABASEMENT 1
ABASHED 13
ABATE 2
ABBREVIATIONS 1
ABDICATE 1
ABDOMEN 2
ABDOMENS 2
ABDUCTION 3
ABDUCTORS 1
ABHORRENCE 1
ABIDE 1
ABILITY 4
ABLAZE 2
ABLE 107
ABNEGATION 2
ABNORMAL 1
ABNORMALLY 1
ABODES 1
ABOLISHING 1
ABOLITION 1
ABOMINABLE 7
ABOMINABLY 1
ABOUNDING 1
ABOUT 1019
ABOVE 146
ABOVEBOARD 1
...
```

2 Entrega

A solução deve ser enviada pelo Moodle dentro de um arquivo .zip, contendo os seguintes arquivos:

- **integrantes.txt**: coloque o nome dos integrantes do grupo (até 2 pessoas) , com um nome por linha
- *frankenstein_ordenado.txt*: lista ordenada de palavras e suas ocorrências, no exato formato listado acima
- *war_and_peace_ordenado.txt*: lista ordenada de palavras e suas ocorrências, no exato formato listado acima
- código fonte correspondente a solução