

Trabalho no 4 v2 – Palavras mais procuradas Algoritmos e Estruturas de Dados 2019/2020 – 2º Semestre

Upload: (link a disponibilizar no infoestudante)

Data Limite: até 15mn depois do fim da respetiva 4ª aula pratica

O RELATÓRIO E LISTAGEM DO CÓDIGO DESENVOLVIDO DEVEM SER SUBMETIDOS NUM ÚNICO DOCUMENTO PDF

Nome:	Dário Filip	e Torres Félix	nº 2018275530	PL: PL2
-------	-------------	----------------	---------------	---------

№ de horas de trabalho: Aulas Práticas de Laboratório: <u>6</u> H Fora de Sala de Aula: <u>5</u> H

CLASSIFICAÇÃO:

(A Preencher pelo Docente)

Análise Empírica de Complexidade

- Faça o download dos 4 textos. Preencha a Tabela I com os valores para cada texto. <u>Use o tempo médio de 20 execuções</u>. Considere essa caracterização na análise qualitativa mais adiante.
- Preencha a Tabela 2 com informação sobre as estruturas de dados e algoritmos que escolheu.
- Preencha a Tabela 3 com os tempos de ordenamento para cada algoritmo.

#	Operação	Texto A	Texto B	Texto C	Texto D
1	Carregamento (tempo em: <u>Milissegundo [ms]</u>)	0,387	2 431,561	3 038,701	2 077,105
2	Núm. palavras distintas	75	1354	1354	1354
3	Núm. utilizadores distintos	2	632	632	632
4	Núm. Pares (palavra,utilizador) distintos	150	4 159	4 159	4 159
5	Núm. Total de palavras (Entradas)	156	140 000	140 000	140 000
6	Núm. Total de utilizadores (Entradas)	156	140 000	140 000	140 000
		Poucas palavras.	Muitas palavras	Muitas palavras,	Palavras e IDs
7	Faz algum ordenamento? Caracterize.	Palavras e IDs	Palavras e IDs	repetidas e	repetidas e
		ordenados.	aleatórios.	ordenadas.	ordenadas.

#	Tarefa	Estrutura de dados	Algoritmo de ordenamento
8	A1	Arrays	Quicksort (e Insertion Sort)
9	A2	Arrays	Radix Sort

#	Tempos de Ordenamento em (Milissegundo (ms))	AI Pesq_global	AI Peso_utilizadores	A2 Pesq_global	A2 PESO_UTILIZADORES
10	Texto A	0,001	0,034	0,090	0,050
11	Texto B	2,414	2,767	3,423	1,155
12	Texto C	4,497	4,044	2,436	1,166
13	Texto D	1,962	2,020	3,671	1,394

Reflexão sucinta sobre os resultados obtidos

(Formato de referência: Arial lOpt; texto para além do número de linhas não é considerado e desvaloriza o relatório)

I. Comente como as estruturas de armazenamento definidas na Tabela 2 influenciaram os tempos de armazenamento e processamento.

Sim. Sendo a estrutura de armazenamento (vários) arrays e não havendo qualquer tipo de ordenação por inserção, (e por isso) não havendo, por exemplo, pesquisa binária, a cada inserção, a função fica limitada a percorrer todo o array (maior O).

2. Analise os resultados dos algoritmos para as tarefas AI e A2 e os resultados. Os resultados foram os esperados? O input mostrou-se relevante? Justifique.

Para o texto A, o QuickSort (A1) mostrou-se mais competente do que o Radix Sort (A2), pois por serem contagens pequenas, ao trabalhar ao nível da base fica mais demoroso do que uma simples comparação de chaves. Para os textos B, C e D, na Pesq_Utilizadores, entre o A1 e o A2, o A2 teve melhor desempenho, pois as contagens aqui são menores ou iguais do que na Pesq_Global, e que o Quicksort tem complexidade O(nlog(n)) e o Radix Sort tem complexidade O(kn), em que k é o número de dígitos da chave, e n o tamanho do array. O Quicksort não tem o melhor desempenho quando as chaves não são únicas, ou seja, quando há muitas chaves iguais.

3. Compare as duas formas de ordenamento: por registo vs. por endereço. Quando será apropriada a escolha de cada uma?

É usado ordenamento por registo ao invés de ordenamento por endereço, pois o tamanho dos registos neste trabalho não é relativamente grande que justifique ordenamento por endereços.

Os registos contêm um número inteiro indicativo do número de vezes da ocorrência de uma determinada palavra.

Bom trabalho, os Docentes da Disciplina,

Carlos L Bento e Catarina Silva

```
CMD_IN_GLOBAL = "PESQ_GLOBAL\n"
CMD_IN_UTILIZADORES = "PESQ_UTILIZADORES\n"
CMD_IN_TERMINADO = "TCHAU\n'
CMD_IN_TERMINADO2 = "TCHAU"
CMD_IN_PALAVRAS = "PALAVRAS\n"
CMD_IN_FIM = "FIM.\n"
CMD_OUT_GUARDADO = "GUARDADAS"
# #### FUNCOES ####
    array_palavras = []
    array_count_global = []

        array_count_utilizadores = []
        # [2, 1, 5, 2] > [[Count, Indice], ...]

        array_ids_utilizadores = []
        # [[109, 114], [109], [455,677,232,124,345], [098,345]] , IDs - Diferentes

        nome_fich = "./StdinsCalculaTempos/StdinTexto" + n_texto + "RelatorioF4.txt"
         acumula_texto = 0
             array_palavras = [] # [omg, xd, a, ahah] | Input "palavra + ID"
             array_count_global = [] # [3, 1, 10, 2] ou [[3, 0], [1, 1], [10, 2], [2, 3]] [Count, Indice]
             my_file = open(nome_fich, "r")
             my_stdin = my_file.read()
             sys.stdin = StringIO(my_stdin)
             if sys.stdin.readline() == CMD_IN_PALAVRAS:
                  array_palavras, array_count_global, array_count_utilizadores, array_ids_utilizadores, n_total = \
                  tempo_texto = end_texto - start_texto
                  acumula_texto = acumula_texto + tempo_texto
                  sys.exit("Erro - Sem Comando Incial: " + CMD_IN_PALAVRAS)
```

A0 – Parte 2/3

```
acumula_texto = (acumula_texto / 20.0) * 1000
       print("***** Tempo Medio em MS - Carregamento/Input TEXTO = " + str(acumula_texto) + " *****")
       n_palavras_distintas = len(array_palavras)
       for i in range(len(array_ids_utilizadores)):
           n_pares_distintas = n_pares_distintas + len(array_ids_utilizadores[i])
       array_utilizadores_distintos = []
       for i in range(len(array_ids_utilizadores)):
                if not array_ids_utilizadores[i][j] in array_utilizadores_distintos:
                    array_utilizadores_distintos.append(array_ids_utilizadores[i][j])
       n_utilizadores_distintos = len(array_utilizadores_distintos)
       print("***** Numero de Palavras Distintas = " + str(n_palavras_distintas) + " *****")
       print(array_palavras)
       print(array_ids_utilizadores)
       input_cmd(array_palavras, array_count_global, array_count_utilizadores)
def input_palavras(array_palavras, array_count_global, array_count_utilizadores, array_ids_utilizadores):
       if linha == CMD IN FIM:
       if palavras[0] in array_palavras:
           indice = array_palavras.index(palavras[0])
           array_count_global[indice][0] += 1
               array_ids_utilizadores[indice].append(int(palavras[1]))
           array_palavras.append(palavras[0])
           array_count_global.append([1, indice])
           array_count_utilizadores.append([1, indice])
   print(CMD OUT GUARDADO)
    return array_palavras, array_count_global, array_count_utilizadores, array_ids_utilizadores, n_total
```

```
def input_cmd(array_palavras, array_count_global, array_count_utilizadores):
        if linha == CMD_IN_TERMINADO2:
        elif linha == CMD_IN_TERMINADO:
       elif linha == CMD_IN_GLOBAL:
           array_count_global = ordenacao(array_count_global)
           valor = array_count_global[-1][0]
            start = len(array_palavras) - 1
            for i in range(len(array_palavras)-1, -1, -1):
                if valor == array_count_global[i][0]:
               indice = array_count_global[i][1]
               alvo.append(array_palavras[indice])
            print(string[:-1])
       elif linha == CMD_IN_UTILIZADORES:
            string =
           start = len(array_palavras)-1
            for i in range(len(array_palavras)-1, -1, -1):
            for i in range(start, len(array_palavras)):
                alvo.append(array_palavras[indice])
               string = string + str(alvo[i]) + " "
           print(string[:-1])
def insertion_sort(array):
```

```
CMD_IN_GLOBAL = "PESQ_GLOBAL\n"
CMD_IN_UTILIZADORES = "PESQ_UTILIZADORES\n"
CMD_IN_TERMINADO = "TCHAU\n'
CMD_IN_TERMINADO2 = "TCHAU"
CMD_IN_PALAVRAS = "PALAVRAS\n"
PARAGEM CORTE = 30
        acumula_global = 0
         sys.stdin = StringIO(my_stdin)
         if sys.stdin.readline() == CMD_IN_PALAVRAS:
             array_palavras, array_count_global, array_count_utilizadores = input_palavras(array_palavras, array_count_global,
             sys.exit("Erro - Sem Comando Incial: " + CMD_IN_PALAVRAS)
             temp_array_count_global = []
                  {\tt temp\_array\_count\_global.append(array\_count\_global[j])}
             tempo_global, tempo_utilizadores = input_cmd(array_palavras, temp_array_count_global, temp_array_count_utilizadores)
             acumula_global = acumula_global + tempo_global
         acumula_global = (acumula_global / 20.0) * 1000
         print("***** Tempo MEDIO em MS - PESQUISA GLOBAL = " + str(acumula_global) + " *****")
print("***** Tempo MEDIO em MS - PESQUISA UTILIZADORES = " + str(acumula_utilizadores) + " *****")
```

A1 – Parte 2/3

```
def input_palavras(array_palavras, array_count_global, array_count_utilizadores):
          array_count_global.append([1, indice])
   start cmd = time.time()
  array_count_global = ordenacao(array_count_global)
   tempo_global = end_cmd - start_cmd
   print("*Tempo em MS - CMD-PesquisaGlobal = " + str(tempo_global * 1000) + " ||| Start Vs End: " + str(start_cmd) +
       alvo.append(array_palavras[indice])
   return tempo_global, tempo_utilizadores
```

A1 – Parte 3/3

```
def insertion_sort(array, indice_baixo, indice_alto):
      for i in range(indice baixo, indice alto + 1):
 def quick_sort(array, indice_baixo, indice_alto):
          if array[indice_meio][0] < array[indice_baixo][0]:</pre>
         pivot = array[indice_meio][0]
          ptr_baixo = indice_baixo
             while array[ptr_alto][0] > pivot:
  def ordenacao(array):
| if __name__ == '__main__':
```

```
def radix_sort(array, tamanho):
    max_digitos = len(str(max(array[:][0]))) # Numero Maior no Array -> Quantos Digitos?
  array contador = []
    for j in range(max_digitos):
                                                     # Contar as ocorrencias de cada digito
       temp = temp_anterior = 0
                                                                                            # este digito no array semiordenado
           array_contador[i] = array_contador[i-1] + temp_anterior # Ver Exemplo/Explicacao nas Notas-iPad
           temp anterior = temp
       array_contador[0] = 0
            array_semiordenado[array_contador[digito]] = array[i]
        array_semiordenado = []
def ordenacao(array):
if __name__ == '__main__':
```