

Disciplina de Introdução à Computação - 2021.1  
Tarefa - Arquivos  
Professor Fernando Antonio Mota Trinta

**Questão 1 (2,0 pontos)** Escreva um programa em Python que seja capaz de analisar o arquivo 'hashtags.txt' e informar a hashtag mais popular e menos popular, além da quantidade de vezes que cada uma aparece no arquivo. Uma hashtag é considerada popular se ela possui o maior número de ocorrências no arquivo.

**Link para o arquivo hastags.txt**

<<https://drive.google.com/file/d/11jrtZTHNaYPGU4pe0KsOGaRij549toLJ/view>>

**Questão 2 (2,0 pontos)** Considere um programa em Python que gera relatórios individuais dos alunos de uma disciplina de um colégio. O programa deve ler um arquivo texto composto com por linhas conforme exemplo abaixo:

```
00030:Fernando Trinta:9.5:10.0:8.7
00031:Raimundo Vale:9.5:10.0
00011:José Maria:7.5:6.0:9.8
00023:Maria José:5.5:5.0:5.7
00001:Cláudio Santos:9.5:10.0:9.7
00100:Saulo Diniz:9.5
00100:Graccho Bolivar:9.5:8.0:7.7
```

O formato de cada linha representa a matrícula do aluno, nome e um conjunto de notas. Cada aluno deve ter três notas associadas. O programa deve gerar um novo arquivo para cada aluno (cada linha) representado no arquivo fornecido com o nome formado pelo número da matrícula + ".txt". Este arquivo deve ser formado conforme formato a seguir:

Ex:

Arquivo 00030.txt

```
Universidade Federal do Ceará
Disciplina de Introdução à Computação
Aluno(a): FERNANDO TRINTA (00030)

Nota de Cálculo 1: 9.50
Nota de Língua Portuguesa: 10.00
Nota de Metodologia Científica: 8.70

Média do Aluno : 9,40
Média da Turma : 8,21
Posição na Turma: (2/5)
```

#### Arquivo 00011.txt

```
Universidade Federal do Ceará  
Disciplina de Introdução à Computação  
Aluno(a): JOSÉ MARIA (00011)
```

```
Nota de Cálculo 1: 7.50  
Nota de Língua Portuguesa: 6.00  
Nota de Metodologia Científica: 9.80
```

```
Média do Aluno : 7,70  
Média da Turma : 8,21  
Posição na Turma: (4/5)
```

Além disso deve ser gerado um arquivo chamado 'erros.txt', em que linhas em que estejam faltando notas de alunos sejam reportadas.

#### Arquivo erros.txt

```
Universidade Federal do Ceará  
Disciplina de Introdução à Computação  
Alunos com falta de informações
```

```
1 - RAIMUNDO VALE  
2 - SAULO DINIZ
```

```
Total: 2 alunos
```

OBS: A posição do aluno é dada de acordo com a ordem das médias dos alunos que possuam as três notas. Alunos com ausência de notas não devem ser contabilizados.

**Questão 3 (3 pontos):** O controle o sistema de medalhas da olimpíada é composto por uma série de arquivos. Um primeiro arquivo chamado 'países.txt' é formado por diversas linhas, onde cada linha representa um par (código; país), como no modelo abaixo:

#### países.txt

```
1; Brasil  
2; Argentina  
3; Estados Unidos  
4; Japão
```

Um segundo arquivo chamado 'esportes.txt' representa os diversos esportes que fazem parte do programa olímpico e é organizado de forma que cada linha representa um esporte e seu respectivo código, conforme exemplo abaixo:

esportes.txt

```
1; Futebol
2; Voleibol
3; Judô
4; Natação
```

Um terceiro arquivo chamado 'provas.txt' representa os diversos resultados dos esportes, indicando os medalhistas em diversas categorias. Cada linha do arquivo é organizada conforme a seguinte disposição:

<codigo\_esporte:nome da prova: pais\_medalha\_ouro: pais\_medalha\_prata:pais\_medalha-bronze>

Um exemplo seria conforme quadro abaixo:

3:60Kg	# representa Judô 60Kg Masc: Ouro(Japão), Prata(Brasil) e Bronze(Estados Unidos)
Masculino:4:1:3	# representa Futebol Masc: Ouro(Brasil), Prata(Argentina) e Bronze(Estados Unidos)
1:Masculino:1:2:3	# representa Judô 48Kg Fem: Ouro(Japão), Prata(Japão) e Bronze(Brasil)
3:48Kg Feminino:4:4:1	# representa Judô 66Kg Masc: Ouro(Argentina), Prata(Japão) e Bronze(Estados Unidos)
3:66Kg	# representa Futebol Fem: Ouro(Estados Unidos), Prata(Argentina) e Bronze(Brasil)
Masculino:2:4:3	
1:Feminino:3:2:1	

Faça um programa que leia estes arquivos e produza dois novos arquivos. O primeiro deve conter o quadro de medalhas das olimpíadas, conforme exemplo abaixo:

Olimpiadas 2020				
Quadro de Medalhas				
Pais	Ouro	Prata	Bronze	Total
Japão	2	2	0	4
Argentina	1	2	0	3
Brasil	1	1	2	4
Estados Unidos	1	0	3	4

OBS: Note que a ordem de apresentação dos países segue a classificação do número de medalhas de ouro, prata e bronze.

O segundo arquivo deve apresentar os campeões por Esporte, conforme exemplo a seguir:

Olimpiadas 2020  
Campeões por Modalidade

Modalidade: Futebol  
Campeão: Brasil - Estados Unidos (1 Ouro - 0 Prata - 1 Bronze)

Modalidade: Judô  
Campeão: Japão (2 Ouro - 2 Prata - 0 Bronze)

OBS: Note que na categoria futebol, há um empate entre Brasil e Estados Unidos, pois cada um teve uma medalha de ouro e outra de bronze.

**Questão 4 (3 pontos)** Um **dataset** é um conjunto de dados finitos, normalmente criados usando linguagens de programação de alto nível. Os datasets são o principal insumo dos processos e análise de dados. Com base nessas informações, você tem a tarefa desafiadora e divertida de analisar um dataset com assuntos de 2021. O conjunto de dados em questão é composto por **2056** arquivos txt, todos seguindo o mesmo padrão, como ilustrado abaixo:

```
conteudo: Centro de Santos tem ruas interditadas para gravações de filme
autor: Hugo Sousa
assunto: filmes
publicado em: 1/2021
```

Seu trabalho é analisar esse dataset e extrair dele as seguintes informações:

- Quais foram os autores e a quantidade de publicações de cada um. Para isso, deve ser gerado um novo arquivo conforme formato abaixo.

```
AUTORES
=====
Nome: João Pedro
Publicações: 300

Nome: Maria Clara
Publicações: 800

...

=====
```

- A quantidade de publicações em cada mês
- Os 3 meses com a maior quantidade de publicações
- Os 3 autores com a maior quantidade de publicações
- As palavras com mais de 5 letras que mais aparecem nas publicações

**Link para o dataset:**

[https://drive.google.com/file/d/10mI9WjUnp18xBX4\\_oCuoBEyxpJ6GkZHj/view?usp=sharing](https://drive.google.com/file/d/10mI9WjUnp18xBX4_oCuoBEyxpJ6GkZHj/view?usp=sharing)