

Iniciado em	quarta, 20 mai 2020, 02:23
Estado	Finalizada
Concluída em	quarta, 20 mai 2020, 02:25
Tempo empregado	1 minuto 18 segundos

Questão 1

Completo

Vale 2,00 ponto(s).

Armazene os documentos disponíveis no diretório DocCol2 em um dicionário onde a chave é o nome do arquivo e o valor é a string contida no arquivo. O documento contendo a string com o maior número de caracteres é:

Escolha uma:

- ☐ a. au8
- ☒ b. au2
- ☐ c. ch5
- ☐ d. gr7

Questão 2

Completo

Vale 2,00 ponto(s).

Crie um dicionário chamado docsXwords onde as chaves são os nomes dos arquivos e os valores são as listas de palavras do documento correspondente. As palavras em cada uma das listas devem ser constituídas apenas por letras do alfabeto, estarem lexicalmente normalizadas e conterem mais que 1 caracter. Qual o documento cuja lista de palavras resultante possui o **maior** número de palavras repetidas:

Escolha uma:

- ☐ a. gr22
- ☐ b. ch30
- ☒ c. au2
- ☐ d. au8

Questão **3**

Completo

Vale 2,00 ponto(s).

Utilizando as listas de palavras do dicionário docsXwords , quais as três palavras que mais aparecem na coleção de documentos:

Escolha uma:

- ☐ a. to, is, of
- ☒ b. the, of, to
- ☐ c. that, is, of
- ☐ d. the, is, of

Questão **4**

Completo

Vale 2,00 ponto(s).

Qual o documento cuja lista de palavras possui o **menor** número de "stop words"? Quantas "stop words" aparecem neste documento:

Dica: Crie um dicionário onde a chave é o nome do documento e o valor o número de stop words no documento.

Escolha uma:

- ☒ a. gr5 com 47 "stop words"
- ☐ b. gr17 com 37 "stop words"
- ☐ c. gr5 com 37 "stop words"
- ☐ d. gr17 com 47 "stop words"

Questão **5**

Completo

Vale 2,00 ponto(s).

Utilize o dicionário docsXwords para construir uma matriz Documentos Palavras para a coleção de documentos do diretório DocCol2 . Utilizando a distância cosseno, qual é o documento mais parecido com o documento 'ch7':

Escolha uma:

- ☒ a. ch16
- ☐ b. ch5
- ☐ c. ch8
- ☐ d. au8

◀ [Material de Apoio - zip](#)

Seguir para...

[Avaliação - ipynb](#) ▶