

Especificação do Trabalho Prático Hadoop
Desenvolvimento de Software para Nuvem
Profs.: Fernando Trinta e Paulo Antonio Rego

Questão 2) Dado um dataset com tweets relacionados à campanha eleitoral presidencial de 2014, responda:

- a) Quais foram as hashtags mais usadas pela manhã, tarde e noite? (1,0 ponto)
- b) Quais as hashtags mais usadas em cada dia? (1,0 ponto)
- c) Qual o número de tweets por hora a cada dia? (1,0 ponto)
- d) Quais as principais sentenças relacionadas à palavra “Dilma”? (1,0 ponto)
- e) Quais as principais sentenças relacionadas à palavra “Aécio”? (1,0 ponto)

Questão 3) Dado um dataset com tweets relacionados à visita da Torre Eiffel em Paris, responda:

- a) Encontre as palavras mais utilizadas nas avaliações. (1,0 ponto)
- b) Encontre as expressões mais usadas. Considere uma expressão um conjunto de palavras na sequência. O tamanho da sequência pode ser determinado por você. (1,0 ponto)
- c) Encontre os principais tópicos relacionados às revisões. (1,0 ponto)
- d) Mapeie a distribuição temporal das revisões. (1,0 ponto)

Material para execução da tarefa:

- Link para os datasets a serem utilizados:
<https://drive.google.com/drive/mobile/folders/0BzwVBj1heLoReXVLaU51bTAxZ2M?usp=sharing>

Formato da Entrega:

Deve ser um relatório contendo:

- a) Descrição do ambiente utilizado para execução das aplicações Hadoop;
- b) Cada item deve ser respondido e devidamente explicado como o resultado foi encontrado em um ou dois parágrafos;