Especificação do Trabalho Prático Haddop
Sistemas de Informação Distribuídos / Desenvolvimento para Nuvem
Profs.: Fernando Trinta e Paulo Antonio Rego

Questão 1) Dado um *dataset* de arquivos texto com o conteúdo de vários livros, deseja-se distribuir a tarefa para se descobrir:

a) Qual ou quais são as palavras com maior número de letras?
b) Qual a média do tamanho das palavras?

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?
b) Quais as hashtags mais usadas em cada dia?
c) Qual o número de tweets por hora a cada dia?
d) Quais as principais sentenças relacionadas à palavra "Dilma"?
e) Quais as principais sentenças relacionadas à palavra "Aécio"?

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

a) Encontre as palavras mais utilizadas nas avaliações.
b) Encontre as expressões mais usadas.
c) Encontre os principais tópicos relacionados às revisões.
d) Mapeie a distribuição temporal das revisões.

Material para execução da tarefa:

- Link para os datasets a serem utilizado: 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;