

Exercício 1:

No Google Colab é possível executar o código linha a linha e analisar seu resultado passo a passo tornando o processo de codificação mais didático, podendo, assim, confirmar a importância do pré-processamento para a análise e correção de erros.

Quanto as stopwords, são palavras que não possuem valor semântico para a busca, normalmente são artigos e preposições.

Já as diferenças entre stemming e lemmatization é que a primeira reduz palavras para sua forma básica cortando sufixos/prefixos, resultando em stems imprecisos. Já lemmatization reduz palavras para sua forma canônica, considerando sua classe gramatical e contexto, resultando em lemas semanticamente corretos. Logo, podemos concluir que a lematização é mais precisa, porém mais complexa.