EXERCÍCIOS DE PROCESSAMENTO DE LINGUAGEM NATURAL

AUTOR: FABRÍCIO GALENDE MARQUES DE CARVALHO

1. INTRODUÇÃO

TERMINOLOGIA E CONCEITOS

TC.1.1. Selecione uma obra literária de domínio público (ex. livros tais como Vinte mil léguas submarinas (de Júlio Verne), a Bíblia, etc.) e ilustre a variedade de dados presente. Considere, por exemplo a construção de frases, orações etc. e compare com expressões de uso corrente.

TC.1.2. Exemplifique uma sentença, escrita na língua portuguesa, que pode surgir em um site de pré-atendimento em uma concessionária, que potencialmente seja difícil de ser interpretado por um *chatbot*. Explique sua resposta em termos de estruturação da sentença e suponha que ela esteja gramaticalmente correta.

TC. 1.3. Sistemas de PNL são geralmente compostos por modelos que são treinados utilizando corpora de texto. Por que modelos que são válidos hoje podem não mais ser adequados daqui a dois anos?

TC.1.4. Por que a utilização de emojis ou outros símbolos não presentes na linguagem textual formal podem dificultar a operação de um sistema de PNL?

TC.1.5. Dê um exemplo de sentença em um processo comunicativo onde a os referentes considerados pelo transmissor e pelo receptor podem ser distintos caso não haja adequada contextualização do processo comunicativo.

TC.1.6. Exemplifique uma a saída para o processo de lematização e stemização. Considere a seguinte sentença:

"Assim que amanheceu, os estudantes, apressados, acordaram e saíram correndo para fazer a prova".

TC.1.7. Cite dois possíveis usos das ***tags*** do tipo POS. Forneça exemplos com sentenças simples, expressas na língua portuguesa ou inglesa.

PRÁTICA DE PROGRAMAÇÃO

PP.1.1 Baseando-se no código-fonte fornecido pelo professor, exemplifique o carregamento da biblioteca NLTK, em Python e efetue a tokenização de um texto em português pertencente a alguma obra literária de domínio público. Utilize um texto de pelo menos 2000 caracteres. Mostre o funcionamento do seu programa e descreva ao menos 5 *POS tags*.

PP.1.2. Exemplifique a stemização e a lematização de um texto, em língua portuguesa. Ilustre um caso onde textos diferentes conduzem a uma mesma saída através do stemming ou lemmatization. Considere como saída um vetor ordenado contendo lemas e stems.

PP.1.3. Repita PP.1.1. considerando a língua inglesa.

PP.1.4. Repita PP.1.2. considerando a língua inglesa.