

**EXERCÍCIO:** Chatbot é um programa de computador que simula uma conversa humana e é capaz de entender as dúvidas do usuário e responder de maneira automatizada\*. Para isso, o chatbot usa diversas tecnologias distintas, dentre elas os bigramas e as medidas de similaridade entre palavras. Nesse exercício, você utilizará o que aprendeu na disciplina Introdução à Programação para implementar essas tecnologias.

**PARTE 1: Bigramas.** Um bigrama (ou 2-grama) é uma sequência de duas letras adjacentes que ocorre uma string. Por exemplo: os bigramas da string 'banana' são: ['ba', 'an', 'na', 'an', 'na'].

Crie um programa capaz de gerar os bigramas de todas as palavras presentes em uma tupla. Mais especificamente, a partir de uma tupla P contendo  $n$  palavras, o seu programa deverá gerar uma lista B contendo os bigramas de cada palavra em P. Veja o exemplo:

Tupla com as palavras	Lista 2d que deverá ser gerada, contendo os bigramas
P = ('banana', 'ananás', 'abacaxi', 'bola', 'bala', 'caxias')	B = [['ba', 'an', 'na', 'an', 'na'], ['an', 'na', 'an', 'ná', 'ás'], ['ab', 'ba', 'ac', 'ca', 'ax', 'xi'], ['bo', 'ol', 'la'], ['ba', 'al', 'la'], ['ca', 'ax', 'xi', 'ia', 'as']]

**PARTE 2: Similaridade de Jaccard (SJ).** A SJ é uma medida usada para avaliar o quanto duas listas de bigramas  $s$  e  $t$  são similares, sendo calculada pela fórmula:  $SJ(s, t) = |s \cap t| \div |s \cup t|$ . O exemplo a seguir mostra os passos que você deve seguir para realizar o cálculo em um programa.

Sejam:  $s = ['ba', 'an', 'na', 'an', 'na']$  (bigramas de “banana”)  
 $t = ['an', 'na', 'an', 'ná', 'ás']$  (bigramas de “ananás”).

Para calcular  $SJ(s, t)$ , basta executar os passos a seguir:

- PASSO 1: ordenar os bigramas
  - $s \rightarrow ['an', 'an', 'ba', 'na', 'na']$
  - $t \rightarrow ['an', 'an', 'ás', 'na', 'ná']$
- PASSO 2: remover repetições
  - $s \rightarrow ['an', 'ba', 'na']$
  - $t \rightarrow ['an', 'ás', 'na', 'ná']$
- PASSO 3: por fim, calcular  $SJ(s, t) = |s \cap t| \div |s \cup t|$ 
  - Ou seja, o módulo da interseção sobre o módulo da união. Para o nosso exemplo:
  - $SJ(s, t) = 2 / 5 = 0.4$

De acordo com a explicação apresentada, crie um programa capaz de calcular a SJ entre um par de listas de bigramas. Você pode testar usando um par de exemplos da lista B, criada no exercício anterior.

**\* \* \* IMPORTANTE:** você **não** poderá utilizar recursos ou pacotes que não foram mostrados na disciplina. Sua solução deve usar apenas conteúdo que tenha sido apresentado na matéria.

\* <https://www.ibm.com/br-pt/topics/chatbots#:~:text=IBM,e%20responder%20de%20maneira%20automatizada.>