

Daniel Cu Sánchez

A01703613

Actividad final 01

```
index.py X
C:\Users\> daniel > OneDrive > Escritorio > metodos-y-simulacion > actividad_final_01 > index.py > ...
21 palabras = list(set(texto1.split() + texto2.split()))
22 vector1 = [texto1.split().count(palabra) * np.log(2 / (texto1.split().count(palabra) + texto2.split().count(palabra))) for palabra in palabras]
23 vector2 = [texto2.split().count(palabra) * np.log(2 / (texto1.split().count(palabra) + texto2.split().count(palabra))) for palabra in palabras]
24 return vector1, vector2
25
26 # Textos de ejemplo
27 texto1 = "Internet efectivo en comunidades rurales, somos ALTERNET"
28 texto2 = "Internet en comunidades rurales, rompemos la brecha digital somos ALTERNET"
29
30 # Vectorización utilizando la técnica de Bolsa de Palabras
31 vector_bolsa1, vector_bolsa2 = bolsa_de_palabras(texto1, texto2)
32
33 # Vectorización utilizando la técnica de tf-idf
34 vector_tf1, vector_tf2 = tfidf(texto1, texto2)
35
36 # Cálculo de la similitud utilizando la medida del coseno
37 similitud_bolsa = similitud_coseno(vector_bolsa1, vector_bolsa2)
38 similitud_tf1 = similitud_coseno(vector_tf1, vector_tf2)
39
40 print("Similitud utilizando bolsa de palabras: 🍌", similitud_bolsa)
41 print("Similitud utilizando tf-idf 🍌:", similitud_tf1)
42

PROBLEMAS TERMINAL
▼ TERMINAL
daniel@AL MINGW64 ~
$ C:\Users\daniel\AppData\Local\Microsoft\WindowsApps\python3.11.exe c:/Users/daniel/OneDrive/Escritorio/metodos-y-simulacion/actividad_final_01/index.py
Similitud utilizando Bolsa de Palabras: 0.7171371656006361
Similitud utilizando tf-idf: 0.0
```

La técnica que tiene más precisión en este caso fue la bolsa de palabras. En resumen, las bolsas de palabras son más simple y rápido, mientras que tf-idf es más preciso al considerar la importancia relativa de las palabras. La bolsa de palabras cuenta las palabras, mientras que tf-idf pondera las palabras según su relevancia. La elección entre ambas técnicas depende del contexto y los requisitos específicos del análisis de texto.