



# INSTITUTO POLITECNICO NACIONAL



Escuela Superior de Cómputo

Practica 4: IDENTIFICACIÓN DE FRASES, PALABRAS  
Y DOCUMENTOS SIMILARES

Unidad de aprendizaje: Tecnologías de Lenguaje  
Natural

Alumno:

Flores Lara Alberto

Profesor: Flores

Estrada Ituriel

Enrique

Grupo

5BV1



## INDICE

• Introducción.....	3
• Objetivos de la practica.....	3
• Desarrollo punto por punto de la practica	
○ Punto 1.....	4
○ Punto 2.....	4
○ Punto 3.....	4
○ Punto 4.....	6
○ Punto 5.....	7
○ Punto 6.....	8
• Conclusión General.....	9

## INTRODUCCION

En esta práctica de Tecnologías de Lenguaje Natural, analizamos y comparamos textos para identificar similitudes entre palabras, frases y documentos.

### OBJETIVO DE LA PRÁCTICA

El objetivo es crear un programa en Python para identificar palabras, frases y documentos similares. Aquí están los pasos principales:

#### 1. Generación del Cuerpo de Documentos:

- Seleccionaremos 10 libros del portal Gutenberg que sean de géneros o temas similares.
- Usaremos el primer capítulo de cada libro para formar nuestro conjunto de documentos.

#### 2. Preparación del Texto:

- Segmentaremos cada documento en oraciones y luego en tokens (palabras).
- Etiquetaremos cada token con su categoría gramatical (por ejemplo, sustantivo, verbo, etc.).

#### 3. Similitud Semántica de Palabras y Documentos:

- **Palabras:** Identificaremos el verbo y el sustantivo más comunes en cada documento. Usaremos las métricas `wup_similarity` y `Path_similarity` para encontrar los términos más similares a estos.
- **Documentos:** Extraeremos la frase más representativa del primer capítulo de cada libro y compararemos estas frases entre sí usando `Path_similarity`.

#### 4. Similitud Sintáctica de Palabras y Documentos:

- **Palabras:** Utilizaremos el modelo pre-entrenado de GloVe para encontrar los términos más similares al verbo más frecuente en cada documento, usando similitud de coseno.
- **Documentos:** Compararemos las frases representativas de los libros utilizando el modelo pre-entrenado de BERT, que es muy efectivo para medir similitudes de texto.

## DESARROLLO PUNTO POR PUNTO DE LA PRACTICA

Antes de comenzar daré una breve descripción de las librerías que utilizamos:

- **nltk:** Utilizamos esta librería para las funciones de tokenización, de etiquetado gramatical (identificar sustantivos, verbos, etc.), y también utilizamos el recurso de WordNet, que es una base de datos de sinónimos, para los synsets.

- re: Nos ayudó a trabajar con expresiones regulares como una forma de buscar y reemplazar patrones para poder hacer una buena limpieza del texto.
- torch y transformers: Estas librerías las utilizamos para trabajar con el modelo de BERT. Las utilizamos para tokenizar el texto y obtener los embeddings que podemos usar para medir la similitud entre textos.
- numpy: Esencial para manejar matrices y operaciones matemáticas. Aquí la usamos para calcular las distancias entre los embeddings.
- scipy: La utilizamos para poder calcular la similitud entre vectores.

### **Punto #1**

Para esta práctica seleccionamos 10 textos de crimen de ficción del portal Gutenberg. Primero procedimos a leer y organizar los documentos de texto para su posterior análisis. A continuación, se presenta una breve descripción de las funciones utilizadas:

- PlaintextCorpusReader: Es una herramienta de nltk que utilizamos para leer y organizar un conjunto de documentos de texto.

### **Punto #2**

Posteriormente, procedimos a limpiar, preprocesar y etiquetar gramaticalmente el texto de los documentos utilizando las librerías nltk, re y stopwords. Esta es una breve descripción de las funciones utilizadas:

- clean\_text: Limpia el texto de caracteres no deseados.
- aplicar\_procesamiento: Aplica varias funciones de preprocesamiento al texto, como convertir a minúsculas, eliminar caracteres especiales y eliminar stopwords.
- pos\_tag\_text: Etiqueta gramaticalmente el texto, identificando la categoría gramatical de cada palabra.

### **Punto #3**

Posteriormente procedimos a encontrar dentro de cada documento la palabra más frecuente en cada categoría gramatical y calcular sus similitudes con otras palabras usando nltk y synsets de wordnet. Esta es una breve descripción de las funciones utilizadas:

- most\_frequent\_word: Encuentra la palabra más frecuente de una categoría gramatical específica (como verbos o sustantivos).
- word\_similarity: Calcula la similitud entre palabras usando sinónimos (synsets) de WordNet.

Estos fueron los resultados obtenidos:

```

Documento: Libro 1.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: rnine
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

Documento: Libro 2.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: door
Similitudes WUP (sustantivo): [(Synset('door.n.05'), 0.75), (Synset('door.n.04'), 0.75), (Synset('door.n.04'), 0.7058823529411765), (Synset('door.n.01'), 0.70588235294117
Similitudes Path (sustantivo): [(Synset('door.n.05'), 0.2), (Synset('door.n.04'), 0.2), (Synset('door.n.04'), 0.1666666666666666), (Synset('door.n.01'), 0.1666666666666666

Documento: Libro 3.txt
Verbo más frecuente: came
Similitudes WUP (verbo): [(Synset('come.v.16'), 0.6666666666666666), (Synset('issue_forth.v.01'), 0.5), (Synset('derive.v.05'), 0.5), (Synset('arrive.v.01'), 0.5), (Synse
Similitudes Path (verbo): [(Synset('come.v.16'), 0.5), (Synset('arrive.v.01'), 0.5), (Synset('come.v.11'), 0.3333333333333333), (Synset('issue_forth.v.01'), 0.333333333333
Sustantivo más frecuente: man
Similitudes WUP (sustantivo): [(Synset('world.n.08'), 0.9655172413793104), (Synset('homo.n.02'), 0.9655172413793104), (Synset('man.n.03'), 0.7058823529411765), (Synset('m
Similitudes Path (sustantivo): [(Synset('world.n.08'), 0.5), (Synset('homo.n.02'), 0.5), (Synset('man.n.06'), 0.3333333333333333), (Synset('man.n.08'), 0.3333333333333333

Documento: Libro 4.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: lupin
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

```

### Resultados del tercer punto de la practica (Libros 1-4)

```

Documento: Libro 5.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: man
Similitudes WUP (sustantivo): [(Synset('world.n.08'), 0.9655172413793104), (Synset('homo.n.02'), 0.9655172413793104), (Synset('man.n.03'), 0.7058823529411765), (Synset('m
Similitudes Path (sustantivo): [(Synset('world.n.08'), 0.5), (Synset('homo.n.02'), 0.5), (Synset('man.n.06'), 0.3333333333333333), (Synset('man.n.08'), 0.3333333333333333

Documento: Libro 6.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: irvin
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

Documento: Libro 7.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: raffles
Similitudes WUP (sustantivo): [(Synset('raffle.n.01'), 0.1111111111111111), (Synset('raffles.n.01'), 0.1111111111111111)]
Similitudes Path (sustantivo): [(Synset('raffle.n.01'), 0.058823529411764705), (Synset('raffles.n.01'), 0.058823529411764705)]

Documento: Libro 8.txt
Verbo más frecuente: fleming
Similitudes WUP (verbo): []
Similitudes Path (verbo): []
Sustantivo más frecuente: rand
Similitudes WUP (sustantivo): [(Synset('witwatersrand.n.01'), 0.4), (Synset('rand.n.02'), 0.4), (Synset('witwatersrand.n.01'), 0.14285714285714285), (Synset('rand.n.01'),
Similitudes Path (sustantivo): [(Synset('witwatersrand.n.01'), 0.1), (Synset('rand.n.02'), 0.1), (Synset('witwatersrand.n.01'), 0.07692307692307693), (Synset('rand.n.01'),

Documento: Libro 9.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: lupin
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

Documento: Libro 10.txt
Verbo más frecuente: captain
Similitudes WUP (verbo): []
Similitudes Path (verbo): []
Sustantivo más frecuente: denison
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

```

### Resultados del tercer punto de la practica (Libros 5-8)

```

Documento: Libro 9.txt
Verbo más frecuente: said
Similitudes WUP (verbo): [(Synset('say.v.11'), 0.5333333333333333), (Synset('say.v.07'), 0.5333333333333333), (Synset('say.v.09'), 0.5), (Synset('pronounce.v.01'), 0.5),
Similitudes Path (verbo): [(Synset('say.v.08'), 0.3333333333333333), (Synset('say.v.09'), 0.3333333333333333), (Synset('state.v.01'), 0.3333333333333333), (Synset('pronou
Sustantivo más frecuente: lupin
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

Documento: Libro 10.txt
Verbo más frecuente: captain
Similitudes WUP (verbo): []
Similitudes Path (verbo): []
Sustantivo más frecuente: denison
Similitudes WUP (sustantivo): []
Similitudes Path (sustantivo): []

```

### Resultados del tercer punto de la practica (Libros 9-10)

Podemos interpretar que la mayoría de los documentos tienen "said" como el verbo más frecuente, y sus similitudes se encuentran consistentemente altas con otras formas del verbo "decir". Para los sustantivos, algunos como "man" y "door" mostraron altas similitudes, mientras que nombres propios o términos no

reconocidos no generaron resultados útiles.

#### Punto #4

A continuación, procedimos a extraer la frase más representativa de cada documento y calcular la similitud entre documentos utilizando nltk, wordnet y modelos pre-entrenados de BERT. Esta es una breve descripción de las funciones utilizadas:

- doc\_to\_synsets: Convierte un documento en una lista de synsets de WordNet.
- document\_path\_similarity: Calcula la similitud entre documentos utilizando la métrica "Path\_similarity".

A continuación, se presentan los resultados obtenidos:

```
Similitud entre Libro_1.txt y Libro_2.txt: 0.78
Similitud entre Libro_1.txt y Libro_3.txt: 0.77
Similitud entre Libro_1.txt y Libro_4.txt: 0.11
Similitud entre Libro_1.txt y Libro_5.txt: 0.66
Similitud entre Libro_1.txt y Libro_6.txt: 0.66
Similitud entre Libro_1.txt y Libro_7.txt: 0.14
Similitud entre Libro_1.txt y Libro_8.txt: 0.10
Similitud entre Libro_1.txt y Libro_9.txt: 0.12
Similitud entre Libro_1.txt y Libro_10.txt: 0.85
Similitud entre Libro_2.txt y Libro_3.txt: 0.55
Similitud entre Libro_2.txt y Libro_4.txt: 0.11
Similitud entre Libro_2.txt y Libro_5.txt: 0.62
Similitud entre Libro_2.txt y Libro_6.txt: 0.44
Similitud entre Libro_2.txt y Libro_7.txt: 0.14
Similitud entre Libro_2.txt y Libro_8.txt: 0.10
Similitud entre Libro_2.txt y Libro_9.txt: 0.12
Similitud entre Libro_2.txt y Libro_10.txt: 0.63
Similitud entre Libro_3.txt y Libro_4.txt: 0.10
Similitud entre Libro_3.txt y Libro_5.txt: 0.44
Similitud entre Libro_3.txt y Libro_6.txt: 0.44
Similitud entre Libro_3.txt y Libro_7.txt: 0.13
Similitud entre Libro_3.txt y Libro_8.txt: 0.09
Similitud entre Libro_3.txt y Libro_9.txt: 0.10
Similitud entre Libro_3.txt y Libro_10.txt: 0.62
Similitud entre Libro_4.txt y Libro_5.txt: 0.73
Similitud entre Libro_4.txt y Libro_6.txt: 0.10
Similitud entre Libro_4.txt y Libro_7.txt: 0.11
Similitud entre Libro_4.txt y Libro_8.txt: 0.91
Similitud entre Libro_4.txt y Libro_9.txt: 0.79
Similitud entre Libro_4.txt y Libro_10.txt: 0.61
Similitud entre Libro_5.txt y Libro_6.txt: 0.33
Similitud entre Libro_5.txt y Libro_7.txt: 0.13
Similitud entre Libro_5.txt y Libro_8.txt: 0.82
Similitud entre Libro_5.txt y Libro_9.txt: 0.72
Similitud entre Libro_5.txt y Libro_10.txt: 0.93
Similitud entre Libro_6.txt y Libro_7.txt: 0.25
Similitud entre Libro_6.txt y Libro_8.txt: 0.09
Similitud entre Libro_6.txt y Libro_9.txt: 0.32
Similitud entre Libro_6.txt y Libro_10.txt: 0.51
Similitud entre Libro_7.txt y Libro_8.txt: 0.10
Similitud entre Libro_7.txt y Libro_9.txt: 0.25
Similitud entre Libro_7.txt y Libro_10.txt: 0.14
Similitud entre Libro_8.txt y Libro_9.txt: 0.89
Similitud entre Libro_8.txt y Libro_10.txt: 0.70
Similitud entre Libro_9.txt y Libro_10.txt: 0.60
```

Resultados del cuarto punto de la practica

Al comparar las frases más representativas de cada documento, observamos que algunos libros tienen una alta similitud, como Libro\_1.txt con Libro\_2.txt (0.78) y Libro\_1.txt con Libro\_10.txt (0.85). Por lo cual podemos deducir que los temas y el lenguaje utilizados en estos libros son muy parecidos. Por otro lado, hay libros con bajas similitudes, como Libro\_1.txt con Libro\_4.txt (0.11) y Libro\_1.txt con Libro\_8.txt (0.10), indicando que estos textos tienen menos en común en términos de contenido y estilo.

## Punto #5

Posteriormente, procedimos a encontrar las palabras más similares al verbo más frecuente en cada documento utilizando embeddings de GloVe. Esta es una breve descripción de las funciones utilizadas:

- `load_glove_embeddings`: Carga las embeddings de GloVe desde un archivo.
- `find_closest_embeddings`: Encuentra las palabras más similares a una palabra dada utilizando las embeddings de GloVe.

Estos fueron los resultados:

```
Documento: Libro_1.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_2.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_3.txt
Verbo más frecuente: came
Términos más similares: ['came', 'took', 'after', 'saw', 'when']

Documento: Libro_4.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_5.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_6.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_7.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_8.txt
Verbo más frecuente: fleming
Términos más similares: ['fleming', 'stuart', 'clarke', 'healy', 'fletcher']

Documento: Libro_9.txt
Verbo más frecuente: said
Términos más similares: ['said', 'told', 'says', 'spokesman', 'saying']

Documento: Libro_10.txt
Verbo más frecuente: captain
Términos más similares: ['captain', 'skipper', 'seaman', 'indies', 'cole']
```

Resultados del quinto punto de la practica

La mayoría de los documentos tienen "said" como el verbo más frecuente, y los términos más similares encontrados utilizando embeddings de GloVe son consistentes con esta palabra, incluyendo variantes como "told", "says" y "saying"; así mismo, en otros casos, como "came" en el Libro\_3 y "captain" en el Libro\_10, las palabras similares están relacionadas con su contexto. Esto indica que los embeddings de GloVe capturaron de buena manera las relaciones semánticas entre estas palabras. Sin embargo, tenemos un caso en particular el cual es "fleming" en el Libro\_8, donde los términos más similares son nombres propios, lo que indica que hubo un error en donde un nombre propio fue mal etiquetado.

## Punto #6

Finalmente, procedimos a calcular la similitud entre documentos usando embeddings de BERT. Esta es una breve descripción de las funciones utilizadas:

- `bert_similarity`: Calcula la similitud entre documentos utilizando embeddings de BERT.

Estos fueron los resultados:

```
Similitud BERT entre Libro_1.txt y Libro_2.txt: 0.62
Similitud BERT entre Libro_1.txt y Libro_3.txt: 0.65
Similitud BERT entre Libro_1.txt y Libro_4.txt: 0.67
Similitud BERT entre Libro_1.txt y Libro_5.txt: 0.65
Similitud BERT entre Libro_1.txt y Libro_6.txt: 0.61
Similitud BERT entre Libro_1.txt y Libro_7.txt: 0.68
Similitud BERT entre Libro_1.txt y Libro_8.txt: 0.59
Similitud BERT entre Libro_1.txt y Libro_9.txt: 0.69
Similitud BERT entre Libro_1.txt y Libro_10.txt: 0.65
Similitud BERT entre Libro_2.txt y Libro_3.txt: 0.89
Similitud BERT entre Libro_2.txt y Libro_4.txt: 0.64
Similitud BERT entre Libro_2.txt y Libro_5.txt: 0.67
Similitud BERT entre Libro_2.txt y Libro_6.txt: 0.93
Similitud BERT entre Libro_2.txt y Libro_7.txt: 0.63
Similitud BERT entre Libro_2.txt y Libro_8.txt: 0.64
Similitud BERT entre Libro_2.txt y Libro_9.txt: 0.69
Similitud BERT entre Libro_2.txt y Libro_10.txt: 0.91
Similitud BERT entre Libro_3.txt y Libro_4.txt: 0.66
Similitud BERT entre Libro_3.txt y Libro_5.txt: 0.67
Similitud BERT entre Libro_3.txt y Libro_6.txt: 0.92
Similitud BERT entre Libro_3.txt y Libro_7.txt: 0.69
Similitud BERT entre Libro_3.txt y Libro_8.txt: 0.61
Similitud BERT entre Libro_3.txt y Libro_9.txt: 0.69
Similitud BERT entre Libro_3.txt y Libro_10.txt: 0.87
Similitud BERT entre Libro_4.txt y Libro_5.txt: 0.81
Similitud BERT entre Libro_4.txt y Libro_6.txt: 0.65
Similitud BERT entre Libro_4.txt y Libro_7.txt: 0.78
Similitud BERT entre Libro_4.txt y Libro_8.txt: 0.89
Similitud BERT entre Libro_4.txt y Libro_9.txt: 0.88
Similitud BERT entre Libro_4.txt y Libro_10.txt: 0.69
Similitud BERT entre Libro_5.txt y Libro_6.txt: 0.66
Similitud BERT entre Libro_5.txt y Libro_7.txt: 0.75
Similitud BERT entre Libro_5.txt y Libro_8.txt: 0.85
Similitud BERT entre Libro_5.txt y Libro_9.txt: 0.88
Similitud BERT entre Libro_5.txt y Libro_10.txt: 0.72
Similitud BERT entre Libro_6.txt y Libro_7.txt: 0.64
Similitud BERT entre Libro_6.txt y Libro_8.txt: 0.63
Similitud BERT entre Libro_6.txt y Libro_9.txt: 0.70
Similitud BERT entre Libro_6.txt y Libro_10.txt: 0.91
Similitud BERT entre Libro_7.txt y Libro_8.txt: 0.72
Similitud BERT entre Libro_7.txt y Libro_9.txt: 0.82
Similitud BERT entre Libro_7.txt y Libro_10.txt: 0.66
Similitud BERT entre Libro_8.txt y Libro_9.txt: 0.88
Similitud BERT entre Libro_8.txt y Libro_10.txt: 0.68
Similitud BERT entre Libro_9.txt y Libro_10.txt: 0.72
```

Resultados del sexto punto de la practica



Utilizando embeddings de BERT, observamos las siguientes similitudes entre los documentos:

- Alta Similitud: Documentos como Libro\_2.txt y Libro\_3.txt (0.89), así como Libro\_6.txt y Libro\_10.txt (0.91), muestran una alta similitud. Esto podría indicar que estos libros comparten temas, estilo y estructura de lenguaje muy similares.
- Moderada Similitud: La mayoría de las comparaciones entre documentos muestran similitudes moderadas, con valores en el rango de 0.60 a 0.70. Esto indica que los documentos comparten algunas características comunes como lo pueden ser el tema y la estructura, pero también tienen diferencias significativas.
- Baja Similitud: Comparaciones como Libro\_1.txt y Libro\_8.txt (0.59) muestran una similitud más baja, lo que sugiere diferencias más marcadas en temas o estilo entre estos libros.

## **CONCLUSION GENERAL**

La capacidad de identificar similitudes entre documentos y palabras no solo es importante para el análisis de textos literarios, sino que también tiene aplicaciones prácticas en sistemas de recomendación, clasificación de textos y análisis de contenido.

Este ejercicio proporcionó una mayor comprensión de cómo aplicar técnicas procesamiento de lenguaje natural para el análisis y comparación de textos, destacando la importancia de las herramientas como WordNet, GloVe y BERT en la obtención de resultados significativos. La metodología aplicada en esta práctica ha demostrado ser útil para aplicaciones prácticas en diversos campos relacionados con el procesamiento y análisis de textos.