

## Ejercicio 1

Aplique `spacy` para detectar las frases sustantivas o nominales de la siguiente oración:  
"El desayuno es la oportunidad de empezar el día con un festín de colores, sabores y nutrientes."

## Ejercicio 2

1. Aplique similitud del coseno para obtener las frases cercanas en la lista a la primera de la misma

```
texto = ['¿Qué es el coseno y para qué  
sirve?', '¿Cuál es el coseno de un  
ángulo?',  
'¿Cuál es el valor del coseno?', '¿Qué es  
el coseno?', 'Esta es una introducción a  
NLP', 'Es probable que sea útil para las  
personas', 'Machine learning es la nueva  
electricidad', 'Habrá menos exageración  
sobre la IA y más acción en  
adelante', '¡Python es la mejor  
herramienta!', 'Python es un buen  
lenguaje', 'Me gusta este libro', 'Quiero  
más libros como este']
```

2. Recordemos que la **Similitud de Dice** es generalmente más sensible a las coincidencias (elementos en común) que la **Similitud de Jaccard**, ya que el numerador en la fórmula de Dice se multiplica por 2. Esto significa que, para dos conjuntos con el mismo número de coincidencias y diferencias, la Similitud de Dice generalmente dará un valor más alto que la Similitud de Jaccard.

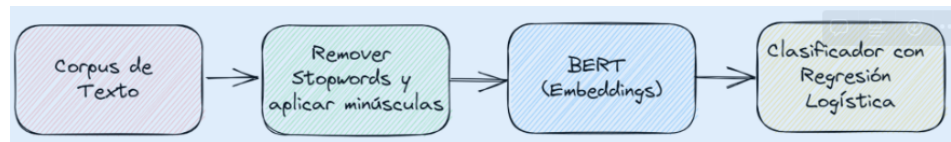
Verifique esta afirmación realizando la misma comparación que en el punto anterior, es decir, comparando la primera oración con las restantes.

3. Calcule las **distancias de Levenshtein** entre los siguientes apellidos:

```
apellidos = [  
    ("Gómez", "Gómez"),  
    ("Fernández", "Hernández"),  
    ("López", "Lopes"),  
    ("Pérez", "Perez"),  
    ("García", "González"),  
    ("Ramírez", "Rodríguez"),  
    ("Sánchez", "Sanchez"),  
    ("Torres", "Torres"),  
    ("Díaz", "Diaz"),  
]
```

## Ejercicio 3

Utilizando la misma secuencia del apartado 4 de la unidad 3 :



Entrenar un clasificador con **TF-IDF** como vectorizador, para cuatro diferentes "labels", el texto del [link](#)

```
!wget  
https://raw.githubusercontent.com/aleoncavallo/tuia\_nlp\_2023/main/texto\_train.txt
```

Obtener las métricas del modelo y analícelas.

Luego definir un grupo de frases y realizar el proceso de inferencia.

Luego graficar las palabras más influyentes.

Realizar el mismo proceso con **Bert** y comparar resultados.

## Ejercicio 4

Mediante el uso de las técnicas de Análisis de sentimientos. Explore los textos de reseñas de una cafetera de cápsulas que se encuentran en el siguiente archivo:

[https://raw.githubusercontent.com/aleoncavallo/tuia\\_nlp\\_2023/main/texto\\_sentimental.txt](https://raw.githubusercontent.com/aleoncavallo/tuia_nlp_2023/main/texto_sentimental.txt)

Analice el valor de cada comentario y también genere una descripción de las estadísticas principales del conjunto. Utilice algún análisis gráfico para su resolución.

## Ejercicio 5

Utilice el ejemplo de capítulo de Resumen de Texto para realizar un resumen del siguiente texto: [Tecnología y Sociedad](#). Analice la salida. Luego inténtelo de a 2 párrafos y verifique si encuentra mayor coherencia.

---

ejercicios unidad 3

Actualizado automáticamente cada 5 minutos

---