

Análisis de Textos
Diciembre de 2021

Ejercicio 1 (10 puntos)

Indique si las siguientes afirmaciones son Verdaderas o Falsas. Justifique en caso de que sea Falsa.

- a) Una stop word es una palabra que no contiene significado léxico.
- b) La lematización de una palabra x es el proceso que genera todas las posibles palabras derivadas a partir de ella.
- c) En la frase *La casa de papel* la palabra **de** tiene categoría adverbio.
- d) Una oración siempre es un enunciado pero un enunciado puede no ser una oración.

Ejercicio 2 (10 puntos)

Sea la oración *La casita de la esquina es de Juan Ignacio*.

- a) Indique cuántas palabras, tipos y tokens hay en ella.
- b) Identifique todos los sintagmas contenidos en ella, indicando núcleo y categoría del núcleo.

Ejercicio 3 (5 puntos)

Calcule la distancia de mínima edición de Levenshtein entre las palabras AMAR y TOSER.

Ejercicio 4 (5 puntos)

Dé una expresión regular que permita:

- a) encontrar todas las ocurrencias de las palabras que comienzan con el string '**Par**', luego viene al menos un dígito y terminan con una secuencia de pares – eventualmente vacía - de secuencias **ab**.
- b) sustituir los símbolos # o & por una X

Ejercicio 5 (10 puntos)

- a) ¿A qué nos referimos con *word embeddings* en el área PLN?
- b) Enuncie la hipótesis distribucional.

Ejercicio 6 (20 puntos)

Se desea contar con una herramienta que clasifique comentarios de usuarios de un hotel en forma automática. Se muestran algunos ejemplos de comentarios:

- *El servicio es muy bueno, pero los precios son demasiado altos.*
- *La restaurante del hotel ofrece comida poco variada y de mala calidad.*
- *El hotel es agradable, resolvió mis necesidades de alojamiento.*

- a) ¿Qué clases definiría para clasificar los comentarios? Justifique.
- b) ¿Qué herramientas utilizaría para construir su clasificador, si no contara con un corpus anotado?
- c) ¿Qué herramientas utilizaría para construir su clasificador, si contara con un corpus anotado?