

Procesamiento del Lenguaje Natural

Luis Gabriel Moreno Sandoval

Taller NLP 1:

El taller podrá ser resuelto en Jupyter Python d de Python, para esto se creará un método que resuelva cada una de las preguntas.

Plazo de entrega del taller es una semana.

1. Crear un programa en el que se genere un archivo (fichero_suma.txt) que sea el resultado de la suma de los valores contenidos en dos archivos (fichero1.txt, fichero2.txt). Estos ficheros contendrán 5 valores (1 por línea).
2. Utilizar el sistema de ayuda de Python para obtener información sobre las librerías “nltk” y “sklearn” que serán utilizadas en las siguientes prácticas.
3. Producir un diagrama de dispersión de los cuatro protagonistas Elinor, Marianne, Edward y Willoughby en “text2: Sense and Sensibility by Jane Austen 1811”:
 - a. ¿Qué puedes observar sobre los diferentes roles que juegan los hombres y las mujeres en esta novela?
 - b. ¿Puedes identificar a las parejas?
4. En la primera oración del texto3 se crea la variable sent3 que la contiene. Ahora en el índice 1 de este arreglo sent3 nos da 'the', se expresaría sent3[1].
 - a. ¿Cuáles son los índices de las otras dos ocurrencias de esta palabra en sent3?
 - b. ¿Qué otros índices nos servirían para recuperar la misma información?
5. ¿Cuál es la diferencia entre las líneas a y b? ¿Cuál dará un mayor valor? ¿Este caso es generalizable?
 - a. `>>> sorted(set(w.lower() for w in text1))`
 - b. `>>> sorted(w.lower() for w in set(text1))`

Tutoriales:

- Tutorial sobre Spyder: <https://pythonhosted.org/spyder/>
- Tutorial sobre Python 3.5: <https://docs.python.org/3.5/tutorial/>
- Tutorial Jupyter: <https://notebooks.azure.com/help/libraries/library-view>
- Tutorial Anaconda: <https://www.anaconda.com/community/>