

Laboratorio 1: Nubes de palabras y estadísticas del lenguaje

Trabaje en parejas.

1. (0%) Abra un jupyter notebook (ejecute el comando “jupyter notebook” en una terminal). Cree un nuevo notebook usando New->Python 3.
2. (40%) Siga los pasos en el siguiente sitio ([Word Cloud](#)) para crear una nube de palabras a partir de un documento pdf. Escoja un texto de su agrado en formato pdf, que se encuentre en inglés y que cuente con de 3000 a 15,000 palabras. El pdf debe ser de máximo 5MB. Modifique el ejemplo para que funcione con su pdf y despliegue el tamaño de las palabras basado en su frecuencia en el documento. Es decir, que las palabras que aparecen más veces tengan mayor tamaño. El código debe pasar de pdf a txt, quitar puntuación y caracteres especiales y generar la nube de palabras.
3. Añada a su programa las siguientes funcionalidades:
 - a. (40%) El programa determina el número de apariciones para cada palabra del texto, que se guardan en una estructura de datos relevante. El programa despliega las N palabras más frecuentes, de acuerdo con un número entero ingresado por el usuario.
 - b. (10%) El programa recibe una palabra por parte del usuario y despliega la probabilidad de que ocurra esta palabra.
 - c. (10%) Modifique el programa para que las palabras de la nube de palabras tengan varios colores.

Entregue su documento jupyter notebook (extensión ipynb), indicando claramente los cambios realizados y justificando con comentarios las respuestas a la pregunta 3. Incluya también el pdf utilizado. Suba sus archivos comprimidos en un .zip en el espacio correspondiente en Mediación Virtual. La fecha límite de entrega es el viernes 22 de marzo a las 23:55 horas. La entrega será considerada tardía después de las 6:00am del día siguiente. Por cada día de atraso perderá 20% y luego de 5 días no será contemplada la evaluación.