**Título de la práctica**

Apellido1, Nombre1., Apellido2, Nombre2., y ApellidoN, NombreN.

Benemérita Universidad Autónoma de Puebla

Nombre de la Materia

Fecha:

**Resumen:** En esta práctica se instaló Python con la librería NLTK el cual es una herramienta para el procesamiento del lenguaje natural, con el fin de aprender a preparar un texto, en este caso se usó el e-book: “Around the World in Eighty Days, by Jules Verne”. Para llevar a cabo el trabajo de alistar el texto, se leyó el archivo y después se separó el texto por palabras, se eliminaron los signos de puntuación usando “re”, se cambió las mayúsculas a minúsculas y finalmente se eliminaron las palabras vacías. Esta práctica aportó un entendimiento más claro en el uso de Python con NLTK para preparar textos, ya que se pudo analizar con cada paso como cambiaba el texto.

**Palabras clave:** Tokens, Python, NLTK, texto, palabras vacías.

**I. INTRODUCCIÓN**

En esta práctica usamos el texto “Around the World in Eighty Days, by Jules Verne” para comprender y analizar el uso de Python con NLTK para preparar un texto, llevando a cabo la separación de Tokens, la eliminación de puntuación, las mayúsculas y las palabras vacías. Es importante limpiar el texto que puede contener palabras no deseadas o sin importancia, si no se hace este proceso, puede que tengamos resultados que no brinden una precisión eficiente y dificulten su comprensión y análisis.

El trabajo está conformado por un resumen, la introducción, el objetivo del problema, el desarrollo del problema, los resultados, las conclusiones y la bibliografía.

**II. OBJETIVO Y PLANTEAMIENTO DEL PROBLEMA**

Aprender a preparar los textos para que sean de utilidad en el proceso de recuperación de información. Para ello deberás separar el texto en tokens, eliminar los tokens inútiles (signos de puntuación, números, palabras vacías y convertir a minúsculas.

**III. DESARROLLO EXPERIMENTAL**

En esta sección deben describirse los materiales y métodos (herramientas de software y pasos seguidos) utilizados para el desarrollo de la práctica. Puede especificarse en una lista:

Elemento 1

Elemento 2

En esta sección debe describirse cómo se llevó a cabo el desarrollo y pruebas del software.

**IV. DISCUSIÓN Y RESULTADOS**

Conjunto de datos obtenidos experimentalmente y tratados estadísticamente. Use tablas para organizar y resumir los resultados. Incluya solo los datos importantes y relevantes, pero suficientes para justificar sus conclusiones. Usa gráficos y figuras. Si cuentas con datos teóricos no olvides obtener el error producido. Esta información es relevante para justificar sus conclusiones.

Cuando construyas tablas no olvides:

1. Título de la tabla

2. Distinción clara de celdas.

3. Títulos de columnas.

4. Notas al pie de la tabla

5. Numeración de tablas

**V. CONCLUSIONES**

Las conclusiones en un reporte deberán recoger los puntos primordiales del trabajo realizado, que se deduce a partir de los resultados obtenidos, de los pasos seguidos. Son parte importante del informe.

**VI. BIBLIOGRAFÍA**

En esta sección se anotan todas las fuentes bibliógrafas consultadas para el desarrollo de la práctica. Siguiendo el formato de los ejemplos a continuación.

1.Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197. doi:10.1016/0022-2836(81)90087-5

2. May P, Ehrlich H-C, Steinke T (2006) ZIB structure prediction pipeline: composing a complex biological workflow through web services. In: Nagel WE, Walter WV, Lehner W

(eds) Euro-Par 2006. LNCS, vol 4128. Springer, Heidelberg, pp 1148–1158. doi:10.1007/11823285\_121

3. Foster I, Kesselman C (1999) The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco

4. Czajkowski K, Fitzgerald S, Foster I, Kesselman C (2001) Grid information services for

distributed resource sharing. In: 10th IEEE international symposium on high performance

distributed computing. IEEE Press, New York, pp 181–184. doi:10.1109/HPDC.2001.945188

5. Foster I, Kesselman C, Nick J, Tuecke S (2002) The physiology of the grid: an open grid

services architecture for distributed systems integration. Technical report, Global Grid Forum

6. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov