

Trabajo Práctico Nº 2

Tema: Introducción – 2da parte

Fecha Inicio: 09/04/2024 **Fecha de Entrega:** 23/04/2024

Actividades:

El texto a continuación pertenece a la introducción de un trabajo (**Texto1**):

Los sistemas recomendadores son herramientas enfocadas a ayudar a los usuarios a obtener aquella información que mejor se corresponda con sus intereses y preferencias. Mientras que un buscador habitual se centra en encontrar aquello que el usuario solicita, un sistema recomendador ayuda al usuario a tomar una decisión, que puede ser la compra de un producto en un portal de comercio electrónico, la lectura de un libro, la revisión de un artículo científico, el acceso a una página web en específico, o el estudio de determinado recurso educativo en una plataforma virtual de aprendizaje.

La clasificación más popular de los sistemas recomendadores está asociada al algoritmo que emplean para realizar la tarea de minería correspondiente y divide a los métodos de recomendación en métodos de filtrado basado en el contenido, métodos de filtrado colaborativo, métodos de filtrado demográfico y métodos híbridos [1-4]. En adición, la literatura ha desarrollado tanto sistemas recomendadores para la sugerencia de ítems para usuarios individuales, como enfocados en grupos de usuarios []. Así, los sistemas recomendadores enfocados en grupos de usuarios se centran en la sugerencia de determinados tipos de ítems que tienden a ser consumidos en grupos y no por usuarios individuales, tales como programas de televisión y paquetes turísticos [].

De manera general, los dominios iniciales de aplicación de los sistemas recomendadores han sido el e-commerce[] y el e-learning[,], aunque en los últimos tiempos estos sistemas están siendo aplicados a escenarios cada vez más diversos []. Así, son relevantes las aplicaciones de los sistemas recomendadores en escenarios de e-health[] y de e-tourism[], como dos contextos relevantes de particular importancia.

Específicamente, resulta importante en los últimos años el desarrollo de sistemas recomendadores en el dominio del turismo[]. En este dominio existe mucha información en línea disponible y por tanto los sistemas recomendadores juegan un papel muy importante con vistas a ayudar a los usuarios en la toma de decisiones sobre qué paquete turístico comprar, qué instalación hotelera visitar, o qué recorrido turístico elegir, entre otras decisiones similares a tomar con vistas a lograr la satisfacción final del cliente [].

- 1) Empleando la librería [NLTK](#) de Python, elimine las *stop_words* empleando el idioma español, *tokenize* el texto anterior, y muestre el resultado con la frecuencia de cada término, ordenado por frecuencia descendente (además del listado, muestre un gráfico con los 20 términos/tokens más frecuentes). Tenga en cuenta el hecho de que existen oraciones y

párrafos, asegúrese de que el conjunto de términos está bien formado y no se omiten oraciones.

- 2) Del texto a continuación, aplique el proceso de eliminación de *stop_words* en inglés y *tokenización*, a continuación emplee el proceso de *Stemming* con los algoritmos de *Porter* y *Lancaster*, comparando los resultados de los dos procesos encolumnados(**Texto 2**):

Based on historical behaviors, sequential recommendation endeavors to predict what a user prefers next. The recent efforts are mainly devoted to modeling the user's interests evolution process or mining multi-interests for recommendation. However, it is largely overlooked that the interest trend (i.e., the evolution of the main interest) and the interest diversity (i.e., the scattered potential interests) could complement each other for better performance. Specifically, the interest trend reveals the user's basic interest and its evolution, which is satisfied by similarity recommendations. Nevertheless, interest diversity covers the various interests caused by some external environmental influence, e.g., fashion trends and advertisements, exploring users' potential interests or interest diversity will facilitate the model for diversity and serendipity recommendation. In a way, these two factors have conflicting aims, we argue that they should be disentangled in modeling first and recombined when making personalized recommendation.

- 3) Aplique el proceso de *Stemming* para el **Texto 1** y muestre el resultado. Advierta si los algoritmos de *Porter* y *Lancaster* en NLTK poseen la implementación para el idioma español, sino es así, aplique otro algoritmo que si la posea.
- 4) Del primer párrafo del **Texto 1**, obtenga 2-gramas y 3-gramas de palabras, muestre los resultados en cada caso.