

Grado en Ingeniería Informática

Gestión del Conocimiento en las Organizaciones

Sistemas de recomendación. Modelos basados en el contenido

Objetivos

El objetivo de esta práctica es implementar un sistema de recomendación siguiendo el modelo basados en el contenido.

Pasos

- Estudiar los modelos basados en el contenido. En las diapositivas de clase hay ejemplos sobre este campo.
- Crear un software que reciba lo siguiente:
 - Conjunto de ficheros de texto plano con extensión `.txt`. Algunos ejemplos se encuentran [aquí](#).
 - Fichero con palabras de parada, *stop words*, a descartar durante el proceso de recomendación. Ejemplos de ficheros con estas palabras se pueden encontrar [aquí](#).
 - Fichero de lematización de términos. Ejemplos de ficheros de lematización se pueden encontrar [aquí](#).

El software a desarrollar puede ser:

1. Una aplicación web realizada con HTML, CSS y JavaScript. Opcionalmente puede utilizar algún framework de su elección: [Vue.js](#), [React](#), [Svelte](#), [Angular](#), etc.
2. Una aplicación por línea de comandos. El programa debe recibir las entradas siguiendo el estilo POSIX. Puedes encontrar una descripción de las guías POSIX para el paso de argumentos a programas mediante la línea de comandos en [este enlace](#).

El software debe proporcionar como salida lo siguiente:

- Para cada documento, tabla con las siguientes columnas:
 - Índice del término.
 - Término.
 - TF.
 - IDF.
 - TF-IDF.



- Similaridad coseno entre cada par de documentos.

Entrega

Enlace a repositorio de [GitHub](#) donde se incluya al menos lo siguiente:

- Readme .md donde se recoja:
 - Instrucciones de instalación de dependencias, despliegue, etc. del software creado.
 - Descripción del código desarrollado.
 - Ejemplo de uso.
- Código fuente desarrollado.
- Un informe en PDF describiendo el análisis realizado y las conclusiones extraídas. En <https://github.com/ull-cs/gestion-conocimiento/tree/main/recommeder-systems/examples-documents> se pueden encontrar diversos ejemplos de ficheros de documentos que pueden ser empleadas.
 - Muestre los resultados obtenidos en los documentos disponibles en el citado repositorio.
 - Al mismo tiempo, proponga un conjunto de 10 documentos nuevos e ilustre cómo funciona el software creado sobre ellos.

Observaciones

- La práctica puede ser realizada por grupos de hasta 4 personas. Si éste es su caso, incluya el nombre de los integrantes del grupo en la entrega y realicen el envío todos a través de la actividad habilitada en el campus virtual de la asignatura.
- Si el repositorio de [GitHub](#) es privado invite a cexposit@ull.edu.es para poder consultarlo.