

Practica 5

Instrucciones

1. Utilice los archivos pre-procesados que obtuvo en la práctica 3, mismos que almacenan la colección CACM y los queries.
2. Si en la práctica tres generó el vocabulario sin ordenarlo alfabéticamente ordénelo, tanto el normal como el reducido.
3. Represente cada documento empleando el modelo vectorial y el esquema de pesado tf. Recuerde que la dimensión de los vectores será igual al tamaño del vocabulario que obtuvo en la práctica 4. Trabajen con el vocabulario sin reducir y después reducido. Guarda las matrices generada, que representa los documentos de la colección, para utilizarla en prácticas posteriores.
4. Utiliza el archivo de queries y represéntelos también empleando el modelo vectorial y esquema de pesado tf. Con vocabulario normal y reducido. Guarda las matrices generadas.
5. Repita los pasos 3 y 4 con esquema de pesado tf-idf.
6. Coloque en Teams el reporte correspondiente. En la sección de resultados muestre las pantallas donde se aprecie una término(palabra). Recuerden que el vocabulario deberá estar ordenado alfabéticamente. Busquen la palabra elegida en un documento(pre-pocesado) en el que aparezca más de una vez. En su matriz tf, muestre el renglón correspondiente al documento, donde deberá aparecer con la frecuencia correspondiente.
7. Coloque en Teams el código escrito.