

Laboratorio 4

Instrucciones

1. Utilice el archivo generado en el laboratorio 3 con los documentos de la colección de CACM preprocesados.
2. Obtengan el vocabulario de la colección utilizando truncamiento. Impriman la longitud del vocabulario. Después de observar el vocabulario resultante, proponga mecanismos para reducirlo. Codifique los mecanismos propuestos. Escriba la longitud del vocabulario reducido. Las longitudes deberán escribirse a pantalla
3. Escriban los vocabularios finales a archivos `vocabulario.txt` y `vocabulario_reducido.txt`
4. Documenten su código.
5. Suban a Teams el código y reporte de su práctica en el formato establecido. Reporte las longitudes obtenidas y mecanismos que utilizó para reducir el vocabulario.