



CURSO: CMP 0575 - TÓPICOS 2 (DATA MINING)

COLEGIO: POLITÉCNICO

Semestre: Primer Semestre 2018/2019 NRC: 1068

Tarea 3: Ejercicio usando la técnica *TextMining*

Problema:

1. Se tiene una colección de documentos a los cuales se le quiere hacer consultas por términos (o conjunto de términos). Confeccione un programa para listar de forma ordenada los documentos que mejor ajustan la consulta realizada (similitud entre el *query* y la colección).

Para dar solución a este ejercicio, se considerará lo siguiente:

- La colección tendrá un máximo de 10 documentos escritos en Inglés.
- Cada documento tendrá al menos 300 palabras distribuidas de forma tal que las *stop words* o *plurals* ocupen menos del 60 % (no más de 180 *stop words* o *plurals* por cada 300 palabras en el documento).
- Es obligatorio mostrar la trazabilidad (por cada fase) de la técnica *TextMining* durante la ejecución del programa.
- Cargar al D2L los códigos implementados (fichero compactado que incluye el ejecutable ej: el .JAR de java) dentro del plazo de entrega.