

# Informe TP2

## *Programa para procesar tweets*

Funcionamiento:

Se utiliza un count-min sketch para aproximar la cantidad de apariciones de  $n$  líneas. Luego se procede a utilizar un heap de mínimos para obtener los  $k$  tags de mayor aparición en esas  $n$  líneas (teniendo en cuenta las apariciones en los bloques de  $n$  líneas anteriores). Se repite esto hasta que se termina el archivo.

Tiempo:

Operación (Para un segmento de $n$ líneas)	Tiempo $N \rightarrow$ Líneas a leer   $K \rightarrow$ Tags a imprimir
Leer $n$ líneas	$O(n)$
Count-min sketch	$O(n)$
Obtener $k$ mayores (heap)	$O(n \log(k))$
Imprimir $k$ mayores	$O(\log(k))$

Total:  $O(n + n + n \log(k) + \log(k)) \rightarrow O(2n + n \log(k) + \log(k)) \rightarrow O(n \log(k))$

## *Programa para procesar usuarios*

Funcionamiento:

Se guarda en un diccionario las líneas del archivo (clave: usuario, valor: cantidad de tags). Luego se invierte el diccionario (clave: cantidad de tags, valor: lista de usuarios). Finalmente se imprimen todos los usuarios para cada número de tags.

Operación (Para un segmento de $n$ líneas)	Tiempo $U \rightarrow$ Usuarios   $T \rightarrow$ Tags
Leer el archivo	$O(u)$
Pasar a diccionario	$O(u)$
Invertir el diccionario	$O(u)$
Imprimir el diccionario	$O(t)$

Total:  $O(u + u + u + t) \rightarrow O(3u + t) \rightarrow O(u + t)$