



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC2233 - Programación Avanzada
2° semestre 2015

Actividad 21

Manejo de Strings

Instrucciones

La agencia de publicidad “SolucionesSuper” quiere analizar los tweets realizados durante el partido Chile-Perú de la Copa América y así evaluar el impacto de las campañas publicitarias.

Tienen pocos recursos para invertir en una solución real, por lo que le han pedido a usted, alumno del fantástico curso de Programación Avanzada, que implemente un prototipo de análisis de palabras frecuentes. Específicamente quieren saber cuáles son las palabras más utilizadas durante ciertos eventos del partido. Además quieren saber cuántas veces se menciona a los jugadores, ya sea por un apodo, nombre, apellido o twitter.

Ellos le han entregado tres archivos:

- eventos.csv que tiene los minutos del partido en los cuáles ocurrió un evento (tarjeta, cambio, gol, ...)
- jugadores.json que tiene los nombres, apellidos, usuarios de twitters y apodos de los jugadores de ambos equipos
- tweets.csv que tiene la lista de tweets, la hora en que fueron realizados y el usuario

Se han definido las siguientes reglas:

- Una palabra es una cadena de caracteres de largo 4 o más
- Una palabra es considerada frecuente si se encuentra 5 o más veces
- Para cada evento solo debe analizar los tweets de 3 minutos antes y 5 minutos después del evento. Solo estos tweets deben ser considerados en las palabras frecuentes.

Requerimientos

- Obtener las palabras frecuentes para cada evento
- Obtener la frecuencia de la aparición de un jugador en los eventos a partir de su nombre, twitter y apodos
- El output del análisis es:

