



# Sistemas Distribuidos I (75.74)

## Stack Overflow Questions Analysis

TP2: Middleware y Coordinación de Procesos

### Docentes

- Pablo D. Roca
- Ezequiel Torres Feyuk
- Guido Albarello
- Ana Czarnitzki
- Cristian Raña



# Requerimientos Funcionales

- Se solicita un sistema distribuidos que procese un subconjunto de preguntas y respuestas de Stack Overflow
- La información será provista en archivos CSV
- El procesamiento de los datos debe brindar la siguiente información:
  - Porcentaje de respuestas con score mayor a 10 que posea un sentiment analysis negativo
  - Top 10 de usuarios según score total (preguntas + respuestas), que tengan un puntaje promedio de preguntas mayor a la media general y un puntaje promedio de respuestas mayor a la media general
  - Top 10 de tags con mayor score (incluye preguntas y respuestas) de cada año
- Como origen de datos se definen los archivos de ingreso registrados en
  - <https://www.kaggle.com/ezetowers/tp2-stackoverflow-data-analysis>



# Requerimientos No Funcionales

- Se permite el uso de librerías de procesamiento de sentimientos tales como Stanford CoreNLP (Java), TextBlob (Python), NLTK (Python)
- El sistema debe estar optimizado para entornos multicomputadoras
- El sistema debe ser invocado desde un nodo que transmite los datos a ser procesados
- Se debe soportar el escalamiento de los elementos de cómputo
- De ser necesaria una comunicación basada en grupos, se requiere la definición de un *middleware*
- El diseño debe permitir una fácil adaptación a otros datasets de stack overflow similares
- Debido a restricciones en el tiempo de implementación, se permite la construcción de un sistema acoplado al modelo de negocio. No es un requerimiento la creación de una plataforma de procesamiento de datos



Se espera del alumno:

- Empleo del tiempo de consultas en clase para resolver dudas y clarificar el negocio del sistema a construir previo a su diseño
- Exposición y verificación en clase de la arquitectura propuesta antes de iniciar su implementación
- Empleo del grupo de correos para realizar consultas que no pudieran ser resueltas en clase
- Consideración de prácticas distribuidas según lo estudiado en clase para elaborar una arquitectura flexible, escalable y robusta
- Aprobación del cuerpo docente para el uso de cualquier librería
- Demo del sistema en funcionamiento previamente ensayada



- Fecha de entrega:
  - 02/11/2021
- Fecha de re-entrega:
  - 16/11/2021
- Formato de entrega:
  - Demostración del sistema
  - Entrega digital mediante correo personal
  - Entrega impresa:
    - Carátula
    - Documento 4+1 *Views* o C4Model incluyendo al menos un diagrama de: robustez, despliegue, actividades, paquetes y secuencia.
    - Código impreso