

Práctica Introducción a Apache Kafka

Sentweet

El objetivo de la práctica es diseñar e implementar un sistema de análisis de sentimiento de comentarios en Twitter.

Nota: Se pueden tomar ideas para la realización del análisis de sentimiento de este post del blog de Hugging Face <https://huggingface.co/blog/sentiment-analysis-python>

Entregables

- **Código fuente**
 - Puede publicarse en un repositorio público de github o similar
 - Nota: el resto de entregables deberán incluirse como documentación dentro del proyecto (por ejemplo en la subcarpeta `/doc`)
- **Documento de diseño**
 - Diagrama de arquitectura de componentes
 - Recomendado utilizar iconos oficiales y/o una herramienta como Draw.io / Diagrams.net (<https://app.diagrams.net/>)
 - Explicación del por qué de los distintos componentes que conforman la solución, justificación de la elección de tecnología (si procede), tareas que realizan, interdependencias...
 - Otros diagramas explicativos: de arquitectura del sistema software, de flujo de peticiones... Ejemplos: <https://c4model.com/>
- **Manual de operación**
 - Instrucciones para desplegar todo el sistema desde cero a partir del repositorio de con el código fuente
 - En caso de que la configuración se haya de hacer por el operador, incluir documentación de configuración
- **Manual de uso**
 - Si el sistema es configurable por el usuario, documentar la configuración
 - Instrucciones detalladas para, una vez desplegado el sistema y éste se encuentre correctamente configurado, realizar las acciones manuales necesarias (que no se realicen automáticamente) para capturar y procesar/analizar la información
 - Instrucciones detalladas para, una vez capturada y procesada/analizada la información, obtener resultados agregados sobre los datos (proveer al menos 1 ejemplo de resultado agregado, como la obtención de la opinión promedio)

Requisitos mínimos

- Utilizar Kafka (y posiblemente otros servicios de su ecosistema)
- Captura de datos desde fichero
- Análisis del sentimiento en tiempo real (a medida que se capturan los datos)
- Mecanismo de consumo de la información (obtención de resultados agregados)

Requisitos opcionales (sin orden particular)

- Captura de datos desde Twitter
- Captura de datos en tiempo real
- Configuración sencilla (por ejemplo, credenciales en un *secret* y parámetros de captura de datos en un *config*)
- Interfaz de configuración
- Interfaz de consumo de resultados
- Análisis/procesamiento/enriquecimiento de datos distintos
- Capacidad de configurar los análisis (modularidad)
- Securitización del sistema
- Otros orígenes de datos
- Orígenes de datos configurables
- Tests (unitarios, integración, aceptación)
- One-click deployment (despliegue con un único comando)

Criterios de valoración

Se valorará positivamente

- Uso racional de tecnologías vistas
 - Decisiones de diseño razonadas en la documentación
- Escalabilidad de la solución
- Automatización de todo tipo de tareas
- Minimizar el soporte requerido para la operación de la solución
- Explicaciones de diseño de requisitos opcionales que no haya dado tiempo a implementar
- Modularidad de la solución

Se valorará negativamente

- Uso de otro software cuando exista un elemento apropiado en el ecosistema Kafka
- Ausencia de documentación solicitada (especialmente documento de diseño) y de razonamiento de las decisiones
- Volúmen innecesario de documentación (“meter paja para hacer bulto”)

No se valorará / no será tenido en cuenta

- Atractivo visual de la interfaz