

Contents

Introduccion	1
Organización del laboratorio	2
primera etapa individual	2
segunda etapa grupal	2
Objetivos del programa	2
Funcionalidades de la entrega individual	2
Funcionalidades de la entrega grupal	2
Requisitos propios de paradigma	3
Estructura de los informes	3
Estructura del informe individual	3
objetivo del informe	3
se recomienda que el informe conteste las siguientes preguntas . .	3
casos	3
implementación de ejemplo	3
inteligencia artificial,	3
Estructura del informe grupal	4
Puntos Extras	4
Entrega y Defensa	4
Fecha límite de entrega individual:	4
Fecha límite de entrega grupal:	4

Introduccion

En este laboratorio vamos a re-implementar el lector automático de feeds del Laboratorio 2 sobre una arquitectura distribuida, a través del framework Spark, utilizando como lenguaje de programación Java1.

vamos a aplicar diferentes aproximaciones para hacerlo:

programación asistida por inteligencia artificial, ejemplos de proyectos tipo o consultas a sitios de referencia.

Organización del laboratorio

primera etapa individual

cada miembro del equipo documentará un proceso diferente de desarrollo del proyecto, usando alguna de las siguientes estrategias:

- chatGPT u otra inteligencia artificial
- una inteligencia artificial específicamente orientada a código,
- algún traductor de código,
- código de ejemplo
- alguna notebook de colab, kaggle o alguna referencia del mismo proyecto Spark.
- búsquedas en Stack Overflow,

segunda etapa grupal

el conjunto del equipo producirá un documento en el que compararán los tres procesos,

compararán los tres códigos obtenidos y justificarán la elección de uno de los tres códigos como mejor opción para el proyecto.

Objetivos del programa

mismas funcionalidades que el programa del Laboratorio 2,

tiene que ofrecer la funcionalidad de recuperar documentos por palabras clave (como lo haría un motor de búsqueda en la web).

Funcionalidades de la entrega individual

Obtener diferentes feeds de noticias, conformando una colección de documentos, cada uno con una noticia recibida de algún feed.

Parsear cada uno de los documentos de la colección,

Imprimir las diferentes noticias por consola,

Imprimir el listado de entidades nombradas

Imprimir el número de ocurrencias de cada entidad nombrada, y de cada clase y subclase de entidad nombrada,

Funcionalidades de la entrega grupal

Recuperar los documentos que contengan una determinada palabra o entidad nombrada,

ordenados desde el documento que tiene la mayor cantidad de ocurrencias de la misma al documento que tiene la menor cantidad de ocurrencias.

Requisitos propios de paradigma

Es imprescindible que usen las funciones map y reduce provistas por el framework

Es imprescindible que usen las funcionalidades propias de Spark para las consultas interactivas.

Estructura de los informes

Estructura del informe individual

objetivo del informe

es que describan las diferentes etapas para pasar del código de lectura de feeds y conteo de entidades nombradas al código en el que esta funcionalidad está implementada sobre Spark.

se recomienda que el informe conteste las siguientes preguntas

¿Cómo se instala Spark en una computadora personal?

¿Qué estructura tiene un programa en Spark?

¿Qué estructura tiene un programa de conteo de palabras en diferentes documentos en Spark?

¿Cómo adaptar el código del Laboratorio 2 a la estructura del programa objetivo en Spark?

¿Cómo se integra una estructura orientada a objetos con la estructura funcional de map-reduce?

casos

implementación de ejemplo

identificar las partes del código de ejemplo que cubren las respuestas a cada una de las preguntas, y pegarlas como parte de la respuesta.

inteligencia artificial,

deberán pegar como parte de la respuesta capturas de pantalla de sus interacciones con el sistema predictivo o inteligencia artificial.

Estructura del informe grupal

El objetivo del informe grupal es que describan las ventajas y desventajas de las diferentes soluciones individuales, y los motivos por los que eligen una de ellas.

También deberán describir la forma en que se incorpora la funcionalidad de recuperar documentos por palabras clave,

Puntos Extras

Integrar en estructura map-reduce también la obtención de feeds.

Usar una librería específica para hacer un histograma de la frecuencia de todas las palabras de los feeds, incluyendo las entidades nombradas.

Entrega y Defensa

Fecha límite de entrega individual:

Jueves 1 de Junio de 2023 a las 23:59:59 hs.

Entrega del código fuente + informe mediante push en bitbucket en el repositorio asignado por la cátedra al estudiante.

Fecha límite de entrega grupal:

Jueves 15 de Junio de 2023 a las 23:59:59 hs.

Entrega del código fuente + informe mediante push en bitbucket en el repositorio asignado por la cátedra al grupo.