

Prueba práctica JAVA

Misión I

APLICACIONES EN INFORMÁTICA AVANZADA

GRUPO **ATA**

ALGORITMOS PARA UN MUNDO MEJOR



APLICACIONES EN
INFORMÁTICA
AVANZADA

BIENVENIDO A AIA

En grupo AIA tenemos una misión, resolver con calidad cualquier problema de nuestros clientes por pequeño que sea, por eso nos hacemos llamar **Problem Solvers**.

Nuestros equipos de Data Scientists y Técnico están trabajando en un proyecto en el que necesitan extraer información de Wikipedia.

Necesitamos tu ayuda para completar esta misión con éxito...



APLICACIONES EN
INFORMÁTICA
AVANZADA

LA MISIÓN

Crear una funcionalidad que implemente y ejecute un proceso de E&T (Extracción y Transformación) minimizando el tiempo de ejecución tanto como sea posible.

Detalles:

1. Leer y cargar todos los nombres del archivo `first10000.txt`, adjunto (omitir la primera línea, es un header)

2. Por cada nombre en el archivo (row) tomar la correspondiente página de Wikipedia siguiendo la siguiente construcción de URL:

`https://en.wikipedia.org/wiki/" + NOMBRE`

De tal forma que, si el primer nombre que hay en el fichero es "GRUP0AIA" la URL sería:

`https://en.wikipedia.org/wiki/ GRUP0AIA`

3. Por cada página de Wikipedia, extraer y analizar:

- Título de la página
- Última fecha de modificación (la fecha de modificación se encuentra al final de cada página de Wikipedia)

4. Generar un archivo CSV que contenga las siguientes columnas:



APLICACIONES EN
INFORMÁTICA
AVANZADA

LA MISIÓN

- `url,name,title,lastModification`

dónde:

url -> es la url de la página de Wikipedia

name -> nombre tal y como aparece en el fichero `first10000.txt`

title -> Título de la página de Wikipedia

lastModification -> Última modificación de la página de Wikipedia

¿Que esperamos?

- Artefacto Maven con código testeado y documentado.
- Un test con ejemplo de ejecución.
- Responde a la pregunta: ¿cuántas páginas de Wikipedia fueron modificadas durante este año?

Información Adicional

- El lenguaje de programación deberá ser JAVA.
- No importa tanto el tiempo en hacerlo como la calidad de código y el testeo del mismo.

Resultado

El resultado de la misión la puedes enviar a Óscar Mejías (mejiaso@aia.es) cuando lo acabes. Nuestra intención no es que lo completes en el mínimo tiempo posible, sino que lo puedas hacer con calidad, por eso tómate el tiempo que creas conveniente.