

Rotten Tomatoes

9 de mayo de 2024

1. DESCRIPCIÓN

El objetivo primario de este proyecto es el desarrollo habilidades primaras de investigación formal en ciencias de la computación, tales como:

- Desarrollar conocimiento acerca de **encontrar y utilizar fuentes bibliográficas** diversas.
- Desarrollar **pensamiento critico** y soporte a argumentos mediante evidencia.
- Mejorar sus **habilidades de sintetización** y desarrollo de puntos de vista informados.
- Crear un **artículo de investigación** organizado y conciso.

Para esto, el estudiante debe:

- Desarrollar una **propuesta de solución a un problema de investigación**, mediante estudio del estado del arte, argumentación del tópico, proposición de teoría y solución.
- Soportar todas sus afirmaciones mediante **investigación de literatura relacionada** y relevante.
- Utilizar **practicass generalizadas en la comunidad científica** de ciencias de la computación, tales como uso de lenguajes de marcación, plantillas estandarizadas y sistemas de citación bibliográfica.

2. DESCRIPCIÓN DEL PROBLEMA

En términos comerciales, los sistemas de clasificación son una de las aplicaciones más populares de Inteligencia Artificial, especialmente por la **relativa facilidad de implementación** en relación a otras aplicaciones.

Como estudiante de Inteligencia Artificial, ud. tendrá la oportunidad de demostrar sus habilidades utilizando una captura de las reviews de películas de críticos de cine de la plataforma Rotten Tomatoes.

En este data set publicado por Kaggle se presentan miles de reviews de películas publicadas en Rotten Tomatoes, y se requiere que ud. como estudiante de inteligencia artificial realice las siguientes tareas:

- Normalizar los datos que actualmente se encuentran en formato csv (la organización de los mismos queda a su discreción).
- Elaborar un motor de clasificación simple que tome en cuenta las siguientes características:
 - Arranque en frío y sin interacciones del usuario
 - Registrar la calificación y review de una película por parte del usuario.
 - Identificar basado en los datos si se califica como "Fresh." o "Rotten".

3. ENTREGABLES

Este proyecto se divide en tres entregables:

- GUI, carga y normalización de datos
- Motor de clasificación
- Artículo de investigación

3.1. GUI, CARGA Y NORMALIZACIÓN DE DATOS

El programa a implementar debe permitir cargar las reviews o cualquier otro archivo con la misma estructura.

Posteriormente debe facilitar:

1- la creación de usuarios y su perfil inicial, registrando a que publisher pertenecen, dado los publishers definidos en la carga inicial.

2- definición de lista de películas dada la carga inicial, se debe usar la referencia del archivo, por ejemplo, m/0814255, esta es una referencia al website de rotten tomatoes <https://www.rottentomatoes.com/m/0814255>.

3- Se debe registrar la valoración y review de las películas por parte del usuario, seleccionando alguna de las películas listadas.

4- Invocar al motor de clasificación basado en la valoración y review se debe indicar si es Rotten o Fresh.

Esta aplicación debe ser implementada utilizando algún lenguaje compatible con la máquina virtual de Java (Java, Kotlin, Clojure, Scala, etc.).

Dado que el objetivo del proyecto es la implementación de un motor de clasificación, queda a discreción del estudiante la forma adecuada para la interacción con la aplicación (shell, TUI, GUI Web, GUI Desktop).

3.2. MOTOR DE CLASIFICACIÓN

El motor de clasificación debe ser una implementación propia en lenguaje Python 3, por lo que únicamente se permite que utilice Python stdlib y bibliotecas básicas de análisis y manipulación de datos como NumPy o Pandas. Puede utilizar cualquier técnica de recomendación vista en clase.

Su aplicación debe enviar y recibir recomendaciones y datos hacia el motor de recomendación en Python, se sugiere alguno de los siguientes tres abordajes:

- Comunicación vía REST(Http)
- Invocación del programa en Python desde la JVM mediante procesos de sistema operativo -e.g. ProcessBuilder-
- Uso de una JVM polígota con soporte a Java y Python -e.g. GraalVM-

Si se considera necesario, el grupo de trabajo puede utilizar una base de datos relacional, NoSQL u orientada a grafos.

3.3. ARTICULO DE INVESTIGACIÓN

Para justificar, demostrar y presentar su solución, debe elaborar un artículo científico con un máximo de 10 páginas en formato ACM.

El artículo debe demostrar una extensa investigación del estado del arte, soluciones similares, justificación de los algoritmos a ser utilizados en el motor, argumentación y descripción técnica de su solución, así como resultados obtenidos mediante la creación del motor.

4. DESCRIPCIÓN DEL PROBLEMA

El proyecto debe elaborarse en grupos de 5 personas. Los grupos presentarán un seminario con sus resultados el día posterior a la entrega del proyecto en el cual se evaluará su funcionalidad, así como preguntas individuales para la sección "seminario" del curso.

Fecha de entrega: 30 de Mayo de 2024, 23:55