

Proyecto Investigativo

Sistemas de Recuperación de Información

Curso: 2024

Semestre: 2do

El proyecto tiene como objetivo brindar experiencias prácticas en la resolución de problemas novedosos y prácticos de recuperación de información, el cual consistirá en la implementación de una o más soluciones propias de cada estudiante o equipo de estudiantes sobre el tema seleccionado.

A continuación se describen por secciones aspectos importantes del proyecto.

Información general

Cuando el estudiantes decida comenzar a trabajar en la solución del proyecto, debe de seguir ciertos pasos. Estos son:

1. Leer detenidamente este documento.
2. Formar un equipo con otros estudiantes.
3. Seleccionar el tema a desarrollar.
4. Buscar trabajos e investigaciones actuales vinculados al problema seleccionado.
5. Trabajar sobre la solución propuesta del equipo.
6. Construir un reporte.
7. Presentar y discutir el trabajo.

La calificación final estará basada según los componentes descritos a continuación:

- Implementación de la propuesta (30 %)
- Reporte escrito (30 %)
- Discusión del trabajo (40 %)
 - 30 minutos de exposición de los estudiantes, el cual incluye una demostración del proyecto funcionando para explicar el problema abordado y la solución propuesta.
 - 10 minutos para responder las preguntas del profesor.

Restricciones programáticas

No existe limitante alguna.

Puede usarse cualquier lenguaje de programación, aunque se recomienda utilizar Python puesto que este lenguaje tiene un conjunto de bibliotecas especializadas para el tratamiento, manejo y análisis de datos. También el uso de toda biblioteca/librería o de APIs es permitido, siempre que el desarrollador conozca sobre el funcionamiento que existe por “detrás” en las herramientas externas al código propio.

Confección de los equipos

Se permitirá equipos de a lo sumo 3 personas sin importar el grupo al que pertenezcan, siendo válido los equipos confeccionados de un integrante.

Tenga en cuenta que aunque el desarrollo del proyecto es en equipo, la nota final no tiene porqué ser pareja para todos los integrantes.

Temas de proyecto

El tema a trabajar es decisión de cada equipo. Lea cada tema descrito y analice el “mejor” para usted. La condición “mejor” puede darse a partir de:

- el tema es cercano a trabajos anteriores,
- le llama la atención la temática,
- considera que el problema a abordar aportará a su perfil profesional deseado, o
- simplemente tomó uno de forma aleatoria y que el karma apoye la decisión.

Tenga en cuenta que el listado de los temas no está dado por complejidad u otra variable que cambie a medida que avanza en la lectura. Simplemente es un listado sin orden.

Los temas posibles son:

1. Técnicas híbridas en Sistemas de Recomendación

Los Sistemas de Recomendación son extremadamente utilizados en las sitios virtuales donde confluyen los usuarios. Estos sistemas están bien definidos y cada uno tiene técnicas particulares.

Se requiere la desarrollo de un sistema que implemente una hibridación entre las técnicas existentes de los Sistemas de Recomendación, de forma que se que se complementen en pos a mejorar la información recomendada a cada usuario.

2. Recuperación justa de la información

Los Sistemas de Recuperación de Información proponen un modelo y una función de similitud entre la consulta y los datos pre-procesados del sistema. Luego, el sistema devuelve de forma ordenada la información que responde a la consulta.

Se desea encontrar un modelo y/o métricas que filtren los datos y solo devuelva aquella información que realmente responda a la consulta. Para ello, es necesario implementar un sistema de acuerdo a las necesidades descritas en la oración anterior.

3. Personalización explicable

Los sistemas actuales son muy precisos al brindarle al usuario lo que realmente espera encontrar, pero no existe una transparencia en ese proceso de personalización.

Se desea crear un sistema donde el usuario de la opción de consultar sus decisiones y selecciones a través de una traza y cómo el sistema ha usado esa información. Además, a partir de una búsqueda, el usuario debe de ser capaz de indicar lo que le fue útil, selección que sirve como personalización a futuras búsquedas.

4. Clasificación de reseñas para mejorar la toma de decisiones del comprador

El contenido de las reseñas generado por un usuario es útil para que los compradores tomen decisiones más informadas en sitios web de compras en línea. Sin embargo, debido al gran volumen de reseñas, especialmente de artículos populares, es muy difícil para cualquiera encontrar fácilmente información relevante. A menudo, el sitio web sólo proporciona algunos métodos de clasificación simples, pero no muy útiles, como clasificación por tiempo o calificación por votos de utilidad. Se pueden considerar varios factores para este propósito, incluido el historial de compras reciente del usuario y el de navegación y las propiedades de los artículos.

Proponga un sistema que haga uso de lo antes mencionado, junto a otras características que considere necesarias, para que los compradores sugieran mejores compras a los usuarios.

5. Recomendación secuencial colaborativa

Los algoritmos tradicionales de recomendación adoptan una visión estática del interés de los usuarios, es decir, en un período de tiempo determinado se recomendará repetidamente a los usuarios el mismo conjunto de elementos, ya que se cree que son más relevantes para los usuarios. Esto es claramente subóptimo, por ejemplo: cuando un usuario ha comprado una computadora portátil, no tiene sentido seguir recomendando otra computadora portátil.

Se desea un sistema con un enfoque para recomendar de forma secuencial el siguiente elemento, el cual es dependiente a la secuencia de acción de los usuarios. Esto le aporta al algoritmo una total flexibilidad para ser contextual y adaptable.

Cuidado al elegir

El trabajo no tiene como objetivo reproducir soluciones existentes a los problemas planteados, sino que cada equipo plantee nuevas soluciones o variaciones de la existentes.

A continuación se plantean un conjunto de preguntas a tener en cuenta a la hora de seleccionar el tema a trabajar y la solución a plantear por cada equipo. Sean estas:

- ¿Cuál es exactamente el problema que deseo trabajar?
- ¿El tema está resuelto ya? De ser así, ¿dónde puedo encontrar la solución? ¿La novedad que propondré generará algún beneficio a la solución existente?
- ¿Qué tipo de cambio puedo ofrecer al encontrado en el estado del arte?
- ¿En qué se diferencia mi propuesta a las existentes?
- ¿Cuáles serían los principales desafíos al problema que enfrento? ¿Tiene algún antecedente o recurso específico para resolver el problema identificado?
- ¿Cómo planeo demostrar que mi propuesta realmente es válida?

Referente a la implementación

Como buenas prácticas, se recomienda algunos aspectos a tener en cuenta en el código a presentar. Estas prácticas sentarán bases para su futuro y se temarán en cuenta a la hora de evaluar este aspecto. Luego, se proponen:

- Usar la herramienta Git y tener un repositorio del proyecto en la plataforma GitHub.
- Contar con al menos dos ramas dentro del repositorio: **develop** para el desarrollo progresivo del proyecto y, **main** para tener la entrega final del proyecto en el momento de la exposición.
- Tener un código limpio, documentado (*docstring*), organizado por carpetas lógicas y desacoplado por funcionalidades. Para la correcta definición del *docstring* puede tener como referencia a la usado en los laboratorios o referirse a <https://peps.python.org/pep-0257/>.
- Tener ejemplos que muestren y demuestren la(s) solución(es) propuesta(s), junto a las métricas establecidas.
- Definir los ficheros:
 - **startup.sh** para ejecutar el proyecto desde ese fichero.
 - **README.md** para mostrar una breve descripción sobre
 - los autores;
 - el problema;
 - los requerimientos, a groso modo, para el correcto desempeño del proyecto;
 - APIs utilizadas, en caso de usar alguna; y
 - la forma en que se usa o ejecuta el proyecto.
 - **requirements.txt** para la definicieon de todas
- De ser posible, toda la implementación tiene que estar dentro de la carpeta **src** y dejar solo en la raíz del repositorio esta carpeta, el informe y los ficheros mencionados en el punto anterior.

Sobre el informe

“Un buen proyecto tiene asociado un buen reporte”, motivo suficiente para exigir la confección de uno. Como no es común hasta la fecha la buena redacción de los informes con la calidad que merita, acá se describen ciertos aspectos que facilitará la redacción, aspectos que se evaluarán además.

No use los puntos descritos a continuación como una plantilla estática; solo se describen los puntos imprescindibles a tratar. Sea coherente al redactar, de forma que aborde cada aspecto mencionado y otros si los considera necesarios.

Los aspectos son:

- Buena ortografía.
- Redacción en tercera persona del singular y como tiempo verbal el pasado. Por ejemplo, no se escribe “Yo consideraré ...”, “Yo considero ...” o “Nosotros consideramos ...”, sino “Se consideró ...”.
- Redacción objetiva, concisa y formal.
- Debe de abordar al menos los siguientes aspectos:
 - Url del proyecto donde se encuentra alojado el proyecto dentro de la plataforma GitHub.
 - Autores.
 - Descripción del tema.
 - Antecedentes de la temática seleccionada.
 - Explicación de la(s) solución(es) implementada(s).
 - Consideraciones tomadas a la hora de implementar la propuesta.
 - Evaluación cuantitativa y cualitativa del trabajo.
 - Declaración autocrítica de las insuficiencias de la implementación llevada a cabo y propuestas de cómo mejorarlas.
No afectará de forma negativa a la nota, todo lo contrario, es importante reconocer las limitantes y el alcance del trabajo desarrollado por cada quien y la forma de solucionar las insuficiencias.
- Uso de la plantilla LNCS.
- Formato del documento a entregar: PDF.
- El informe tiene que estar en el repositorio. Puede estar en la raíz de este o dentro de otra carpeta ubicada en la raíz.
- Se recomienda hacer uso de un problema actual que sea “sencillo” para introducir el tema y ayudar al lector a percatarse de la importancia de lo que leerá.
- Referencias a las ideas no propias del equipo.
- Definición correcta de la bibliografía utilizada.

Otros aspectos

La selección de un conjunto de datos durante las pruebas y evaluaciones de los sistemas complejizan al desarrollador porque los buenos conjuntos de datos son privados, obligando a pagar para obtenerlos (opción decantada para este trabajo), los gratis generalmente son pobres en datos y, por raro que parezca pero es perfectamente común que pase, puede no existir un conjunto idóneo de datos para el problema que se desea resolver, obligando al investigador a generar datos para su problema.

Para la nota final del proyecto se considerará la generación de conjuntos de datos particulares a cada problema, por parte de cada equipo. Si el equipo desea generar los datos, debe de reflejar en el informe la forma en que lo crea y brindar ciertas estadísticas e información referente a los datos generados. Además, la creación del conjunto de datos no puede facilitar las métricas, o sea, los datos deben de ser suficientemente diversos.

Este aspecto no es necesario tratarlo, es solo un extra a considerar si tiene el resto de los requerimientos bien.