

Máster SIANI – Ciencia de Datos en Ingeniería

Trabajo de Curso 2023/2024

Recomendador de revistas científicas para enviar una publicación

Objetivo

En la actualidad el número de revistas científicas es del orden de miles, lo que supone que en un determinado campo puedan existir decenas de revistas. Por tanto, un dilema al que se enfrentan los investigadores es decidir a qué revista enviar un trabajo científico. Además de los factores de calidad, un elemento a tener en cuenta es que los trabajos publicados en dicha revista tengan una temática similar ya que aumenta el interés y visibilidad entre los otros investigadores del campo.

Por tanto, el objetivo que se plantea en este trabajo es realizar un proyecto de Ciencia de Datos con el objetivo de realizar un sistema inteligente que recomiende una revista científica para enviar un artículo para su publicación. De esta forma, el investigador dando el título del trabajo, el resumen y las palabras claves, el sistema inteligente le recomendará la revista a la que debería enviar el artículo. Este conocimiento lo habrá aprendido el recomendador a partir de ejemplos de trabajos enviados previamente a cada una de las revistas candidatas.

Para ello se harán uso de las técnicas vistas durante el curso relacionadas con preprocesamiento, obtención de modelos, evaluación de resultados, y más específicamente de procesamiento de lenguaje natural, así como otras que el estudiante considere necesarias.

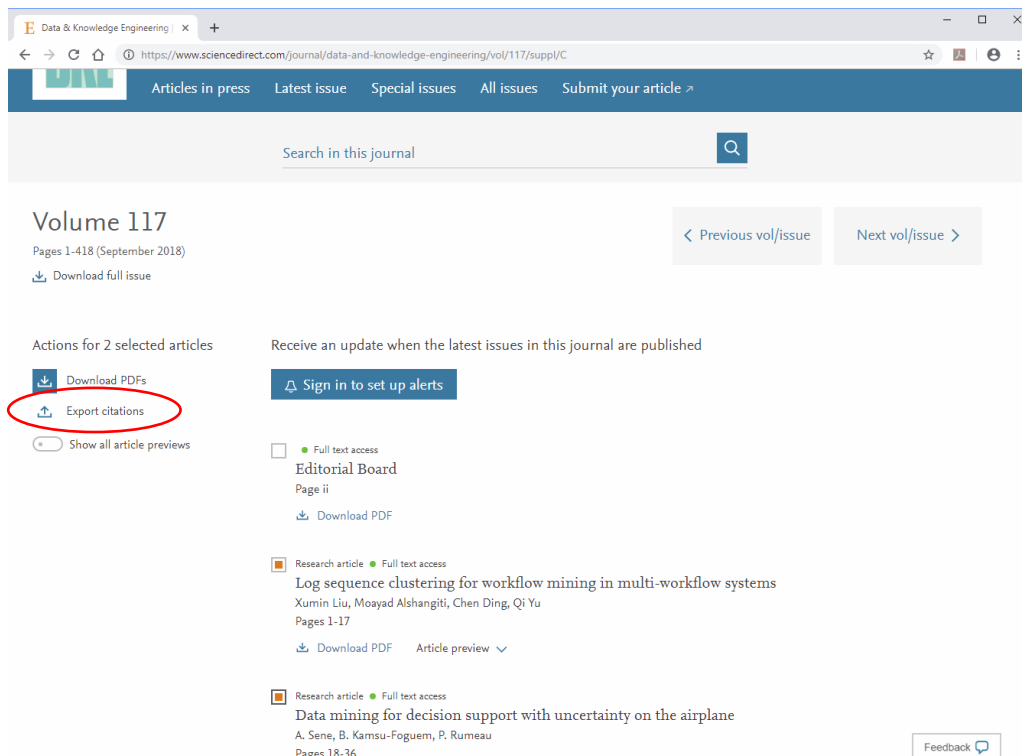
Como resultado del trabajo de curso se deberá redactar una memoria explicando el proceso seguido y justificando las decisiones de diseño tomadas, y un análisis y discusión de los resultados obtenidos. También se deberán entregar los programas realizados en Python para el desarrollo del trabajo y los conjuntos de datos utilizados, así como la presentación que servirá de guía para la defensa del trabajo.

Ámbito

Para la implementación del sistema inteligente se hará uso de un conjunto de revistas de la editorial Elsevier. El corpus o conjunto de documentos con el que se realizará el trabajo serán los artículos publicados los años 2018, 2019, 2020, 2021 y 2022 en las revistas indicadas, utilizando los siguientes datos: título, resumen (*abstract*) y palabras claves.

La obtención de estos datos se realiza a través del portal web de la editorial ScienceDirect (<https://www.sciencedirect.com/>) al que se puede acceder a través de la biblioteca de la universidad.

Para poder descargar los datos básicos antes comentados se puede realizar exportando todas las referencias de cada número a un formato RIS o Bibtex, incorporando el abstract.



Desarrollo

Como se ha indicado anteriormente, el trabajo de curso se realizará utilizando el lenguaje de programación Python. Para la fase de preprocesado del texto aparte del módulo `feature_extraction.text` de Sklearn, también se podrá considerar el paquete *Natural Language Toolkit* (NLTK) disponible como software libre, u otro similar. Para la obtención de los modelos se utilizará igualmente la librería Sklearn. Para la parte optativa se realizará utilizando el framework de desarrollo Keras o Pytorch. No obstante, lo anteriormente indicado no excluye la utilización de otros paquetes que el estudiante considere necesario para la realización del trabajo.

El desarrollo del trabajo consta de una **parte obligatoria** y una **parte optativa**.

- La parte obligatoria se basará en la aproximación clásica que, a partir de los datos obtenidos de las revistas, se obtendrá la matriz de términos-documentos. Una vez obtenida la matriz se procederá a entrenar y validar un/os modelo/s de clasificación de documentos. (80% de la calificación)
- La parte optativa se basará en la aplicación de técnicas conexionistas para la obtención del recomendador, siendo el estudiante el que proponga la solución que considere más adecuada bajo este paradigma. Al igual que en la parte obligatoria, se debe entrenar y validar el/los modelo/s para la clasificación de documentos. (20% de la calificación)

Entrega

La entrega del trabajo se realizará a través del campus virtual de la asignatura. Se deberá entregar en un fichero comprimido lo siguiente:

- Memoria descriptiva.
- Presentación en PowerPoint del trabajo realizado.
- Código en Python generado durante la realización del trabajo.
- Conjuntos de datos utilizados en la realización del trabajo.

Defensa

El trabajo deberá ser defendido en una sesión de tutoría cuya fecha será acordada entre el estudiante y el profesor.

Fechas Importantes

- **Entrega** a través del campus virtual (fecha límite): **30 de diciembre de 2023**
- **Defensa** (fecha límite): **30 de enero de 2024**

Revistas

1. Applied Ergonomics
2. Data & Knowledge Engineering
3. Expert Systems with Applications
4. Journal of Visual Communication and Image Representation
5. Pattern Recognition
6. Robotics and Autonomous Systems