

Proyecto: **ADR**

Ejecutores:

Rolando Sánchez Ramos C-311

David Manuel García C-311

Andry Rosquet Rodríguez C-311

1. INTRODUCCIÓN

Con este informe se tratará de brindar una breve y superficial explicación relacionada al proyecto ADR-Retrieval-System perteneciente a la asignatura System-Recovery-Information. Se abordarán los modelos implementados y algunas otras particularidades. En una futura actualización se explicarán las ideas con mayor profundidad.

2. DESARROLLO

Los modelos implementados para nuestro sistema de recuperación de información son el Booleano, Vectorial y Vectorial Generalizado (el resumen de estos modelos se presentará en el informe definitivo). La implementación de estos se encuentra en el directorio llamado `retrieval_models`, cada uno de ellos se encuentra en una carpeta con su propio nombre. Cada modelo se encuentra dividido en tres archivos cuyos formatos son (NOMBRE será el específico de cada modelo):

- `NOMBRE_document`: define una clase `NOMBREDocument` la cual se encarga de crear el vector del documento deseado con un vocabulario ya definido. Para cada modelo el vector se forma de manera específica, ejemplo: el booleano crea un vector binario pero los vectoriales forman uno basado en la frecuencia de cada término.
- `NOMBRE_query`: define una clase `NOMBREQuery` que mediante sus métodos y propiedades específicas le otorga un formato a la consulta para que sea posible aplicarle la función de similitud con un documento.
- `NOMBRE_model`: define una clase `NOMBREModel` la cual mediante sus métodos internos permite calcular la función de similitud entre un documento y la consulta. Recordar que en función del modelo, la función de similitud y operaciones necesarias serán o no diferentes.

Abordemos otros elementos importantes del proyecto, estos los encontramos en el directorio `utils`. En este se llevaron a cabo implementaciones como:

- El algoritmo de Rocchio que sirve para aplicarle retroalimentación a los modelos vectoriales y otorgarle importancia a la información brindada por unos documentos u otros. Este se encuentra implementado en `vector_feedback.py` donde además de el método `rochio_algorithm`, se implementó retroalimentación básica para modelos vectoriales en el método `classical_vector_feedback`.
- El archivo `query_expansion` como dice su nombre contiene la implementación de expansión de consultas basadas en sinónimos, utilizada mayormente para el modelo booleano.
- También se calculan las métricas basadas en la opinión de un experto, entre estas precisión, recobrado, medida-f, r-precisión y otras. Se pueden encontrar en `metrics.py`.