Simón Alonso Guitiérrez
Pilar Fierro Casanova

El zip entregado se encuentran dos ficheros con la implementación del código en lenguaje python: main.py y graficas.py junto con un fichero txt traditionalEvaluationResults014.txt con el resultado de ejecutar dicho código con el resultado de la práctica anterior.

En main.py se encuentra el código fuente para calcular las métricas y las escribe en un fichero .txt siguiendo el formato solicitado en el enunciado. Por otro lado gráficas.py genera una gráfica con las medidas totales recogidas en los documentos .txt generados por main.py

La capacidad del programa para obtener las medidas esperadas de los sistemas a y b.

El programa obtiene las siguientes métricas solicitadas: precision, recall, f1, prec@10, MAP, recall_precision, interpolated_recal_precision. Para ello se ha creado una clase cuyos atributos y funciones se encargan de almacenar los valores necesarios y realizar los cálculos respectivamente.

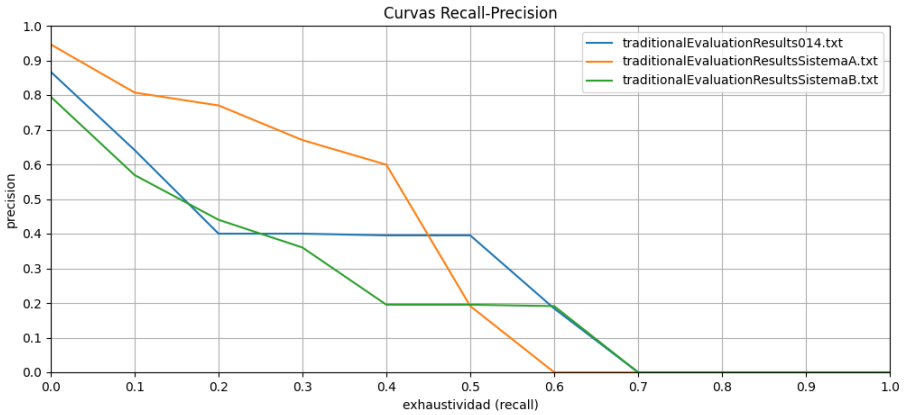
Una reflexión de la medida o medidas que se consideren mejor para reflejar la calidad de un sistema de recuperación.

Una de las medidas más importantes es el F1, ya que combina tanto la precisión como el recall permitiéndonos así tener una visión general de cómo de bueno es el recuperador de información que se ha implementado. Al unir ambas métricas, si el sistema implementado obtiene un F1 más alto que otro, significa que en promedio, el sistema se comportará mejor que el otro.

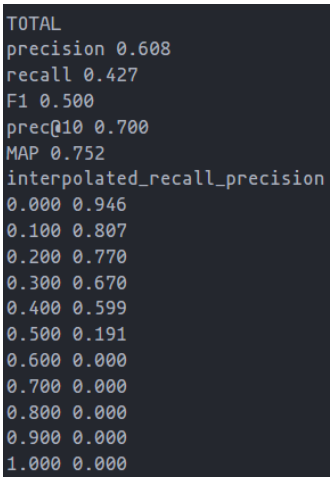
Otra de las medidas importantes es el MAP, ya que evalúa la precisión promedio de una necesidad de información. Esto no lo hace una sola vez, si no que tiene en cuenta dicha precisión promedio.

Comparación de las medidas de nuestro sistema frente a las obtenidas por los sistemas a y b.

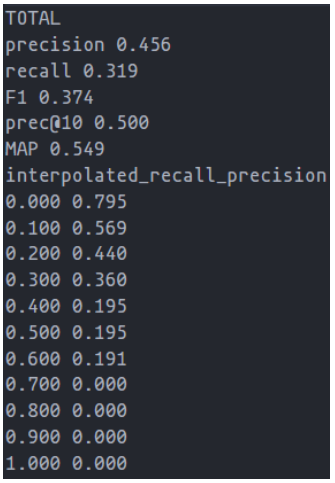
Cuando se emplean como datos de entrada a analizar, los resultados obtenidos en la práctica anterior se aprecia que difieren un poco con los datos esperados, pero en la gráfica que se muestra a continuación se aprecia que se encuentra entre ambas medidas. Siendo la línea azul nuestros datos de entrada, la línea naranja el juez A y línea verde el juez B.



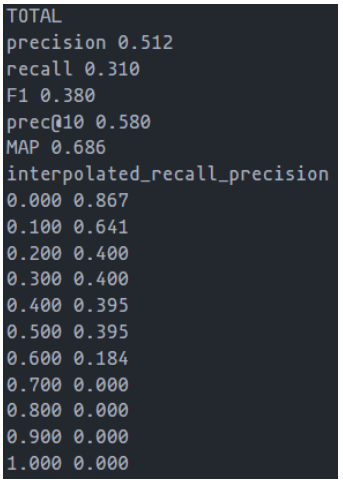
Aquí se muestran con mayor detalle:



Sistema A



Sistema B



Sistema práctica anterior

Como se muestra en las imágenes nuestro sistema se comporta peor que el sistema A pero mejor que el sistema B. Teniendo un F1 y MAP superior a este último.

Para obtener la gráfica mostrada anteriormente se ha de ejecutar lo siguiente:

```Python
python main.py -qrels zaguanRels.txt -results resultados_sistema_a.txt
-output traditionalEvaluationResultsSistemaA.txt

python main.py -qrels zaguanRels.txt -results resultados_sistema_b.txt
-output traditionalEvaluationResultsSistemaB.txt

python main.py -qrels zaguanRels.txt -results equipo014.txt -output
traditionalEvaluationResults014.txt

python graficas.py traditionalEvaluationResults014.txt
traditionalEvaluationResultsSistemaA.txt
traditionalEvaluationResultsSistemaB.txt
```

¿Qué decisiones de diseño en cuanto a campos indexados, elección de analizadores, modelo de ranking, y procesamiento de necesidades de información para la construcción de consultas justifican las medidas obtenidas por vuestro sistema?

Campos indexados: identifier, type, creator, contributor, publisher, title, description, dato, subject, language

Analizador seleccionado: utilizamos el SpanishAnalyzer02 que era una versión modificada en el que se incluyen stopwords para el lenguaje español.

Modelo de ranking: se le daba la misma importancia a todas las cláusulas y los documentos están ordenados por score descendentemente. Esto puede que nos haya perjudicado a la hora de recuperar los mejores documentos, ya que se podría haber especificado que algunas cláusulas tengan más importancia que otras.

Procesamiento de las necesidades de información: en la práctica anterior se realizó un recuperador de información que era más restrictivo de lo normal, ya que necesitábamos que se cumplieran todas las cláusulas de una necesidad de información concreta. Esto nos ha podido servir para justificar la precisión elevada que se ha obtenido, pero también nos ha podido perjudicar a la hora de recuperar todos los documentos necesarios en el caso de que alguna cláusula se haya escrito mal o sea un poco distinta a lo que nosotros esperábamos.

¿Existen otros factores externos que podrían haber influido en estos resultados?

Por ejemplo, que las necesidades de información no estén bien formuladas, o que estas se adecuen mejor a un recuperador de información que a otro.