

# Evaluación del Sistema de Recuperación de Información Tradicional

Jaime Ruiz-Borau Vizárraga  
546751

Alberto Sabater Bailón  
546297

## 1. Informe

**1.- ¿Se corresponden las medidas de evaluación globales de tu sistema con alguna de las columnas que se muestran en el cuadro 2, o las curvas globales de precisión-exhaustividad de la figura 1? ¿Con qué columna/curva?**

Si, se corresponden con las medidas de la curva n.

**2.- ¿Qué decisiones de diseño de tu sistema (campos indexados, analizadores elegidos, modelo de ranking, procesamiento especial de consultas, ...) justifican las medidas obtenidas? ¿Hay otros factores externos que pueden haber afectado a estos resultados?**

Las medidas obtenidas se han visto influenciadas por el uso del Spanish Analyzer, ya que los resultados están en castellano y mejora la búsqueda. El analizador además ha sido modificado para adaptarse a las necesidades de información y se han lematizado todas las entradas.

Los resultados también se han visto influenciados por la Query utilizada por el buscador. En esta Query, se puede indicar el autor que se está buscando, el periodo de publicación (incluyendo los últimos n años), el título y la descripción del trabajo. También se ha de tener en cuenta que se le ha exigido a la búsqueda que los resultados que devolviese debían satisfacer la mayoría de estas restricciones, y por lo tanto se ha producido un mayor filtrado de documentos potencialmente no relevantes. Todos estos campos han sido indexados previamente a la consulta.

Para el ranking se ha usado el que usa Lucene por defecto, que es una combinación del modelo de espacio vectorial (VSM) y el modelo booleano.

**3.- Según tu opinión, ¿Qué medida, o medidas, reflejan mejor la calidad de un sistema de recuperación?**

La precisión@10, porque únicamente tiene en cuenta los primeros resultados. En un buscador, el usuario, por norma general, únicamente mira si los primeros resultados son buenos. También el F1 score, porque es una medida que tiene en cuenta precisión y recall y establece una buena relación entre ellos.

**4.- En base a las medidas del cuadro 2, las curvas de la figura 1 y tu respuesta a la cuestión anterior. ¿Qué sistema que no sea el tuyo recomendarías utilizar?**

Recomendaríamos **el sistema c** en base a los criterios expuestos en la pregunta 3.