

Informe evaluación sistema tradicional equipo 16

Fran Ferraz (737312) y Guillermo Cruz (682433)

¿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?

- Sí, concretamente se corresponden con la columna c del cuadro 2, y con la curva global c de la figura 1.

¿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?

- Se cree que los campos elegidos para indexar son los correctos, pues básicamente se han indexado casi todos los campos que aparecen en el xml. El analizador también, pues se eligió el SpanishAnalyzer con el filtro SnowballFilter que implementa el algoritmo de stemming de Snowball para el idioma español, porque la mayoría de documentos están en español. Por lo tanto el haber conseguido una precisión de 0.164 y una exhaustividad de 0.226, puede deberse a que el procesamiento de las consultas se necesite mejorar, si se considera que los campos indexados y el analizador están bien elegidos.

- La prec@10 que muestra la precisión de los primeros 10 resultados sale casi el doble de mejor que la precisión normal, y el procedimiento elegido para el ranking fue repetir los términos más importantes varias veces en la consulta, para que los documentos en los que aparecieran dichos términos aparecieran en posiciones más altas. Se cree que el procedimiento ha funcionado razonablemente bien si se compara con la precisión normal, pero se pueden buscar otras técnicas que lo mejoren.

- Ha penalizado mucho en el sistema que para la necesidad 11-4, se ejecutó la consulta buscando documentos anteriores a 2009, y hay muy pocos. Solo se devolvieron 2 resultados y además no relevantes, por lo que todas las medidas de esta necesidad están a 0, penalizando muchísimo. Aun así la MAP ha salido de 0.394, y teniendo en cuenta que la precisión promedio de la necesidad 11-4 es 0 por el motivo descrito, se cree que es un buen resultado.

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

- Precisión y exhaustividad son importantes pero se debe promediar sobre conjuntos grandes de documentos y consultas, por lo que para conjuntos pequeños pueden resultar engañosas. Se cree que prec@k es una medida importante si se elige bien el parámetro k, pues refleja la calidad de los primeros documentos que va a ver el usuario, y los más importantes por tanto. Por último se cree que la más importante es la MAP, puesto que a mayor MAP, mejor sistema.

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?

- Se recomendaría utilizar el sistema r, que tiene la MAP y la prec@10 más altas, siendo estas las medidas más significativas. Además, también tiene valores altos de precisión y exhaustividad.