

Sistemas Inteligentes de Acceso a la Información
Problemas y cuestiones
Introducción a los Sistemas Inteligentes
de Acceso a la Información

1. Supongamos que un sistema de IR presenta un ranking de 10 documentos para una consulta sobre una colección, de modo que los relevantes se hallan en las posiciones 2, 3, 5 y 10, y que dos documentos relevantes no se obtienen. Calcular y presentar una tabla *recall-precision* interpolada a 11 niveles para la consulta dada.
2. Supongamos que un sistema de IR presenta los resultados para dos consultas sobre una colección de 100 documentos. Para la primera, recupera 6 documentos, de los cuales 3 son relevantes pero no recupera 2 relevantes. Para la segunda, recupera 5 documentos, de los cuales 1 es relevante pero no recupera 8 relevantes. Calcular los valores de F_1 por macro- y micro-averaging.
3. Discutir la validez de la siguiente afirmación: En un sistema de RI sobre textos jurídicos, es más importante la precisión que la cobertura, por lo que es recomendable utilizar F_2 en lugar de F_1 .
4. ¿Cuáles son las métricas que permiten comparar dos sistemas a través de un solo número? Diferenciar las aplicables en caso de salida con ranking y sin ranking.
5. A la hora de promediar la cobertura y la precisión de un sistema, para obtener un solo número, podemos usar la media aritmética, la media geométrica, o la medida F_β (entre otras). Poner un ejemplo que permita argumentar por qué es preferible la tercera medida a las dos primeras.
6. El cálculo de tablas de cobertura-precisión interpoladas a once niveles, ¿se puede considerar un método optimista (es decir, devuelve valores mejores que los que realmente produce el sistema) o un método pesimista? ¿Por qué?