

Organización de Datos – Curso Servetto

Evaluación Sistemas de recuperación total de textos (FTRS), 12 de julio del 2006

Tema 1

Para probar es necesario tener 8/13 puntos. Además no debe haber ningún error conceptual grave.

1. Considerar que cada palabra es un término y cada línea un documento.

a) Resuelva la consulta rankeada “H o I” para los siguientes documentos, utilizando el método del coseno.(3)

D1: F F F G H I
D2: G G H H H
D3: F F I
D4: F F F G H

b) Represente los documentos D3 y D4 utilizando representación vectorial booleana y calcule la distancia usando Jaccard.(1)

2. Un término aparece en los siguientes documentos:

Doc 4, Doc 12, Doc 15, Doc 18, Doc 25, Doc 26 y Doc 29

Dicho término representa bastante bien al resto de los términos que existen en el sistema, por lo que se quiere analizar en base a él si conviene utilizar códigos unarios, delta o gamma. ¿Cuál de estas tres alternativas conviene?. ¿Cómo se explica que esa alternativa sea mejor que las otras dos?(3)

3. Explicar cómo puede resolverse la siguiente consulta (*ASA) utilizando léxico rotado suponiendo que se tienen los siguientes términos:

CASO
MASA
TASA(3)

4. Mostrar el ahorro que podría efectuarse en un índice si se utiliza Front Coding para el almacenamiento de los términos del punto anterior. Calcular el ahorro que se produciría si se usara front coding parcial. (3)