

Organización de Datos – Curso Servetto

Evaluación Sistemas de recuperación total de textos (FTRS), 19 de julio del 2006

Tema 1

Para probar es necesario tener 10/16 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 “C o D” para los siguientes documentos, utilizando el método del coseno.(3)

D1: A B C C D
D2: C D
D3: A A B
D4: D E

Cómo hubiera cambiado la solución si la consulta hubiera sido “C y D”? Qué consideraciones es necesario tener en el almacenamiento si se deseaba poder contestar consultas sobre “frases”? (3)

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 3, Doc 8, Doc 11, Doc 19, Doc 21, Doc 29 ,Doc 33 y Doc 35

Dicho término representa bastante bien al resto de los términos que existen en el sistema. Calcular el espacio necesario para almacenar estas distancias utilizando 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 (p*s) utilizando digramas suponiendo que se tienen los siguientes términos:

PECES
PECAS
PEPE
QUESOS(3)

4. Mostrar el almacenamiento de los términos del punto anterior si se utiliza front coding parcial con N=3 (3)