



Recuperación de Información

2 de Febrero de 2015

1. (2 puntos) Describe brevemente los pasos en un proceso típico de indexación en un Sistema de Recuperación de Información (1 página).
2. (2 puntos) Cómo se calcula la medida coseno cuando la información (los distintos pesos) los obtenemos a partir de un índice invertido. Ilustras los pasos para la consulta con los términos aaa y ccc del Cuadro 1.

Cuadro 1: Ejemplo de colección documental

término	d1	d2	d3	d4	d5	d6	d7	d8
aaa	4	0	0	0	2	8	0	0
bbb	0	4	2	1	0	2	4	0
ccc	1	1	2	0	0	4	2	1
ddd	0	0	0	1	2	0	4	4
eee	0	1	4	0	0	0	4	4
fff	1	0	2	0	0	2	0	1
ggg	0	0	0	2	4	0	8	8

3. (2 puntos) Diferencia entre evaluar al eficacia y la eficiencia de un sistema de Recuperación de Información. Define las métricas Precisión y MAP. ¿Qué mide cada una?
4. (2 puntos) ¿Qué es el PageRank? ¿Cómo se calcula?. Qué problemas tendríamos para calcularlos utilizando el grafo representado por la matriz 4.

Cuadro 2: Representación matricial del grafo web

	a	b	c
a	1	0	1
b	0	1	1
c	0	0	1

5. (2 puntos) Cómo construirías un filtro de correo basura (spam). Pon un ejemplo teniendo en cuenta los documentos de la Tabla 2, donde los documentos 5,6,7 y 8 son clasificados como spam

Notas

Tiempo para realizar el examen: 3 horas.