

Recuperación de Información

Curso 2013/14. Convocatoria de Febrero

1. (1 punto) Describe brevemente los pasos en un proceso típico de indexación en un Sistema de Recuperación de Información.
2. (0.5 puntos) Describe la medida Coseno y la medida Jaccard.
3. (1 punto) Supongase que tenemos una colección de cuatro documentos descrita por la siguiente matriz término/documento, donde las celdas indican el número de apariciones de los términos en cada documento.

	$d1$	$d2$	$d3$	$d4$
bahia	4	1	-	-
brasil	1	2	-	4
carnaval	2	1	2	8
rio	-	1	1	1
tropico	-	4	-	2

- a) Calcular el peso tf-idf de los términos en cada documento.
- b) Utilizando la similitud por coseno, calcular el valor de recuperación (score) de los documentos $d1$ y $d4$ para la consulta *carnaval bahía*.
4. (1 punto) ¿Cómo calcula el score el Language Model (Modelado del Lenguaje)? Pon un ejemplo (considerando los datos del Ejercicio3 con valor $\lambda = 0.5$) para el documento $d4$ sobre la consulta *Carnaval en el trópico*.
5. (1 punto) ¿Qué representaría el valor de PageRank de un sitio web? ¿Cómo se calcula?
6. (1 punto) Enumera al menos cinco aplicaciones de la Clasificación en Recuperación de Información.
7. (1 punto) Para una consulta se han recuperado la siguiente secuencia de documentos

N R R N N R R N N R N N R N N R N N N N

Suponiendo que en la colección hay un total de 10 documentos relevantes, calcular la exhaustividad, la precisión, $P@10$, $AP@20$ y la curva precisión/exhaustividad del ranking devuelto.

Notas

Tiempo para realizar el examen: 3 horas.