

## 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	1	1
brasil	1	1	1	1
carnaval	1	1	1	1
rio	1	1	1	1

- a) Calcular el peso tf-idf de los términos en los documentos d1 y d4 para la consulta *Carnaval*.
  - b) Utilizando la similitud coseno, calcular la similitud entre los documentos d1 y d4 para la consulta *Carnaval*.
4. (1 punto) ¿Cómo calcula el ejemplo (considerando la consulta *Carnaval*) la similitud entre los documentos d1 y d4?
  5. (1 punto) ¿Qué similitud se utiliza en el ejemplo?
  6. (1 punto) Enumera al menos tres términos que se recuperan en el ejemplo.
  7. (1 punto) Para una consulta se han recuperado los documentos d1, d2, d3 y d4. ¿Cuál es el documento más relevante?

Suponiendo que en la exhaustividad, se devuelven los documentos d1, d2, d3 y d4.

Notas