



Universidad de Granada

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

ETSII Informática y de Telecomunicación, C/Periodista Daniel Saucedo Aranda s/n - 18071 - Granada (España)

## Recuperación de Información

Curso 2016/17. Convocatoria de Septiembre

1. ¿Qué es un índice invertido? ¿Qué componentes tiene?

En el caso de "Evaluación Única", esta pregunta se cambia por

- Describe un índice Lucene, indicando cada una de sus partes.
2. Diferencia entre un modelo de recuperación booleano, vectorial y probabilístico. Pon un ejemplo de cada uno
3. Supóngase 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.

|                 | <i>d1</i> | <i>d2</i> | <i>d3</i> | <i>d4</i> |
|-----------------|-----------|-----------|-----------|-----------|
| <i>bahia</i>    | 4         | 1         | -         | -         |
| <i>brasil</i>   | 1         | 2         | -         | 4         |
| <i>carnaval</i> | 2         | 1         | 2         | 8         |
| <i>rio</i>      | -         | 1         | 1         | 1         |
| <i>tropico</i>  | -         | 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. Qué es la clasificación documental. Describe el algoritmo de clasificación de *Rocchio*.
5. ¿Qué representaría el valor de PageRank de un sitio web?. ¿Cómo se calcula?. ¿Cómo es utilizado dicho valor en un Sistema de Recuperación de Información?

## Notas

Tiempo para realizar el examen: 2:30 horas.