

Recuperación de Información

24 Enero de 2017

■ Teoría

1. (2 puntos) Describe el proceso de indexación distribuida: Map Reduce
2. (2 puntos) Describe el modelo de RI basado en Modelado de Lenguaje.
3. (2 puntos) Ante una consulta, un SRI encuentra 20 documentos para los que tenemos los siguientes juicios de relevancia

nr r nr r r nr nr nr r nr

Suponiendo que en la colección hay un total de 10 documentos relevantes, calcular

- la curva de precisión interpolada para los 11 puntos de exhaustividad.
 - la NDCG@5
4. (2 puntos) Define el proceso de Crawling en Recuperación de Información web.
 5. (2 puntos) Algoritmo de Clasificación de Rocchio. Calcular el centroide para la clase A del siguiente ejemplo, donde asumimos una frecuencia binaria de los términos

terminos	clase
t7, t8, t10	C
t5, t6, t7	B
t1, t2, t3, t4	A
t2, t3, t8	A
t9, t10	C
t3, t4, t6	B
t1, t4, t6	A

■ Prácticas

1. Implementar en Lucene un método que fuese capaz de devolver el título y precio de los 5 elementos más relevantes a la siguiente necesidad de Información

Busco libros que contengan las palabras *java* y *programming* en el título a una distancia no mayor de 3 y cuyo autor sea *Barry Bud*

Notas: Tiempo para realizar el examen

Preguntas Teoría 1 a 3: 75 minutos

Preguntas Teoría 4 y 5: 50 minutos

Pregunta Práctica: 25 minutos.