



Examen Julio, preguntas

Recuperación da Información (Universidade da Coruña)

Examen **Recuperación de Información** Julio 2016

Apellidos: _____ Nombre: _____

Examen sin libros, apuntes, ni dispositivos electrónicos.

Tiempo: 2h

1. (1.5 puntos)

Pseudocódigo, estructuras de datos y funciones, explicación. Debe contestar a cada apartado en el espacio dedicado.

a) (0.6) Pseudocódigo del hilo de crawler estudiado en clase.

b) (0.2) Explicación de estructuras de datos y funciones principales usadas en a)

c) (0.2) Explicación en lenguaje natural

d) (0.5) ¿Puede adaptarse ese código para que el crawler visite las páginas con una estrategia primero en anchura y de forma que la profundidad en cada sitio web esté limitada a un valor de 3? Marque la respuesta correcta y razónelo.
Si es posible diga como, si no es posible diga por qué.

-Es posible

-No es posible

2. (1.5 puntos) Pseudocódigo, estructuras de datos y funciones, explicación. Debe contestar a cada apartado en el espacio dedicado.

a) (0.6) Pseudocódigo del algoritmo Term-At-A-Time de procesamiento de consultas estudiado en clase.

b) (0.2) Explicación de estructuras de datos y funciones principales usadas en a)

c) (0.2) Explicación en lenguaje natural

d) (0.5 puntos) ¿Puede adaptarse ese código para procesar consultas en el modelo de RI basado en modelos de lenguaje Query Likelihood con suavización de Dirichlet? Marque la respuesta correcta y razónelo. Si es posible diga como, si no es posible diga por qué.

-Es posible

-No es posible

3. (1 punto) Una query q tiene como documentos relevantes los documentos con docID: 1, 3, 6, 7, 10, 12, 13, y un sistema de RI recupera el siguiente ranking de documentos: 2, 6, 5, 8, 3, 12, 11, 14, 7, 13, donde de nuevo los números son los docID de los documentos.

a) (0.4) Rellene la tabla mostrando el recall y la precisión calculada para cada documento recuperado.

Ranking	Recall	Precisión
d2		
d6		

b) (0.4) Rellene la tabla con la precisión interpolada a los niveles estándar de recall.

Nivelse estándar de Recall	Precisión Interpolada

c) (0.2) Compute el MAP

4. (0.5 puntos) Considere una colección de documentos que contiene estos documentos

d1: scary green crocodile

d2: scary green big

d3: small crocodile

y la query: big green crocodile

Considere el esquema de pesado rawtf x idflog, donde los log son base 2, tanto en la query como en los documentos.

Compute la similaridad de coseno para esa query de los tres vectores y el ranking producido

5. (0.5 puntos) Explique la fórmula de Rocchio para Relevance Feedback. Dada una query expresada por el vector $q = \langle 2, 0, 3, 1, 0 \rangle$ y los documentos devueltos por el sistema y marcados como relevantes d_1, d_2, d_3 y como irrelevantes d_4, d_5 , compute la query modificada supuesto el mismo peso para la query original, los documentos relevantes y los irrelevantes.

$$d_1 = \langle 3, 1, 2, 1, 0 \rangle$$

$$d_2 = \langle 4, 1, 3, 2, 2 \rangle$$

$$d_3 = \langle 1, 0, 5, 0, 3 \rangle$$

$$d_4 = \langle 1, 3, 0, 1, 2 \rangle$$

$$d_5 = \langle 0, 4, 0, 2, 2 \rangle$$