

Control 1 CC52D 2004/01

Ricardo Baeza-Yates - Carlos Castillo

7 de septiembre de 2005

Tiempo: 2.00 horas, con apuntes. Todas las preguntas valen lo mismo.

Pregunta 1

Indique 5 medidas de calidad de un sistema de recuperación de información y ordenelas de la más importante a la menos importante, justificando su orden.

Pregunta 2

Suponga que tenemos un sistema de RI que usa solamente el modelo vectorial usando TF-IDF como se vio en clases. Tomemos ahora una consulta dada q .

¿Cambiará la respuesta a q si eliminamos stopwords?

Analice los casos en que q contiene/no contiene stopwords.

Si su respuesta es positiva, indique si esto favorece a documentos que tienen pocas o muchas stopwords

Pregunta 3

Para cada uno de los siguientes casos, dibuje una Web con 5–10 nodos y marque uno de ellos con una cruz, de modo que el nodo marcado tenga:

- Pagerank alto **pero grado interno bajo**.
- Authority score alto **pero grado interno bajo**.
- Hub score alto **pero grado externo bajo**.

¿Cuál es la diferencia más importante entre Pagerank y Authority score?

Pregunta 4

¿Cómo se puede utilizar XML Schema en un sistema de búsqueda ...

- ... para mejorar la precisión ?
- ... para mejorar la velocidad ?

Pregunta 5

En un encuentro de desarrolladores de sitios Web, se llega al acuerdo de que todos los dueños de sitios Web ejecutarán diariamente el siguiente comando:

```
% ls -laR $WEBHOME > $WEBHOME/listing.txt
```

En que `$WEBHOME` representa el directorio base de sus sitios Web. El listado incluye el nombre de archivo, tamaño y fecha de última modificación solamente.

Describa cómo ésta información puede ser usada por crawlers de máquinas de búsqueda y cuales serían los beneficios para la máquina de búsqueda y para los sitios Web.

Pregunta 6

Tanto XPath como XSLt pueden comenzar a entregar resultados a una consulta incluso sin haber terminado de leer el documento XML que está siendo consultado. ¿Cómo se puede aprovechar esta característica en un buscador Web?

Dibuje un fragmento de árbol XML que contenga un nodo que calce con la siguiente consulta XPath:

```
/A[B[2]]/B[C[3]/D]/C[2]/E
```