

Test

Primeros dos valen 3, segundos dos valen 2

Parcial 2 de Bases de Datos - 2024

Ejercicio 1: Sea la BD de una pizzería con las siguientes tablas:

cliente(cid, nombre, teléfono, dirección, edad)
pizzas(zid, znombre, tamaño, precio)

pedido(cid, zid, phora, paño, pmes, pdia, cantidad), ^{Puntero}
Se tiene la siguiente información de las tablas: ^{1 2 3 4 5 6} = 27

No olvidar puntero

Dar todo como STR

cliente	pizzas	pedido
1000 tuplas	40 tuplas	1.200.000 tuplas
	V(pizzas,tamaño) = 4	V(pedido,Zid) = 40
	V(pizzas,znombre) = 10	V(pedido,cid) = 1000
		V(pedido,paño) = 40

Asumir que un bloque tiene tamaño de 4096 bytes. Todos los años se piden todas las pizzas. Puede asumir que el último año es el 2024. Sea la consulta:

$$\prod_{zid, nombre} (\text{cliente} \bowtie \sigma_{paño \geq 2015} (\text{pedido}))$$

Responder:

1. ¿Cuál es la probabilidad de que un pedido tenga paño ≥ 2015 ? Puede usar uniformidad. (hacer la cuenta en no más de dos líneas) – 30%

selección
total

$$fs(\sigma_{paño \geq 2015}, \text{pedido}) = \frac{2024 - 2015}{40} = \frac{9}{40}$$

2. ¿Cuántas tuplas tiene el resultado intermedio de la selección? (hacer la cuenta en una línea) – 20%

$$r = \sigma_{paño \geq 2015} (\text{pedido}) = |\text{pedido}| \cdot fs = 1.200.000 \cdot \frac{9}{40} = 270\,000$$

3. ¿Cuántos bloques tiene el resultado intermedio de la selección? (hacer la cuenta en no más de 3 líneas y en no menos de 2 líneas) – 50%

pedido(cid, zid, phora, paño, pmes, pdia, cantidad), ^{Puntero}
e la siguiente información de las tablas: ^{1 2 3 4 5 6} = 27

$$\text{cantBloq} = \left\lceil \frac{\text{cantTup} \cdot \text{tamTup}}{\text{tamBloq}} \right\rceil = \left\lceil \frac{270\,000 \cdot 27}{4096} \right\rceil = 1780$$

CantTup se saca de resultado anterior.

Ejercicio 2: Responder: (Sea breve, no más de dos líneas por pregunta)

1. Asumir que se tiene un sistema de retorno de la información que usa modelo de espacio vectorial, que trabaja en base a una colección de archivos en una PC que no son páginas web, los términos son palabras individuales. Los resultados de las consultas se ordenan por relevancia. ¿Cuáles son las estructuras de información almacenadas que usa una consulta para poder ser procesada eficientemente? No se pide describir estas estructuras, solo dar sus nombres. (50%)

✓

vocabulario, índice invertido

2. Supongamos que tenemos un sistema de retorno de información (SRI) para páginas web de noticias. Hay que seleccionar los elementos del vocabulario con que va a trabajar el SRI. Indicar los tipos de elementos que debería contener el vocabulario para ser útil para este tipo de documentos (responder en una oración, no justificar). No suma puntaje poner: raíces de palabras no Stop Word (eso se usa en general y quiero una respuesta más específica). (50%)

- ✓ Entidades nombradas: nombre, lugar, fecha, etc.
- ✓ Sinónimos (wordnet)
- ✓ Palabras clave.

Ejercicio 3: Considerar el siguiente texto:

"En las Sierras de Córdoba, veía con tristeza cómo los incendios arrasaban el paisaje. Reflexionaba sobre la frecuencia creciente de estos desastres debido a errores humanos. Recordé grandes incendios históricos y su devastación, y cómo forman parte de la tradición oral y cultural de la región, mientras hablábamos sobre la ayuda necesaria y las experiencias vividas."

Indicar tres tipos de contexto dando su nombre y un ejemplo de cómo aparece en el texto. (No más de dos líneas por cada tipo de contexto.)

Ejercicio 4: crear un prompt de no más de 5 líneas sobre: aprender componentes de sistema de retorno de la información. Quiero que haya modificadores de rol, audiencia objetivo, de tipología de texto y de estructura.

Como si fueses un profesor de ciencias de la computacion, quiero que elabores un resumen sobre los componentes del sistema de retorno de la informacion, enfocado a estudiantes de tercer año de ciencias de la computacion.
Quiero que el resumen que elaboras tenga forma de lista.