

CC4102 - Control 1

Profs. Pablo Barceló y Gonzalo Navarro

15 de Abril de 2019

P1 (3.0 pt)

Considere los siguientes problemas:

- **Colinearidad:** Dados n puntos en el plano con coordenadas enteras en tres líneas horizontales $y = 0$, $y = 1$, e $y = 2$, determinar si hay línea no-horizontal que contenga a tres de ellos.
- **Separación:** Dados m segmentos de recta en el plano, ¿existe una recta que divida a estos segmentos en dos conjuntos no vacíos?

Demuestre que existe una reducción del primer problema en el segundo que toma tiempo $O(n \log n)$ (2pts). ¿Qué puede concluir acerca del costo que toma resolver **Separación**? (1pto).

P2 (3.0 pt)

Compararemos dos técnicas clásicas para calcular el join natural de dos tablas en una base de datos relacional. Esta consiste en, fijadas dos columnas c_1 y c_2 de las tablas T_1 y T_2 respectivamente, obtener las filas f_1 y f_2 de T_1 y T_2 tal que $f_1[c_1] = f_2[c_2]$.

Nested loop: Para cada tupla de una relación, se recorre la otra, emitiendo todo par de columnas que satisfaga la igualdad.

Merge join: Se ordenan las filas de ambas tablas por la columna deseada y luego se recorren ambas secuencialmente, buscando los pares que satisfagan la igualdad.

Se le pide lo siguiente:

1. Indique el costo, en el modelo de memoria externa, de ambos métodos. Suponga que los tamaños de las tablas son $N_1 > N_2 \gg M$.
2. Indique en qué caso usaría cada una, según los tamaños de las tablas, en términos asintóticos.
3. Muchos manejadores de bases de datos indexan las tablas, preparándolas para ciertos joins. Indique nuevamente las complejidades del merge join en caso de que (1) ambas columnas ya vengan ordenadas, (2) sólo la de la tabla mayor venga ordenada, (3) sólo la de la tabla menor venga ordenada. ¿Cuál es ahora la mejor estrategia, en cada caso?

Tiempo: 2.0 horas

Con una hoja de apuntes

Responder en hojas separadas