

Concorrenca e Paralelismo. Bloque II Paralelismo

Práctica 3: Descomposición de dominio: vector de similitud de conjuntos de ADN

Primavera 2020



Vector de similitud de conjuntos de ADN

Datasets de secuencias de ADN

- M secuencias de texto
- Cada una con N bases (caracteres A,C,G,T,N)

Vector de similitud

- Dados dos conjuntos de secuencias de ADN
- Medida de lo similares que son la secuencia i del primer conjunto y la secuencia i del segundo conjunto
 - M resultados: Los dos conjuntos deben tener el mismo número M de secuencias
 - SIMPLIFICACION: Todas las secuencias tienen el mismo número N de bases

Vector de similitud de conjuntos de ADN

Código secuencial

```
int main(int argc, char *argv[]){
    int *data1, *data2, *result;
    data1 = (int *) malloc(M*N*sizeof(int));
    data2 = (int *) malloc(M*N*sizeof(int));
    result = (int *) malloc(M*sizeof(int));

    /* Initialize Matrices */
    ...

    for(i=0;i<M;i++) {
        result[i]=0;
        for(j=0;j<N;j++) {
            result[i] += base_distance(data1[i*N+j], data2[i*N+j]);
        }
    }

    free(data1); free(data2); free(result);
}
```

Vector de similitud de conjuntos de ADN

Código secuencial: `similarity.c`

Descomposición de dominio

- Dividid las matrices entre p procesos, cada uno con $rows = M/p$ filas (por simplicidad, empezad con el caso en que consideraremos que el número de procesos es múltiplo de M : $M \bmod p = 0$).
- Cada tarea se encargará de calcular M/p elementos del vector resultado.
- En la entrega considerad cualquier número de procesos, no solo múltiplos de M .

Vector de similitud de conjuntos de ADN

Paralelización

- Implementación SPMD.
- La inicialización de la matriz la hace el proceso 0.
- Distribuir datos a todos los procesos con operaciones colectivas
- Recolección del vector resultado con operaciones colectivas
- La E/S (printf) la hace el proceso 0.
- Imprimid por separado tiempo de comunicaciones y tiempo computación de cada proceso

Vector de similitud de conjuntos de ADN

- Puntuación: 1
- Deadline: 17 Mayo
- Realización en mismas parejas que P1 y P2
- Entrega en tarea de Moodle con mismas instrucciones que P1 y P2:
 - Código fuente. Al inicio en comentarios nombre, apellidos y DNI de los miembros de la pareja.
 - Captura de pantalla que incluya:
 - Nombre de usuario (echo \$USER)
 - Compilación
 - Ejecución con tres procesos (muestra por pantalla los tiempos, no los resultados)
 - Pequeño resumen (máximo una página) con las claves del trabajo realizado
 - Diferente para cada miembro de la pareja