

INTEGRATIVA:



Problema:

Se trata de resolver una situación real de una empresa de Biotecnología. La empresa se enfrenta a la disyuntiva de seleccionar ciertos marcadores biológicos para poder realizar un diagnóstico preciso para detectar una enfermedad.

Se tienen 2 equipos que realizan una detección de marcadores moleculares, cada uno entrega los resultados almacenados en archivos de textos. Los datos de estos marcadores se encuentran almacenados en un archivo con una lista de posibles marcadores que se organizaron en forma de matriz. Cada archivo de datos contiene fijo 5 columnas y X filas. Donde X es la cantidad de filas del archivo y depende de la cantidad de datos que tenga el archivo que se desee procesar. En la columna 1 se encuentra el código del ID de un gen. En la columna 2 se encuentra el código del ID de un posible marcador biológico detectado para ese gen. En la columna 3 se encuentra la coordenada de inicio del marcador en ese gen. En la columna 4 se encuentra la coordenada de término del marcador en ese gen. En la columna 5 se encuentra la energía calculada para ese marcador.

Usted ha sido contratado para diseñar un programa que sea capaz de leer 2 archivos de texto con datos de 2 bases de datos distintas. Estos archivos tienen distinto tamaño de filas. Su programa debe realizar una comparación entre ambos archivos y seleccionar los marcadores óptimos a partir de los marcadores de ambos archivos de datos.

Un marcador es óptimo si cumple con las siguientes características:

- Debe tener el código del ID del gen igual en ambos archivos.
- Y debe tener el código del ID del marcador igual en ambos archivos.
- Considerando las 2 anteriores condiciones, puede permitir una diferencia de hasta N unidades de diferencia entre “inicios” y “términos” de cada marcador. Por ejemplo, si un “inicio” (en uno de los archivos) es mayor o menor hasta N dígitos al “inicio” del otro archivo y sí, además, el “término” de ese marcador es mayor o menor hasta N dígitos al “término” del otro archivo se debe considerar como marcador óptimo.
- Donde N es un parámetro ingresado por el usuario para determinar el nivel de sensibilidad en la selección de cada marcador. Si el N ingresado por el usuario:
- Es igual a 0 el nivel de sensibilidad es: **Estricto**.

Esto implica que si el “inicio” en un archivo A es igual al del otro archivo B, entonces al comparar los “términos” pueden tener un “término” de distinto valor. O si el “término” es igual en ambos archivos, entonces el “inicio” puede ser distinto.

- Si se ingresa un 1 es intensidad: **Mediano**.

Esto implica que si el “inicio” en un archivo A es menor en 1 unidad al “inicio” del archivo B, entonces se pueden dar dos casos: primer caso, al comparar “los términos”, el “término” en el archivo A (del “inicio” menor) puede ser igual o menor a 1 unidad del “término” del archivo B. Segundo caso el “término” del archivo B puede ser mayor o igual a 2 unidades respecto al “término” del archivo A.

O también puede implicar que si el “inicio” en un archivo A es mayor en 1 unidad al “inicio” del archivo B, entonces se pueden dar dos casos: primer caso, el “término” en el archivo A (del “inicio” mayor) puede ser igual o menor a 1 unidad del “término” del archivo B. Segundo caso, el “término” del archivo B puede ser mayor o igual a 2 unidades respecto al “término” del archivo A.

- Si se ingresa un 2 es: **Relajado** (el nivel menos estricto).
 Esto implica que si el “inicio” de A es menor en 2 unidades al “inicio” de B, entonces se pueden dar dos casos: primero caso, el “término” A puede ser menor o igual a 2 unidades al “término” de B. Segundo caso, “término” de B puede ser mayor o igual a 2 unidades al “término” de A. También implica que si el “inicio” de A es mayor en 2 unidades al “inicio” de B, entonces se pueden dar dos casos: primero caso, el “término” A puede ser menor o igual a 2 unidades al “término” de B. Segundo caso, “término” de B puede ser mayor o igual a 3 unidades al “término” de A.

Por ejemplo, son considerados como marcadores óptimos para **Estricto**:

N	INICIO A	TÉRMINO A	INICIO B	TÉRMINO B	ESTADO
0	42	65	42	65	óptimo
0	42	65	42	64	óptimo
0	42	65	42	66	óptimo
0	42	65	42	67	óptimo
0	41	65	42	65	óptimo

Por ejemplo, son considerados como marcadores óptimos para **Mediano**:

N	INICIO A	TÉRMINO A	INICIO B	TÉRMINO B	ESTADO
1	42	65	42	66	óptimo
1	42	65	42	64	óptimo
1	41	65	42	65	óptimo
1	43	65	42	65	óptimo
1	41	65	42	66	óptimo
1	41	65	42	64	óptimo
1	41	65	42	67	óptimo
1	41	65	42	68	NO
1	41	65	42	63	óptimo
1	41	65	42	62	NO
1	43	65	42	64	óptimo
1	43	65	42	66	óptimo
1	43	65	42	62	NO
1	43	65	42	67	óptimo
1	43	65	42	68	NO

Por ejemplo, son considerados como marcadores óptimos para **Relajado**:

N	INICIO A	TÉRMINO A	INICIO B	TÉRMINO B	ESTADO
2	42	65	42	67	óptimo
2	42	65	42	63	óptimo
2	42	65	42	68	óptimo
2	42	65	42	69	NO
2	40	65	42	66	óptimo
2	40	65	42	62	óptimo
2	41	65	42	67	óptimo
2	41	65	42	68	NO
2	41	65	42	63	óptimo
2	41	65	42	62	NO
2	43	65	42	64	óptimo
2	43	65	42	66	óptimo
2	43	65	42	62	NO
2	43	65	42	67	óptimo
2	43	65	42	68	NO

- En caso de seleccionar un marcador como óptimo, debe generar una nueva matriz con el mismo formato de los archivos de entrada y guardarlo en otro archivo de salida llamado: marcadores_seleccionados.txt
- Además, debe mostrar en pantalla cuántos datos totales (marcadores clasificados) fueron seleccionados como óptimos.
- Su programa requiere como entrada de datos 2 archivos de texto y un valor entero (el nivel de sensibilidad de selección).
- Finalmente se debe comentar las funciones del programa.