



**Universidad
de Concepción**

Facultad de ingeniería

Lenguajes de programación

Tarea semestral

Vicente Alvarado Ulloa
Bioingeniería

Nicolás Pinto Gutiérrez
Bioingeniería

Valentina Sepúlveda Díaz
Bioingeniería

Francisco Gacitúa Fuentes
Ingeniería Ambiental

Nombre curso: Lenguaje de programación 503215-1

Junio de 2022

ANÁLISIS

El laboratorio de Biotecnología de la Universidad de Concepción necesita catalogar distintas secuencias de ADN las cuales constan de las letras “A”, “C”, “G” y “T”, clasificándolas según su medida de desorden.

Para aquello nosotros como grupo debemos crear un programa en python que pueda calcular la medida de desorden de ciertas cadenas de ADN, que serán ingresadas por el usuario desde un archivo de nombre “ADN.dat”, para en base a esto ordenarlas de forma creciente según su desorden. Lo haremos seleccionando las cadenas como más ordenadas primeramente a aquellas que siguen un orden más cercano al alfabético, es decir, que pueden alcanzar el estado de orden alfabético con el menor número de “movimientos”, para determinar esta medida, se toma en cuenta la cantidad de inversiones (cada inversión es un desplazamiento en un espacio hacia la derecha de una letra), este valor, la medida de desorden, la calcularemos matemáticamente asignándole un valor a cada letra, para así darles una jerarquía. Posteriormente contaremos el número de inversiones sumando la cantidad de letras que están hacia la derecha de cada carácter y que estén antes en el orden alfabético (se le dará un valor a cada letra y se evaluará cada posición a la derecha, sumando 1 a la medida de desorden si existen cifras que representen un valor menor). Una vez obtenida la medida de desorden, serán más ordenadas aquellas cadenas con mayor longitud e igual desorden, y como tercer factor si tienen igual desorden, e igual longitud será más ordenada la que sea más alfabética (lo determinaremos con la función “sort()” ya existente en python).

Ejemplo:

AACATGAAGG

Para calcular la medida de desorden de esta cadena:

Letra A:

Toda letra A independiente de su posición en la cadena no posee ninguna letra que esté antes en el abecedario por lo que siempre se le asignará un valor de desorden igual a 0.

Letra C:

La letra C en este caso posee 3 letras a su derecha que se encuentran antes en el abecedario las cuales serían la letras marcadas en rojo AACATGAAGG por lo que se le asignará un valor de desorden igual a 3.

Letra T:

Esta posee 5 letras que están antes en el abecedario y que se encuentran a su derecha AACATGAAGG por lo que su valor de desorden será 5.

Primera letra G:

AACATGAAGG

Posee un desorden de 2.

Segunda letra G:

Siguiendo el mismo método ésta tendrá un valor de desorden igual a 0

AACATGAAGG

Tercera letra G:

No posee letras a su derecha por lo que su valor de desorden será 0.

Una vez hecho este paso se sumarán todos los valores calculados dándonos un valor de 10 el cual será la magnitud de la medida de desorden, el cual luego será comparado con las otras cadenas para así ordenarlas de menor a mayor.

Secuencia borrador del programa

Para comenzar con el programa se lee un archivo de nombre "ADN.dat" que tiene su estructura de la forma siguiente:

- Primera línea es la cantidad de casos (número mayor que 0 y menor que 50).
- La segunda línea nos da dos números, el tamaño menor y mayor posible para cada una de las cadenas (siendo el primer número menor o igual al segundo, y siendo los valores posibles del 5 al 50).
- La tercera línea es la cantidad de cadenas de ADN a analizar (mayor que 0 y menor que 50).
- Luego una línea por cada cadena de ADN (verificar el rango de cada cadena y si tiene letras diferentes de: ["A","C","G","T"].)

** Cada vez que no se cumpla la estructura en la entrada el programa debe terminar e indicar el error

Luego de validar la entrada, se almacena en una lista los strings de cadenas de ADN en el mismo orden luego se procesa cada cadena para obtener su medida de desorden, y generamos una nueva lista donde estén ordenadas de menor a mayor desorden, en esta lista evaluamos si se repite la misma medida de desorden en diferentes cadenas por lo que ordenaremos estos grupos uno por uno, dentro de ellos irán primero las cadenas de longitud mayor, y si coinciden en longitud, se deja primero la más alfabética. Finalmente se imprimen de forma ordenadas todas las cadenas de menor a mayor desorden.

Algunas funciones a implementar:

- Comprobación de entrada válida
- Obtención del valor de medida de desorden de un string
- Entrada: Una lista de strings de cadenas de ADN
 - Le damos a cada valor de la cadena ordenada ["A","C","G","T"] un valor numérico del 0 al 3, para darle niveles de jerarquía alfabética.
 - Luego debemos analizar mediante un algoritmo de python, variable por variable, asignándole un valor a la cantidad de números menores a él que se encuentran situados a su derecha en la cadena.
 - Posteriormente se sumarán todos los valores dentro de una misma cadena y el resultado de esta será la "medida de desorden"

- Salida: Una lista con una cifra del desorden en el mismo orden de las cadenas de strings dadas de input
- Ordenar cadenas con igual medida de desorden.