

Examen Unidad 3 Estructura de Datos

Exam, Form: A

Name: _____

Student Number: _____

TA: _____

Date: _____

Section 1. Utilización de lenguaje C++

En algunos problemas del área de bioinformática se requiere la utilización de secuencias de ADN, ARN o proteínas para la realización de sus investigaciones. Debido a esto se han creado diferentes formatos de archivos para representar esta información. Uno de estos formatos, es el formato phylip, el cual en el primer renglón indica el número de secuencias y su longitud, en los renglones subsecuentes se encuentran cada una de las secuencias, nombre del individuo, seguido por su secuencia biológica. un ejemplo es el siguiente:

```
5 8
Nicotiana GTTGACAA
Galphimia GTTGACAG
Oenothera GTTGACAA
Victoria GTTGACAA
Cypirapea ????????
```

1. Elabore un programa que obtenga la información del archivo de texto que se le proporcionará, y que realice una representación de esta información para su posterior utilización, por ejemplo arrays, estructuras, string, etc.

Tener en cuenta lo siguiente:

- El archivo de texto se enviará como argumento para la ejecución del programa
- Validar que el archivo exista y se pueda abrir
- Lo no previsto en esta redacción se aclarará en el examen

Section 2. Listas

Una vez que logró leer la instancia que contiene las secuencias de ADN, ARN o proteínas, realizar lo siguiente:

2. Realizar los ajustes necesarios al programa anterior con el fin de incluir los siguientes requerimientos:
 - (a) Incluir a su programa un método o función de tal manera que con las secuencias del archivo se formen pares y se evalúen de la siguiente manera:
 - Se genera un nodo resultante del tipo *node
 - Se comparan posición por posición las dos secuencias con el fin de generar la información del nodo resultante, donde para cada posición si los caracteres son diferentes en esa misma posición en el nodo resultante ponemos la unión de los dos caracteres e incrementamos el score en 1, si son iguales, ponemos el carácter coincidente y el score no se incrementa
 - los nodos resultantes van a quedar unidos a través de una lista simple
 - (b) Agregar un método o función que recorra la lista resultante y que al mismo tiempo calcule el score total de la lista y muestre el resultado

Section 3. Pilas y Colas

Modificar la lista resultante del ejercicio anterior de tal modo que el nodo de la lista contenga la siguiente información:

3. El nodo se debe estructurar de la siguiente manera, siendo opción del estudiante utilizar la clase nodo adaptado o la siguiente estructura:
4. En esta ocasión se va a generar un nodo para cada secuencia, es decir ya no se van a calcular por pares, con la diferencia que la secuencia (seq) deberá convertirse a binario, utilizando el siguiente código
5. Convierta la lista anterior en un árbol binario bajo las siguientes consideraciones:
 - Cada secuencia de la instancia de la secuencia se deberá representar en una hoja del árbol
 - los nodos internos se construirán evaluando pares de secuencias tomando en cuenta las consideraciones de evaluación correspondientes a la unidad 2
 - Cada nodo (raíz, nodo interno, hojas) deberá tener 3 apuntadores: padre, izquierdo, derecho
 - Los nodos hojas correspondientes a las secuencias de la instancia tienen score 0
 - Los nodos internos deberán almacenar el score local y total: es decir el resultado de la comparación de sus descendientes izquierdo y derecho más el acumulado de sus descendientes
 - los nodos hojas tienen NULL en izquierdo y derecho
 - El nodo raíz tiene NULL en padre
6. Cada nodo se va a insertar a la lista utilizando los criterios de inserción y eliminación de la cola, su programa debe mostrar que realiza las operaciones.