

Final Introducción a la Programación 1 - 05/07/2024

Consigna: Crear un E.T.L para análisis de secuencia de ADN.

Objetivo: El objetivo de este ejercicio es crear un proceso E.T.L. de Extracción (Extract), Transformación (Transform) y Carga (Load) para analizar una secuencia de ADN. Su tarea consiste en calcular el complemento de ARN de una secuencia de ADN dada. Tanto las hebras de ADN como las de ARN son secuencias de nucleótidos.

- Los cuatro nucleótidos que se encuentran en el ADN son adenina (A), citosina (C), guanina (G) y timina (T).
- Los cuatro nucleótidos que se encuentran en el ARN son adenina (A), citosina (C), guanina (G) y uracilo (U).
- Dada una cadena de ADN, su cadena de ARN transcrita se forma reemplazando cada nucleótido por su complemento:
 - G -> C
 - C -> G
 - T -> A
 - A -> U

Método Extract:

Definición: El método Extract se encarga de recibir una secuencia de ADN en formato String.

Acciones a realizar: Separar la secuencia de ADN en fragmentos de 3 caracteres cada uno. Si no se logra armar un fragmento de 3 caracteres, se deberá descartar los caracteres restantes.

Retorna: La lista de fragmentos extraídos.

Método Transform:

Definición: El método Transform se encarga de tomar los fragmentos de ADN extraídos y aplicar una transformación.

Acciones a realizar:

- Reemplazar cada nucleótido con su complemento correspondiente.
 - Las reglas de reemplazo son las siguientes:
 - G -> C
 - C -> G
 - T -> A
 - A -> U
- Además, se debe contar la cantidad de cada nucleótido presente en el ARN calculado.

Retorna: La una lista compuesta de la siguiente manera:

- En la primera posición de la lista, la lista con los ARN.
- En la segunda posición de la lista, la cantidad de adenina.
- En la tercera posición de la lista, la cantidad de citosina.

- En la cuarta posición de la lista, la cantidad de guanina.
- Y por último, en la quinta posición de la lista, la cantidad de uracilo.

Método Load:

Definición: El método Load se encarga de recibir la lista dada por el método transform y imprimirlos en pantalla y guardarlos en sus respectivas variables.

Acciones a realizar:

- Imprimir en pantalla el siguiente mensaje:
 - ARN obtenido: {ARN} # Donde {ARN} es el ARN obtenido y calculado.
 - Cantidad de adenina: {a} # Donde {a} es la cantidad de adenina calculada.
 - Cantidad de citosina: {c} # Donde {c} es la cantidad de citosina calculada.
 - Cantidad de guanina: {g} # Donde {g} es la cantidad de guanina calculada.
 - Cantidad de uracilo: {u} # Donde {u} es la cantidad de uracilo calculada.
- Guardar en las variables globales todo lo recibido en el método load (guardar el ARN y todas las cantidades de nucleótidos contados)

Armado del ETL:

Además de crear los métodos anteriores y codificarlos, también **deberán usar el ETL**, para que dado un ADN original, logre imprimir el ARN.

Ejemplo:

Dado el ADN "ATCGGACTAATCGTT" el sistema ETL imprime lo siguiente:

```
ARN obtenido: ['UAG', 'CCU', 'GAU', 'UAG', 'CAA']
Cantidad de adenina: 5
Cantidad de citosina: 3
Cantidad de guanina: 3
Cantidad de uracilo: 4
```

Lo único provisto y no se puede cambiar es lo siguiente:

```
# variables globales
arn = []
count_a = 0
count_c = 0
count_u = 0
count_g = 0

# E.T.L start
adn_secuencia = "ATCGGACTAATCGTT"

# TO DO, IMPLEMENT
```