

UNIDAD TEMÁTICA 2 – ARBOLES GENÉRICOS, TRIES y ARBOLES B

Prácticos Domiciliarios Individuales - 6

Referencia: Tries de Sufijos, Trabajos de Aplicación 2,3, 4 y 5, Prácticos domiciliarios 1-5.

EJERCICIO 1

Se desea utilizar un árbol de sufijos para reconocer patrones en una secuencia de ADN.

En este caso, el “alfabeto” está integrado por los caracteres **“a”, “c”, “g” y “t”**

Se desea:

- 1- Diseña una estructura de datos apropiada para representar un nodo de este trie, para este alfabeto.
- 2- construye un **árbol de sufijos** para la siguiente secuencia de ADN :

“ccttgacttgc”

- 3- Escribe en pseudocódigo un algoritmo que permita indicar si un cierto patrón (una string de caracteres del alfabeto definido) se encuentra contenido en el texto
- 4- Demuestra la ejecución del algoritmo manualmente, utilizando el trie de sufijos creado en **2.** , para el patrón **“ct”**
- 5- Extiende el algoritmo para indicar **todas** las posiciones del texto en que comience el patrón.
- 6- Demuestra la ejecución (manualmente), para el patrón **“tg”**
- 7- **¿Cuál es el orden del tiempo de ejecución de este algoritmo? (justifícalo)**
- 8- **Desarrolla y dibuja el TRIE COMPRIMIDO correspondiente.**
- 9- **¿CÓMO EXTENDERÍAS LA ESTRUCTURA DEFINIDA, PARA SOPORTAR CUALQUIER ALFABETO ALFANUMÉRICO?**

EJERCICIO 2

Se desea determinar la ocurrencia de cierto marcador genómico (patrón) en una secuencia de ADN determinada.

La secuencia en cuestión se encuentra adjunta en el documento **“secuencia.txt”**

En este caso, el “alfabeto” correspondiente al problema tiene los siguientes caracteres: **“a”, “c”, “g” y “t”**

Se debe:

- 1- Desarrollar **un programa java** que, partiendo de la estructura diseñada en el **Ejercicio 1** :
 - A partir de la secuencia, construya todos los sufijos correspondientes
 - Inserte todas las cadenas resultantes en un trie
 - Indique las posiciones en que aparecen los siguientes fragmentos (marcadores, patrones) en la secuencia de ADN:
 - a. ccta
 - b. ggtca
 - c. aacct