

Biblioteca de Algoritmos de String-Matching para Investigadores.

Resumen:

Dentro del campo de la bioinformática es necesario analizar y comparar largas secuencias de ADN para encontrar patrones que permitan sacar conclusiones acerca de el o los organismos que se este estudiando y contribuyan así a una mejor entendimiento de los mismos. Desde un punto de vista computacional, las secuencias de ADN pueden ser vistas de un modo más abstracto como cadenas de caracteres, se desea aplicar algoritmos estas cadenas para extraer información útil y relevante. Estos algoritmos de manipulación de cadenas son la herramienta necesaria para el análisis de secuencias.

Algunos problemas podrían parecer sencillos en un principio, por ejemplo buscar un patrón en un texto se puede resolver fácilmente por fuerza bruta, pero esta no es una estrategia plausible para aplicaciones serias con cadenas demasiado largas. Es por ello que a lo largo de los años se han desarrollado y perfeccionado algoritmos de procesamiento de cadenas que son útiles en la bioinformática. Mi propósito con este proyecto es implementar algunos de estos algoritmos y ponerlos a disposición de especialistas en la materia, sin importar si tienen conocimientos de programación o no.

Antecedentes:

Justificación:

Este proyecto se llevará a cabo con la finalidad de integrar un entorno de trabajo que ponga a disposición de los especialistas y estudiantes del campo de la bioinformática herramientas, en este caso algoritmos de procesamiento de cadenas, que faciliten su trabajo y que además pueda ampliarse y proporcione una base para implementar otros otros proyectos.

Objetivos:

- a) **Objetivo general:** Desarrollar una biblioteca de funciones con utilidades especialmente diseñadas para trabajar con secuencias de ADN así como un entorno de trabajo que con la capacidad de hacer uso de dichas funciones de manera intuitiva y visual.
- b) **Objetivos específicos:** La biblioteca de funciones será escrita con el propósito de que se utilizada para la implementación de nuevos programas de análisis y alineamiento de secuencias. Por otro lado el entorno de trabajo visual tiene como objetivo facilitar las tareas cotidianas de aquellos que trabajen usualmente con los algoritmos proporcionados por la biblioteca.

c) Productos:

- El primer producto es la biblioteca de funciones junto con su documentación correspondiente. La biblioteca estará escrita en el lenguaje C++ para mantener un diseño intuitivo orientado a objetos y al mismo tiempo conseguir un manejo eficiente de los recursos computacionales disponibles. Se procurará que la biblioteca esté lo más completa posible y los algoritmos sean eficientes tanto en tiempo como en memoria. Además para facilitar la ampliación y mejora de esta API, esta será de código abierto.
- El segundo producto es un entorno de trabajo visual basado en el IDE “Eclipse”, el cuál tendrá las funcionalidades necesarias para hacer un uso completo de las funciones en la API antes mencionada de una manera intuitiva y que permita trabajar con distintos formatos de archivos como *.gb* y *.fasta*. Este programa será también de código abierto.

Metodología:

El primer paso es investigar los algoritmos y técnicas de programación que se utilizan en el alineamiento de secuencias y otros problemas similares del campo al que se dedica este proyecto, para luego identificar aquellos que deban codificarse e integrarse en nuestra API.

Una vez identificados los algoritmos que se deben incluir es estudiar cada uno de ellos para entender su complejidad computacional y ser capaz de implementarlos.

El siguiente paso es hacer el diseño y la documentación de la biblioteca, pensando en las interfaces y las clases que serán necesarios, a partir de este punto tendremos en cuenta los detalles de la implementación.

Después de eso se procederá a la codificación, es decir, a la implementación de las interfaces. Con este paso realizado solamente resta publicar la biblioteca en Internet.

En el caso del entorno de trabajo para investigadores el primer paso es el diseño del sistema teniendo en cuenta la comunicación con API y pensando en cuál es la forma de hacer uso de ella de una manera sencilla e intuitiva.

A partir del diseño se llevará a la fase de desarrollo del entorno de trabajo y será publicado también en Internet.

Referencias: