

Universidad de Buenos Aires
Laboratorio de Sistemas Embebidos
Especialización en Inteligencia Artificial

Computación, algoritmos y estructuras de datos

Docente: Camilo Argoty

TP 3: Alineación de Secuencias Genéticas mediante el Algoritmo de Needleman–Wunsch

Descripción General

En este ejercicio se aplicarán conceptos fundamentales de algoritmos y estructuras de datos en el contexto de la bioinformática, específicamente en la **alineación global de secuencias de nucleótidos**. El objetivo es comprender cómo los algoritmos permiten comparar secuencias biológicas para identificar similitudes, mutaciones y relaciones evolutivas.

Parte 1: Conceptos Teóricos

El estudiante deberá investigar y redactar un documento (1–2 páginas) explicando:

1. Qué es una secuencia de nucleótidos y por qué su comparación es importante en biología molecular.
2. Qué es una alineación de secuencias, diferenciando entre:
 - Alineación global y alineación local.
 - Coincidencias (matches), desajustes (mismatches) y huecos (gaps).
3. El modelo de puntuación usado para alinear secuencias:
 - Reglas de puntuación para coincidencias y desajustes.
 - Penalización por apertura y extensión de huecos.
4. Una descripción conceptual del algoritmo de **Needleman–Wunsch**, incluyendo:
 - Construcción de la matriz de programación dinámica.
 - Ecuación de recurrencia.
 - Proceso de *traceback* para recuperar el alineamiento óptimo.

El documento debe incluir al menos una referencia bibliográfica confiable.

Parte 2: Implementación del Algoritmo

El estudiante deberá implementar en el lenguaje de programación de su preferencia (se recomienda Python) un programa que:

1. Reciba varias parejas de secuencias de nucleótidos, por ejemplo:

```
("GATTACA", "GCATGCU")
("ACGT", "ACCT")
("ATGCT", "AGCT")
```
2. Implemente desde cero el algoritmo de Needleman–Wunsch utilizando el siguiente esquema de puntuación:
 - Match = +1
 - Mismatch = -1
 - Gap = -2
3. Genere como salida, para cada pareja de secuencias:
 - La matriz de puntuación completa.
 - El alineamiento global óptimo entre ambas secuencias.
 - El puntaje final del alineamiento.

Entrega

El estudiante deberá entregar:

- Un **informe en PDF** con la parte teórica y una explicación breve del código implementado.
- El **código fuente** debidamente comentado.
- Una captura o impresión de los alineamientos generados por el programa.