**Estructura de datos y algoritmos**

**Taller 3**

**Ejercicio 2.** *Distancia de Hamming.*

**2.** Dé la complejidad en tiempo y la complejidad espacial del algoritmo.

La complejidad en tiempo de la función Cargar\_secuencia\_desde\_Archivo es O(n), es lineal donde ‘n’ es el número de caracteres en el archivo.  
   
 La complejidad en tiempo de la Distancia\_Hamming es O(n) donde ‘n’ es la longitud de la secuencia y la complejidad espacial de las funciones es O(1) porque solo almacena la distancia entre las secuencias.

**3.** ¿Cuál es el principal inconveniente del uso de esta distancia en el contexto de secuencias ADN?

Teniendo en cuenta que Hamming no cuenta inserciones, eliminación o sustitución, En el contexto del ADN estas operaciones deben estar en cuenta. Por esto no es la mejor opción.

**Ejercicio 3.** *Descomposición en k-mers y función hash.*

**2.** ¿Cuántos *k*-mers se generan por una secuencia de longitud |*S*| ?

Para una secuencia de longitud |s| se generarán |s|-k+1 k-mers donde K es la longitud del k-mer.

**3.** ¿ Cual es la complejidad en tiempo del algoritmo de descomposición en *k*-mers de una secuencia *S* dada ?

La complejidad en tiempo del algoritmo de descomposición en k-mers de una secuencia S dada es O(|S| - k + 1), donde |S| es la longitud de la secuencia y k es la longitud del k-mer. Esto se debe a que recorre la secuencia una vez y para cada posición se extrae un k-mer de longitud k.

**Ejercicio 4.** *Similitud genética y índice de Jaccard.*

Dé los resultados sobre el conjunto de datos dado en forma de matriz de similitud para K = 7 y K = 11. Indique el tiempo de ejecución para cada K.

|  |  |  |
| --- | --- | --- |
| dataset\k | 7 | 11 |
| dataset xylella | 3.1529338359832764 segundos | 12.15759825706482 segundos |