**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 del algoritmo es O(n), donde n es la longitud de las cadenas a comparar.

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

El principal inconveniente del uso de esta distancia en el contexto de secuencias de ADN es que no toma en cuenta la relación entre los caracteres. Por ejemplo, si se tienen dos secuencias de ADN que difieren en una sola posición, pero la diferencia es entre dos bases nitrogenadas complementarias, la distancia de Hamming será 1, a pesar de que la diferencia sea mínima.

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

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

|S| - k + 1 porque se generan todas las subcadenas de longitud k

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

O(|S|) porque se recorre la secuencia una vez

**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 | 99.18897581100464 seconds | 158.47192883491516 seconds |