**Estructura de datos y algoritmos**

**Taller 3**

**Jonatan David Vergara Suarez**

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

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

Su complejidad es de O(3n) = O(n) esto debido a que su complejidad y tiempo depende de la longitud de las cadenas que se están comparando.

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

Que por mucho que el resultado de distancia de la secuencia se pequeño, esta con esos cambios esta podría codificar cosas completamente diferentes

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

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

Por una secuencia de S con k-mers de longitud k, se generan S-k+1 k-mers

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

O(n-k+1) = O(n)

La longitud del algoritmo depende enteramente de la longitud de la cadena que se quiere descomponer en k-meros y entre el tamaño de los k-meros

**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 xylefa8416( secuencia 2 y 3) |  |  |