**Estructura de datos y algoritmos**

**Taller 3**

* **Santiago Torres**
* **Juan David Orozco**

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

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

En tiempo O(n), y en complejidad O(n).

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

No todas las secuencias tienen la misma distancia y por ello el algoritmo tiende a fallar. Además no tiene en cuenta información como mutaciones o estructura.

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

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

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

**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 |  |  |