**Estructura de datos y algoritmos**

**Taller 3**

**Rafael Andrés Salcedo Valdivieso**

**Tomas Barrios Guevara**

**Juan Pablo Aponte Veloza**

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

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

El algoritmo de distancia de Hamming recorre las dos secuencias de entrada una vez, comparando cada posición. Como la longitud de ambas secuencias es la misma, la complejidad en tiempo es lineal, es decir, O(n), donde n es la longitud de las secuencias de entrada.

El algoritmo de distancia de Hamming no tiene que almacenar información adicional que dependa de la entrada, aparte de las secuencias de entrada en sí. Por lo tanto, la complejidad espacial es constante O(1).

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

Solo cuenta el número de posiciones en las que los nucleótidos difieren entre las secuencias, sin tener en cuenta inserciones, eliminaciones o sustituciones de nucleótidos. En la realidad, las mutaciones genéticas pueden implicar tales cambios

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

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

Se generan ∣S∣−k+1 k-mers

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

Como se deben generar ∣S∣−k+1 k-mers, la complejidad en tiempo sería O(∣S∣⋅k). Sin embargo, si consideramos k como un valor constante, la complejidad en tiempo sería O(∣S∣).

**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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| k=7 | xylefa8416 | xylefa8417 | xylefaco6c | xylefaco32 | xylefaco33 | xylefacodi |
| xylefa8416 | 1,0000 | 0,9892 | 0,9909 | 0,9909 | 0,9813 | 0,9909 |
| xylefa8417 |  | 1,0000 | 0,9983 | 0,9983 | 0,9881 | 0,9983 |
| xylefaco6c |  |  | 1,0000 | 1,0000 | 0,9897 | 1,0000 |
| xylefaco32 |  |  |  | 1,0000 | 0,9897 | 1,0000 |
| xylefaco33 |  |  |  |  | 1,0000 | 0,9897 |
| xylefacodi |  |  |  |  |  | 1,0000 |

Tiempo de ejecución: 8.555184125900269 segundos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| k=11 | xylefa8416 | xylefa8417 | xylefaco6c | xylefaco32 | xylefaco33 | xylefacodi |
| xylefa8416 | 1,0000 | 0,5131 | 0,5337 | 0,5317 | 0,4793 | 0,4321 |
| xylefa8417 |  | 1,0000 | 0,4628 | 0,4617 | 0,4951 | 0,4828 |
| xylefaco6c |  |  | 1,0000 | 0,8584 | 0,4662 | 0,4444 |
| xylefaco32 |  |  |  | 1,0000 | 0,4691 | 0,4445 |
| xylefaco33 |  |  |  |  | 1,0000 | 0,4537 |
| xylefacodi |  |  |  |  |  | 1,0000 |

Tiempo de ejecución: 40.57579827308655 segundos