Analisis de algoritmos de ordenamiento

Bejar Merma Ángel Andrés
Universidad Nacional de San Agustin
andresbjar97@gmail.com
Ciudad de Arequipa

Resumen

El trabajo consiste en un análisis del rendimiento de un conjunto de algoritmos en función de su tiempo de ejecución con diferentes conjuntos de datos tanto para c++ /c, como para Python. Se prueban, comparan y concluye qué algoritmo es mejor para conjuntos de datos pequeños, promedio, veremos el peor caso, caso promedio y mejor.

1. Introducción

La ordenación es una de las cuestiones importantes en computación el problema de ordenación es importante porque de eso depende otros algoritmos los que los hacen mas o menos eficientes uno de de estos algoritmos es el algoritmo de búsqueda .

El concepto clave es algoritmo ,un algoritmo se puede definir como una secuencia de pasos bien definidos para resolver un problemas . El algoritmo toma una entrada y proporciona una salida.[2] Se han propuesto varios algoritmos de ordenación con diferentes restricciones, p. Ej. número de iteraciones (bucle interno, bucle externo), complejidad y problema de consumo de CPU.

2. Marco Teórico

Los algoritmos de ordenamiento se dividen en dos categorías bien diferenciadas:el Bubble Sort, Insertion Sort ,y Selection Sort en la categoría de O(n^2), mientras que Quick Sort y Merge Sort O(n log n).La descripción de cada uno de ellos a continuación.

a) **Bubble sort**

El algoritmo de ordenación básico es Bubble Sort. Compara dos elementos adyacentes y realiza una operación de intercambio . Esto también se denomina algoritmo de ordenación por comparación.Una de sus ventajas es sus fácil implementación .

b) **Heap sort**

La ordenacion de montón en español , es una técnica de ordenación basada en comparaciones ,basada en la estructura de datos binary heap. Es similar a ordenamiento por selección donde primero encontramos el elemento mínimo y colocamos el elemento mínimo al principio. Repetimos el mismo proceso para el resto de elementos[5].

c) **Insertion sort**

El ordenamiento por inserción (insertion sort en inglés) es una manera muy natural de ordenar para un ser humano, y puede usarse fácilmente para ordenar un mazo de cartas numeradas en forma arbitraria. Requiere O(n^2) operaciones para ordenar una lista de n elementos.

Inicialmente se tiene un solo elemento, que obviamente es un conjunto ordenado. Después, cuando hay k elementos ordenados de menor a mayor, se toma el elemento k+1 y se compara con todos los elementos ya ordenados, deteniéndose cuando se encuentra un elemento menor (todos los elementos mayores han sido desplazados una posición a la derecha) o cuando ya no se encuentran elementos (todos los elementos fueron desplazados y este es el más pequeño). En este punto se inserta el elemento k + 1 debiendo desplazarse los demás elementos[6].

d) **Shell sort**

ShellSort es principalmente una variación de Insertion Sort. En la ordenación por inserción, movemos los elementos solo una posición adelante. Cuando un elemento debe moverse mucho más adelante, se involucran muchos movimientos. La idea de shellSort es permitir el intercambio de elementos lejanos. En shellSort, hacemos que la matriz esté ordenada por h para un valor grande de h. Seguimos reduciendo el valor de h hasta que se convierte en 1. Se dice que una matriz está ordenada por h si todas las sublistas de cada elemento h están ordenadas[4].

e) **Merge sort**

Este es un algoritmo de divide y conquista, con la ventaja de fusionar listas con nuevas listas ordenadas. La complejidad del peor caso de la ordenación por fusión es O (nlog n), ya que podría usarse para conjuntos de datos grandes y peores. La ordenación por fusión utiliza los siguientes tres pasos [8]

1) **Divide:**Si el tamaño de la matriz es mayor que 1, divídalo en dos subarreglos iguales de la mitad del tamaño.
2) **Conquista:** Ordenar ambas subarreglos por recursividad
3) **Fusiona:** Combine ambas arreglos ordenadas en uno de tamaño original. Esto le dará un arreglo ordenada completo.

Merge Sort es más adecuada para entradas grandes ,pero usa más memoria en comparación con otros algoritmos de divide y conquista.

f) **Quick sort**

Este método es una mejora sustancial del método de intercambio directo y recibe el nombre de Quick Sort, por la velocidad con la que ordena los elementos del arreglo. Quicksort es un algoritmo basado en la técnica de divide y vencerás, que permite, en promedio, ordenar n elementos en un tiempo proporcional a n log n.

3. Metodología

Lo que se hizo fue probar los algoritmos de ordenamiento en Python utilizando Google colab y el entorno de ejecución GPU. La maquina que nos asigno google fue una Tesla k80 se intento reiniciar el entono para obtener una tesla t 40 , pero fue inútil.

Para la pruebas en c++ La computadora que utilizo es una intel i5 de quinta generación con 2 cores ,8 gb de memoria ram

Los algoritmos escogidos son los siguientes:

- Bubble sort
- Heap sort
- Insertion sort
- Selection sort
- Shell sort
- Merge sort
- Quick sort

En python se hizo una prueba con 5 000 elementos

En el lenguaje de c se hizo una prueba con 50 000 elementos

En el lenguaje de c++ se hizo una prueba con 5 0000 elementos

Y finalmente se comparo el tiempo total que toma los algoritmos en c y python al tener una entrada de 1000 001 de elementos

algorithm	¿stable?	best time	average time	worst time	extra memory
selectionsort	no	$O(n^2)$	$O(n^2)$	$O(n^2)$	$O(1)$
insertionsort	yes	$O(n)$	$O(n^2)$	$O(n^2)$	$O(1)$
shellsort	no	$O(n*log(n))$	$O(n^{1.25})^{\dagger}$	$O(n^{1.5})$	$O(1)$
quicksort	no	$O(n*log(n))$	$O(n*log(n))$	$O(n^2)$	$O(log(n))$
mergesort	yes	$O(n*log(n))$	$O(n*log(n))$	$O(n*log(n))$	$O(n)$
heapsort	no	$O(n)$	$O(n*log(n))$	$O(n*log(n))$	$O(1)$

Figura 1: Tabla complejidades de algoritmos[7]

4. Resultados

```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 470.74       Driver Version: 460.32.03    CUDA Version: 11.2     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla K80           Off  | 00000000:00:04.0 Off |                    0 |
| N/A   40C    P0    58W / 149W |    121MiB / 11441MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
```

Figura 2: Gpu TeslaK80

```
Architecture:        x86_64
CPU op-mode(s):      32-bit, 64-bit
Byte Order:          Little Endian
CPU(s):              2
On-line CPU(s) list: 0,1
Thread(s) per core:  2
Core(s) per socket:  1
Socket(s):           1
NUMA node(s):        1
Vendor ID:           GenuineIntel
CPU family:          6
Model:               63
Model name:          Intel(R) Xeon(R) CPU @ 2.30GHz
Stepping:            0
CPU MHz:             2299.998
BogoMIPS:            4599.99
Hypervisor vendor:   KVM
Virtualization type: full
L1d cache:           32K
L1i cache:           32K
L2 cache:            256K
L3 cache:            46080K
NUMA node0 CPU(s):   0,1
Flags:               fpu vme de pse tsc msr pae mce cx8
```

Figura 3: Especificaciones Entorno Colab

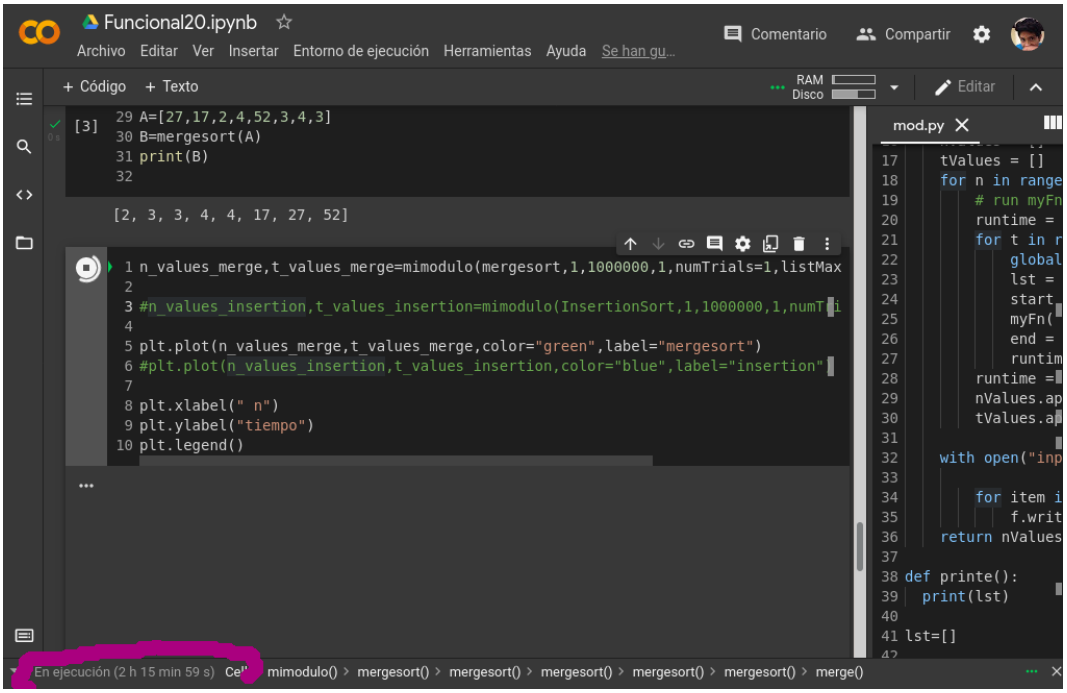

Figura 4: Momento ejecutar codigo en python

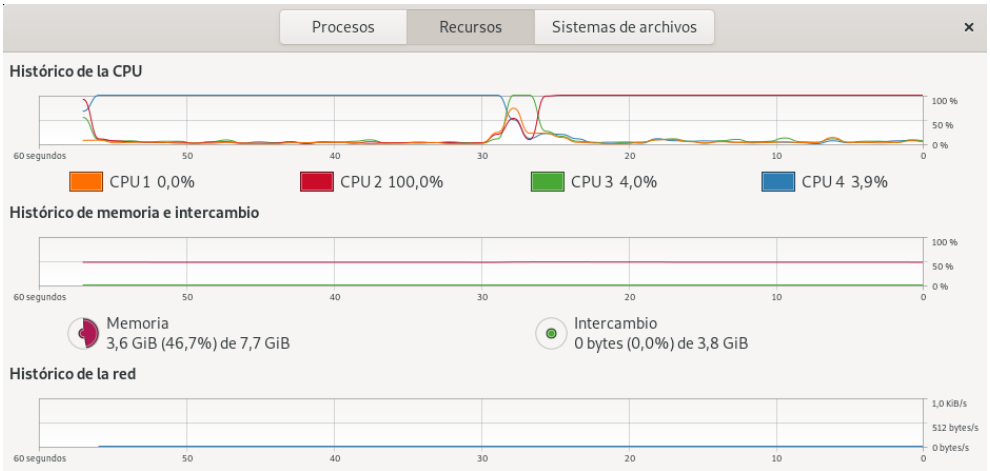

Figura 5: Uso de una core al 100 por ciento

cuaderno de trabajo [1]

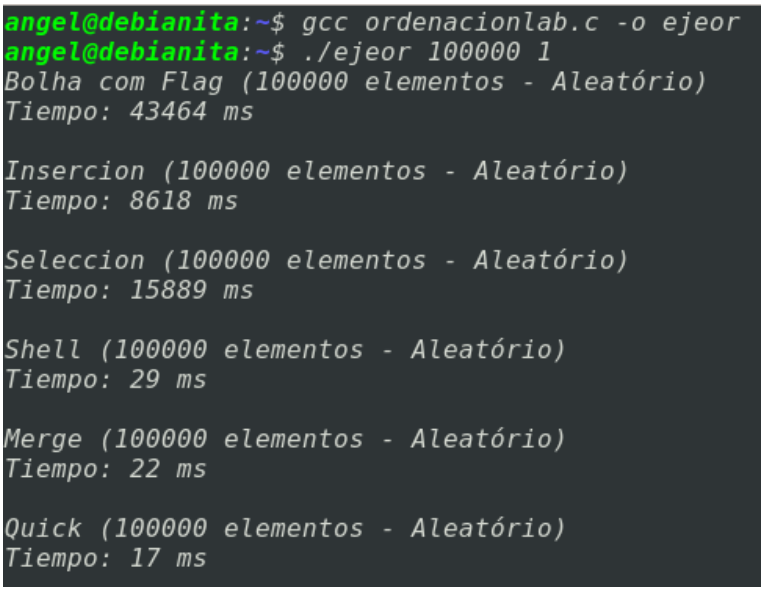

Figura 6: Ejecutado en c con 100 mil elementos aleatorios

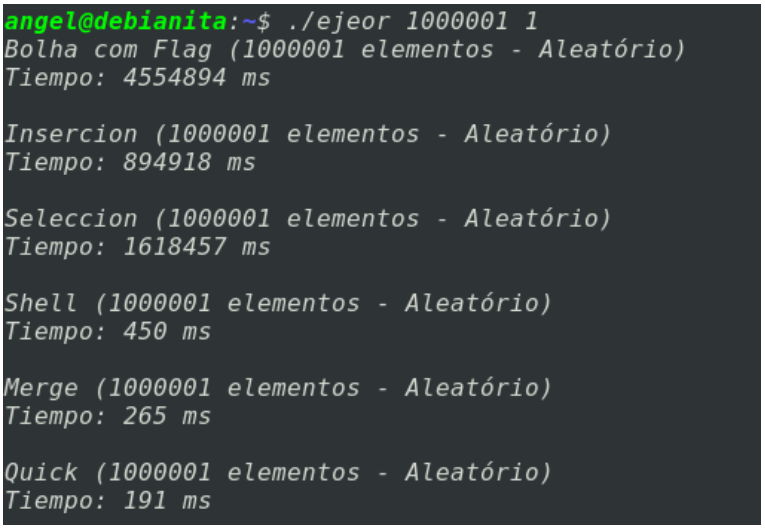

Figura 7: Con 1 millon de elementos

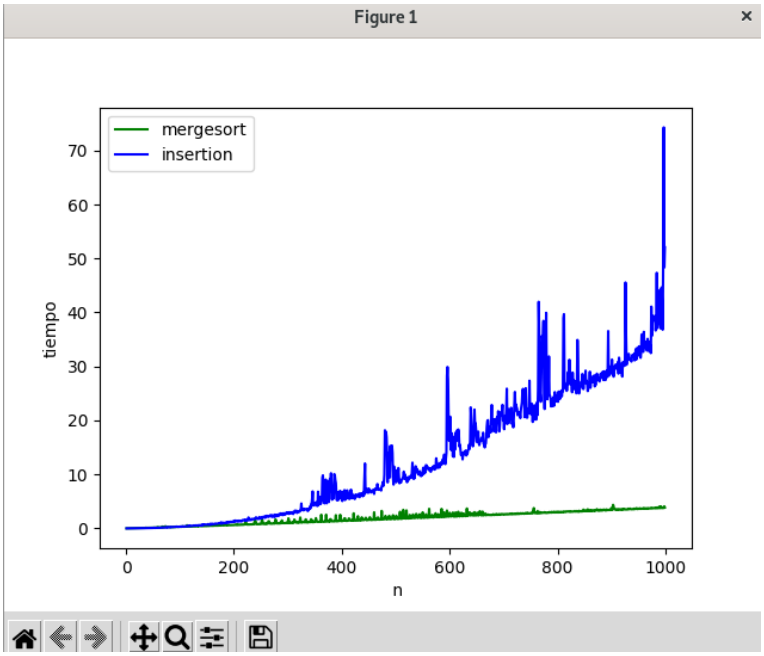

Figura 8: Con 1000 elementos

5. Conclusiones

La ejecución de los algoritmos de python en Google Colab el proceso tardo mas de 8 horas, Al terminar de ejecutarse se hizo la prueba en la Pc y tardo 1 hora menos ,al ordenar 1000 000 elementos de un arreglo.

El lenguaje de programación c++ el mas optimo en la ordenación a diferencia de python en c++ utilizo 1 hora con 57 minutos siendo el algoritmo que tomo mas tiempo Bubble Sort y el de menor Quick sort,En python tardo 7 horas aproximadamente.

Es importante darse cuenta que la toma de tiempo extra en python se debido gran medida a la creación de un archivo con los elementos de entrada . El ordenamiento por selección realiza menos cambios (intercalación) que la inserción, ya que solo hay un cambio por iteración, es decir, en el Selection Sort hace n intercambios en total.

La ordenación por inserción, por otro lado, realiza menos comparaciones que la ordenación por selección, ya que el elemento que se va a insertar de forma ordenada no siempre tiene que llegar hasta el final. De hecho, esto solo ocurre en el peor de los casos, donde la matriz se ordena en orden inverso. La ordenación por selección necesita comparar todos los elementos restantes cada vez para determinar quién es el más pequeño.

En teoría, ambos pertenecen a la misma clase de complejidad, O (n^2).En la práctica, la ordenación por inserción funciona mejor que la ordenación por selección.

Finalmente, la ordenación por selección no se considera un algoritmo eficiente para entradas grandes. Hay alternativas O (nlogn) , como Quick Sort y Merge Sort, así como alternativas lineales como Counting Sort.

La ordenación por fusión o Merge Sort es más adecuada para casos grandes y peores, pero usa más memoria en comparación con otros algoritmos de división y conquista.

[3]

La complejidad de Quick Sort en los casos mejores y medios es O(nlogn) y en el peor de los casos O(n^2) . La complejidad media del tiempo de caso es O(nlogn) que es un poco costosa si se utiliza para grandes conjuntos de datos, también en el peor de los casos la complejidad es O(n^2), lo que lo hace impráctico para los peores conjuntos de datos grandes.

6. Discusión

Algunos algoritmos de ordenacion son mejores para una entrada pero no para otra. Quick sort es mejor para ordenar datos no ordenados (no necesariamente datos aleatorios). Bubble Sort , por

ejemplo, puede superar Quick Sort cuando los datos están ordenados o casi ordenados.

De los algoritmos que comparten la misma clase de orden, es de considerar el valor de la constante de orden big-O. Aunque puede tener un tiempo cuadrático en el peor de los casos, Quicksort a menudo se considera el algoritmo más rápido en matrices aleatorias; la implicación es que tiene la constante de orden menor que, por ejemplo, Heapsort.

El algoritmo Shellsort utiliza los llamados incrementos de Hibbard. Aunque Shellsort no es teóricamente tan rápido como otros, sigue siendo comparable en velocidad cuando las matrices no son enormes y se emplean incrementos de ordenación mejorados.[7].

Referencias

[1] Bejar Merma Angel Andres. Cuaderno de trabajo. https://colab.research.google.com/drive/1FZjcnchMuRvmnAxCTwf54DTL1DiJmsNm?usp=sharing, 2021. [Online; accessed read].

[2] Varinder Kumar Bansal, Rupesh Srivastava, and Pooja. Indexed array algorithm for sorting. In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pages 34–36, 2009.

[3] Arthur Brunet. Estructura de datos. https://joaoarthurbm.github.io/eda/posts/selection-sort/, 2021. [Online; accessed read].

[4] geekforgeek. geekforgeek. https://www.geeksforgeeks.org/shellsort/, 2021. [Online; accessed read].

[5] heapsort. geekforgeek. https://www.geeksforgeeks.org/heap-sort/, 2021. [Online; accessed read].

[6] insertionsort. wikipedia. https://es.wikipedia.org/wiki/Ordenamiento_por_inserci%C3%B3n, 2021. [Online; accessed read].

[7] Robert M. Kline. Comparacion de algoritmos en java. https://www.cs.wcupa.edu/rkline/ds/shell-comparison.html#comparisons, 2021. [Online; accessed read].

[8] Wikipedia. Definición de merge sort. https://es.wikipedia.org/wiki/Ordenamiento_por_mezcla, 2021. [Online; accessed].