

Trabajo Práctico 1

[75.29/95.06] Teoría de Algoritmos I
Primer cuatrimestre de 2018

Grupo MinMax

Alumno	Padrón	Mail
del Mazo, Federico	100029	delmazofederico@gmail.com
Djeordjian, Esteban Pedro	100701	edjeordjian@gmail.com
Kristal, Juan Ignacio	99779	kristaljuanignacio@gmail.com
Raveszani, Nicole	101031	nraveszani@gmail.com

Índice

1. Cálculo Empírico de Tiempos de Ejecución	1
1.1. Complejidad Teórica	1
1.1.1. Selección	1
1.1.2. Inserción	1
1.1.3. Mergesort	1
1.1.4. Quicksort	2
1.1.5. Heapsort	2
1.2. Comparación	3
1.2.1. Peores tiempos	3
1.2.2. Tiempos promedio	3
1.3. Peores sets	3
1.3.1. Selección	3
1.3.2. Inserción	3
1.3.3. Quicksort	4
1.3.4. Heapsort	4
1.3.5. Mergesort	4
1.4. Tiempos de ejecución	4
1.5. Comparación con valores teóricos	4
2. Algoritmo Gale-Shapley	5
2.1. Algoritmo	5
2.2. Demostraciones	5
2.2.1. Tiempo polinómico	5
2.2.2. Matching estable	5
2.3. Ejecución	6
A. Ejecución y compilación	7
B. Ejecución principal	7
B.1. Generación de sets aleatorios	7
B.2. Pruebas de ordenamientos	7
B.3. Medición de tiempos	7
B.4. Stable matching	7
C. Tablas de ejecución	7

1. Cálculo Empírico de Tiempos de Ejecución

1.1. Complejidad Teórica

1.1.1. Selección

El ordenamiento de seleccion consiste en....

De acuerdo a nuestra implementación¹ del ordenamiento de selección, la complejidad teóricas es:

$$T_{seleccion}(n) = O(1 + 1 + 1 + n * (1 + n + (1 + 1 + n(2) + 1) + 4)) = O(2n^2 + 9n + 3) \in O(n^2)$$

Esto viene de analizar linea por linea teniendo en cuenta:

- Hacer una asignación y comparar variables: $O(1)$
- Operadores aritméticos básicos: $O(1)$
- Sacar la longitud de un arreglo: $O(1)$
- Hacer un ciclo definido que itera por todo el arreglo: $O(n * f(n))$ siendo $f(n)$ lo que hay dentro del ciclo.
- Hallar el indice de un elemento: $O(n)$

Este ordenamiento itera siempre por el arreglo sin importar su disposicion inicial. Por lo tanto, su peor caso y su caso promedio son iguales al tiempo analizado... :

$$PeorT_{seleccion}(n) \in O(2n^2 + 9n + 3) \in O(n^2)$$

$$PromedioT_{seleccion}(n) \in O(2n^2 + 9n + 3) \in O(n^2)$$

1.1.2. Inserción

El ordenamiento de insercion consiste en... Al igual que el de seleccion, el algoritmo consiste de dos ciclos anidados que recorren el arreglo entero, por lo tanto, el tiempo esperado sera cuadratico

Este ordenamiento usa las mismas herramientas usadas en el ordenamiento de selección. Por lo tanto, con las consideraciones anteriores tomadas, queda:

$$O(1 + 1 + 1 + n(3 + 2 + 2 + 1 + n(1 + 3))) = O(3n^2 + 8n + 3) \in O(n^2)$$

Siendo la condicion de corte que en la iteracion el elemento actual no sea menor a lo iterado descendientemente, el peor caso del algoritmo es donde el arreglo inicial esta dispuesto de manera descendiente, de tal forma que el ciclo sea ejecutado sin cortes.

$$PeorT_{insercin}(n) \in O(3n^2 + 8n + 3) \in O(n^2)$$

Como es claro que encontrar en un arreglo un intervalo ordenado (suficientemente grande) es mucho menos probable que no hacerlo (es decir, para un arreglo de n elementos, de las $n!$ posibles distribuciones de los elementos, solo una estara ordenada, y con un n no demasiado grande, se cumple que $n! - 1 \approx n$). Luego, en el caso promedio se tiene que el algoritmo tambien tiene orden cuadratico.

$$PromedioT_{insercin}(n) \in O(3n^2 + 8n + 3) \in O(n^2)$$

1.1.3. Mergesort

El algoritmo Mergesort consiste en... El metodo usado es el de Division y Conquista, y por lo tanto, para calcular su complejidad teorica va a ser utilizado el Teorema Maestro

En cuanto a código se introducen nuevos metodos:

- Append y pop: $O(1)$
- Extend: $O(k)$ siendo k la cantidad de elementos agregados

¹Las complejidades de los métodos nativos de las listas de Python 3 fueron sacadas de: <https://wiki.python.org/moin/TimeComplexity>

También es de notar que este es el primer algoritmo donde usamos llamadas recursivas, por su naturaleza de división y conquista.

Aplicando el Teorema Maestro:

$$T_{mergesort}(n) = a.T_{mergesort}(\frac{n}{b}) + f(n^c)$$

La cantidad de llamadas recursivas (a) son dos. Cada llamada recursiva divide la entrada inicial a la mitad (b=2). Por último, el tiempo de las llamadas no recursivas es

$$O(1 + 1 + 1 + 4 + 1 + 2 + n(1 + 1 + 1) + 1 + \frac{n}{2} + 1) = O(\frac{9n}{2} + 11)$$

(asumiendo un n lo suficientemente grande como para que se requieran intercalar aproximadamente n elementos la mayoría de las veces), por lo que d=1. Se tiene entonces:

$$T_{mergesort}(n) = 2.T_{mergesort}(\frac{n}{2}) + f(n)$$

Como $(a = b^c)$ $(2 = 2^1)$, se tiene que:

$$T_{mergesort}(n) \in O(n \log n)$$

Texto sobre peor tiempo y tiempo promedio de mergesort

$$PeorT_{mergesort}(n) \in O(n \log n)$$

$$PromedioT_{mergesort}(n) \in O(n \log n)$$

1.1.4. Quicksort

El algoritmo Quicksort consiste en.... Al igual que en Mergesort, siendo Division y Conquista el paradigma utilizado, se usa el Teorema Maestro para calcular su complejidad.

Considerando las complejidades analizadas hasta ahora (en código no hay nada nuevo) y aplicando el Teorema Maestro se tiene:

$$T_{quicksort}(n) = a.T_{quicksort}(\frac{n}{b}) + f(n^c)$$

La cantidad de llamadas recursivas (a) son 2. Para analizar b, se debe considerar dos casos. Por un lado, en el caso promedio, b será igual a 2, ya que en un arreglo arbitrariamente grande, es mucho más probable tomar en cualquier posición a un elemento alejado de las cotas superiores e inferiores que muy próximo a ellas, por lo que la recursión dividirá el problema en “dos partes iguales” (probablemente nunca sean exactamente iguales, pero con un n arbitrariamente grande, las diferencias que puedan existir son despreciables en tanto el pivote elegido no esté lo suficientemente próximo a una cota superior o inferior). El tiempo de las llamadas no recursivas es $O(4n + 4)$, por lo que $c = 1$

Para este caso, como $(a = b^c)$ $(2 = 2^1)$, se tiene que:

$$PromedioT_{quicksort}(n) \in O(n \log n)$$

Pero existe un peor caso en donde el pivote elegido es una el menor o el mayor elemento del arreglo, dividiendo el arreglo en dos partes de 1 y $n - 1$ elementos respectivamente. Por lo tanto, usamos como peor caso un arreglo ordenado descendientemente y tomando de pivote el primer elemento del arreglo. Para este peor caso se tiene que b=1, ya los elementos quedan “a la derecha o a la izquierda” del pivote. Para este caso, donde no se puede aplicar el teorema del maestro (porque no hay división y conquista propiamente dicha) queda:

$$PeorT_{quicksort}(n) = O(n(4n + 4)) = O(4n^2 + 4n) \in O(n^2)$$

1.1.5. Heapsort

El algoritmo Heapsort consiste en....

Se introduce el concepto de cola de prioridad / Heap (explicar apenas). Los tiempos del heap de minimos implementado son...

- Heapify: $O(x)$
- Upheap: $O(x)$
- Downheap: $O(x)$
- Heappop: $O(x)$
- Heapush: $O(x)$

Considerando todo esto queda:

$$O(n + n + 1 + 1 + 1 + n(1 + \log n)) = O(3n + 3 + n \log n) \in O(n \log n)$$

Texto sobre peor tiempo y tiempo promedio de mergesort

$$PeorTheapsort(n) \in O(3n + 3 + n \log n) \in O(n \log n)$$

$$PromedioTheapsort(n) \in O(3n + 3 + n \log n) \in O(n \log n)$$

1.2. Comparación

1.2.1. Peores tiempos

En orden ascendente de eficiencia:

$$PeorTinsercion(n) \in O(4n^2 + 9n + 2) \in O(n^2)$$

$$PeorTselecccion(n) \in O(3n^2 + 8n + 3) \in O(n^2)$$

$$PeorTquicksort(n) \in O(4n^2 + 4n) \in O(n^2)$$

$$PeorTheapsort(n) \in O(3n + 3 + n \log n) \in O(n \log n)$$

$$PeorTmergesort(n) \in O(n \log n) \in O(n \log n)$$

1.2.2. Tiempos promedio

En orden ascendente de eficiencia:

$$PromedioTinsercion(n) \in O(4n^2 + 9n + 2) \in O(n^2)$$

$$PromedioTselecccion(n) \in O(3n^2 + 8n + 3) \in O(n^2)$$

$$PromedioTheapsort(n) \in O(3n + 3 + n \log n) \in O(n \log n)$$

$$PromedioTmergesort(n) \in O(n \log n) \in O(n \log n)$$

$$PromedioTquicksort(n) \in O(n \log n) \in O(n \log n)$$

Para poder terminar de ubicar en la escala a quicksort y mergesort, se pone el foco en la complejidad de su parte no recursiva: por parte de quicksort es $O(4n+4)$, y en el caso de mergesort es $O(\frac{9n}{2} + 11)$. Es decir, la diferencia fundamental en ambos casos es que en el peor caso quicksort será mucho menos eficiente.

1.3. Peores sets

1.3.1. Selección

Como se ve en el análisis, el comportamiento de este algoritmo no depende de las características del arreglo. En término más específicos, el algoritmo de selección no es un algoritmo “adaptativo”: su conducta no se ve afectada por ninguna característica del array, por lo cual realizará el mismo número de comparación entre elementos tanto en el peor caso, en el caso promedio y en el mejor caso (considerando una longitud de arreglo arbitrariamente grande).

1.3.2. Inserción

Este algoritmo sí presenta un peor caso: cuando el arreglo de entrada se encuentra ordenado de forma descendente (y no al revés, según la implementación dada). El bucle interno del algoritmo deberá comparar cada elemento hasta llegar a la primera posición del conjunto, con lo cual realizará el máximo número de comparaciones posibles entre elementos. Esto también está contemplado dentro de su comportamiento en promedio: se destaca el caso en que, si el arreglo estuviera ordenado al revés, el algoritmo terminaría en un orden de $O(n)$ (siendo para este caso trivial, mejor que cualquier ordenamiento).

1.3.3. Quicksort

Quicksort presenta un caso en que su comportamiento resulta cuadrático: cuando todos (o casi todos) los elementos son menores que el pivote seleccionado. En otras palabras, el pivote es una cota del set. En nuestra implementación, Quicksort elige siempre al primer elemento del arreglo como pivote, por lo que un set que cumpla con estas características es uno ordenado de forma descendente.

1.3.4. Heapsort

Para este algoritmo, el numero de comparaciones entre los elementos del arreglo puede variar en poca proporción dependiendo orden en que se presentan los mismos (según su ubicación en el heap). Aún así, tanto el caso promedio como el peor, su tiempo de ejecución será $O(n \log n)$, lo cual significa que asintóticamente no hay diferencia entre ambos, aunque pueden diferir en un valor constante.

1.3.5. Mergesort

Como se analizó con el Teorema Maestro, Mergesort es de orden $O(n \log n)$, por lo que no presenta un peor caso para ningún set en general.

1.4. Tiempos de ejecución

Se calcularon los promedios de cada tiempo para cada cantidad de elementos y cada algoritmo:

TABLA DE PROMEDIOS

GRAFICOS

Es claro que el caso de orden solo afecta descendente en quicksort e insertionsort. Se ve en la comparación que para pocos elementos el comportamiento de los ordenamientos es similar. Pasando los mil elementos se ve como quicksort e insertsort empeoran notablemente. Para el set mayor, se ve que insertionSort tarda cuatro veces más, y quicksort tarda más de 200 veces lo que tardaba en promedio. El resto de los ordenamientos no se ve afectado notablemente. En el gráfico es más que notorio como quicksort tomó un orden cuadrático al igual que los ordenamientos por inserción y selección.

En el siguiente gráfico se muestra la comparación explícita del insertion sort: y a continuación el análogo de quicksort

TABLA DE PEORES SETS

GRAFICOS DE PEORES SETS

La tabla entera con todas las ejecuciones puede verse en el anexo al final.

1.5. Comparación con valores teóricos

De acuerdo a las complejidades analizadas para el peor caso, se predice que los algoritmos de selección inserción y quicksort tardarán mucho más en ordenar arreglos que heapsort o mergesort. Esto se cumple, con la salvedad que en el análisis teórico para el peor caso se preveía que el ordenamiento por selección tendría peor desempeño que quicksort, lo cual no ocurrió, por ser el set especialmente diseñado para que quicksort se comporte de forma poco eficiente.

Con respecto al caso promedio, los tiempos se cumplieron de forma exactamente coherente con el orden de eficiencia predicho (en orden ascendente) en el análisis teórico: inserción, selección, heapsort, mergesort y quicksort. En síntesis, excepto por el peor caso de quicksort que no fue considerado de forma explícita, el análisis teórico se condice con los valores visualizados en la práctica.

2. Algoritmo Gale-Shapley

2.1. Algoritmo

El algoritmo de asignación propiamente dicho se reduce a la función homónima:

```
1  codigocodigocodigo
```

2.2. Demostraciones

2.2.1. Tiempo polinómico

Para facilitar la notación de esta sección, J será la cantidad de jugadores y E la cantidad de equipos. Para la justificación detallada de cada sentencia del algoritmo, ver los comentarios de la complejidad en el código.

En el algoritmo el primer método que se llama es `jugadoresEquiposGS`. En caso en que se especifique la creación de los archivos `.rtf`, el método `generarArchivos` tiene una complejidad es de orden $O(E \cdot J)$, ya que por cada equipo se debe crear un archivo de longitud igual a la cantidad de jugadores, y lo análogo ocurre a la recíproca.

La carga de archivos con el método `cargaDeArchivos` naturalmente también de orden $O(J \cdot E)$ ya que por cada jugador que se quiera crear, se debe recorrer un archivo cuya longitud es igual a la cantidad de equipos, y viceversa.

La asignación, donde se genera el matching establece, tiene como condición de corte que no haya más vacantes en los equipos. Considerando esto, el hecho de que a continuación en el algoritmo se itere por cada equipo, y que se verifique para cada uno si todos tiene vacantes, es solo una especificación para lograr cubrir todas las vacantes, por lo que no agregan un orden diferente: simplemente es la forma de implementar que el algoritmo avance a partir de que se cubren distintas vacantes en distintos equipos. Esto implica que cada equipo deberá completar todos sus lugares libres, y como hay igual cantidad de jugadores que de vacantes, el peor caso se considera cuando cada equipo debe preguntarle a todos los jugadores si desean ocupar una vacante: esto es $O(J)$. Como hay E equipos, el orden de la asignación también es $O(J \cdot E)$.

Finalmente, almacenar el archivo de la asignación tiene un orden línea $O(J + E)$. Esto se debe a que por cada equipo, se guarda su número seguido de los números de jugadores que tiene. La cantidad de jugadores que tiene un equipo es igual a sus vacantes (ver demostración en el próximo item de que cada equipo llena todas sus vacantes) y como la cantidad de vacantes totales es igual al número de jugadores, se tiene que se guardó un archivo con $J + E$ números.

El orden cuadrático del algoritmo, entonces, es evidente, ya que es el peor caso de complejidad a lo largo de las funciones del algoritmo. Como $O(J \cdot E)$ es un orden polinomial, el algoritmo es de orden polinomial.

2.2.2. Matching estable

Para esta demostración, se simplificará la estructura original de la solución implementada. Considerando el primer método llamado `jugadoresEquiposGS`, se excluyen las características de la generación de los archivos `.prf`, su carga, y el correspondiente almacenamiento del archivo donde se guarda la asignación estable. Entonces, el método se reduce a la asignación propiamente dicha.

Por una cuestión de comodidad, se puede traducir esto a un pseudocódigo basado en la documentación del método de asignación: `"Mientras haya un equipo con vacantes."` `"El equipo actual ofrece una vacante a su jugador favorito actual"` `"Si el jugador esta libre, acepta la vacante"` `"Si no esta libre, pero prefiere más al equipo actual, acepta la vacante y el otro equipo pierde a ese jugador."`

Como hay igual cantidad de jugadores que de vacantes totales, se ve que todos los equipos podrán cubrir todos su puestos. Supongamos que hubiera un equipo `.a` que ya ha `.ofrecido` una vacante.^a todos los jugadores de acuerdo con su favoritismo, y ninguno la hubiera aceptado. Esto

es decir que todos los jugadores ya han aceptado una vacante en un equipo previamente. Pero esto es una contradicción, porque existe la misma cantidad de jugadores que de vacantes totales, y por hipótesis hay una (al menos) vacante que ^ano puede cubrir. Justamente, la contradicción viene de suponer que un equipo no ha podido satisfacer una de sus vacantes ofreciéndole a todos los jugadores, por lo que esto es falso. Con esto se ve además que la máxima cantidad de "proposiciones" que un equipo puede hacer es igual a la cantidad de vacantes disponibles en total, es decir, con la cantidad de jugadores.

Ahora se demostrar que una vez que cada equipo haya cubierto sus vacantes, no habrá parejas conflictivas (inestabilidades). La misma podría darse entre (j_1, a) y (j_2, b) , donde j_1 es un jugador que quiere estar en el equipo b , j_2 es un jugador que quiere estar en el equipo a , a es un equipo que prefiere más a j_2 sobre j_1 y b es un equipo que prefiere más a j_1 sobre j_2 .

Supongamos que a ^{elige primero}. Si j_1 aceptó la vacante en a , es porque j_1 estaba libre, o bien, porque prefería estar en a que en otro equipo c (no necesariamente $c = b$).

Supongamos que era porque estaba solo. Luego b prefiere más a j_2 que a j_1 , porque le ofreció su vacante antes. Si esto no fuera así, b le hubiera ofrecido antes su vacante a j_1 , y el mismo se hubiera cambiado de equipo, pero esto no sucedió. En este caso se llega a un absurdo, porque por hipótesis b prefiere más a j_1 (en otro caso la pareja no es conflictiva y no hay inestabilidad).

Entonces debemos suponer que j_1 aceptó la vacante porque estaba en otro equipo c . Luego, se llega al mismo absurdo anterior, porque al fin y al cabo b termina con j_2 .

Luego, debe ser que b ^{elige primero}. Si j_2 aceptó la vacante en b , es porque j_2 estaba libre, o bien, porque prefería estar en b que en otro equipo d (no necesariamente $d = b$ o $d = c$). En cualquiera de los dos casos se llega a un absurdo similar al anterior. Todos los caminos a los que conduce la hipótesis "hay una inestabilidad en la asignación final" hacen llegar a una contradicción, por lo que queda probado que la asignación es estable.

No existe posibilidad que quede un equipo con alguna vacante libre, o un jugador sólo, como se probó al principio.

2.3. Ejecución

En el anexo junto con el código del algoritmo se ve la posibilidad que brinda de generar los archivos prf. El resultado de una de las ejecuciones fue la siguiente asignación:

```
1: 150 120 119 11 22 175 28 15 78 130
2: 182 200 8 191 44 36 60 134 17 159
3: 189 34 1 138 142 45 190 100 197 187
4: 88 9 132 46 62 91 188 103 145 70
5: 48 20 178 27 168 97 30 176 124 98
6: 125 129 25 86 180 165 106 5 177 12
7: 155 158 31 54 195 93 146 29 24 170
8: 164 63 110 127 109 140 96 160 149 143
9: 154 169 198 161 61 99 84 179 33 151
10: 173 69 35 117 163 65 112 118 104 43
11: 166 76 58 137 185 18 77 94 2 186
12: 51 4 171 199 87 162 19 181 85 80
13: 49 40 55 101 26 123 131 92 114 72
14: 193 115 39 37 135 6 172 38 141 174
15: 10 147 167 73 156 14 133 57 74 152
16: 126 192 136 148 71 23 52 153 111 32
17: 116 196 67 82 64 21 59 68 107 113
18: 105 56 90 66 47 183 7 75 108 16
19: 41 81 102 184 3 122 194 79 157 139
20: 42 83 13 128 53 50 144 89 121 95
```


A. Ejecución y compilación

B. Ejecución principal

Todo el código fue hecho en Python 3.6.1. Para ejecutar la parte de ordenamientos.. para la de stable matching...

B.1. Generación de sets aleatorios

El script `generarSetRandom.py` permite generar los sets aleatorios de n números (en este caso, $n=10000$) en un rango determinado. Ejemplo de ejecución:

```
>>> from generarSetRandom import generarSet
>>> generarSet("set1.txt", 1, 1000000, 10000)
```

B.2. Pruebas de ordenamientos

Para verificar que los ordenamientos anden correctamente se realizaron varias pruebas unitarias que verifiquen distintos comportamientos desde los casos triviales hasta los casos borde y demás. Todas las pruebas son iguales, e incluyen asserts con ordenamientos aleatorios.

```
>>> python3 heapTest.py -v
```

B.3. Medición de tiempos

Los tiempos de ejecución fueron calculados con el módulo `time` de Python, en particular su función `process_time` ².

B.4. Stable matching

Para generar un matching estable, se debe importar el archivo `matchinGS.py` y ejecutar la siguiente línea: `matchingGS('ruta de los archivos de jugadores y su correspondiente nombre sin número', 'extensión de los archivos jugadores', 'ruta de los archivos de equipos y su correspondiente nombre sin número', 'extensión de los archivos de equipos', 'nombre completo del archivo donde se guardará la asignación', cantidad de jugadores, cantidad de equipos, cantidad de vacantes por equipo, 'y' si se quieren generar los archivos o cualquier otro caracter en otro caso)`

Por ejemplo, para cumplir con la consigna se requieren los siguientes comandos:

```
>>> from matchingGS import matchingGS
>>> matchingGS('archivos/jugador_', '.prf', 'archivos/equipo_', '.prf', 'asignacion.txt', 200, 20)
```

C. Tablas de ejecución

²https://docs.python.org/3/library/time.html#process_time