Tarea 1

Trinidad Hernández Norma Verónica
Vargas Muñoz José de Jesús
Vilchis Domínguez Miguel Alonso

26 de agosto de 2015

1. Ejercicios

1. **Explicar la complejidad de los problemas tipo NP, NP-Completos y NP-Duros para panaderos.**
Reducciones
A un panadero se le encomienda la tarea de cocinar galletas con forma de cuadrado, pero sabe que si le da dicha forma a las galletas, estas no se cocinarán del centro. Por el contrario, sabe que si cocina galletas de forma circular estas se cocinarán de una manera uniforme. ¿Cómo entrega el panadero galletas cuadradas y cocidas del centro? Haciendo uso de una reducción.

Definición 1.1. Sean P y Q dos problemas. Supóngase que un ejemplar arbitrario E de P puede solucionarse convirtiendo al ejemplar E en un ejemplar E' de Q, resolviendo para E' y reduciendo la solución S' en la solución S para E. Si existen algoritmos C_{Ej}, para convertir un ejemplar E en otro E', y C_{Sol} para convertir la solución S' en S, se dice entonces que existe una reducción de P en Q, la cual se denota como $P \propto Q$.

El panadero toma la masa de galletas (E) y utiliza un cortador (C_{Ej}) para darles forma circular (E'). Cocina esta masa con formas redondas en el horno y obtiene galletas perfectas (S). Cuando las galletas se enfrían, el panadero utiliza un cuchillo para cortarlas (C_{Sol}) y obtener ejemplares cuadrados (S'). Nuestro panadero acaba de aplicar el *Algoritmo de Reducción (o de Simplificación)* para resolver su problema. Tal algoritmo es usado cuando P es muy difícil de resolver de manera directa y Q es más fácil.

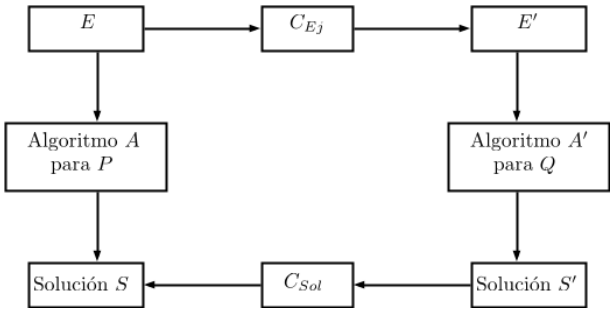

Figura 1: Reducción de P en Q: $P \propto Q$

Supongamos que el panadero tarda 2 segundos en cortar una porción de masa en forma de una galleta circular, y supongamos que le toma 4 segundos cortar con cuchillo una galleta circular para obtener una cuadrada. Cortar n galletas circulares le tomará al panadero $T(n) = 2n$ segundos, y cortarlas después en galletas cuadradas le tomará $T(n) = 4n$ segundos. Se dice que ambas transformaciones son *polinomiales*, ya que el tiempo que toma convertir el ejemplar de un problema a otro ($T(n) = 2n$), así como convertir la solución de un problema en la solución de otro ($T(n) = 4n$), es polinomial. Por desempeño computacional de orden polinomial nos referimos a que es posible expresar el tiempo de transformación como un polinomio que está en función del tamaño de la entrada. En este caso, del número de galletas.

Nuestro panadero puede no saber cómo cocinar galletas cuadradas en un horno, pero sí cómo cocinarlas circulares y después cortarlas. El panadero le ocultará a los demás que desconoce la manera

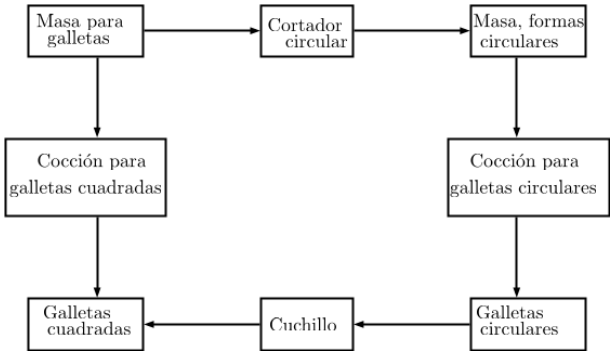

Figura 2: Reducción de P en Q para panaderos

de cocinar galletas cuadradas en el horno y esto no le angustia: el panadero puede obtener galletas cuadradas de otra forma, en tiempo polinomial, y eso es lo importante. Nuestra idea se resume en el siguiente teorema.

Teorema. Si L_1 es polinomialmente reducible a L_2 y existe un algoritmo polinomial para L_2, entonces existe un algoritmo polinomial para L_1. □

El No Determinismo

Nuestro panadero estricto sigue recetas al pie de la letra para sus galletas, lo cual le garantiza exactamente las mismas galletas cada vez, siempre y cuando use los mismos ingredientes. Este panadero sigue algorimos deterministas cada vez que cocina. En términos de computación, decimos que un algoritmo es *determinista* si para una entrada particular, siempre produce la misma salida, con la máquina subyacente siempre pasando por la misma secuencia de estados. Volviendo a la cocina, el algoritmo determinista del panadero es la receta que sigue, la entrada particular son los ingredientes de sus galletas, la salida es el producto final, la máquina subyacente está representada por sus manos y los utensilios de los que se vale, y la secuencia de estados es la forma que va tomando la masa.

Ahora bien, supongamos que a nuestro panadero se le encarga crear una galleta que sea crujiente en su exterior y suave en su interior. El panadero sabe que este requerimiento particular para la galleta depende de la manera de batir la masa, pero no sabe en específico que método al aplicarse resultará en la galleta deseada. El panadero modifica la receta como sigue: comienza la preparación de la galleta como sabe hacerlo y cuando llega el momento de batir tira un dado, el cual le indicará qué técnica de batido usar.

Un *algoritmo no determinista* tiene todas las operaciones cláiscas de uno determinista, con la adición de una primitiva no determinista llamada *nd-choice*, la cual es usada para manipular selecciones.

Nuestro panadero usó una primitiva nd-choice en su receta (algoritmo): el dado. Entonces, dependiendo del resultado del dado se obtiene una galleta distinta. Él prueba la galleta una vez salida del horno; esto para determinar si resultó según los requerimientos. Nuestro panadero acaba de crear, sin saberlo, un algoritmo no determinista de tiempo polinomial. El algoritmo es el siguiente:

a) Aplicar cierto número de pasos de la receta como suele hacerlo.

b) Utilizar una primitiva *nd-choice* para hacer un cambio aleatorio a la receta.

c) Seguir con los pasos de la receta.

d) **Verificar** si el producto obtenido es como se deseaba. De ser así, reportar **sí**, y **no** en caso contrario.

Es importante notar que esta estructura de algoritmo no determinista tiene desempeño polinomial porque tanto la aplicación de la primitiva *nd-choice* como la verificación toman tiempo polinomial.

Definición 1.2. La clase de problemas para los cuales existe un algoritmo no determinista cuyo desempeño computacional es polinomial (en función del tamaño de la entrada) es llamada **Clase NP**.

Supongamos para efectos de nuestro ejemplo que la Clase NP se restringe a la panadería de nuestro personaje.Esto es, la Clase NP se compone de cierto número de galletas. Digamos que el panadero elige una galleta X y descubre que todas las demás recetas de galletas donde usa su dado (los demás problemas de la clase NP) pueden reducirse a la galleta X. Decimos entonces que la galleta es **NP-Dura.**

Definición 1.3. Un problema X es llamado *Problema NP-Duro*, *NP-Hard*, si cada problema en NP es polinomialmente reducible en X.

Ahora supongamos que nuestro panadero conoce una receta de galletas X que es de la clase NP, es decir, utiliza un dado en su preparación. Supongamos además que la galleta es NP-Dura, es decir, todas las demás recetas galletas de la Clase NP pueden reducirse en tiempo polinomial a la receta de la galleta X. Entonces decimos que nuestra galleta es NP-Dura.

Definición 1.4. Un problema X es llamado *Problema NP-Completo*, *NP-Complete*, si

a) el problema X pertenece a NP y

b) el problema X es NP-Duro

2. ¿Para qué valores de n el algoritmo de inserción directa es mejor que mezcla directa?

$$Inserción\ directa: \quad 8(n^2)$$

$$Mezcla\ directa: \quad 64n * log(n)$$

- **Respuesta en forma analítica:**
  Para saber que algoritmo es mejor que otro nos basamos en diferentes métricas en las que podemos cuantificar cual es la cantidad de recursos que son requeridos para resolver la tarea. Estos recursos principalmente se dividen en dos, el tiempo de ejecución del algoritmo y la cantidad de memoria que el algoritmo requiere.
  Implícitamente podemos medir el tiempo de ejecución si conocemos cuantas instrucciones ejecutará el algoritmo, debido a que cada una de nuestras computadoras puede ejecutar un número determinado de instrucciones por segundo, es por ello que para conocer el número de segundos que tomará un algoritmo en finalizar queda determinado por:

$$Tiempo(s) = \frac{\#\ instrucciones}{\#\ instrucciones\ por\ segundo\ que\ la\ computadora\ ejecuta} \quad \left(\frac{instrucciones}{instrucciones/s}\right)$$

  Por esa razón, para saber que algoritmo es mejor que otro basta con fijarse en el número de instrucciones que cada uno de ellos requieren, asumiendo que la misma computadora ejecutara los algoritmos.Es importante notar que el número de instrucciones es una función que depende de los datos de entrada que se le dan al algoritmo.
  Por lo ya descrito, para conocer cuándo el algoritmo de inserción directa es mejor que el algoritmo de mezcla directa se hace una análisis sobre la entrada comparando la cantidad de instrucciones que cada uno de estos requiere.
  Con el objetivo de discernir para que valores de n el número de instrucciones del algoritmo de inserción directa es menor que el algoritmo de mezcla directa, si definimos I como la función que calcula el número de instrucciones por cada algoritmo basta con resolver la siguiente desigualdad:

$$I(Inserción\ directa) \leq I(Mezcla\ directa)$$

  Es decir:

$$8 * n^2 \leq 64n * log(n)$$

  $\iff$

$$n \leq 8 * log(n)$$

  Trabajando solo con la igualdad encontraremos el valor de n en el que se deja de cumplir la desigualdad estricta. Es decir:

$$n - 8 * log(n) = 0$$

Dicho algoritmo en realidad está en base 2, debido a que el análisis de complejidad de merge sort depende de un árbol binario completo es por esa razón que la altura queda representada por $log_2(n)$. Si graficamos esta igualdad obtenemos:

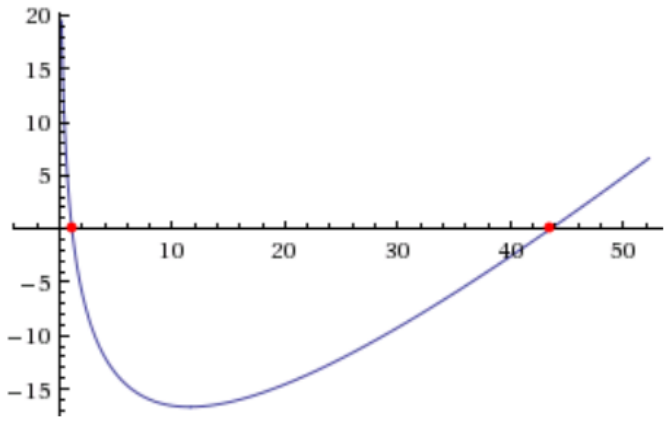

Figura 3: Gráfica $y = n - 8 * log_2(n)$

Por lo que, la solución a la igualdad es cuando $n = 1{,}0\overline{999}$ y $n = 43$, con lo que concluimos que si $2 \leq n \leq 43$, entonces el algoritmo de inserción directa es más mejor que el algoritmo de mezcla.

- **Respuesta de manera experimental:**
  Al poner en ejecución el programa se tomo como precaución el no trabajar con interfaz gráfica ni con otro programa ejecutandose al mismo tiempo para evitar, en medida de lo posible, la intervención de ruido en nuestras mediciones.
  Realizamos 3 versiones para cada algoritmo, una escrita en java, otra escrita en C y una escrita en python. Al hacer la ejecución de nuestro código y graficar el tiempo que tardó cada programa, obtenemos los siguientes resultados:

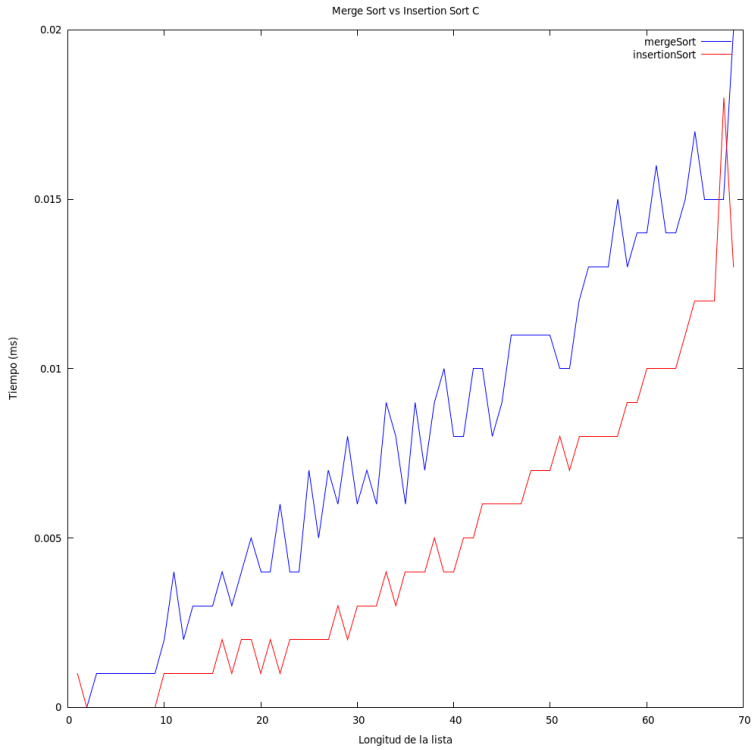

Figura 4: Gráfica que muestra tiempo vs longitud de la lista para el lenguaje C

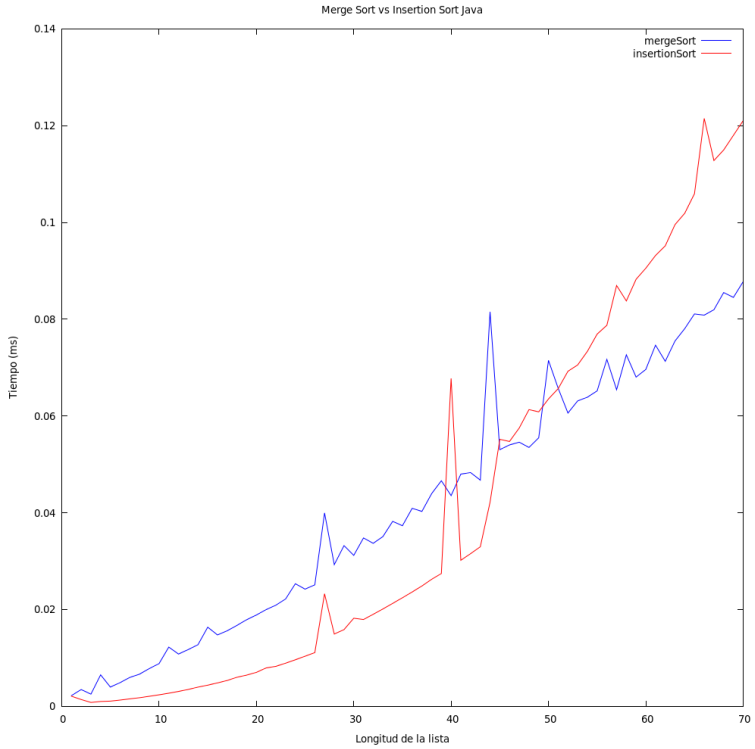

Figura 5: Gráfica que muestra tiempo vs longitud de la lista para el lenguaje Java

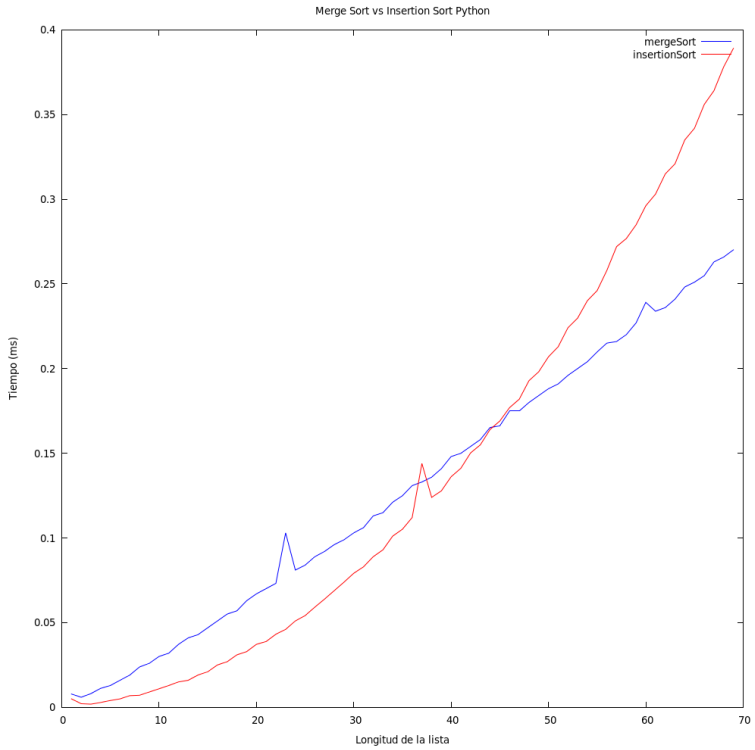

Figura 6: Gráfica que muestra tiempo vs longitud de la lista para el lenguaje Python

En las gráficas podemos apreciar que para el lenguaje de programación Python la cota teórica se respetó totalmente, sin embargo, para C esto no ocurrió, las causas pueden ser diversas, pero nos aventuramos a conjeturar que una de las posibles razones fue que al generar números aleatorios para ordenar, estos pudieron comportarse de tal manera que no explotaran el peor caso del algoritmo de Inserción directa, dicho caso se presenta cuando los números están acomodados de manera inversa al orden natural.

3. ¿Cuál es el valor más pequeño de n, tal que un algoritmo cuyo tiempo de ejecución es

x corre más rápido cuyo tiempo de ejecución es y?.
Para este ejercicio nos apoyamos en el software Mathematica, la solución se obtuvo mediante el uso de gráficas para las funciones, y encontrando los puntos críticos.

a) $x = 100n^2 yy = 2^n$

b) $x = 6n^2 yy = 32n\log(n) + 5n$

4. Ejecutar el algoritmo de Inserción directa y de Mezcla directa para grandes valores de n
Para este ejercicio se tomaron las mismas precauciones que para el ejercicio 2, las iteraciones se realizaron sin interfaz gráfica y sin ejecutar otro programa para evitar, hasta donde es posible, el ruido en nuestros datos.
La presentación de los datos la realizamos en forma de gráfica para que los datos sean más fáciles de interpretar.

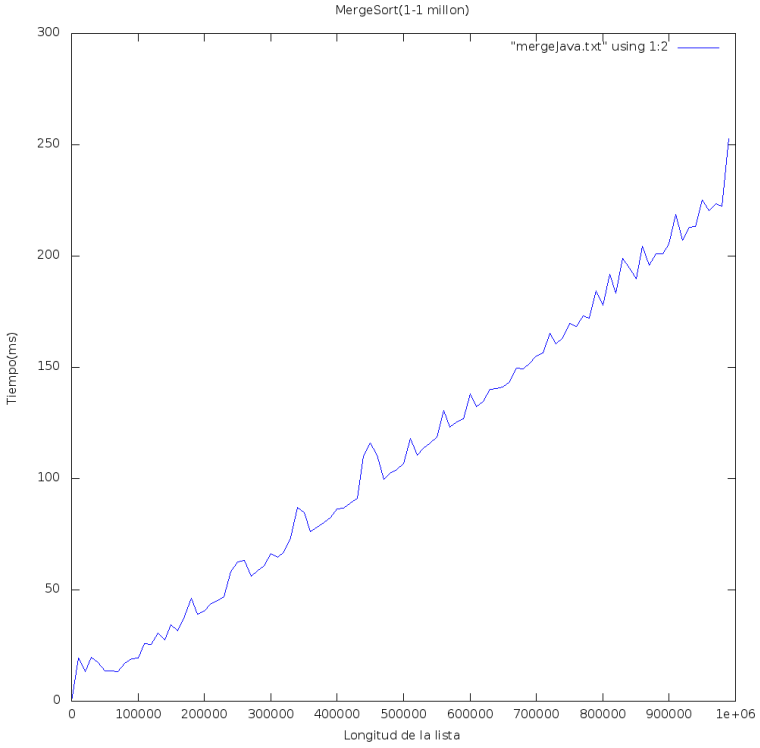

Figura 7: Gráfica para MergeSort en el lenguaje Java de 1 a 1 millon

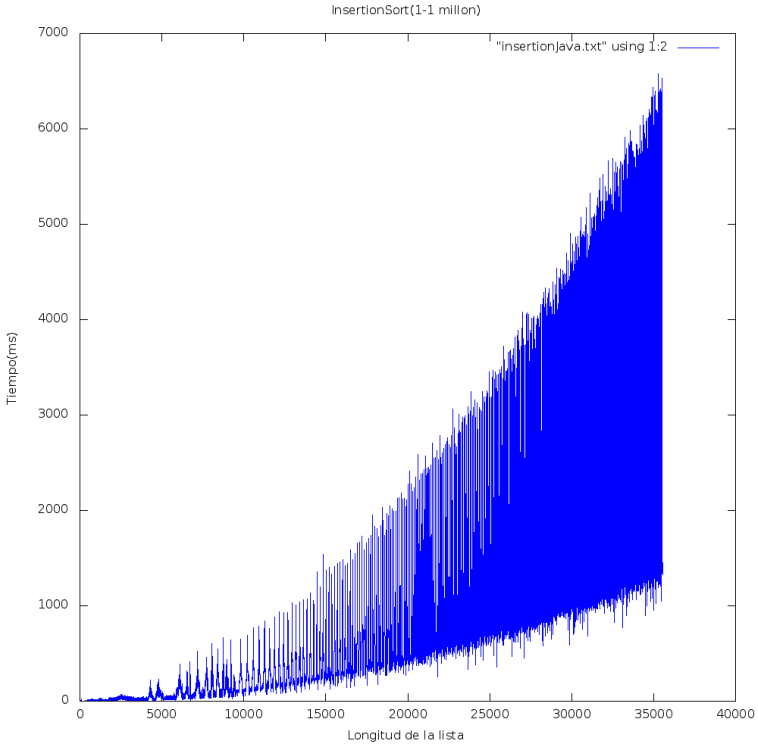

Figura 8: Gráfica para InsertionSort en el lenguaje Java de 1 a 1 millon

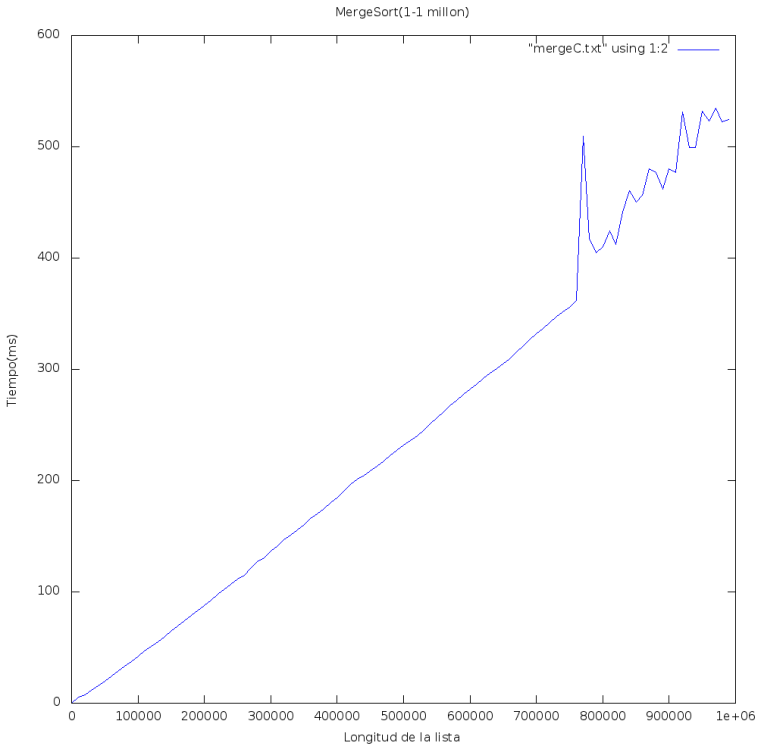

Figura 9: Gráfica para MergeSort en el lenguaje C de 1 a 1 millon

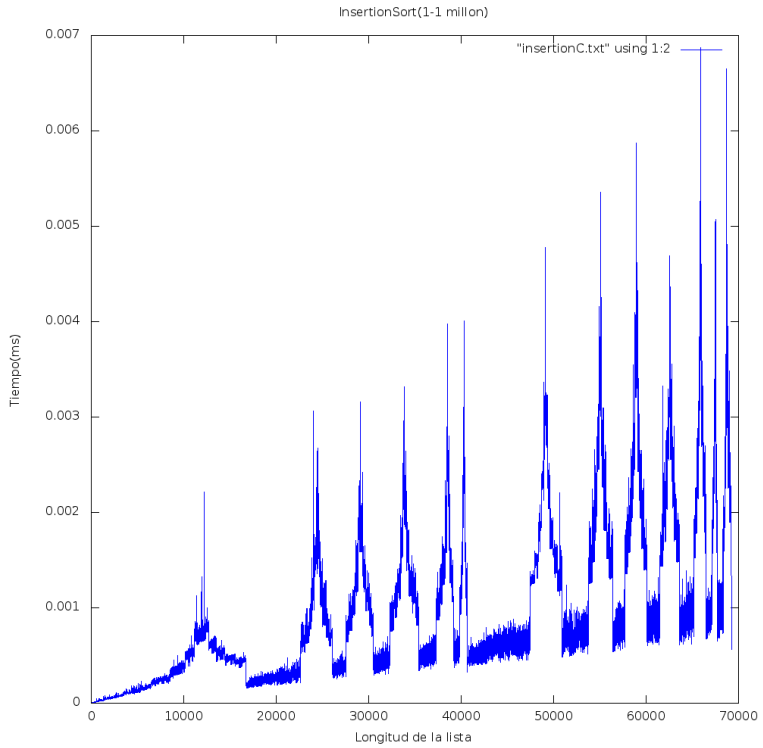

Figura 10: Gráfica para InsertionSort en el lenguaje C de 1 a 1 millon

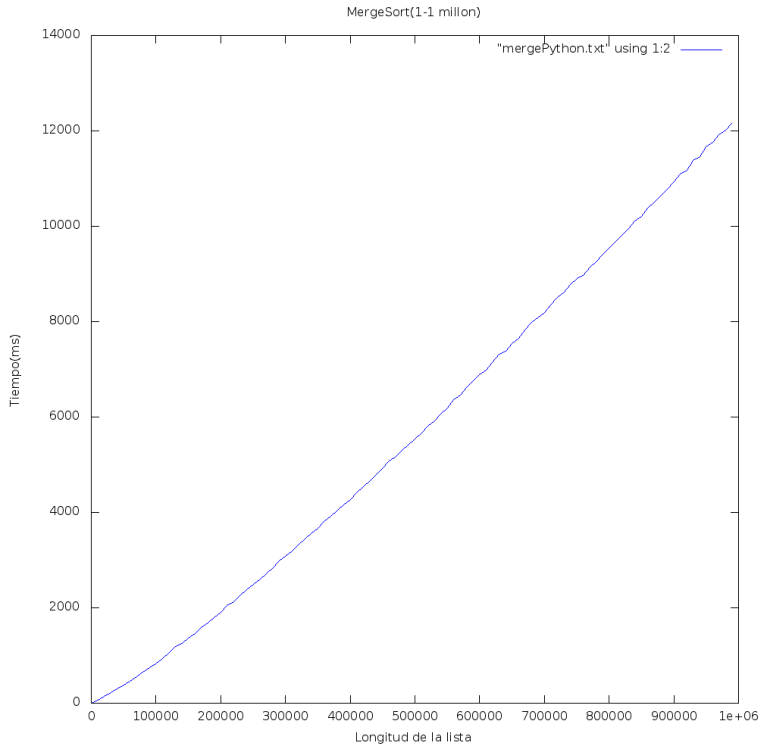

Figura 11: Gráfica para MergeSort en el lenguaje Python de 1 a 1 millon

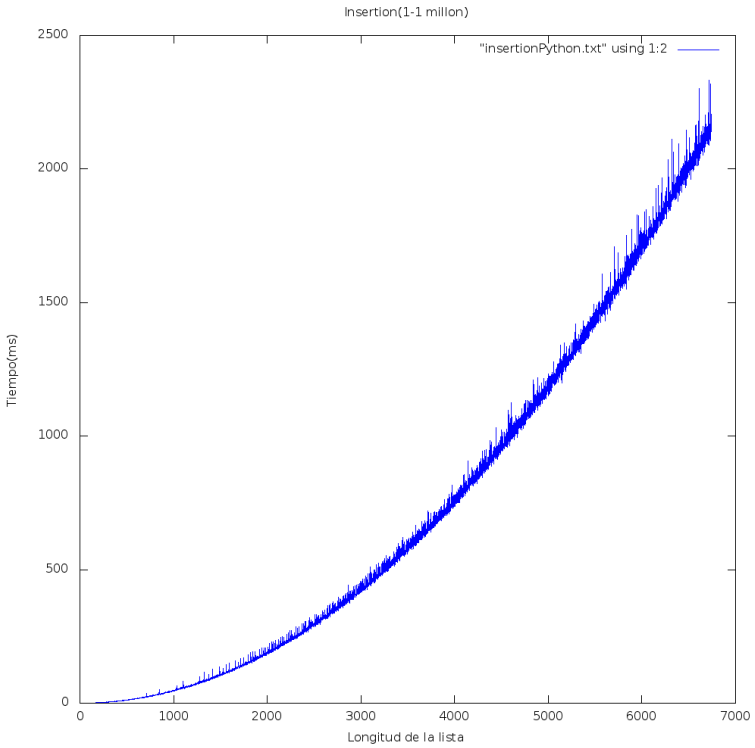

Figura 12: Gráfica para InsertionSort en el lenguaje Python de 1 a 1 millon

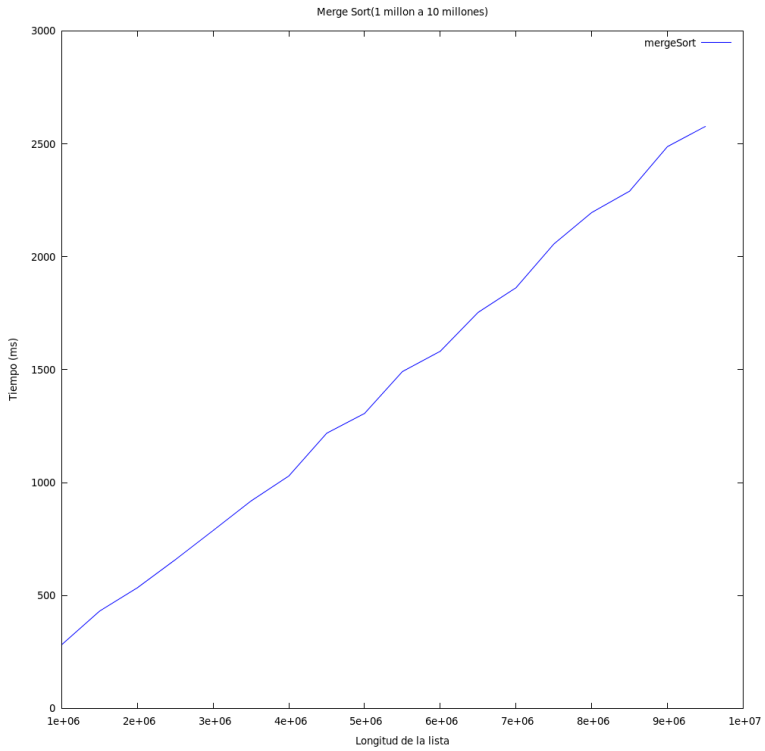

Figura 13: Gráfica para MergeSort en el lenguaje Java

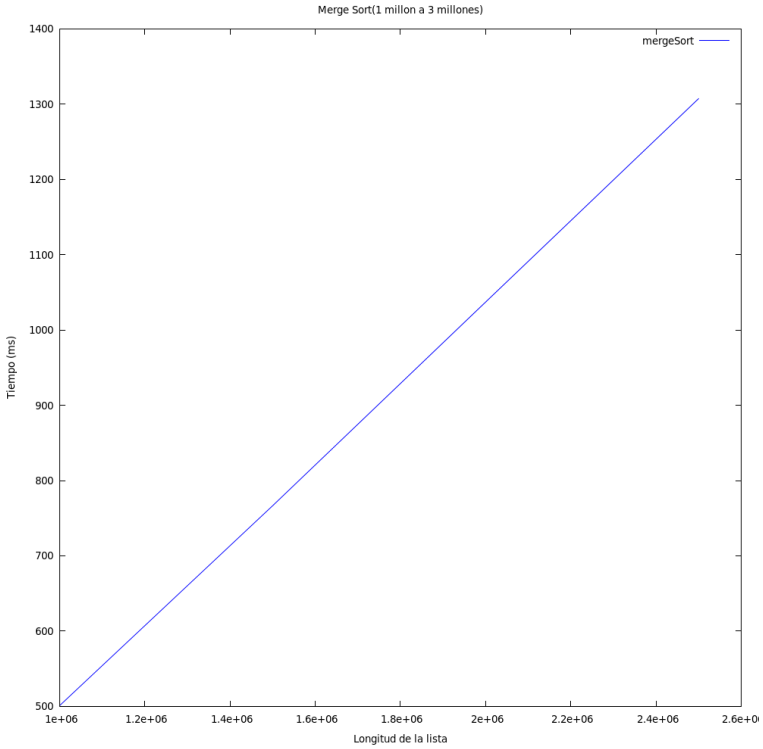

Figura 14: Gráfica para MergeSort en el lenguaje C

Nota: No se agregó gráfica para InsertionSort ni para los programas en python para cuando n vale más de un millón. Ya que, después de mucho tiempo ejecutandose o el sistema operativo mataba el thread o simplemente no acababa. Máquina en la que se ejecuto el programa:

- Equipo de escritorio
- Fabricante: Lenovo, de 64 bits
- CPU: Pentium(R) Dual-Core CPU E5400 @ 2.70GHz,
- fabricante del cpu: Intel Corp
- Capacidad: 2700MHz
- Reloj: 200MHz
- cache 0: capacidad 64KiB
- cache 1: capacidad 2MiB
- memoria Ram: tamaño 3GiB