**INFORME DE INGENIERÍA**

**Paso 1: Identificación del problema**

**Definición del problema**

Encontrar los tres algoritmos de ordenamiento más eficientes para la implementación de un programa que tiene como base ordenar números enteros o reales.

**Justificación**

En busca de un programa más eficiente las empresas naturalmente usan coprocesadores cuya función es descargar trabajo del procesador principal ya que poseen una tarea especializada. Los coprocesadores pueden realizar operaciones nativas como lo es la función de ordenamiento. Por lo tanto, después de estudiar los costos de implementación de dicho algoritmo se ha decidido encontrar tres algoritmos de ordenamiento más eficientes dependiendo del caso de ordenamiento dado (real o entero, y el numero de elementos a ordenar).

**Requerimientos funcionales**

|  |  |
| --- | --- |
| **R1** | Ordenar datos ingresados |
| **Descripción** | Ordena datos (reales o enteros) ingresados por el usuario, y muestra el resultado. |
| **Entradas** | Datos por ordenar. |
| **Salidas** | Datos ordenados. |

|  |  |
| --- | --- |
| **R2** | Generar valores aleatorios ya ordenados |
| **Descripción** | Generar el número de valores que el usuario quiera y que estos estén ordenados |
| **Entradas** | Número de valores a generar |
| **Salidas** | Lista de valores ordenados |

|  |  |
| --- | --- |
| **R3** | Generar valores aleatorios de acuerdo a un % de desorden. |
| **Descripción** | Generar valores aleatorios con base en el tamaño de la secuencia y el % de desorden se obtiene un número k de cuantas posiciones deben estar desordenadas. |
| **Entradas** | Tamaño de la secuencia y % de desorden |
| **Salidas** | Lista de valores con k posiciones desordenadas |

|  |  |
| --- | --- |
| **R4** | Generar valores intercambiados |
| **Descripción** | Generar valores aleatorios ordenados y con base en el tamaño de la secuencia y el % de desorden se generan k/2 pares de posiciones diferentes y se intercambian los valores entre cada par de ellas. |
| **Entradas** | Tamaño de la secuencia y % de desorden |
| **Salidas** | Lista de valores con k/2 posiciones intercambiadas. |

|  |  |
| --- | --- |
| **R5** | Mostrar el tiempo de ordenamiento |
| **Descripción** | Mostrar el tiempo en que se ordenan los valores. |
| **Entradas** | Datos por ordenar |
| **Salidas** | Tiempo de ordenamiento. |

|  |  |
| --- | --- |
| **R6** | Restringir o permitir la ejecución de los algoritmos de ordenamiento. |
| **Descripción** | Dependiendo del número y tipo de datos a ordenar se debe restringir o permitir la ejecución de alguno de los tres algoritmos de manera que resulte ser el más eficiente. |
| **Entradas** | Datos por ordenar |
| **Salidas** | Opciones de algoritmo que pueda ser usado para ordenar esa secuencia. |

**Paso 2: Recopilación de la información**

*Algoritmo*

Es una secuencia de pasos bien definidos que buscan resolver un problema computacional.

*Eficiencia*

Medida del uso de los recursos computacionales requeridos por la ejecución de un algoritmo en función del tamaño de las entradas. <http://elvex.ugr.es/decsai/algorithms/slides/2%20Eficiencia.pdf>

*Complejidad temporal*

Función que describe el comportamiento (en tiempo) de un algoritmo conforme se incrementa el tamaño de la entrada.

*Complejidad espacial*

Uso de recursos de espacio que requiere un algoritmo en cuestión.

*Notación asintótica*

Son aquellas notaciones utilizadas para describir el tiempo de ejecución asintótico de un algoritmo. Conocemos la notación O para el peor caso, Ω para el mejor caso y Ɵ para el caso promedio.

Según Kenneth H. Rosen en su libro *Matemática Discreta y sus aplicaciones*, los algoritmos más eficientes ordenados de manera descendente son aquellos que:

|  |  |
| --- | --- |
| Complejidad | Terminología |
| O(l) | Complejidad constante |
| O(log n) | Complejidad logarítmica |
| O(n) | Complejidad lineal |
| O(n log n) | Complejidad n log n |
| O( | Complejidad polinómica |
| O() | Complejidad exponencial |
| O(n!) | Complejidad factorial |

Por lo tanto, debemos tener como criterio de selección la complejidad temporal del algoritmo con base en esa tabla.

**Un análisis ya existente:**

Encontramos en un blog de *pereiratechtalks.com* perteneciente a Sergio Andrés Flórez un análisis hecho para encontrar el algoritmo de ordenamiento más eficiente:

Sergio toma siete algoritmos conocidos: Burbuja, inserción, selección, Heapsort (ordenamiento por montículos), conteo, merge sort y Quicksort, los implementa y los ejecuta en los maquinas con diferentes especificaciones para encontrar cuál es el algoritmo más eficiente para ordenar hasta 1.000.000.000 datos.

Después de un análisis de las gráficas de los siete algoritmos, llegó a la conclusión de que en definitiva los perdedores eran burbuja, inserción y selección. Por lo tanto, se decide analizar más a de cerca los cuatro restantes, y se llega a la conclusión de que el algoritmo más eficiente fue el de conteo con una complejidad de O(n+k).

**Paso 3: Búsqueda de soluciones creativas**

Decidimos hacer una lluvia de ideas de posibles soluciones:

* Crear un árbol binario e ir insertado los valores ingresados y luego imprimirlos de acuerdo al recorrido inorden.
* Ordenamiento burbuja, selección o inserción que hemos implementado antes.
* Bogosort: consiste en tomar el conjunto de valores ingresados y reorganizarlos de manera aleatoria en el conjunto, hasta que logre organizarlos de manera ascendente o descendente.
* Ordenamiento por panqueques: Ordena un arreglo de números enteros mediante rotaciones del arreglo en entre el valor más alto y el más bajo; este va reduciendo el tamaño del arreglo rotado mientras que se ordenan los valores en orden ascendente.
* Ordenamiento Radix: Es un método de ordenamiento que consiste en organizar valores de un arreglo en base a ciertas claves o categorías que el algoritmo aplica para separar los valores mediante la comparación de números en sus posiciones de cifra significativas.
* Listar algoritmos que no hemos implementados pero que son conocidos por su eficiencia: ordenamiento por mezcla, montículos, conteo y ordenamiento rápido.

**Paso 4: Transición de ideas a los diseños preliminares**

A primera vista, la opción de crear un árbol y luego imprimir su recorrido inorden queda descartada debido a su consumo de recursos de almacenamiento y de tiempo.

Los algoritmos inestables como el bogosort y el ordenamiento por panqueques también quedan descartadas porque no garantiza una respuesta segura y no tienen la capacidad de organizar arreglos grandes de manera eficiente.

Investigando más a fondo el algoritmo Radix queda descartado porque este algoritmo depende mucho de la eficiencia de las instrucciones es un interior y tiene dificultades con diferentes tipos de datos (enteros y decimales).

Por otro lado, ya hemos implementado los algoritmos de ordenamiento por inserción, selección y burbuja. Hemos concluido que el ordenamiento burbuja es el menos eficiente de ellos tres debido al número de comparaciones y recorridos que hace al arreglo. De esta manera queda descartado.

Por lo tanto, las ideas más factibles son: algoritmo de selección, inserción, por montículos, conteo, mezcla y rápido.

|  |  |  |  |
| --- | --- | --- | --- |
| Ordenamiento | Explicación | Observaciones | Complejidad conocida |
| Selección | Se ordena un arreglo de valores mediante la búsqueda del valor más bajo dentro del arreglo, este valor se reemplaza con el primero; luego se busca el siguiente valor más bajo y se reemplaza con el segundo y se repite hasta que se ordene todo el arreglo. | Es sencillo de implementar, pero poco eficiente con una lista grande de valores | O() |
| Inserción | La principal característica de este algoritmo es que intercambia valores para “abrirle” espacio al valor que quedará en su posición correcta. Es sencillo de implementar, no necesita almacenamiento temporal adicional. | En sencillo de implementar, pero poco eficiente con una lista grande de valores. | O() |
| Montículos | Se ordena una estructura de datos específica (binaria) mediante la cual se transforma un arreglo en un árbol binario, en este proceso se toma el valor máximo en el árbol y se retira, este proceso se repite hasta que se acaben todos los valores dentro del árbol y queden ordenados de menor a mayor en el arreglo. | Tienen un buen desempeño con listas grandes, se comporta mejor que el Quicksort en los peores casos.  Complejo de implementar. | O(n log n) |
| Mezcla | Se ordena un arreglo mediante la división del arreglo en mitades y luego esas mitades se dividen en la mitad hasta que quedan valores individuales y luego se juntan para tener un arreglo ordenado. | Está definido recursivamente. Puede ocupar el doble de espacio que ocupan los datos inicialmente. Es estable. | O(n log n) |
| Conteo |  | Requiere memoria adicional, pero resulta muy eficiente en tiempo. No realiza comparaciones, solo ordena enteros, bueno con los números repetidos. | O(n+k) |
| Rápido | Es un método recursivo que consiste en tomar un pivote dentro del arreglo de valores y se encarga de separar los valores más altos que el pivote en un lado y los más bajos en el otro. Por último, se repite este proceso hasta que queden ordenados los valores | Eficiente en tiempo y en memoria. Implementación requiere recursión. | O(n log n) |

**Paso 5: Evaluación o selección de la mejor solución (Criterios y selección)**

Criterio A: Implementación

* [3] Implementación iterativa
* [2] Implementación recursiva
* [1] Implementación compleja

Criterio B: Almacenamiento temporal adicional

* [3] No usa almacenamiento temporal adicional
* [2] Usa almacenamiento temporal adicional, pero es estable.
* [1] Usa almacenamiento temporal adicional, y es inestable.

Criterio C: Complejidad temporal

* [5] Complejidad lineal
* [3] Complejidad n log n
* [1] Complejidad polinómica

Criterio D: Comparaciones

* [3] No realiza comparaciones
* [2] Realiza pocas comparaciones
* [1] Realiza numerosas comparaciones

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Criterio A | Criterio B | Criterio C | Criterio D | Total |
| Selección | 3 | 1 | 1 | 2 | 7 |
| Inserción | 3 | 2 | 1 | 2 | 8 |
| Montículos | 1 | 2 | 3 | 1 | 7 |
| Mezcla | 2 | 2 | 3 | 2 | 9 |
| Conteo | 3 | 2 | 5 | 3 | 10 |
| Rápido | 2 | 2 | 3 | 2 | 9 |

Los tres algoritmos que elegir son: Mezcla (Merge sort), conteo (Counting sort) y rápido (Quick sort).

**Paso 6: Preparación de informes**

**Diseño del diagrama de clases de la solución**

**Pseudocódigo de los tres algoritmos**

**Counting sort:**

min <- list[0]

max <- min

i <- 0

while(i<list.length)

if(list[i] < min)

min = list[i]

if(list[i] > max)

max = list[i]

i <- i+1

aux <- array(max-min+1)

i <- 0

while(i<list.length)

value <- list[i]-min

aux[value] <- aux[value]+1

i <- i+1

output <- new list

i <- 0

while (i < aux.length)

if(aux[i] > 0)

for j <- 1 to aux[i]

output.add(i+min)

i <- i+1

return output

**Diseño de casos de pruebas unitarias**

*Diseña para cada uno de los 3 algoritmos relevantes al menos tres casos de prueba para verificar su correcto funcionamiento* ***(uno estándar, uno que pruebe casos límite y otro interesante)*** *cumpliendo con el formato donde se especifica la clase, el método a probar, el escenario, las entradas y la salida esperada.*

*Ordenados de mayor a menor y deba ordenarlos de menor a mayor*

*Todo ordenado*

*Normal*

*Que toda la secuencia sea de los mismos datos.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prueba 1: Algoritmo | | | | |
| Clase | Método | Escenario | Entrada | Resultado |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Análisis de complejidad temporal de los tres algoritmos**

Counting sort. Sea n el tamaño del arreglo de entrada y k su rango (valor máximo – valor mínimo +1)

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | min <- list[0] | c1 | 1 |
| 2 | max <- min | c2 | 1 |
| 3 | i <- 0 | c3 | 1 |
| 4 | while(i<list.length) | c4 | n+1 |
| 5 | if(list[i] < min) | c5 | n |
| 6 | min = list[i] | c6 | n |
| 7 | if(list[i] > max) | c7 | n |
| 8 | max = list[i] | c8 | n |
| 9 | i <- i+1 | c9 | n |
| 10 | aux <- array(max-min+1) | c10 | 1 |
| 11 | i <- 0 | c11 | 1 |
| 12 | while(i<list.length) | c12 | n+1 |
| 13 | value <- list[i]-min | c13 | n |
| 14 | aux[value] <- aux[value]+1 | c14 | n |
| 15 | i <- i+1 | c15 | n |
| 16 | output <- new list | c16 | 1 |
| 17 | i <- 0 | c17 | 1 |
| 18 | while (i < aux.length) | c18 | k+1 |
| 19 | if(aux[i] > 0) | c19 | k |
| 20 | for j <- 1 to aux[i] | c20 | n+1 |
| 21 | output.add(i+min) | c21 | n |
| 22 | i <- i+1 | c22 | k |
| 23 | return output | c23 | 1 |

Sea T(n,k) el tiempo que tarda en ejecutarse el algoritmo en función de n y k.

1. T(n,k) = c1+c2+c3+c4n+c4+c5n+c6n+c7n+c8n+c9n+c10+c11+c12n+c12+c13n+c14n +c15n+c16+c17+c18k+c18+c19k+c20n+c20+c21n+c22k+c23

Sea A una contante tal que A = c1+c2+c3+c4+c10+c11+c12+c16+c17+c18+c20+c23 , B una constante tal que B = c4+c5+c6+c7+c8+c9+c12+c13+c14+c15+c20+c21 y C una constante tal que C = c18+c19+c22

1. T(n.k) = A+Bn+Ck

Asi, como A, B y C son coeficientes constantes, T(n,k) es O(n+k). Supongamos que n > k, con lo que el algoritmo counting sort es O(n).

**Análisis de complejidad espacial de los tres algoritmos**

Entradas, auxiliares y salidas

**Bibliografía**

Matemática Discreta y sus aplicaciones. Kenneth H. Rosen. Quinta edición.

<https://pereiratechtalks.com/analisis-de-algoritmos-de-ordenamiento/>

<https://techlandia.com/ventajas-desventaas-algoritmos-ordenamiento-info_181515/>

<https://jona83.wordpress.com/unidad-1/algoritmos-de-ordenamiento/>

<https://prezi.com/rifcpibcg--b/counting-sort/> Counting sort