

Métodos Externos de Ordenamiento

Ordenamiento externo es un término genérico para los algoritmos de ordenamiento que pueden manejar grandes cantidades de información. El ordenamiento externo se requiere cuando la información que se tiene que ordenar no cabe en la memoria principal de una computadora (típicamente la RAM) y un tipo de memoria más lenta (típicamente un disco duro) tiene que utilizarse en el proceso.

Un ejemplo de ordenamiento externo es el algoritmo de ordenamiento por mezcla. Supongamos que 900 MB de información deben ser ordenados utilizando únicamente 100 MB de RAM.

1. Leanse 100 MB de información en la memoria principal y ordenense utilizando un algoritmo tradicional (típicamente quicksort).
2. Escribase la información ordenada en el disco.
3. Repítanse los pasos 1 y 2 hasta que toda la información esté ordenada en pedazos de 100 MB. Ahora se deben mezclar todos los pedazos ordenados.
4. Léanse los primeros 10MB de cada pedazo ordenado a la memoria principal (total de 90 MB) y destínense los 10 MB restantes para el buffer de salida.
5. Ordénese los nueve pedazos mezclandolos y grábense el resultado en el buffer de salida. Si el buffer de salida está lleno, escribase al archivo destino final. Si cualquiera de los 9 buffers leídos queda vacío, se llena con los siguientes 10 MB de su pedazo original de 100 MB o se marca este como completado si ya no hay registros remanentes.

Otro ejemplo es el algoritmo de ordenamiento por mezcla equilibrada, que es una optimización del anterior.

Intercalación

Ordena el vector tomando cada número e insertándose en la posición que toma su valor, es decir, si tengo un cinco en el arreglo; lo pongo en la posición cinco, es decir, lo que vale el elemento es la posición en la que se coloca. Por supuesto, no se podrán ordenar los arreglos que tengan valores repetidos y el vector necesita estar del tamaño del número más grande que se encuentre en él.

Mezcla Directa

La ordenación de archivos se lleva a cabo cuando el volumen de los datos a tratar es demasiado grande y los mismos no caben en la memoria principal de la computadora.

Al ocurrir esta situación no pueden aplicarse los métodos de ordenación interna, de modo que debe pensarse en otro tipo de algoritmos para ordenar datos almacenados en archivos.

Por ordenación de archivos se entiende, entonces, la ordenación o clasificación de éstos, ascendente o descendentemente, de acuerdo con un campo determinado al que se

denominará campo clave. La principal desventaja de esta ordenación es el tiempo de ejecución, debido a las sucesivas operaciones de entrada y salida.

Mezcla Natural

Entre los métodos de ordenamiento externos más comunes se encuentran el de Mezcla Directa (o Merge Sort) y el de Mezcla Natural (o Natural Merge Sort). El método de Mezcla Natural consiste en aprovechar la existencia de secuencias ya ordenadas dentro de los datos de los archivos. A partir de las secuencias ordenadas existentes en el archivo, se obtienen particiones que se almacenan en dos archivos o ficheros auxiliares. Las particiones almacenadas en estos archivos auxiliares se fusionan posteriormente para crear secuencias ordenadas cuya longitud se incrementa arbitrariamente hasta conseguir la total ordenación de los datos contenidos en el archivo original.

Para ilustrar la Mezcla Natural es muy conveniente utilizar números enteros como los datos almacenados en un archivo, de tal manera que este método funciona como se ilustra en el siguiente ejemplo (se utilizan corchetes o paréntesis cuadrados para separar las secuencias en las particiones):

Nombre del archivo original: Origen

Archivos auxiliares: Auxiliar1 y Auxiliar2

Origen: [10,17],[5,8,9],[3,22,50],[4],[2,20,30],[11,16,19],[6,21,23]

Primera Iteración: Las secuencias que se encuentren dentro del archivo origen se almacenan de forma alternada en los archivos auxiliares:

Partición

Auxiliar1: [10,17],[3,22,50],[2,20,30],[6,21,23]

Auxiliar2: [5,8,9],[4,11,16,19]

Aquí vale la pena resaltar lo que sucede con la secuencia [4] del archivo origen: cuando se realizan las particiones (aún sin entrar al proceso de fusión) los elementos 11, 16 y 19 se convierten en una sola secuencia junto con el 4.