

Renteria Magaña Rayni Damian

Algoritmos externos de ordenación

En la actualidad es muy común procesar tales volúmenes de información que los datos no se pueden almacenar en la memoria principal de la computadora. Estos datos, organizados en archivos, se guardan en dispositivos de almacenamiento secundario, como cintas, discos, etcétera.

El proceso de ordenar los datos almacenados en varios archivos se conoce como **fusión o mezcla**; se entiende por este concepto a la combinación o intercalación de dos o más secuencias ordenadas en una única secuencia ordenada. Se debe hacer hincapié en que sólo se colocan en la memoria principal de la computadora los datos que se pueden acceder en forma directa.

Por **intercalación de archivos** se entiende la unión o fusión de dos o más archivos, ordenados de acuerdo con un determinado campo clave, en un solo archivo. Analicemos el siguiente ejemplo.

Suponga que se tienen dos archivos, $F1$ y $F2$, cuya información está ordenada de acuerdo con un campo clave:

$F1$: 06 09 18 20 35
 $F2$: 10 16 25 28 66 82 87

Se debe producir, entonces, un archivo $F3$ ordenado, como resultado de la mezcla de $F1$ y $F2$. Sólo se pueden acceder en forma directa dos claves, la primera del archivo $F1$ y la segunda del archivo $F2$. Las comparaciones que se realizan para producir el archivo $F3$ son:

- | | |
|-----------|--------------------------------------------------------------------------------------------------------------------|
| (06 < 10) | sí se cumple la condición
Se escribe 06 en el archivo de salida $F3$ y se vuelve a leer otra clave de $F1$ (09) |
| (09 < 10) | sí se cumple la condición
Se escribe 09 en el archivo de salida $F3$ y se vuelve a leer otra clave de $F1$ (18) |
| (18 < 10) | no se cumple la condición
Se escribe 10 en el archivo de salida $F3$ y se vuelve a leer otra clave de $F2$ (16) |

El estado de los archivos $F1$, $F2$ y $F3$, hasta el momento, es como se muestra más abajo. La flecha señala el último elemento leído de los archivos que se están intercalando.

$F1$:	06	09	18	20	35		
		↓					
$F2$:	10	16	25	28	66	82	87
		↓					
$F3$:	06	09	10				

El proceso continúa hasta que en uno u otro archivo se detecte su final; en tal caso sólo se tendrá que copiar la información del archivo no vacío al archivo de salida $F3$. El resultado final de la intercalación entre los archivos $F1$ y $F2$ es:

$F3$: 06 09 10 16 18 20 25 28 35 66 82 87

Mezcla Directa

El método de ordenación por **mezcla directa** es probablemente el más utilizado por su fácil comprensión. La idea central de este algoritmo consiste en la realización sucesiva de una partición y una fusión que produce secuencias ordenadas de longitud cada vez mayor. En la primera pasada, la partición es de longitud 1 y la fusión o mezcla produce secuencias ordenadas de longitud 2. En la segunda pasada, la partición es de longitud 2 y la fusión o mezcla produce secuencias ordenadas de longitud 4. Este proceso se repite hasta que la longitud de la secuencia para la partición sea:

$$\text{Parte entera } ((n + 1)/2)$$

Donde n representa el número de elementos del archivo original.

Supongamos que se desea ordenar las claves del archivo F . Para realizar tal actividad se utilizan dos archivos auxiliares a los que se les denominará $F1$ y $F2$.

F : 09 75 14 68 29 17 31 25 04 05 13 18 72 46 61

PRIMERA PASADA

Partición en secuencias de longitud 1.

Fusión en secuencias de longitud 8.

F : 09 14 17 25 29 31 68 75' 04 05 13 18 46 61 72'

$F1$: 09' 14' 29' 31' 04' 13' 72' 61'

$F2$: 75' 68' 17' 25' 05' 18' 46'

Fusión en secuencias de longitud 2.

F : 09 75' 14 68' 17 29' 25 31' 04 05' 13 18' 46 72' 61'

SEGUNDA PASADA

Partición en secuencias de longitud 2.

F1: 09 75' 17 29' 04 05' 46 72'
F2: 14 68' 25 31' 13 18' 61'

Fusión en secuencias de longitud 4.

F: 09 14 68 75' 17 25 29 31' 04 05 13 18' 46 61 72'

TERCERA PASADA

Partición en secuencias de longitud 4.

F1: 09 14 68 75' 04 05 13 18'
F2: 17 25 29 31' 46 61 72'

Fusión en secuencias de longitud 8.

F: 09 14 17 25 29 31 68 75' 04 05 13 18 46 61 72'

CUARTA PASADA

Partición en secuencias de longitud 8.

F1: 09 14 17 25 29 31 68 75'
F2: 04 05 13 18 46 61 72'

Fusión en secuencias de longitud 16.

F: 04 05 09 13 14 17 18 25 29 31 46 61 68 72 75

El método de ordenación por **mezcla equilibrada**, conocido también como **mezcla natural**, es una optimización del método de mezcla directa.

La idea central de este algoritmo consiste en realizar las particiones tomando secuencias ordenadas de máxima longitud en lugar de secuencias de tamaño fijo previamente determinadas. Luego se realiza la fusión de las secuencias ordenadas, en forma alternada, sobre dos archivos. Aplicando estas acciones en forma repetida se logrará que el archivo original quede ordenado. Para la realización de este proceso de ordenación se necesitarán cuatro archivos. El archivo original F y tres archivos auxiliares a los que se denominará $F1$, $F2$ y $F3$. De estos archivos, dos serán considerados de entrada y dos de salida; esto, de manera alternada, con el objeto de realizar la fusión-partición. El proceso termina cuando en la realización de una fusión-partición el segundo archivo quede vacío.

Supongamos que se desea ordenar las claves del archivo F utilizando el método mezcla equilibrada.

F : 09 75 14 68 29 17 31 25 04 05 13 18 72 46 61

Los pasos que se realizan son:

PARTICIÓN INICIAL

$F2$: 09 75' 29' 25' 46 61'

$F3$: 14 68' 17 31' 04 05 13 18 72'

PRIMERA FUSIÓN-PARTICIÓN

F : 09 14 68 75' 04 05 13 18 25 46 61 72'

$F1$: 17 29 31'

SEGUNDA FUSIÓN-PARTICIÓN

$F2$: 09 14 17 29 31 68 75'

$F3$: 04 05 13 18 25 46 61 72'

TERCERA FUSIÓN-PARTICIÓN

F : 04 05 09 13 14 17 18 25 29 31 46 61 68 72 75

$F1$:

Observe que al realizar la tercera fusión-partición el segundo archivo queda vacío; por lo tanto, se puede afirmar que el archivo ya se encuentra ordenado.

References

Cairó, O., & Guarati, S. (2006). *Estructuras de datos (3a. ed.)*. McGraw-Hill Interamericana.