

CAPÍTULO X: ESTRUCTURA DE FICHEROS

CONCEPTO DE ARCHIVO.

Los datos que se encuentran en memoria masiva suelen organizarse en archivos. Un archivo (o fichero) es un conjunto de información sobre un mismo tema, tratada como una unidad de almacenamiento y organizada de forma estructurada para la búsqueda de un dato individual. Un archivo está compuesto de registros homogéneos que contienen información sobre el tema. Cada registro a su vez está compuesto por campos.

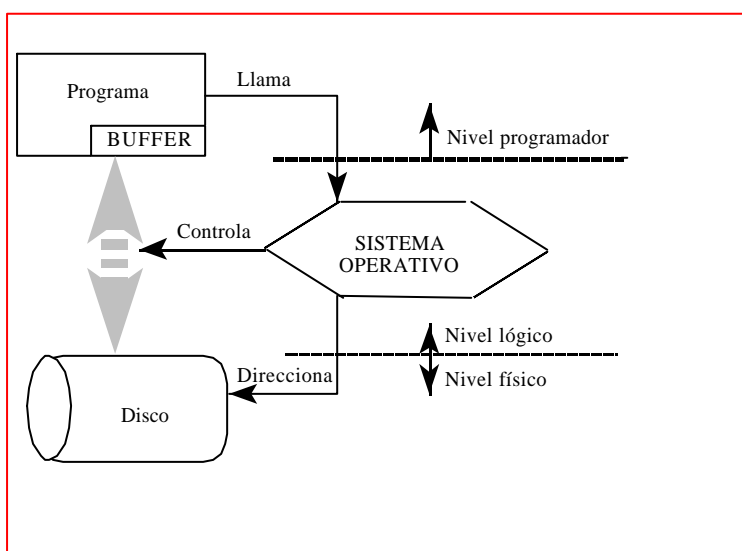
La vida de todo archivo comienza cuando se crea y acaba cuando se borra. Durante la vida del archivo se suelen realizar sobre él determinadas operaciones de recuperación o consulta y de mantenimiento o actualización. Estas operaciones las realizan programas específicos, que actúan a nivel de registro. Las operaciones de actualización que se pueden realizar sobre el archivo son:

- modificación de un registro
- eliminación o borrado de un registro
- inserción de un registro

La mayor parte de las operaciones de recuperación y actualización implican la realización de una localización o búsqueda de un registro concreto para luego actuar sobre él (leerlo, escribir o cambiar parte de él, borrarlo, etc..).

Los sistemas informáticos disponen en su software de programas de utilidad para efectuar operaciones básicas con archivos (creación, borrado, copia o duplicado entre dispositivos de igual o distinta naturaleza, etc..). Estos programas, al igual que los módulos del sistema operativo y utilidades generales del software de un computador, se utilizan mediante un lenguaje de control.

El sistema operativo transporta, cada vez que accede al dispositivo, una



cantidad fija de información (bloque o registro físico) que depende de las características hardware o físicas de éste. En un bloque puede haber varios registros del archivo o puede que un registro ocupe varios bloques. Un factor importante en el diseño de archivos es la longitud de bloque o el factor de blocaje (o simplemente blocaje), que se define como el número de registros del archivo que entran

en un bloque. Cuanto mayor sea éste, menor será el número de accesos al dispositivo necesarios para procesar el archivo.

El sistema operativo realiza, además, la transformación de la dirección lógica usada en los programas de usuario, en la dirección física con la que se direcciona en el soporte. La dirección lógica es la posición relativa que ocupa el registro en el archivo, tal y como se ve éste desde el programa del usuario. La dirección física es la posición real o efectiva donde se encuentra el registro en el soporte de información (dirección "hardware"). En el archivo los registros aparecen al usuario en secuencia lógica, es decir, ordenados linealmente.

El archivo es una estructura de datos externa al programa. El programa accede al archivo para leer, modificar o escribir un registro. Al leer se transfiere de bloque en bloque la información del archivo a un área o zona de la memoria principal asociada a las entradas/salidas del archivo (buffer), de forma que esta información sea procesable por el programa. El programa puede igualmente transferir información de este área al archivo, modificando su contenido añadiéndole información.

Los archivos se guardan o almacenan en dispositivos de memoria masiva, estando limitados en tamaño tan solo por el de los dispositivos que los albergan. Los dispositivos o soportes de memoria auxiliar, pueden ser de tipo secuencial o no direccionables o de acceso directo o direccionables. En los soportes no direccionables si el último acceso se ha efectuado al registro físico m , para acceder al registro $s > m$, hay que leer o pasar por los registros situados entre el m y el s . En los soportes direccionables, por el contrario, se puede leer o escribir directamente un registro físico sin más que dar su dirección física, sin necesidad de recorrer o leer otros registros.

Una clave o identificativo es un campo (o conjunto de campos) que identifica cada registro del archivo. En un archivo puede haber, una, varias o ninguna clave. Las claves se pueden utilizar para la localización rápida de los registros en archivos con determinadas organizaciones. Cuando una clave se utiliza como campo de localización en el archivo la denominaremos llave.

Los registros que componen el archivo pueden no tener la misma longitud. Esto puede suceder o bien por existir campos de longitud variable o bien por haber campos que se repiten un número variable de veces (o ambas cosas simultáneamente). Por lo que respecta a la longitud de sus registros, los archivos pueden estar formados por registros de uno de los siguientes tipos:

- ✍ **Longitud fija.** Todos los registros tienen la misma longitud.
- ✍ **Longitud variable.** El sistema reserva una palabra al comienzo de cada registro para anotar su longitud.
- ✍ **Delimitados.** El sistema incluye un carácter especial, normalmente el carácter de separación de líneas al final de cada registro. En este caso se dice que el archivo es de tipo texto.
- ✍ **Indefinido.** En este caso, el sistema operativo no realiza ninguna gestión sobre la longitud de los registros del archivo. Es el programa de usuario el que se encarga de localizar el principio y final de cada registro.

TIPOS DE ARCHIVOS.

✍ **Permanentes:** Contienen información relevante para una aplicación, es decir, los datos necesarios para el funcionamiento de la misma. Su vida es larga y generalmente no pueden generarse de una forma inmediata a partir de otros archivos. Se clasifican en:

✍ Maestros: Contienen el estado actual de los datos susceptibles de ser modificados en la aplicación. Son el núcleo central de las aplicaciones. Todos los procesos están, en general, orientados a actualizar el archivo maestro o a obtener resultados de él.

✍ Constantes: Son aquellos que contienen datos fijos para la aplicación. En ellos las modificaciones son infrecuentes, accediendo normalmente a ellos para consultar.

✍ Históricos: Contienen datos que fueron actuales en tiempos anteriores. Se conservan para reconstruir la situación actual o situaciones anteriores.

✍ **Temporales:** Contienen información relevante para un determinado proceso o programa, pero no para el conjunto de la aplicación. Se generan a partir de los datos de archivos permanentes o para actualizar éstos, y su vida es generalmente muy corta. Se clasifican en:

✍ Intermedios: Se utilizan para almacenar resultados de un programa que han de ser utilizados por otro, dentro de una misma tarea.

✍ de maniobra: Se utilizan para almacenar datos propios de un programa que éste no puede conservar en memoria principal, por falta de espacio en ésta. Se encuentran normalmente en programas de cálculo numérico, compiladores y editores. Su vida es siempre menor que el tiempo de ejecución del programa.

✍ de resultados: Se utilizan para almacenar datos elaborados que van a ser transferidos a un dispositivo de salida, como por ejemplo un archivo de impresión, que contiene datos que van a ser transferidos a una impresora.

ORGANIZACIÓN DE ARCHIVOS.

✍ **Secuencial.** Los registros se encuentran en cierto orden yuxtapuestos consecutivamente. Los registros han de ser leídos necesariamente según este orden.

✍ **Indexada.** Se utiliza un índice para obtener la ubicación de la zona del archivo donde se encuentra el registro buscado. Esto permite localizar un registro sin leer previamente todos los que le preceden (sólo los de su zona).

✍ **Encadenada.** Cada registro contiene un puntero que permite localizar el siguiente registro. El archivo tiene la misma estructura que una lista lineal.

✍ **Directa.** La ubicación de registro en el soporte se obtiene a partir del valor de la llave (mediante un algoritmo de transformación de ésta).

✍ ✍ ✍ ✍ ✍ ✍ ✍