

Preparador Informática

www.preparadorinformatica.com

TEMA 14. INFORMÁTICA / S.A.I.

UTILIZACIÓN DE FICHEROS SEGÚN SU ORGANIZACIÓN.

TEMA 14 INF / SAI: UTILIZACIÓN DE FICHEROS SEGÚN SU ORGANIZACIÓN

1. INTRODUCCIÓN

2. FICHEROS

2.1. DEFINICIÓN

2.2. CARACTERÍSTICAS

2.3. OPERACIONES CON FICHEROS

2.4. TIPOS DE FICHEROS

2.4.1. FICHEROS PERMANENTES

2.4.2. FICHEROS TEMPORALES

3. ORGANIZACIONES

3.1. ORGANIZACIÓN SECUENCIAL

3.2. ORGANIZACIÓN RELATIVA

3.3. VARIANTES DE LA ORGANIZACIÓN SECUENCIAL

3.3.1. ORGANIZACIÓN SECUENCIAL ENCADENADA

3.3.2. ORGANIZACIÓN SECUENCIAL INDEXADA

3.3.3. ORGANIZACIÓN SECUENCIAL INDEXADA-ENCADENADA

4. PARÁMETROS DE UTILIZACIÓN DE UN FICHERO

5. CONCLUSIÓN

6. BIBLIOGRAFÍA

1. INTRODUCCIÓN

Los sistemas de almacenamiento de la información han ido evolucionando con el paso del tiempo. En un principio, la información se gestionaba mediante sistemas de ficheros manuales. Por ejemplo, en los colegios se guardaban archivados en un conjunto de carpetas etiquetadas los expedientes de cada uno de los alumnos con sus datos personales y académicos.

Pero a medida que el archivo manual aumentaba y que la información que se gestionaba era más compleja fue necesario sustituir este sistema por otro informatizado. De este modo, se pasó de tener los datos en formato papel a guardar los datos en ficheros informatizados estructurados en campos y registros, permitiendo acceder a ellos de manera mucho más rápida.

A finales de los años 60 surgen las bases de datos, que son conjuntos de datos relacionados entre sí, organizados y estructurados, con la ventaja de que los programas que utilizan dichos datos no tienen que preocuparse del almacenamiento físico de los mismos.

Actualmente existen sistemas de análisis de datos y big data que trabajan no solamente con datos estructurados sino también con datos no estructurados. Esto ha supuesto un desafío para poder organizar y homogeneizar datos de diferentes fuentes como ayuda a una toma de decisiones más eficiente por parte de las organizaciones.

Hoy en día los sistemas de ficheros están superados, pero es importante conocer sus características y organizaciones, por lo que el presente tema está dedicado a estudiar los primeros sistemas basados en ficheros para el almacenamiento y gestión de la información.

2. FICHEROS

2.1. DEFINICIÓN

Para almacenar la información de modo permanente se utilizan dispositivos de almacenamiento masivo denominados memoria secundaria, ya que los datos guardados en la memoria principal desaparecen al desconectar el ordenador.

La información contenida en los dispositivos de almacenamiento se estructura en unidades denominadas ficheros.

Es decir, un fichero es un conjunto de información relacionada de forma lógica y organizada de forma estructurada a la que se le da un nombre para identificarlo de forma unívoca

2.2. CARACTERÍSTICAS

Los ficheros están formados por registros lógicos que contienen datos relativos a un mismo elemento u objeto (por ejemplo, los datos de usuarios de una plataforma educativa). A su vez, los registros están divididos en campos que contienen cada una de las informaciones elementales que forman un registro (por ejemplo, el nombre del usuario, teléfono, email).

Terminología básica:

- **Datos:** son las características que se almacenan para obtener, a partir de ellas, alguna información.
- **Campo:** contiene cada una de las informaciones elementales que forman un registro. Es la mínima unidad de información dentro de un registro.
- **Registro:** es un conjunto de campos lógicamente relacionados que describen un mismo elemento u objeto. Es también la unidad de tratamiento de los ficheros de datos.
- **Fichero:** es un conjunto de registros relacionados.

Ejemplo: Fichero con los datos de usuarios de una plataforma educativa. Cada registro está consta de 5 campos.

	Usuario	Nombre	Apellidos	Email	Ciudad
Registro 1:	jgarfer	Juan	García Fernández	jgarfer@jgarfer.es	Granda
Registro 2:	asanper	Alberto	Sánchez Pérez	asanper@asanper.es	Sevilla
Registro 3:	fmonpad	Fernando	Montes Padilla	fmonpad@fmonpad.es	Madrid
Registro 4:	rhersan	Raúl	Hernández Sanz	rhersan@rhersan.es	Valencia

Como los ficheros suelen ser muy voluminosos, solo se pueden llevar a la memoria principal partes de ellos para poder procesarlos. La cantidad de información que es transferida entre el soporte en el que se almacena el fichero, y la memoria principal, en una sola operación de lectura/escritura, recibe el nombre de registro físico o bloque.

En cada operación de lectura/escritura se transfieren varios registros del fichero, es decir un bloque suele contener varios registros lógicos. Al número de registros que entran en un bloque se le conoce con el nombre de factor de blocaje, y a esta operación de agrupar varios registros en un bloque se le llama bloqueo de registros.

Preparador Informática

2.3. OPERACIONES CON FICHEROS

Las operaciones principales que se pueden realizar sobre un fichero son:

- **Crear:** consiste en crear la estructura de un fichero. Operación previa a cualquier otra.
- **Abrir:** una vez creado el fichero debe abrirse para poder trabajar con él.
- **Consultar:** consiste en recuperar la información contenida en el fichero.
- **Actualizar:** una vez que el fichero ha sido creado, puede surgir la necesidad de modificar la información que contiene. Esto supone tres tipos de operaciones distintas a nivel de registro:
 - Inserción de un registro.
 - Modificación de un registro.
 - Eliminación de un registro.

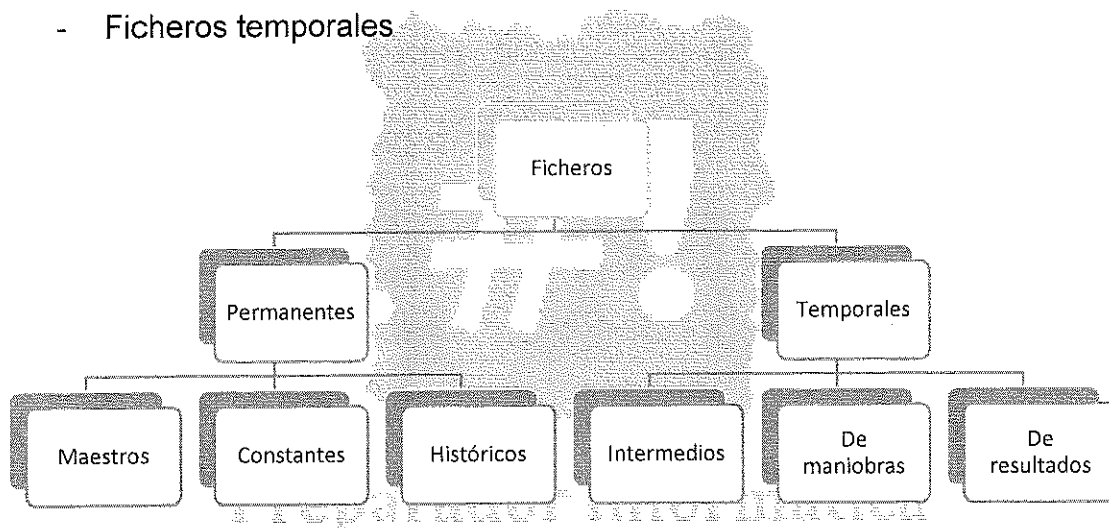
- **Renombrar:** consiste en cambiar el nombre del fichero.
- **Cerrar:** después de terminar de utilizar un fichero se debe cerrar.
- **Borrar:** elimina la información del fichero, así como su estructura.
- **Etc.**

2.4. TIPOS DE FICHEROS

Los ficheros se pueden clasificar atendiendo a diversos criterios, pero uno de los criterios más utilizados es según la función que vayan a desempeñar los ficheros.

De este modo, podemos distinguir entre:

- Ficheros permanentes
- Ficheros temporales



2.4.1. FICHEROS PERMANENTES

Los ficheros permanentes almacenan información relevante para una aplicación. Es decir, los datos necesarios para el funcionamiento de ésta. Tienen un periodo de permanencia en el sistema amplio. Se clasifican en:

- **Ficheros maestros:** almacenan el estado actual de los datos que pueden modificarse desde la aplicación. Es la parte central de la aplicación, su núcleo. Ejemplo: fichero con los datos de los usuarios de una plataforma educativa.
- **Ficheros constantes:** almacenan datos fijos para la aplicación. No suelen ser modificados y se utilizan principalmente como ficheros de consulta. Ejemplo: fichero con códigos postales.

- **Ficheros históricos:** almacenan datos que fueron considerados como actuales en un periodo o situación anterior. Se utilizan para la reconstrucción de situaciones. Ejemplo: fichero con los usuarios que han sido dados de baja en una plataforma educativa.

2.4.2. FICHEROS TEMPORALES

Los ficheros temporales almacenan información relevante para una parte de la aplicación, pero no para el conjunto de la aplicación. Se generan a partir de datos de ficheros permanentes. Tienen un periodo corto de existencia en el sistema. Se clasifican en:

- **Ficheros intermedios:** almacenan datos de una aplicación que serán utilizados por otra aplicación.
- **Ficheros de maniobras:** almacenan datos de una aplicación que no pueden ser mantenidos en memoria principal por falta de espacio.
- **Ficheros de resultados:** almacenan datos que van a ser transferidos a un dispositivo de salida. Ejemplo: fichero de impresión.

3. ORGANIZACIONES

Hay diferentes formas de estructurar u organizar los registros en un fichero para almacenarlos en un soporte de almacenamiento. Al conjunto de reglas utilizadas para introducir información en forma de registros en los soportes de almacenamiento se denomina organización de ficheros.

Básicamente los dos modos de organizar los ficheros son:

- 1) Organización secuencial
- 2) Organización relativa

3.1. ORGANIZACIÓN SECUENCIAL

Un fichero con organización secuencial se caracteriza por almacenar los registros en posiciones físicas contiguas de memoria, de modo que para acceder al registro n es obligatorio recorrer los $n-1$ registros anteriores.

En este tipo de ficheros hay una correspondencia entre el orden físico (orden en el que están grabados los registros) y el orden lógico (orden en el que se han dado de alta y recuperado los registros). Puede utilizarse con soportes tanto secuenciales, como direccionables.

Las **ventajas** de los ficheros con organización secuencial son:

- Rapidez en el acceso a un bloque de registros contiguos
- No se requiere realizar operaciones de compactación del fichero
- No se desperdicia espacio en el dispositivo de almacenamiento porque no hay huecos
- Se pueden utilizar cualquier tipo de registros (de longitud fija, variable o indefinida).

Los **inconvenientes** de este tipo de organización son:

- Para *acceder* al registro n hay que recorrer los $n-1$ registros anteriores. El acceso es secuencial.
- Para realizar una *consulta* hay que crear un proceso en el que se compare el valor del campo que se pretende localizar con el valor del mismo campo correspondiente a cada registro leído del fichero.
- La *adición* de registros se realiza a continuación del último registro ya existente. No se pueden insertar nuevos registros.
- No se pueden *eliminar* registros. Para eliminar un registro se marca de modo que no se muestre, pero el registro existe y ocupa espacio en el dispositivo del almacenamiento. (borrado lógico)
- Para *mantener ordenado* y compactado el fichero, hay que crear un fichero nuevo a partir del existente.

La organización secuencial es adecuada para ficheros en los que no se realizan procesos interactivos, y en los que cada proceso debe acceder a la mayor parte

de los registros del fichero. En cambio, no son la opción más adecuada cuando se necesita procesar frecuentemente registros aislados, con un índice de utilización muy bajo.

3.2. ORGANIZACIÓN RELATIVA

Un fichero con organización relativa se caracteriza porque el almacenamiento físico de los registros se hace a través de una clave que indicará la posición del registro en el fichero. Hay dos tipos de organización relativa:

A. **Organización relativa directa:** la clave coincide con la dirección, debiendo ser numérica y comprendida dentro del rango de valores de las direcciones. La secuencia lógica de almacenamiento de los registros en el fichero coincide con la secuencia física de almacenamiento de los registros sobre el dispositivo.

Las **ventajas** de este tipo de organización de ficheros son:

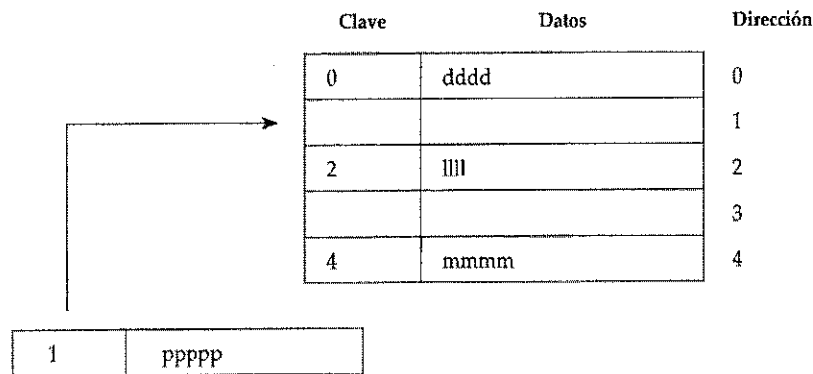
- Acceso directo a los registros. No se necesita un algoritmo de transformación
- Permite realizar operaciones de escritura y lectura simultáneamente, ya que primero se localiza el registro y luego se realiza la operación deseada: inserción, eliminación, consulta, modificación, etc.
- El acceso a los datos se realiza de dos formas diferentes:
 - Directamente, mediante la clave del registro
 - Secuencialmente, a partir del primer registro almacenado en el fichero, por lo que son muy rápidos en el tratamiento individual de registros.

Los **inconvenientes** de este tipo de organización de ficheros son:

- Al realizar un acceso secuencial, en una consulta sobre todos los registros del fichero hay que recorrer todas las direcciones, aunque estén vacías.
- Deja gran cantidad de posiciones libres de memoria dentro del fichero (huecos), debido a que las claves de los registros pueden indicar posiciones de almacenamiento no contiguas, lo que implica una falta de

aprovechamiento del soporte del almacenamiento respecto al número real de registros almacenados.

- Se producen colisiones, ya que puede existir más de un registro con la misma clave, lo que provoca errores.



B. Organización relativa aleatoria. Se utiliza una función matemática (hash), que a partir de la clave y mediante una transformación genera la dirección de cada registro. Es decir, las direcciones lógicas de almacenamiento (claves), no coinciden con las direcciones físicas. Si a partir de diferentes claves se obtiene la misma dirección se produce una colisión. Estos registros se almacenan en lo que se conoce como zona de overflow o sinónimos.

Las **ventajas** de este tipo de organización de ficheros son:

- Acceso inmediato a los registros mediante su clave
- No es necesario ordenar el fichero
- Se pueden realizar operaciones de escritura y lectura a la vez
- Son muy rápidos en el tratamiento individual de registros.
- Se pueden realizar accesos secuenciales.

Los **inconvenientes** de este tipo de organización de ficheros son:

- Las consultas sobre todo el fichero son lentas
- El fichero contiene gran cantidad de huecos o espacios libres
- El algoritmo para la conversión de las claves y el algoritmo necesario para el almacenamiento de sinónimos han de ser creados de modo que dejen el menor número de huecos libres y se genere el menor número de sinónimos.

**Direcciones
de memoria**

1
2
3
4
5
6
7
8

Clave	Datos
FH54	Registro B
ES62	Registro A
CT24	Registro E
AR01	Registro C
DF32	Registro F
JK59	Registro D

Huecos libres

Zona de sinónimos

La organización relativa es adecuada para ficheros en los que el acceso a los datos de un registro se hace siempre empleando la misma clave y la velocidad de acceso a un registro es lo que más importa.

3.3. VARIANTES DE LA ORGANIZACIÓN SECUENCIAL

Con el objetivo de mejorar las prestaciones de la organización secuencial surgen una serie de organizaciones que son una variante de esta y que pueden ser utilizados con soportes direccionables:

- Organización secuencial encadenada
- Organización secuencial indexada
- Organización secuencial indexada - encadenada

3.3.1. ORGANIZACIÓN SECUENCIAL ENCADENADA

Un fichero con organización secuencial encadenada almacena junto a cada registro un puntero con la dirección del registro siguiente, según el orden lógico del fichero. En estos ficheros la secuencia física y la secuencia lógica no coinciden, pudiendo ocurrir que el último registro en secuencia física sea el primero en secuencia lógica y viceversa.

La organización secuencial encadenada es adecuada para ficheros en los que se deben realizar frecuentes inserciones de pocos registros. Si frecuentemente, se deben insertar muchos registros, en comparación con el número de registros del fichero, puede ser preferible utilizar un fichero secuencial puro.

3.3.2 ORGANIZACIÓN SECUENCIAL INDEXADA

Un fichero con organización secuencial indexada cuenta con

- **Área primaria:** contiene los registros ordenados por clave agrupados en segmentos
- **Área de índices:** contiene la clave más alta y la dirección de comienzo para cada segmento.
- **Área de overflow:** contiene los registros añadidos después de crear el fichero.

Una vez que se accede al primer registro del segmento, dentro de él se localiza (de forma secuencial) el registro buscado.

Tabla de índices		Claves	Datos	Direcciones
68	0	12	A	0
138	3	25	B	1
205	6	68	C	2
		89	D	3
		123	E	4
		138	F	5
		198	G	6
		204	H	7
		205	I	8

Zona de índices Zona de registros de datos

La organización secuencial indexada es adecuada para ficheros que deben combinar consultas a registros concretos y el procesamiento secuencial de todo el fichero.

3.3.3. ORGANIZACIÓN SECUENCIAL INDEXADA-ENCADENADA

Un fichero con esta organización aprovecha lo mejor de la organización secuencial encadenada e indexada. Se caracteriza por la utilización de punteros e índices de forma simultánea, lo que implica un aumento del espacio ocupado, pero proporciona una gran rapidez en la búsqueda de registros.

4. PARÁMETROS DE UTILIZACIÓN DE UN FICHERO

En función del uso que se vaya a dar al fichero, serán adecuados unos tipos u otros de organización. Mediante la utilización de **parámetros de referencia**, se puede determinar el uso de un fichero. Estos parámetros son:

- a) **Capacidad:** permite conocer el espacio, en caracteres, que ocupa el fichero. Para determinar la capacidad se debe multiplicar el número previsto de registros por la longitud media de los registros.
- b) **Actividad:** permite conocer la cantidad de consultas y modificaciones que se realizan en el fichero. Para determinar la actividad se deben tener en cuenta:
 - **Tasa de consulta o modificación:** porcentaje de registros consultados o modificados en cada tratamiento del fichero.
 - **Frecuencia de consulta o modificación:** número de veces que se accede al fichero para hacer una consulta o modificación en un periodo de tiempo fijo.
- c) **Volatilidad:** permite conocer la cantidad de inserciones y borrados que se realizan en un fichero. Para determinar la volatilidad se deben tener en cuenta:
 - **Tasa de renovación:** porcentaje de registros renovados en cada tratamiento del fichero.
 - **Frecuencia de renovación:** número de veces que se accede al fichero para renovarlo en un periodo de tiempo fijo.
- d) **Crecimiento:** permite conocer la variación de la capacidad del fichero y se mide con la tasa de crecimiento, que es el porcentaje de registros en que aumenta en cada tratamiento el fichero.

5. CONCLUSIÓN

En este tema se ha presentado una visión global de los ficheros, los cuales permiten organizar y memorizar conjuntos de datos del mismo tipo o naturaleza con una determinada estructura, siendo un medio para el almacenamiento de los datos o resultados de una aplicación específica. Pero si las aplicaciones, al ser diseñadas, deben depender directamente de sus ficheros o archivos, se pierde independencia y surgen serios inconvenientes: como información duplicada, incoherencia de datos, fallos de seguridad, etc.

Para solucionar estos problemas, surge el concepto de base de datos. Una base de datos permite reunir toda la información relacionada en un único sistema de almacenamiento, pudiendo cualquier aplicación utilizarla de manera independiente y ofreciendo una mejora en el tratamiento de la información, así como una evolución para el desarrollo de aplicaciones.

6. BIBLIOGRAFÍA

- Joyanes Aguilar, L. **Fundamentos de programación. Algoritmos, estructuras de datos y objetos**. Editorial McGraw-Hill
- Prieto A., y otros. **Introducción a la informática**. Editorial McGraw-Hill
- Korth y Silberschatz. **Fundamentos de bases de datos**. Editorial McGraw-Hill
- Tanenbaum, A. **Sistemas operativos modernos**. Editorial Prentice-Hall
- www.xataka.com (Web sobre actualidad TIC e informática)