

TEMA 14

UTILIZACIÓN DE FICHEROS SEGÚN SU ORGANIZACIÓN

INDICE:

1. INTRODUCCIÓN	1
2. FICHEROS.....	1
2.1. Estructura de un fichero	1
2.2. Operaciones sobre ficheros	1
3. ORGANIZACIÓN INTERNA	2
4. TIPOS DE ORGANIZACIONES	2
4.1. Secuencial	2
4.2. Secuencial encadenada.....	3
4.3. Secuencial indexada.....	4
4.4. De acceso directo	5
5. CONCLUSIÓN	6
6. BIBLIOGRAFÍA	6
7. NORMATIVA	7

Realizado por Cayetano Borja Carrillo

Tiempo de escritura: 1 hora y 40 minutos

1. INTRODUCCIÓN

Un ordenador es una máquina encargada de procesar información digital, es decir, realiza cálculos con datos codificados en binario y devuelve unos resultados.

Cuando un programa se está ejecutando, es normal que necesite leer o almacenar información de soportes de almacenamiento permanente, como un disco duro. La información que contienen estos soportes no son ceros y unos sin ningún sentido, sino que están organizados en zonas llamados ficheros o archivos.

En este se desarrollan las distintas formas que pueden organizarse los datos en un fichero y los usos para los que se recomienda cada una. Se trata de un tema de gran importancia dentro del campo de estudio de los sistemas operativos ya que, sin los ficheros, un ordenador no podría almacenar los resultados para su posterior consulta.

2. FICHEROS

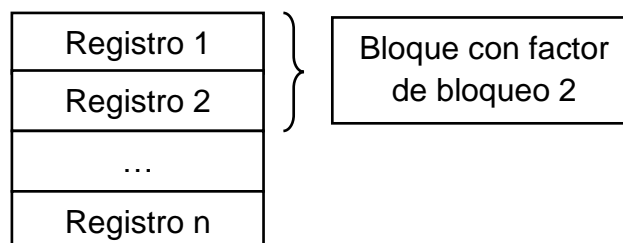
Un fichero es un recurso informático compuesto de paquetes de datos más pequeños llamados registros. Cada registro contiene información de una entidad y se divide en campos de distinto tipo y tamaño.

2.1. Estructura de un fichero

A continuación, se muestra la estructura de un fichero que contiene 3 registros:

	Campo 1	Campo 2	Campo 3
Registro 1	DNI	Nombre	Apellidos
Registro 2	DNI	Nombre	Apellidos
Registro 3	DNI	Nombre	Apellidos

Cuando un programa necesita leer un fichero, el sistema operativo empaqueta los registros en bloques, después los transporta a un buffer de E/S y, finalmente, los traslada a la RAM. Cada bloque puede contener varios registros y su tamaño, llamado factor de bloqueo, depende de las propiedades físicas del ordenador.



2.2. Operaciones sobre ficheros

El sistema operativo permite a los programas realizar una serie de operaciones con los ficheros. Estas operaciones son las siguientes:

- Creación: Se crea la estructura del fichero en un soporte de almacenamiento.
- Apertura: Se localiza y se carga un fichero existente en la memoria RAM para empezar a trabajar con él.
- Cierre: Se “rompe” la conexión entre el programa y el archivo de forma segura.
- Escritura: Guarda en el fichero los cambios realizados en él.
- Borrado: Elimina el fichero del soporte de almacenamiento.
- Renombrado: Permite cambiar el nombre del fichero.
- Copia: Se obtiene un duplicado del fichero.

3. ORGANIZACIÓN INTERNA

La organización interna de un fichero hace referencia a cómo se organizan de forma física y lógica los registros que lo componen.

La organización física se refiere a cómo están realmente almacenados los datos de un fichero en un dispositivo de almacenamiento externo como un disco duro, un pendrive o una unidad de estado sólido (SSD).

La organización lógica se refiere a cómo están organizados los ficheros en la estructura jerárquica de carpetas que ofrece el sistema de archivos. Cada carpeta puede contener un número arbitrario de ficheros y otras carpetas, formándose una estructura con forma de árbol. Al primer directorio de la estructura se le llama “directorio raíz”.

4. TIPOS DE ORGANIZACIONES

Existen dos formas principales de organización de ficheros:

- Secuencial: Para acceder a un registro hay que recorrer todos los registros anteriores.
- De acceso directo o relativo: Se puede acceder directamente a un registro concreto con sólo saber su dirección física.

4.1. Secuencial

En un fichero con esta organización, todos sus registros tienen la misma longitud, el mismo número de campos y cada campo tiene un tamaño fijo. Los registros se van almacenando de forma contigua uno al lado de otro según su orden de llegada, sin dejar huecos entre ellos. La organización física de los registros es idéntica a la organización lógica.

Registro 1	Registro 2	...	Registro n-1	Registro n
------------	------------	-----	--------------	------------

Las operaciones que se pueden realizar sobre ficheros con esta organización son:

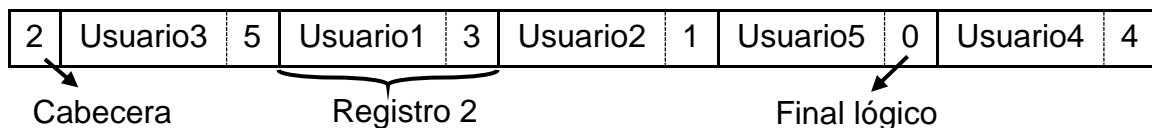
- **Insertión:** El registro se introduce siempre al final del fichero.
- **Búsqueda:** Se recorren secuencialmente todos los registros uno a uno desde el principio hasta llegar al deseado.
- **Modificación:** Se sobrescribe el registro con las modificaciones.
- **Borrado lógico:** El registro no se borra realmente, sino que se escribe un valor o marca en él para indicar que ese registro debe de ser ignorado.
- **Borrado físico:** Se borra el registro totalmente. Para hacerlo, se copian todos los registros en un fichero intermedio sin tener en cuenta los registros borrados y luego se vuelca esa información al fichero original.

Este tipo de organización tiene unas lecturas muy lentas y unas inserciones muy rápidas. Se recomienda utilizar cuando interese almacenar información que no se consulte con frecuencia.

4.2. Secuencial encadenada

Cada registro almacena, además de su propia información, un campo llamado puntero que indica donde se encuentra físicamente el siguiente registro. Los registros están ordenados (alfabéticamente o numéricamente) de forma lógica según su clave, aunque físicamente pueden estar desordenados.

Al comienzo del fichero hay un puntero inicial llamado cabecera que indica donde está el primer registro y el último registro tendrá un puntero que indica el fin lógico.



Las operaciones que se pueden realizar con esta organización son:

- **Insertión:** Se escribe el registro al final del fichero y se reajustan los punteros para que sigan ordenados de forma lógica.
- **Búsqueda:** Se recorren secuencialmente los registros según la posición que indican los punteros empezando por la cabecera. Por ejemplo, para buscar el elemento "Usuario3" en la estructura anterior, primero se lee la cabecera que apunta al registro 2 (Usuario1), el registro 2 apunta al registro 3 (Usuario2) y este al registro 1 (Usuario3).
- **Modificación:** Si el tamaño del registro no cambia, se hace de forma directa sobrescribiendo los campos y reajustando los punteros si es necesario. Si el tamaño del registro cambia hay dos opciones:
 - **Modificación lógica:** Se almacena el registro modificado al final del fichero, se borra de forma lógica el antiguo y se reajustan los punteros.

- **Modificación física:** Se crea un fichero intermedio con todos los registros junto el modificado y se vuelcan esos datos al fichero original.
- **Borrado lógico:** Se modifican los punteros para que el registro sea inaccesible, aunque realmente estará en el fichero ocupando espacio.
- **Borrado físico:** Se utiliza un fichero intermedio.

Debido a los punteros, esta organización ocupa más espacio y las inserciones son más lentas que la anterior, pero las lecturas suelen ser más rápidas.

4.3. Secuencial indexada

Un fichero con esta organización está compuesto por las siguientes 3 áreas o zonas:

- **Área primaria o de datos:** Contiene los registros de datos de forma secuencial, ordenados ascendentemente por el valor de su clave. El sistema divide esta área en segmentos que abarcan varios registros.
- **Área de índices:** Almacena secuencialmente registros con 2 campos: la clave del último registro de un segmento y un puntero que indica donde se encuentra el primer registro de ese segmento.
- **Área de excedentes o de *overflow*:** Lugar donde se almacenan los nuevos registros. Mantienen una organización secuencial encadenada y se trasladarán al área primaria en la siguiente reorganización.

La estructura de un fichero con esta organización con segmentos de 4 registros es la siguiente:

Área de índices		Área primaria			} Segmento
016	1	1	001	Nombre1	
		2	005	Nombre2	
041	5	3	012	Nombre3	
		4	016	Nombre4	
		5	022	Nombre5	
		6	028	Nombre6	
Área de excedentes		7	035	Nombre7	Apellidos
067	Nombre9	Apellidos			
076	Nombre10	Apellidos			

Las operaciones que se pueden realizar con esta organización son:

- **Inserción:** Se inserta en el área de excedentes.
- **Búsqueda:** Se busca en el área de índices el segmento donde debería de estar el registro. Después se busca el registro en ese segmento del área primaria. Si el registro no se encuentra ahí, se busca en el área de excedentes. Todas las búsquedas son secuenciales.

- Borrado: Se utiliza un fichero intermedio.
- Modificación: Se utiliza un fichero intermedio excepto si el registro modificado no cambia su tamaño ni el valor de su campo clave.
- Reorganización: Los registros que hay en el área de excedentes pasan al área primaria y se actualiza el área de índices.

Esta organización desperdicia mucho espacio por todas las zonas extra que tiene que almacenar, pero las búsquedas son muy rápidas. Se recomienda utilizar cuando se vaya a consultar información con mucha frecuencia.

4.4. De acceso directo

El espacio del fichero se fija previamente en función de los datos que hay que almacenar y se divide en celdas numeradas. Cada registro se almacena en una celda y puede ser accedido a través de la dirección de celda.

El número de celda donde se ubica cada registro se calcula teniendo en cuenta el valor de su campo clave. Existen 3 métodos básicos de calcularlo y son:

- Directa: Si la clave es numérica, la dirección de celda del registro será el mismo número que el valor de su clave. Por ejemplo, si la clave es 3, el registro se almacena en la tercera celda.

1	001	Nombre1	Apellidos1
2			
3	003	Nombre2	Apellidos2
4			
5	005	Nombre3	Apellidos3

- Asociada: Se suele usar cuando las claves son alfanuméricas. Se construye una tabla en la que figurarán todas las claves y la dirección física donde se encuentra el registro correspondiente. Al añadir nuevos registros, las claves se colocan al final de la tabla, por lo que normalmente estará desordenada, aunque se puede ordenar. Para localizar las claves en la tabla se hace por lectura secuencial.

Tabla

Matrícula1	1
Matrícula2	2
Matrícula3	3

Celdas

1	Matrícula1	Opel	Rojo
2	Matrícula2	Ford	Azul
3	Matrícula3	Fiat	Blanco
4			

- Calculada (Hashing): La dirección de cada registro se obtiene mediante una función *hash*, que consiste en aplicar un algoritmo matemático al valor de su

clave. Este método produce colisiones, pues distintos cálculos pueden llevar al mismo resultado. Un ejemplo de función *hash* es la “Mitad del cuadrado”, donde el valor de clave (si es numérica) se eleva al cuadrado y se extraen las cifras intermedias. Ejemplo: Clave = 64 $\rightarrow 64^2 = 4096 \rightarrow$ Dirección = 09. Cuando aparece una colisión, se aplica un algoritmo de resolución de colisiones, como el doble *hasheo*, que determinará dónde colocar el registro.

Las operaciones que se pueden realizar con esta organización son:

- Búsqueda e inserción: Se calcula la dirección de celda donde debe de estar y se lee o se introduce el registro.
- Modificación: Se sobrescribe la información si el registro mantiene el mismo tamaño o se utiliza un fichero intermedio si no.
- Borrado: Se utiliza un fichero intermedio.

Esta organización es la que más espacio ocupa, pero las operaciones son muy rápidas. Se recomienda utilizar en ficheros que almacenen pocos registros y se van a hacer muchas operaciones.

5. CONCLUSIÓN

A los primeros ordenadores se le introducían los datos en unas tarjetas perforadas, hacían los cálculos y devolvían los resultados. Cuando aparecieron los primeros dispositivos de almacenamiento capaces de mantener información sin corriente eléctrica supuso un gran avance, pero también un gran reto: ¿cómo almacenar y organizar esa información?

Los ficheros surgieron como respuesta a este reto y su aparición ha sido fundamental para el avance de la informática. Sin ellos, algo tan básico como almacenar una foto o reproducir una canción no sería posible.

6. BIBLIOGRAFÍA

- López Ureña, L. A. et al. (1997). *Fundamentos de Informática (1ª ed.)*. Rama.
- Prieto Espinosa, A. et al. (2006). *Introducción a la informática (4ª ed.)*. McGraw-Hill.
- Brookshear, J. G. (2012). *Introducción a la computación (11ª ed.)*. Pearson Educación.
- Joyanes Aguilar, L. (2020). *Fundamentos de programación (5ª ed.)*. McGraw-Hill.

7. NORMATIVA

Para el desarrollo de este tema, se ha tenido en cuenta la siguiente normativa, donde se especifican los contenidos, competencias y criterios de evaluación de los Ciclos Formativos y Bachillerato en Andalucía:

- Orden 7 de julio de 2009 (SMR). La parte correspondiente al módulo “Sistemas operativos monopuesto”.
- Orden 19 de julio de 2010 (ASIR). La parte correspondiente al módulo “Gestión de bases de datos”.
- Orden 16 de junio de 2011 (DAW/DAM). La parte correspondiente al módulo “Bases de datos” y “Acceso a datos”.
- Instrucción 13/2022 (Bachillerato). La parte correspondiente a la asignatura “Tecnologías de la Información y Comunicación”