
Tema 5: Soportes y Organización de Ficheros



Índice: Parte I

Parte I:

1. Gestión no automatizada de la información
2. Gestión automatizada de la información.
3. Tipos de soportes.
4. Tiempos de acceso.

1. Definición y conceptos.
 - 1.1. Conceptos previos: registros
 - 1.2. Ficheros..
2. Operaciones sobre ficheros.
3. Organización y procesamiento de ficheros.
 - 3.1. Secuencial.
 - 3.2. Secuencial ordenada.
 - 3.3. Secuencial ordenada encadenada.
 - 3.4. Directa.
 - 3.5. Aleatoria o relativa.
 - 3.6. Secuencial indexada. Tipos de índices: comprimido, disperso, multinivel, en árbol balanceado.
 - 3.7. En agrupamiento o cluster.



Parte I



Gestión no automatizada de la información.

¿Qué gestiones tenemos que hacer con un fichero manual?

- Crearlo: preparar el formato de la ficha con sus campos para imprimirlo.
- Abrir el archivo para efectuar el mantenimiento:

Rellenar un ficha.	—————→	Alta.
Suprimir una ficha.	—————→	Baja.
Rectificar una ficha.	—————→	Modificar.
Consultar un ficha.	—————→	Consultar.

- Cerrar el archivo.



Gestión no automatizada de la información.

Organización y método de acceso.

- **Organización de la información en un libro:**
 - **Índice:** permite el acceso directo a un contenido. Formado por:
 - Contenidos
 - Puntero hacia la página del libro donde está la información.
 - Inconveniente: consume un espacio adicional
 - Ventaja: *Más fácil leer secuencialmente el índice y acceder directamente a la información, que buscarla a través del libro.*
 - **Mantener ordenados los datos:** ej. diccionario.
 - Abrimos el diccionario, leemos una palabra y nos movemos hacia delante y hacia atrás.
 - Ventaja: no necesita índice.
 - Inconveniente: Los datos tienen que estar almacenados en orden.

Añadir palabra -> hacerle hueco.

Gestión no automatizada de la información.

- **Organización de la información en una biblioteca**
 - Almacenar fichas ordenadas con un criterio (p.e. Autor).
 - Tener varios archivadores con fichas ordenadas con distintos criterios.
- **Ordenador:** formas similares de almacenamiento y acceso a la información.
 - Necesidad de organizarla para localizarla.
 - Buscar la organización que se ajuste más a las necesidades.

Gestión automatizada de la información.

No Automatizada vs Automatizada

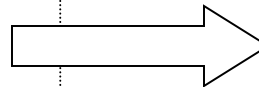
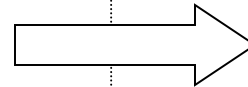
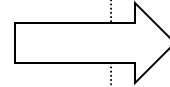
Biblioteca:

- **Archivadores** con fichas de libros.

- **Ficha:** información con datos de libros.

- **Datos:** título, autor, año, editorial, etc.

- Soporte físico de almacenamiento: papel.



Informática.

- **Fichero** o Archivo: conjunto de fichas del mismo tipo.

- **Registro.**

- **Campos:** nombre y longitud.

Gestión Automatizada de la Información. Conceptos de generales. Ficheros y Registros

- La información o datos se encuentra almacenada en los dispositivos de almacenamiento siguiendo una determinada estructura. A esta estructura se le llama registro.
- Un registro es un conjunto de datos que forman una unidad de información y que se manipulan como un bloque. Cada registro está compuesto de unidades más pequeñas elementales de información que se denominan campos.

Ejemplo: Registro: persona
Campos: nombre, dirección, edad, sexo, antigüedad, estado civil...

Registro

Identificador	Campo 1	Campo 2	Campo n
---------------	---------	---------	------	---------

Gestión Automatizada de la Información. Conceptos de generales. Ficheros y Registros

- Es un dato estructurado compuesto de estructuras más pequeñas que se denominan campos.
- Por ejemplo: Una compañía con los datos de los empleados.

- Registro Persona:

Campos:	Nombre	(alfabético)
	Dirección	(alfabético)
	Edad	(numérico)
	Sexo	(alfabético)
	Antigüedad	(numérico)
	Casado	(booleano)

Los registros son la estructura de datos más extendida.

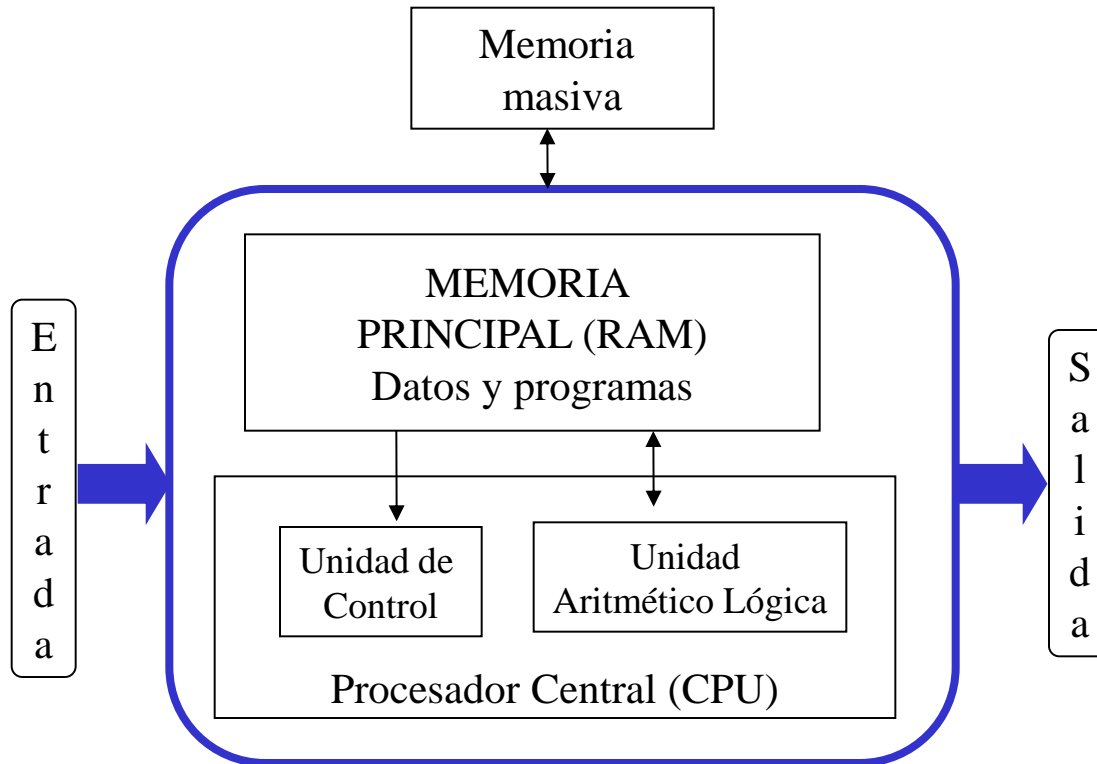
Gestión Automatizada de la Información. Conceptos de generales. Ficheros y Registros

- Definición de fichero:

Conjunto de informaciones sobre un mismo tema tratado como una unidad de almacenamiento y organizado de forma estructurada para la búsqueda de un dato individual

- Finalidad:
 - Almacenamiento permanente de la información en memoria.
 - Gestión de grandes volúmenes de datos.

Gestión Automatizada de la información.Introducción.

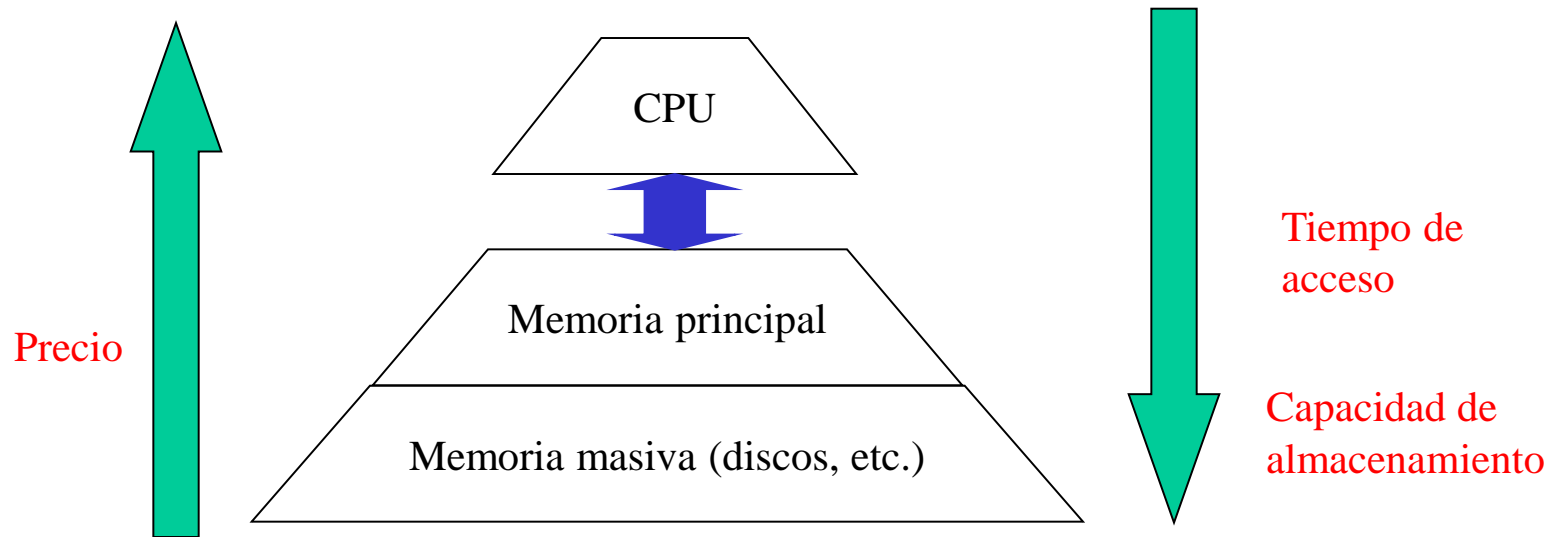


Gestión Automatizada de la Información.

Introducción

- Memoria principal: única memoria directamente accesible por el ordenador.
 - Es cara.
 - Disponibilidad limitada.
 - Es volátil: la información almacenada no es permanente.
- Memorias periféricas:
 - Más baratas.
 - Mayor capacidad.
 - No volátil.
 - Almacenamiento permanente en forma de **archivos o ficheros**.
- Veremos:
 - Tipos de soportes periféricos.
 - Métodos de acceso a los datos almacenados en archivos.
 - Organización de archivos o ficheros

Gestión Automatizada de la Información.Introducción.









Gestión automatizada de la información.

Gestiones a hacer en un fichero informático.

- Crearlo: definir sus campos
 - Nombre. Ej: Año
 - Tipo. Ej: entero.
 - Dominio. Ej: 1900-actualidad.
- Abrirlo para realizar operaciones:
 - Añadir registros.
 - Suprimir registros.
 - Modificar campos de un registro.
 - Consultar información de un registro
- Cerrarlo.

Gestión automatizada de la información.

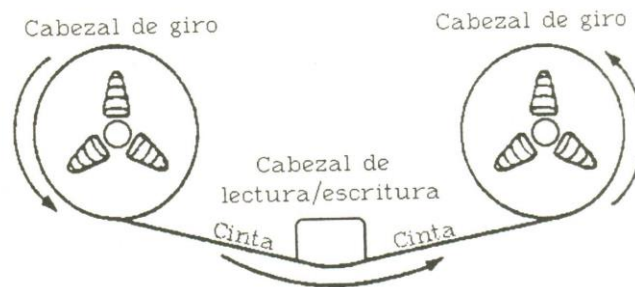
Gestiones a hacer en un fichero informático.

- Abrir fichero implica avisar al sistema de que se va a trabajar con él → solicitar que prepare una zona de memoria donde transferir los datos.
- Cerrar el fichero implica escribir lo que hay en el buffer y liberar la memoria.
- En ocasiones necesitamos **reorganizar** un fichero: optimizar la colocación de los registros para aprovechar los huecos.
- Una vez abierto hacemos:
 - Altas.
 - Bajas.
 - Modificaciones.
 - Consultas de un registro
 - Consultas globales.

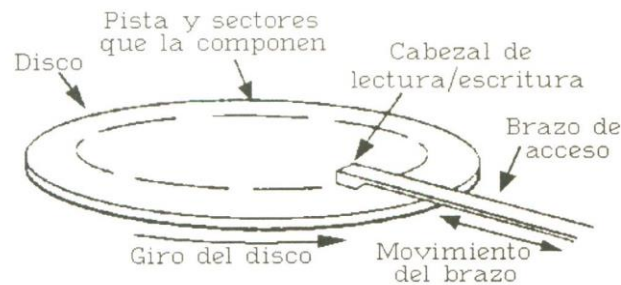
Gestión automatizada de la información.

Relación entre el soporte y el método de acceso.

- Cintas magnéticas: gestión secuencial de la información.



- Discos magnéticos: permite acceso directo **y secuencial**.



Gestión automatizada de la información.

Organización de ficheros y métodos de acceso.

- Acceso: secuencial o directo (dependiendo del soporte).
- Necesidad de organizar los datos (ordenarlos, hacer un índice...).
 - La elección de la forma de ordenación del fichero dependerá de las expectativas de uso.
 - Ej: Si se van a realizar muchas altas buscaremos una organización del fichero que haga altas rápidas.
 - A tener en cuenta en la elección de una organización:
 - La necesidad de reorganización.
 - El consumo de memoria secundaria.
 - La facilidad para desarrollar el sistema.

Gestión automatizada de la información.

- Criterios de eficiencia de la organización de un fichero:
 - Tiempo para leer un registro concreto.
 - Tiempo para leer el siguiente registro físico o todos.
 - Tiempo para insertar un nuevo registro.
 - **Tiempo para modificar un registro existente.**
 - Tiempo para borrar un registro.
 - Tiempo y frecuencia de la reorganización.
 - **Requisitos de espacio.**
 - Simplicidad programación. Reutilización.

Gestión automatizada de la información.

- Análisis de los patrones de uso de una aplicación:
 - Tipo de información y su volumen inicial
 - Aumento o disminución de su volumen en el tiempo
 - Frecuencia (absoluta) de cada operación a realizar con los datos
 - Prioridad (relativa) de cada operación para la organización
 - **Distribución en el tiempo de cada operación (cuello de botella)**
 - Continuidad de servicio: reorganización y copia de seguridad

Gestión Automatizada de la Información. Tipos de soportes.

Tipos de soporte según su conexión al ordenador.

- Memoria de almacenamiento:
 - Secundaria -> continuamente accesible al ordenador (disco duro).
 - Terciaria -> necesidad de colocarla en un dispositivo de lectura para que esté accesible (cintas, discos, etc.).
- Secundaria-> **más rápida y más cara.**
- Terciaria -> **más lenta y más barata.**

Gestión Automatizada de la Información. Tipos de soportes.

Tipos de soporte en cuanto al método de acceso.

- De acceso secuencial: tipo cinta.
 - Más baratos.
 - Más lentos.
 - Más fáciles de romperse.
 - Gran capacidad.
 - Suelen utilizarse para archivos históricos y copias de seguridad.
- De acceso secuencial y acceso directo: discos.
 - Más útiles por su acceso directo.

Gestión Automatizada de la Información. Tipos de soportes.

Tipos de soportes en cuanto a la tecnología de grabación

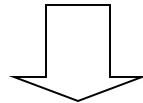
Soportes magnéticos.

Soportes ópticos.

Soportes magneto-ópticos.

Gestión Automatizada de la Información. Soportes magnéticos.

- Formados por un substrato recubierto con partículas de material magnético.
- Partículas magnéticas:
 - Dos estados -> magnetizadas a la izquierda o a la derecha (0 ó 1).
 - Cada partícula es una unidad de información.
- Cabeza del dispositivo con dos posiciones:
 - Lectura: detecta la posición del campo magnético.
 - Escritura: Cambia la polaridad de las partículas.



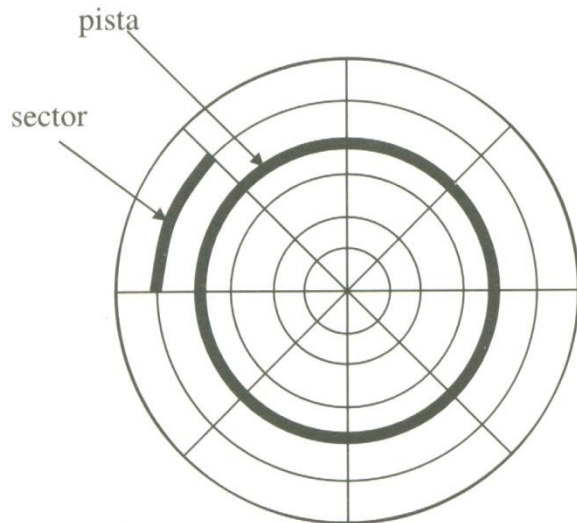
Es regrabable: permite modificar la información.

- Información afectada por : campos magnéticos, y temperaturas $> 40^{\circ}\text{C}$
- Ej: Cintas magnéticas, disquetes y discos duros.

Gestión Automatizada de la Información. Soportes magnéticos.

- Discos magnéticos.

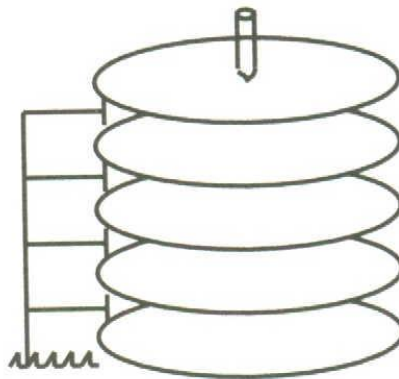
- Se graban en pistas concéntricas.
- Formateados en sectores:



- **Pista exterior:** directorio con el contenido del disco: FAT (File Allocation Table) o TOC (Table of contents)
- **Ficheros:** almacenados en sectores consecutivos.

Gestión Automatizada de la Información. Soportes magnéticos.

- Discos magnéticos.
- Varios discos apilados: aumenta la capacidad del soporte.
- Cada disco tiene su cabeza lectora.





Gestión Automatizada de la Información. Soportes ópticos.

- Características:
 - No permiten modificar información-> una sola grabación.
 - No son afectados por campos magnéticos.
 - Alta resistencia a la humedad y la temperatura.
 - Las cabezas lectoras no lo deterioran.
 - Grabadores caros, lectores baratos -> en la actualidad los dos son asequibles.
- CD - ROM: (Read Only Memory) - Disco compacto -
Memoria de solo lectura.
 - Velocidad de acceso inicial 150 Kb/seg. Actualidad: 8x,16x, 56x.

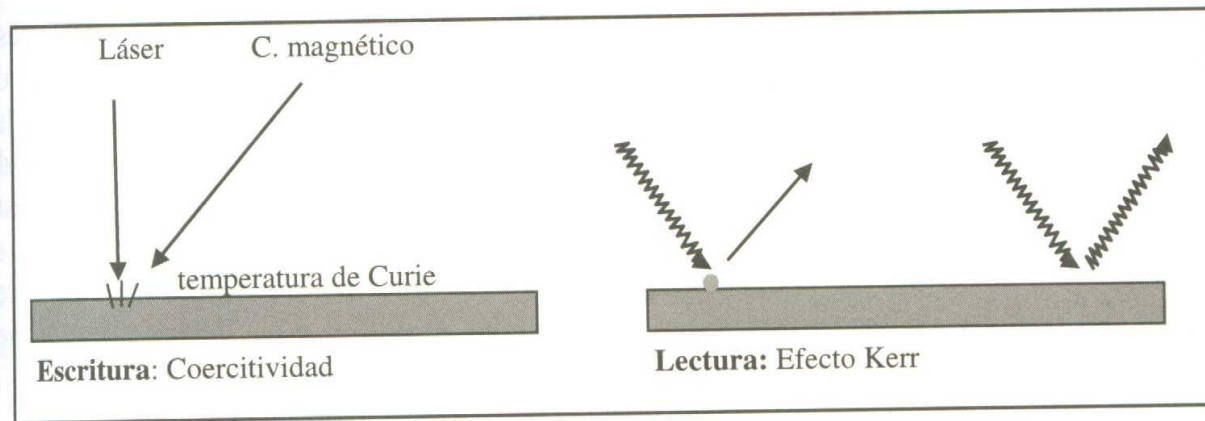
Gestión Automatizada de la Información. Soportes ópticos.

- DVD (Digital Versatil Disk):
 - Velocidad de transferencia 9 **veces mayor** que el CD-ROM.
 - Capacidad de almacenamiento entre 4.7 y 8 GigaBytes, dependiendo de:
 - El tamaño.
 - El formato de grabación.
 - El número de caras útiles.

Soportes magneto-ópticos.

- Escritura: Un rayo láser calienta un punto y un campo magnético graba la información.
 - En frío la información es inalterable.
 - **Volviéndolo a calentar se puede modificar.**
- Lectura: un haz de luz polarizada cambia su polarización con el campo magnético.
- CD- ROM regrabable.

Regrabable



4. Tiempos de acceso.

- Acceso a un periférico \approx 1000 veces mayor que el acceso a memoria principal.
- Aplicaciones con uso intensivo de datos (Bases de Datos, Sistemas de Recuperación de información): **tiempo** del proceso **condicionado** por las operaciones de **entrada / salida** a los periféricos.
- **Más importante el acceso a datos que las operaciones sobre ellos.**



Parte II

Índice.

1. Definición y conceptos.
 - 1.1. Conceptos previos: registros
 - 1.2. Ficheros..
2. Operaciones sobre ficheros.
3. Organización y procesamiento de ficheros.
 - 3.1. Secuencial.
 - 3.2. Secuencial ordenada.
 - 3.3. Secuencial ordenada encadenada.
 - 3.4. Directa.
 - 3.5. Aleatoria o relativa.
 - 3.6. Secuencial indexada. Tipos de índices: comprimido, disperso, multinivel, en árbol balanceado.
 - 3.7. En agrupamiento o cluster.

Conceptos de generales. Ficheros

- **Sistemas de gestión de ficheros:** programas que permiten diseñar ficheros con determinadas estructuras y recuperar y actualizar eficazmente esas estructuras.
- **Base de datos:** Colección de ficheros a los que se puede acceder por un conjunto de programas y que contienen todos ellos datos relacionados.
- **Fichero:**
 - Colección de registros relacionados:
 - Formados por campos (datos individuales, elemento básico tiene un tipo de dato y longitud)
 - Identificación de un registro: mediante un campo **clave primaria**.
 - Si se necesita más de una clave: clave primaria y clave secundaria.
 - Las claves permitirán localizar los registros en el fichero.

Conceptos de generales. Ficheros

Ejemplo de campos de un registro.

Campo1 NOMBRE y APELLIDOS	tipo caracter
Campo2 NÚMERO DE VUELO	tipo caracter
Campo3 FECHA DE VUELO	tipo caracter
Campo4 NÚMERO DE ASIENTO	tipo número
Campo5 FUMADOR	tipo lógico
Campo6 CIUDAD ORIGEN	tipo caracter
Campo7 CIUDAD DESTINO	tipo caracter
Campo8 PRECIO	tipo real

Ej. de clave: nombre y apellidos del pasajero.

Operaciones sobre ficheros.

- Crearlo: definir sus campos
- Abrirlo para realizar operaciones:
 - Añadir registros = altas .
 - Modificar campos de un registro.
 - Suprimir registros = bajas = borrar.
 - Consultar información de un registro.
 - Para estas operaciones necesitamos poder leer y escribir en el fichero:
 - Leer: Transferencia de información desde el archivo a la memoria principal.
 - Escribir: Transferencia de información desde la memoria principal al archivo
- Cerrarlo.

Conceptos Generales. Ficheros

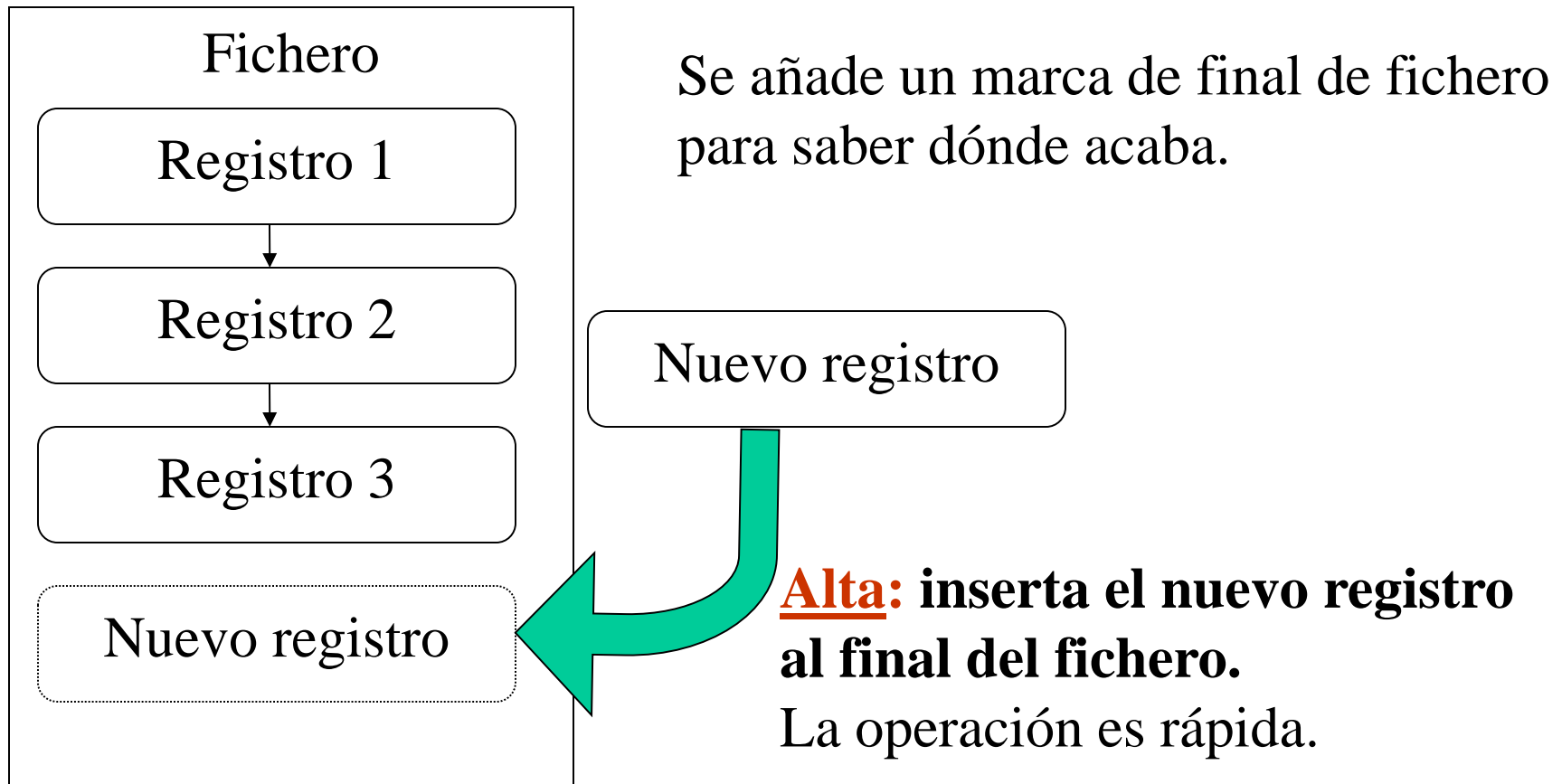
- **Organización de ficheros** → forma de estructurar y almacenar los datos de un fichero en un dispositivo de almacenamiento
- **Objetivo:** maximizar la eficiencia de las operaciones más frecuentes sobre el fichero.

Organización y procesamiento de ficheros.

- 3.1. Secuencial.
- 3.2. Secuencial ordenada.
- 3.3. Secuencial ordenada encadenada.
- 3.4. Directa.
- 3.5. Aleatoria o relativa.
- 3.6. Secuencial indexada. Tipos de índices: comprimido, disperso, multinivel, en árbol balanceado.
- 3.7. En agrupamiento o cluster.

Organización secuencial.

- Los registros se colocan uno a continuación del anterior según se van introduciendo.



Organización secuencial.

- **Consultar:**
 - Leer los anteriores hasta encontrar el deseado.
 - Necesario leerlos todos para garantizar que uno no está.
 - Operación lenta.
 - Consulta de TODOS los datos: rápida, se leen secuencialmente todos.

- Pasos para la consulta:

Abrir fichero.

Mientras queden registros:

Leer siguiente registro.

Si registro = registro_buscado

Mostrar “Datos del registro buscado = ...”

Cerrar Fichero

Acabar

FinDeSi

Fin de Mientras

Mostrar “No existe el registro buscado”

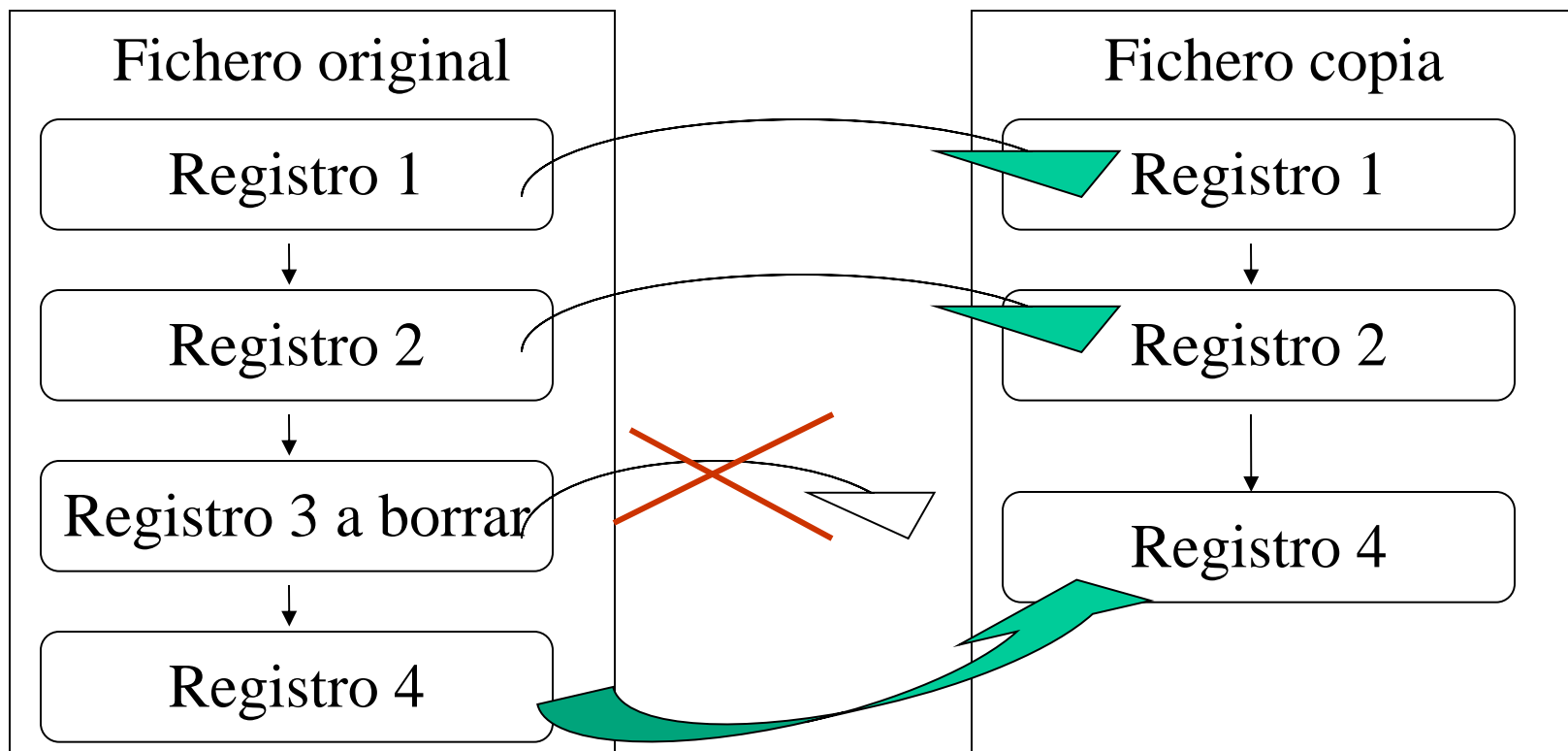
Cerrar Fichero

Acabar

Organización secuencial.

- **Eliminar o borrar un registro.**

- Se copian todos los registros hasta encontrar el deseado que no se copia.
- Al acabar se elimina el fichero original y se renombra el copia.



Organización secuencial.

- **Modificar un registro.**
 - Se copian todos los registros hasta encontrar el deseado que se copia modificado.
 - Si los registros son de longitud fija y el dispositivo de acceso directo no es necesario copiar:
 - Localizar el registro.
 - Modificarlo.
 - Reescribirlo en su misma posición.

Organización secuencial.

- Conclusión:
 - Consulta individual, baja(borrar) y modificación de registros individuales es muy costosa (=muchos accesos al soporte = mucho tiempo).
 - Altas rápidas.
 - Consume poco espacio de almacenamiento (sólo el necesario para los datos).
 - Consulta de todos los datos (global) es buena.

Organización secuencial ordenada.

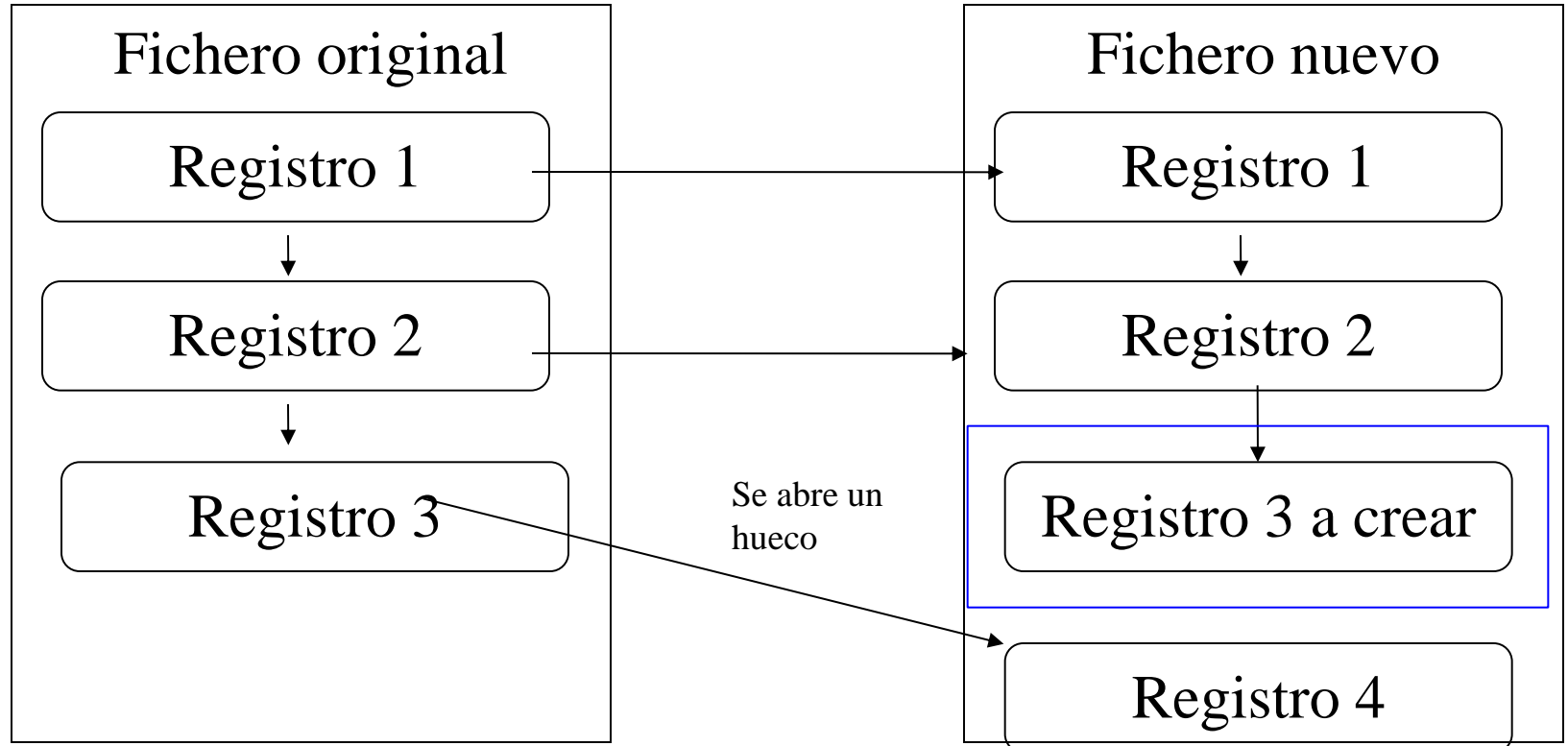
- Se mantienen los registros ordenados por el valor de un campo llamado **campo clave**.
- Consultas globales: aparecerán los datos ordenados.
- Fichero secuencial + ordenado + longitud del fichero fija + soporte de acceso directo → Mejora tiempo de respuesta en consultas individuales respecto a la secuencia.

Organización secuencial ordenada.

- **Búsqueda (consulta)** de un registro: algoritmo de búsqueda binaria
 - Se lee el registro de la mitad del fichero.
 - Se compara su clave con la del registro buscado:
 - Si coincide con la búsqueda: hemos encontrado el registro.
 - Si la búsqueda es mayor → continua la búsqueda en la segunda mitad del fichero.
 - Si la búsqueda es menor → continua la búsqueda en la primera mitad del fichero.
 - Se repite el proceso con la mitad restante hasta que quede 1 registro.

Organización secuencial ordenada.

- **Alta de un registro.**
 - Abrir un hueco para escribir el registro



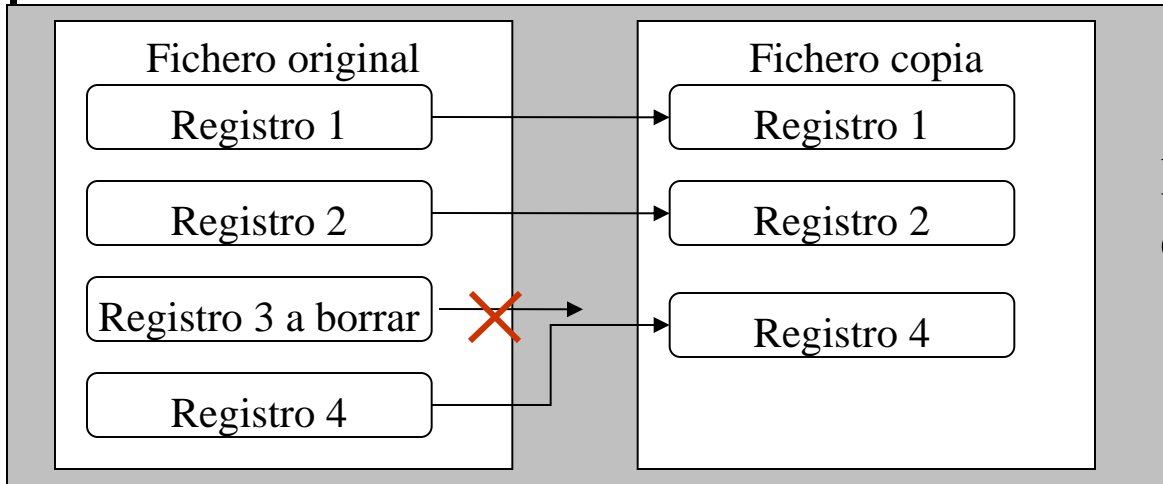
Organización secuencial ordenada.

- **Modificar** un registro, si se modifica el campo NO clave:
 - Búsqueda del registro.
 - Modificación en memoria.
 - Reescritura en el soporte.
 - Es un proceso eficiente si los registros son de longitud fija.

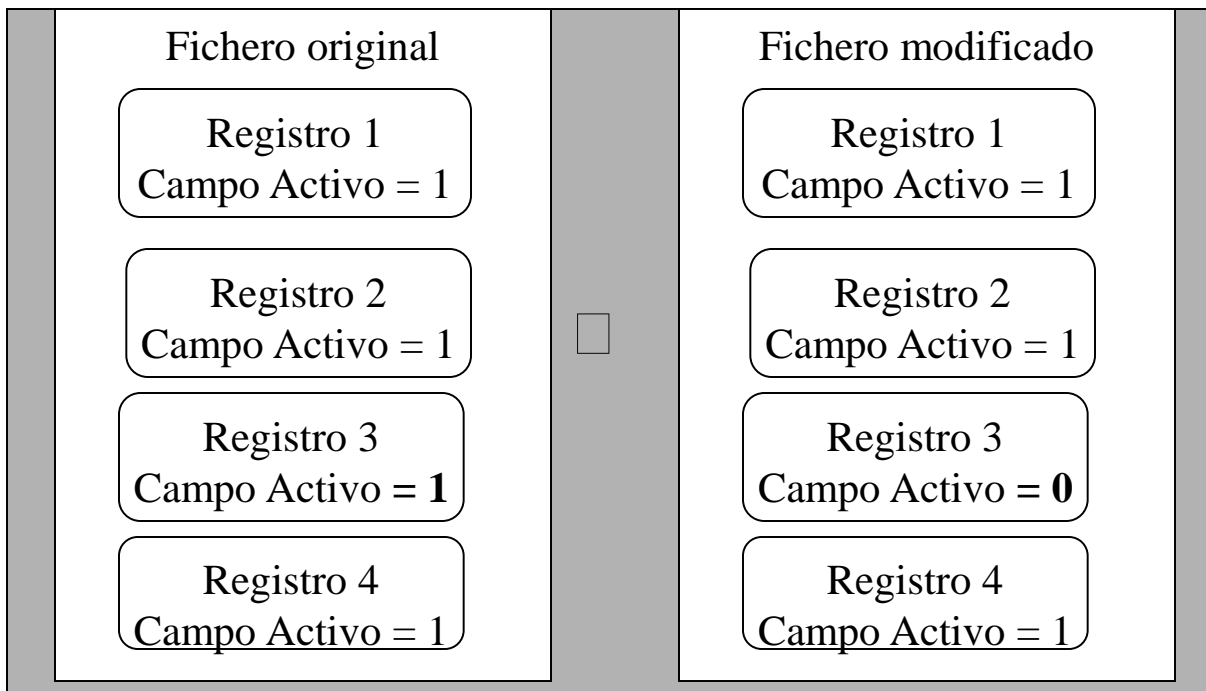
Y se modifica el campo clave??

- **Borrar** un registro, dos posibilidades:
 - Eliminarlo:
 - Buscarlo.
 - Reescribir los siguientes adelantándolos una posición.
 - Borrado ‘lógico’: se incluye un campo en los registros que indique si el registro está activo o no.
 - Buscarlo.
 - Modificar el campo de registro activo.
 - Reescritura en el soporte.

Organización secuencial ordenada.



Borrado físico de un registro eliminándolo.



Borrado lógico de un registro.

Organización secuencial ordenada.

- Problemas del borrado lógico:
 - Implica un pequeño consumo adicional de memoria (para el campo que indica si está borrado).
 - Los registros borrados siguen ocupando espacio: esto hace menos eficiente los procesos de búsqueda o modificación.
- Ventaja: es más rápido porque no hay que reescribir los campos a partir del borrado.

Organización secuencial ordenada.

- El mantenimiento de los ficheros es costoso ya que es necesario tenerlos ordenados por el campo clave.
 - Si se inserta (dar de alta) un nuevo registro hay que buscar la posición en la que va y abrir un hueco.
 - Si se modifica **el campo clave** de un registro hay que cambiarlo de posición.
- Para reducir este coste:
 - Posibilidad de procesarlos periódicamente todos juntos-> proceso por **lotes** o procesamiento ‘batch’:

Se aplazan las operaciones de mantenimiento y en una única pasada se actualizan todas las incidencias desde la última actualización.

- Posibilidad de dejar huecos en la creación inicial del fichero para las altas futuras.

Organización secuencial ordenada encadenada.

- Incluye un campo adicional en cada registro: puntero al siguiente registro.
- Dos áreas de almacenamiento: **área primera y área para posteriores altas.**
- Mejora el proceso de alta respecto a la secuencial ordenada.
- **Alta:**
 - Escritura en el primer registro libre del área adicional.
 - Modificación del puntero.
- **Consultas globales:** más complicado, porque hay que cambiar la posición de la cabeza lectora para seguir la
- Inconvenientes: degenera si hay un gran volumen de altas y de bajas.



Organización directa.

- Fichero: conjunto de ‘cajas’ numeradas en cada una de las cuales cabe un registro de tamaño fijo.
- **Alta:**
 - Se indica la posición exacta en que se almacena el registro.
 - La cabeza lectora del disco se coloca en esa posición y lo graba, independientemente de que otras posiciones estén llenas o vacías.
- **Consulta:**
 - Dado un número de registro se accede directamente a él.
 - Consulta global: leerlos todos en secuencia.
- **Baja:** dejar el registro vacío.
- Necesidad de tener un dispositivo de acceso directo.
- Problema: recordar la posición física de cada registro-> dar un significado a esa posición (ej: nº de expediente de un alumno).

Organización aleatoria o relativa.

- A veces no es posible emplear el valor de un campo como posición del registro en el fichero.
 - Ej: si el DNI fuera el campo clave se desaprovecharía mucho espacio de memoria.
- Se establece una **función de acceso** que distribuye los registros en las posiciones según el valor de un campo determinado → **clave**.
- Área primaria y área de desbordamiento para las colisiones
- **Ejemplo función de acceso**

$\text{funcion_acceso} = \text{resto}(\text{N_expediente} / 111) + 1$

Organización aleatoria o relativa.

Area
Primaria

Area
Desborda
miento

Ficheros.

N Registro	N Expedient	Campo1	Campo N
1				
....				
104	769			
105				
106				
...				
111	...			
112	325			
113				

Organización aleatoria o relativa.

- Proporciona un procesamiento eficiente para altas, bajas, consultas y modificaciones de registros de los que conocemos el valor clave.
- Para consultas individuales de registros de los que no conocemos otro valor distinto del de la clave--> requiere muchos accesos a disco.
- Consume más espacios en área primaria como en la zona de desbordamiento, pero puede compensar si este desperdicio no es excesivo

Ejercicio: Realiza una tabla comparativa

	Secuencia	Secuencia Ordenada	Sec. Ord. encadenada	Directa
Altas				
Bajas				
Modificación				
Cons. Individ				
Cons. Globales				
Espac. Adicional				
Ventajas				
Desventajas				
Ejemplon Optimo de uso				

Organización secuencial indexada.

- Registros colocados en orden de entrada en el fichero de datos.
- Creación de uno o más ficheros adicionales: **directorío o tabla de índices**.
 - Contiene dos campos: valor de la clave (campo indexado) y posición del registro en el fichero.
 - Puede estar ordenado por la clave --> permite entonces **búsquedas binarias**.
- Inconveniente: necesidad de espacio adicional para las tablas de índices.
- Ventaja: fáciles de programar.
- Organización muy utilizada.

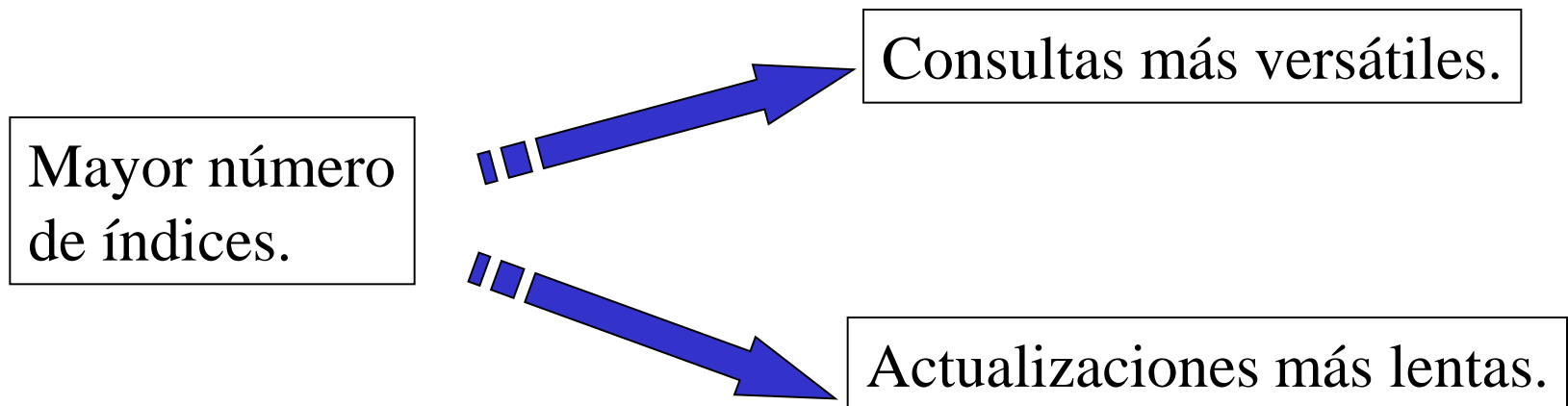


Organización secuencial indexada.

- **Alta:**
 - Escritura al final del fichero de datos.
 - Actualización de la tabla de índices.
- **Consulta:**
 - Búsqueda en el fichero de índices y acceso al fichero de datos.
 - Consultas individuales eficientes.
 - Consulta global siguiendo el orden de un campo indexado:
 - Seguir secuencialmente el índice y acceder directamente a cada uno de los registros.
 - Es más lento que en un fichero secuencial ordenado.
- **Modificación:**
 - Si no afecta a un campo indexado:
 - Consulta.
 - Modificación en memoria principal.
 - Reescritura en el soporte
 - Si afecta a un campo indexado → modificación también de la tabla de índices.

Organización secuencial indexada.

- **Baja:**
 - Búsqueda de un índice.
 - Eliminación en el fichero de datos.
 - Actualización de todas las tablas de índices.



Organización secuencial indexada.

- Posibilidad de distintas organizaciones en el fichero de índices.
- Posibilidad de incluir la tabla de índices en el mismo fichero que los datos.
- Otros tipos de índices.
 - Índice disperso.
 - Índice multinivel.
 - Índice en árbol balanceado.

Ejercicio: comparar el consumo de espacio y la eficiencia de las consultas individuales y globales en orden, en una organización secuencial ordenada y en una secuencial indexada por una clave cuyo tamaño sea prácticamente igual al tamaño del registro.

– Ej: Índice = Apellido, Nombre

Registro = Apellido, Nombre, año-nacimiento, sueldo.

Índice disperso.

- Se mantiene el fichero de datos ordenado (como en la organización secuencial ordenada).
- Cada elemento del índice apunta a un sector del fichero.
- El índice contiene: mayor valor de la clave en ese sector y puntero al sector.
- Búsqueda:
 - Lectura secuencial del índice hasta encontrar una clave mayor que la buscada.
 - Lectura en el sector de datos al que apunta el índice.

Índice disperso.

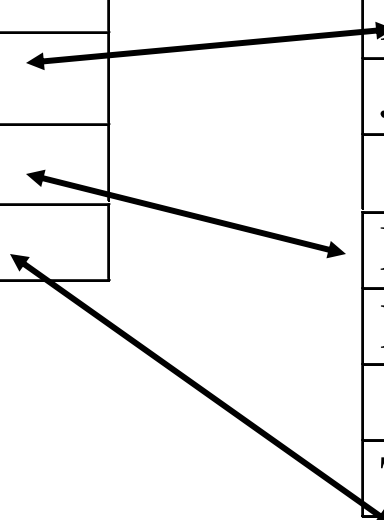
- Ejemplo.

Índice disperso

Nombre	p.reg
Dolores	
María	
Teresa	

Fichero de datos.

Nombre	DNI	Edad
Andrés	66666666	27
....
Dolores	33333333	80
Juan	11111111	13
...
María	22222222	49
Pedro	44444444	36
...
Teresa	55555555	56



Índice multinivel.

- Índice disperso: registros ordenados por el valor del campo de indexación.
- Índice multinivel: índice del índice.
- Se pueden implementar varios niveles de índices.
- Ventaja: Se puede mantener el índice más pequeño en memoria principal.



Organización en agrupamiento o cluster.

- Objetivo de las organizaciones de ficheros: minimizar los tiempos de acceso a los periféricos.
- Organización en cluster: ubicar físicamente juntos los registros con mayor probabilidad de ser recuperados.
- Elección de algún criterio de semejanza → dependerá del tipo de consultas que se realicen.

Índice en árbol balanceado.

- Índices propuestos: se desorganizan al introducir nuevos registros. Necesidad de reorganizaciones periódicas (para el servicio).
- Índice en árbol balanceado:
 - Reorganización dinámica de los índices.
 - Tiempo de acceso equivalente a todos los registros (balanceado).
 - Sectores dedicados a índices (lo llamaremos página de índices) y sectores dedicados a datos (página de datos).

Índice en árbol balanceado.

- Página de datos:
 - Registros ordenados por el valor de la clave.
 - Las páginas de datos están encadenadas: no son consecutivas físicamente dentro del fichero.
 - Contiene un conjunto de registros y un puntero a la siguiente página.

Página de datos 1

Registro_datos1	Registro_datos...	Registro_datosm	Puntero a sgte. pag.
-----------------	-------------------	-----------------	----------------------

Página de datos 2

Registro_datos1	Registro_datos...	Registro_datosn	Puntero a sgte. pag.
-----------------	-------------------	-----------------	----------------------

Estructura de una página de datos.

Índice en árbol balanceado.

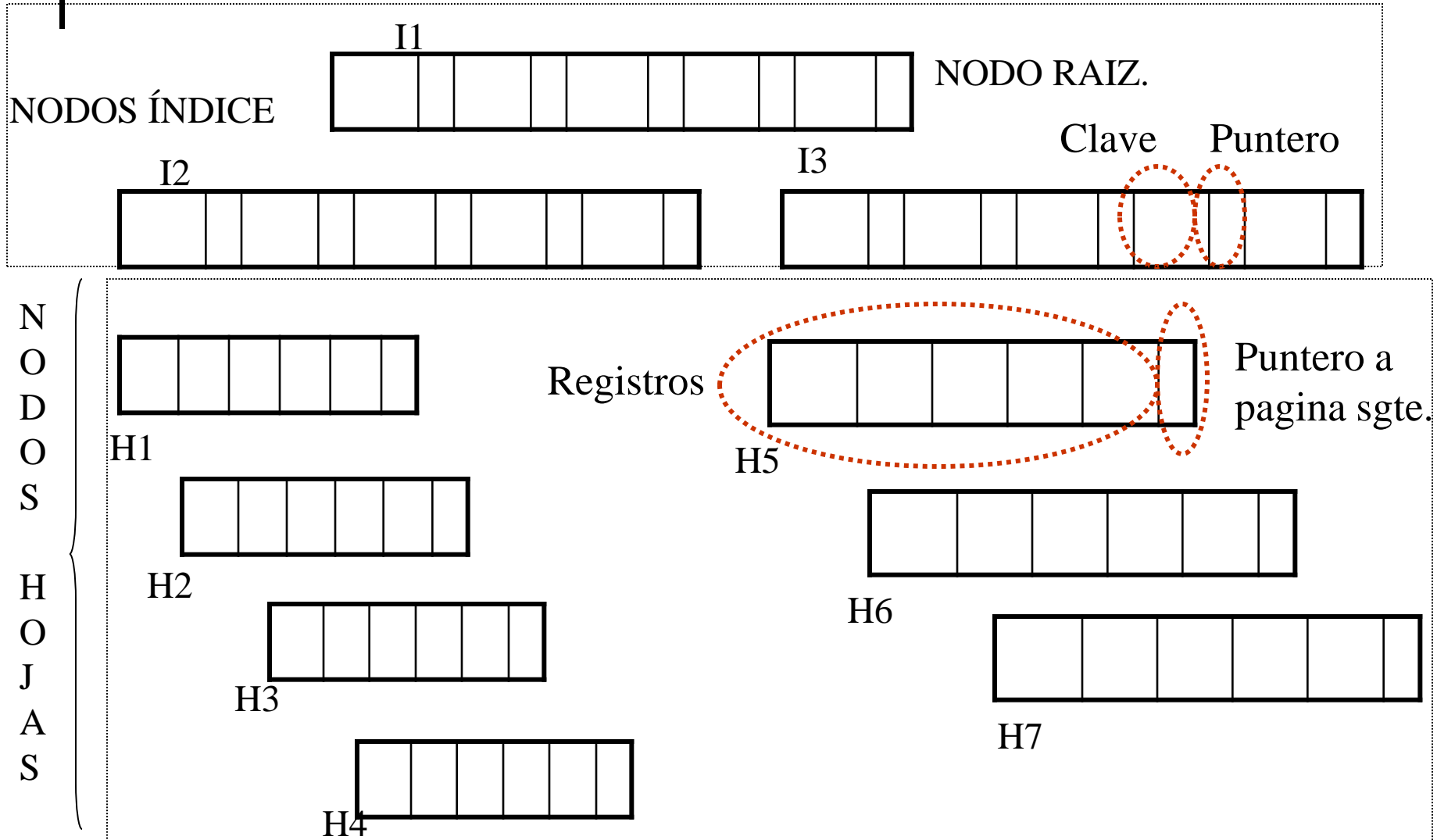
- Página de índices:
 - Índices dispersos.
 - Contienen un apunte por cada página de datos:
 - Mínima clave de una página de datos.
 - Puntero.

Clave 0	Puntero 0
Clave 1	Puntero 1
Clave 2	Puntero 2
....
Clave n	Puntero n

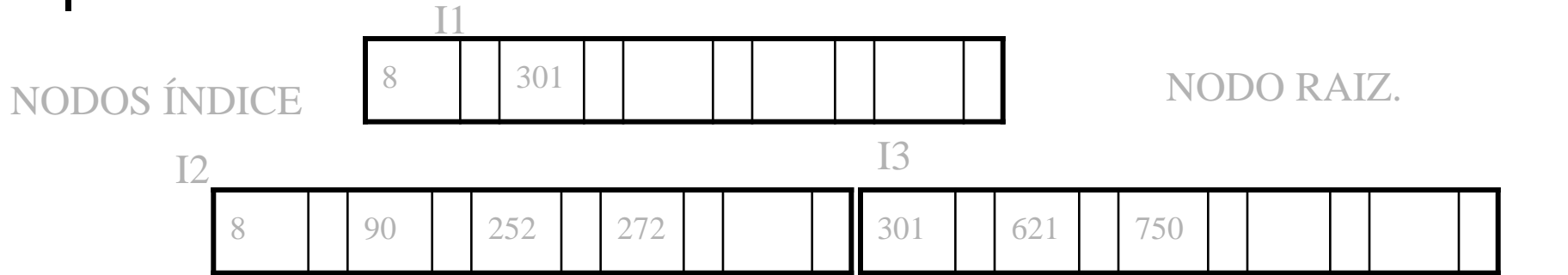
Índice en árbol balanceado.

- Ejemplo:
 - Página de datos con 5 registros.
 - Página de índices con 5 claves.
 - Índice balanceado con dos niveles de índices.
 - I1 → página raíz de índices.
 - I2, I3 → páginas de índice de segundo nivel.
 - Páginas de datos: H1... H7.

Índice en árbol balanceado.



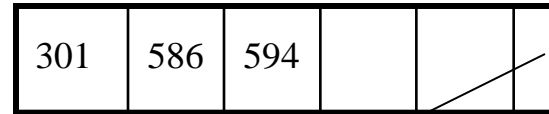
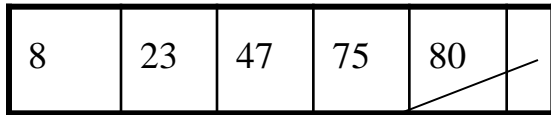
Índice en árbol balanceado.



Conexiones entre las páginas de datos.

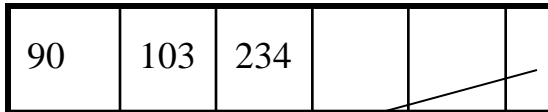
N
O
D
O
S

H
O
J
A
S



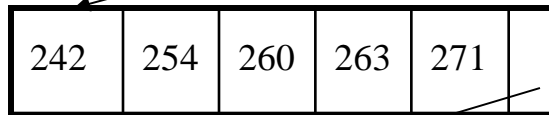
H1

H5



H2

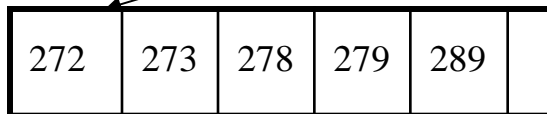
H6



H3



H7



H4

Ficheros.

Informática II.



Índice en árbol balanceado.

- Consulta global:
 - Recorrer las páginas de datos siguiendo los punteros.
- Localizar el registro 234:
 - Ir al raíz: como $234 < 301 \rightarrow$ examinar sector I2.
 - En sector I2: $234 > 90$ y $234 < 252 \rightarrow$ ir a página de datos H2.
 - Coste de la consulta: una lectura de cada nivel del índice y una lectura en el sector de datos.
- Inserción:
 - Insertar registros con valores de la clave: 235, 200, 144.







Índice en árbol balanceado.

- Ejercicio: Crear el árbol balanceado resultante después de cada una de las siguientes inserciones:
 - Insertar 150.
 - Insertar 78.
 - Insertar 274

Bibliografía.

- Gestión Digital de la Información. De bits a bibliotecas digitales. (2002)

Rosalía Peña, Ricardo Baeza-Yates, José. V. Rodríguez.

Ed. Ra-Ma.

Bibliografía.

- Gestión Digital de la Información. De bits a bibliotecas digitales. (2002)

Rosalía Peña, Ricardo Baeza-Yates, José. V. Rodríguez.

Ed. Ra-Ma.