

Tema 1: Índice

1. Introducción
2. Definición de fichero
3. Diseño lógico y físico de ficheros
4. Gestión de ficheros
5. Distintas formas de interactuar con ficheros
6. Tipos de ficheros



@LABDA. Univ. Carlos III

1.1. Introducción

Información: ...|| 7. *Comunic.* Comunicación o adquisición de **conocimientos** que permiten ampliar o precisar los que se poseen sobre una materia determinada.|| 8. *Comunic.* Conocimientos así comunicados o adquiridos. || ...

(Real Academia Española)

Dato: **Representación** de una información de manera adecuada para su tratamiento (informatizado).

1 de Octubre de 2007

Empiezan las clases en UC3M



@LABDA. Univ. Carlos III

1.1. Introducción

Dato: Representación de una información de manera adecuada para su **tratamiento** (informatizado).

Proceso de Datos: *obtención de beneficio a partir de ellos*

almacenamiento necesario: de inmediata disponibilidad

principal

Almacenamiento de Datos: *perdurabilidad y alcance de los datos*

almacenamiento necesario:

- permanente
- masivo
 - coste
 - espacio físico
- portable o accesible

secundario



@LABDA, Univ. Carlos III

1.1. Introducción

Comparativa de requisitos para el almacenamiento principal y secundario

	Principal	Secundario
durabilidad	volátil	permanente
coste	caro	barato
capacidad/espacio	reducida	extensa
velocidad	rápida	lenta
acceso	privilegiado	externo



@LABDA, Univ. Carlos III

1.1. Introducción

Organización de los datos:

- para ser procesados, los datos tienen que **estar en memoria principal**
- por razones de durabilidad, coste, espacio, y alcance, es necesario **contar con almacenamiento auxiliar** en soporte secundario

Distribución en Memoria Principal:

- **el mapa de memoria**

Distribución en Memoria Secundaria:

- **ficheros**



@LABDA. Univ. Carlos III

1.1. Introducción

Distribución en Memoria Principal

Tipos de Organización en Memoria Principal:

- Pilas: la pila de diskettes
- Colas: los apuntes
- Listas: la lista de tareas (direccionada)
- Bolsas y conjuntos: la caja de clips

Determinado por el **tipo de proceso** que se requiere
y por la **estructura física** de la memoria principal



@LABDA. Univ. Carlos III

1.1. Introducción

Distribución en Memoria Secundaria

Tipos de Organización en Memoria Secundaria:

- Serial
- Consecutivo
- Direccionado
- ...

Determinado por el **tipo de soporte** de almacenamiento físico y por el **tipo de proceso**: selectivo, ordenado, multiclave.



@LABDA. Univ. Carlos III

1.2. Ficheros

¿Qué es un fichero?

m. Caja o mueble con cajonería donde se pueden **guardar ordenadamente** las *fichas* ||

2. Infor. Conjunto **organizado** de *información almacenada* en un **soporte** común.

(Real Académica Española)

¿Por qué? Necesidad de **agrupar y organizar** *datos almacenados en soporte secundario*

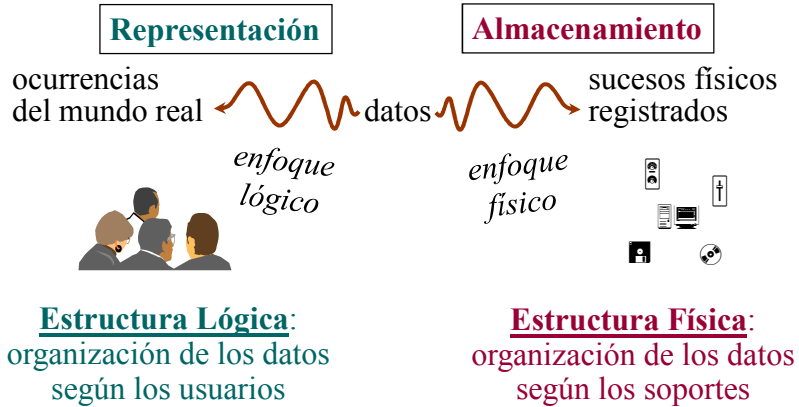
FICHERO: “ *Conjunto organizado y nominado de informaciones estructuradas almacenadas en un soporte no volátil* ”



@LABDA. Univ. Carlos III

1.2. Ficheros

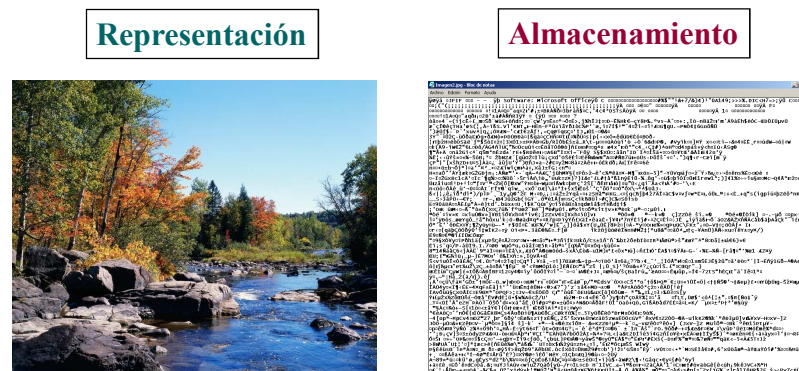
Estructuras de Fichero (Dos visiones



@LABDA, Univ. Carlos III

1.2. Ficheros

Estructuras de Fichero (Dos visiones



@LABDA, Univ. Carlos III

1.2. Ficheros

Contenido de un fichero:

Colección de informaciones (registros)

Registro Lógico: datos referidos a un elemento del fichero.

Es la unidad mínima de proceso de una aplicación.

Tamaño de un registro: caracteres que ocupa.

Registro Físico: caracteres que almacenan y acceden juntos.

Es la unidad mínima de acceso del soporte.

Tamaño de un registro: caracteres que caben. Ese espacio sobre el soporte se denomina **bloque**.



@LABDA. Univ. Carlos III

1.2. Ficheros

Unidades mínimas de almacenamiento (sub-atómicas):

Información mínima interpretable

= elemento de datos o campo

Tipo de un campo es una abstracción que describe sus características (tipo carácter, numérico, imagen...)

Ejemplos: caracteres $C(n)$, número entero $E(n)$, booleano B , ...

Información mínima registrable o accesible

= carácter (Memoria Intermedia)

= registro físico o bloque (soporte)



@LABDA. Univ. Carlos III

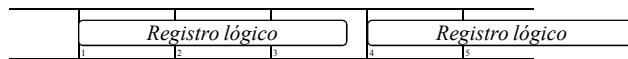
1.2. Ficheros

Correspondencia fisico-lógica a nivel de registro:

consideración de tamaño: R_l y R_f **no suelen coincidir en tamaño**

- Registro Expandido

Cuando el **registro lógico** es mayor que el **físico** (ocupa varios)



- Bloque

Cuando el **registro físico** es mayor que el **lógico** (cabén varios)



Factor de Bloqueo: número de registros lógicos que caben en uno físico



@LABDA, Univ. Carlos III

1.2. Ficheros

Correspondencia fisico-lógica a nivel de operación:

Debe existir una correspondencia (**mapping**) a nivel práctico

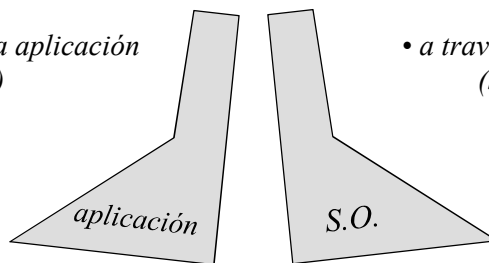
¿cómo se conectan las estructuras **lógica** y **física**?

- mediante la aplicación
(alto nivel)

- a través del S.O.
(bajo nivel)



usuarios



soportes



@LABDA, Univ. Carlos III

1.3: Diseño lógico y físico

Diseño Lógico: “Descripción de la estructura lógica de los registros de un archivo”

Estructura lógica (de un registro): descripción y disposición de los elementos de un registro, que en conjunto definen un individuo.

Diseño Físico: “Determinación de la organización física de un fichero”.

Organización física: disposición de los registros en el soporte en relación a su implementación, orden, direccionamiento, etc.



@LABDA. Univ. Carlos III

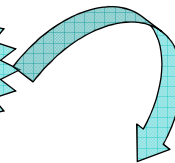
1.3: Diseño lógico y físico

Diseño Físico del registro lógico: “Descripción de las secuencias de bytes que implementan un registro lógico”.

Estructura física del registro: longitud fija, variable, etc.

LÓGICO

Nombre de usuario	C(50),
Contraseña	C(12),
Dir. Email	C(30),
Nombre	C(20),
Apellido	C(15),
Ciudad	C(25),
País	C(20)



FÍSICO-LÓGICO

Crear un nuevo usuario y contraseña para acceder al sistema:
Número de usuario:
Contraseña:

Por favor, escriba algunos datos sobre usted:
(IMPORTANTE: Para finalizar el proceso debe escribir una dirección de correo verdadera)

Dirección de correo:
Correo (de nuevo):
Nombre:
Apellido:
Ciudad:
País:



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Objetivos del Diseño Lógico

- Fiel representación de la realidad
- Organización física transparente al usuario
- Facilitar a los usuarios la interacción con los datos
- Evitar redundancias lógicas:

“representar dos veces la misma información”

E F I C A C I A



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

- **Elemento de datos** (*campo*): unidad mínima e indivisible de datos que interviene en un proceso

Notación: campo tipo(tamaño)
(si es opcional, se encerrará entre corchetes [...])

- **Agregado de datos:** colección de elementos (campos y/o agregados)

- **Vector:** agregado compuesto por un número fijo de elementos cuya interpretación es complementaria (entre todos los elementos definen un concepto).

Notación: (elemento₁; elemento₂; ...; elemento_n)

- **Grupo Repetitivo:** agregado compuesto por un número fijo o variable de elementos cuya interpretación es común (el mismo concepto para definir cada elemento)

Notación: (elemento)* para 0..N elementos
o bien (elemento)+ para 1..N elementos



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Elementos de un Registro. Ejemplos:

DNI	N(8)
Nombre	C(20)
Fecha	(Día N(2); Mes N(2); Año N(4))
Hijos	(DNI N(8); Nombre C(20))*
Coche (opt.)	[Matrícula N(4); Letra C(3)]

Tipo: abstracción de un campo o agregado que describe sus características

Ocurrencia: valor concreto que toma un tipo (referido a un individuo)

Ejemplo: 03555555
Juan Valdés
12-05-1960
(13333333, José Valdés); (13333334, María Valdés)



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Ejemplo: fichero que registre los datos de mis películas en DVD

- Necesitaré registrar su título (el más largo tiene 42 letras)
- El año de su realización y el nombre de la productora (20 letras)
- El género: drama, comedia, suspense, acción, histórica, c.ficción
- Quiero también registrar el nombre del director
- También de los actores y/o actrices si la película los tiene.
- Quiero saber quién ha hecho la banda sonora, y si la tengo necesitaré registrar en que formato : CD, CAS, LP

* el nombre de una persona es una información compuesta de nombre y apellido, 20 car. cada uno



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Ejemplo: solución diseño lógico

Título	C(42)
Año	N(4)
Productora	C(20)
Género	C(9)
Director	(nombre C(20); apellido C(20))
Actores	(nombre C(20); apellido C(20))*
BS	[Autor (nombre C(20); apellido C(20)), Formato C(3)]



@LABDA. Univ. Carlos III

Ejercicio de diseño lógico

Ejemplo 2: fichero que contiene información de estudiantes de ITIG

- Nombre completo (20 nombre + 20 apellido)
- Teléfonos (+34 y número) si los tiene, como máximo 3
- Dirección (50 como máximo)
- Asignaturas aprobadas. Nombre de la asignatura (50 carac. como máximo), curso (1..3) y convocatoria (1..6). En total se sabe que un alumno como máximo tendrá 40 asignaturas aprobadas
- Si ha cursado las 40 asignaturas, también necesitamos saber si ha realizado el proyecto y si es así el título (70 carac.) y el año en que lo presentó.



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Objetivos del Diseño Físico

Se buscará gastar menos recursos para el mismo resultado

Recursos a optimizar:

- Espacio (disminuir la ocupación y el tamaño total)
- Tiempo de Respuesta (minimizar los accesos)
- Mantenimiento (coste de reorganización)
- Desarrollo (coste de desarrollo y modificación)

EFICIENCIA



@LABDA. Univ. Carlos III

1.3. Diseño lógico y físico

Diseño Físico: dependencia del soporte

- Tamaño de bloque
- Tiempo de Acceso (localización y acceso)
- Tasa de Transferencia (caracteres por unidad de tiempo)

Optimizaciones de bajo nivel:

- buscar soportes adecuados (tamaño bloque, tipo de acceso)
- minimizar el tiempo de localización (distribución adecuada)
- introducir redundancia física
 “grabar lo mismo varias veces”
- uso de memorias intermedias

Notación: **O_n**: organización de **qué tipo** y con **qué parámetros**



@LABDA. Univ. Carlos III

1.3: Diseño lógico y físico

Diseño Físico del registro lógico:

Representación eficiente de la información (en cadenas de bytes)

Notación:

Cadenas de bytes de cierta longitud y ocurrencia (fija o variable).

[prefijo]	“+34”
nombre_campo	b(2),
numero	b(1),
otro_campo	b(numero),
y_otro	b(6) ¹⁵
otro_mas	b(9) ^{numero}
el_ultimo_campo	b(65) ^{expresión}

expresión: puede ser condicional (de resultado '0' (falso) ó '1') o exp. aritmética.



@LABDA, Univ. Carlos III

1.3. Diseño lógico y físico

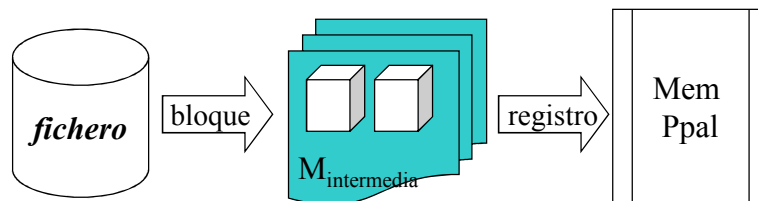
Memorias intermedias:

lectura: anticipar próximas lecturas (para que estén en Mppal)

- recogiendo todo el bloque leído
- realizando más lecturas en el tiempo de proceso

escritura: demorar escrituras (proceso perezoso)

- varios registros escritos en el mismo bloque (ins/mod)
- evitando operaciones que se corrigen o deshacen



@LABDA, Univ. Carlos III


1.3. Diseño lógico y físico

Memorias intermedias paginadas:

Páginas: agrupación de información en memoria intermedia.

Son como *ventanas* sobre el fichero, gestionadas por el S.O.

- **Tamaño de página:** depende de soporte y S.O. (suele coincidir con bloque)
- **Mecanismo de acceso paginado:**
 - al acceder, busca la página y si no está la trae
 - si no hay sitio libre, se hace sitio según su política
 - cuando está ocioso, trae y libera bloques (según el S.O.)
- **Políticas de liberación de páginas:**
 - LRU(Least-Recently Used)/MRU(Most-Recently Used): la que lleva más/menos tiempo sin usarse
 - la de menor tiempo de localización (más cercana)
- **Proximidad de registros:**

 registros más probables en siguiente acceso deben estar en la misma página

@LABDA, Univ. Carlos III

1.4. Gestión de ficheros

¿Quién gestiona los ficheros sobre el soporte?

 **El Sistema Gestor de Ficheros (SGF) del Sistema Operativo**

Sus funciones dependen del tipo de soporte y del propio Sistema Operativo.

Las primitivas que proporciona de modo más generalizado se agrupan como sigue:

- **Iniciales (de todo el fichero)** {
 - CREAR
 - ASIGNAR
 - ABRIR / CERRAR...
- **Acceso (de un elemento)** {
 - LEER
 - ESCRIBIR

- Manejo de memorias intermedias
- Protección
- Mantenimiento
- (Propias del soporte)



@LABDA, Univ. Carlos III

1.5. Interacción con ficheros

Interacción con ficheros:

Operaciones sobre la totalidad:

- Creación / Destrucción
- Apertura / Cierre
- Consulta de todo el fichero

Operaciones sobre algunos registros

- *consulta selectiva*
- actualización: inserción / *modificación* / *borrado*

Algunas operaciones (*selectivas*) requieren localización



@LABDA. Univ. Carlos III

1.5. Interacción de ficheros

Clave: campo (o conjunto de campos) que facilitan la interacción de los usuarios con el fichero.
(posibilitan la identificación y/o localización de registros)

Tipos de clave:

- Clave de **Identificación**:
identifica unívocamente un registro en el fichero
(toma valores diferentes para cada registro)
- Clave de **Direccionamiento**:
localiza la posición del registro (puede ser no unívoca)
- Clave de **Ordenación**:
criterio de ordenación; puede ser físico (fichero) o lógico (proceso)
- Clave de **Búsqueda**:
información utilizada por los usuarios para localizar registros



@LABDA. Univ. Carlos III

1.5. Interacción de ficheros

Ejemplo: claves de un fichero

- Clave de identificación: **Título**
(el título es único, me sirve para identificar un DVD)
- Clave de direccionamiento: **Productora**
(tengo una estantería por productora, y allí se localizan esos DVD)
- Clave de Ordenación: **Año, Título**
(los tengo ordenados por año, y luego alfabéticamente según el título)
- Clave de búsqueda: **Año, Director**



@LABDA. Univ. Carlos III

1.5. Interacción de ficheros

Cualquier clave debe ser mínima:

“No existe ningún subconjunto de la clave que cumpla las mismas funciones que la totalidad (de dicha clave)”

Ejemplos: claves incorrectas

- Clave de identificación: **Año, Productora**
(probablemente tenga dos películas que coincidan en productora y año)
- Clave de direccionamiento: **Formato**
(pocos valores posibles, poca capacidad de direccionamiento)
- Clave de Ordenación: **Título, Año**
(si está ordenado por título, que es único, no tiene sentido el año)
- Clave de Búsqueda: **Título, Director**
(no es mínima, porque el título por sí solo identifica un solo registro)



@LABDA. Univ. Carlos III

1.6. Tipos de ficheros

Hay varias taxonomías, atendiendo a distintos criterios:

unidad básica de información:

binarios: unidad básica, el bit

textuales: unidad básica, el carácter

tipados: unidad básica, el registro

formato de los registros

homogéneos: todos los registros siguen el mismo patrón

heterogéneos: contienen varios tipos de registro

(el primer campo deberá incluir el tipo de registro)



@LABDA, Univ. Carlos III

1.6. Tipos de ficheros

Ejemplo: fichero heterogéneo, de películas (pel) y musicales (mus)

Tipo	b(1) (!!!! Cuidado esto no entra en el diseño lógico)
/* película */	
Título	C(42)
Año	C(4)
Productora	C(20)
Género	C(9)
Director	(nombre C(20), apellido C(20))
/* musical */	
Título	C(20)
Discográfica	C(20)
Género	C(9)
Autor	(nombre C(20), apellido C(20))



@LABDA, Univ. Carlos III

1.6. Tipos de ficheros

Atendiendo a la función del fichero

permanentes: orientados al almacenamiento

temporales: orientados al manejo o proceso

Por la función de su contenido:

maestros (de situación): información diversa (variada)

constantes (de referencia): pocos valores y poco volátiles

Por la oportunidad del contenido

borrador: aún no entrado en uso

vigente: en uso

histórico: de uso pasado



@LABDA, Univ. Carlos III

1.6. Tipos de ficheros

Atendiendo al tamaño de sus registros:

Longitud Fija: los registros fijos y de formato definido

Longitud Variable: registros variables y de formato definido

Longitud Indefinida: registros variables y sin formato definido

Tipos de Registro:

- Longitud fija: siempre se utiliza todo su tamaño
- Longitud variable: sus ocurrencias son de diversa longitud (su tamaño es el del mayor de sus ocurrencias)
- Longitud indefinida: se desconoce la longitud de sus ocurrencias



@LABDA, Univ. Carlos III