

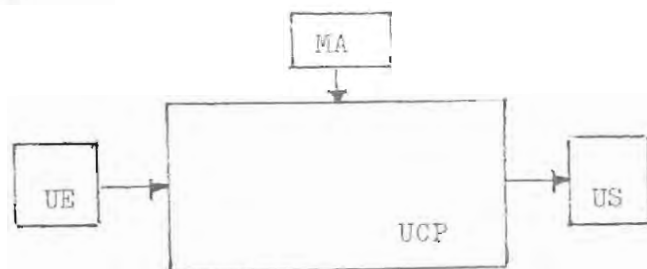
REGISTROS DE INFORMACION

- I. Introducción
- II. Medios y Soportes Físicos: Clasificación y Descripción
- III. Archivos: Organización y Búsqueda

por el Ing. José María P. Ferrara

I. Introducción

El tratamiento de la información por medios automáticos está basado en el uso del computador. Bajo este nombre se sintetiza un conjunto de unidades constituyentes de un equipo de computación, fundamentalmente tres: Unidades de entrada; Unidades de salida y Unidad central de proceso (U.E; U.S. y U.C.P. respectivamente).



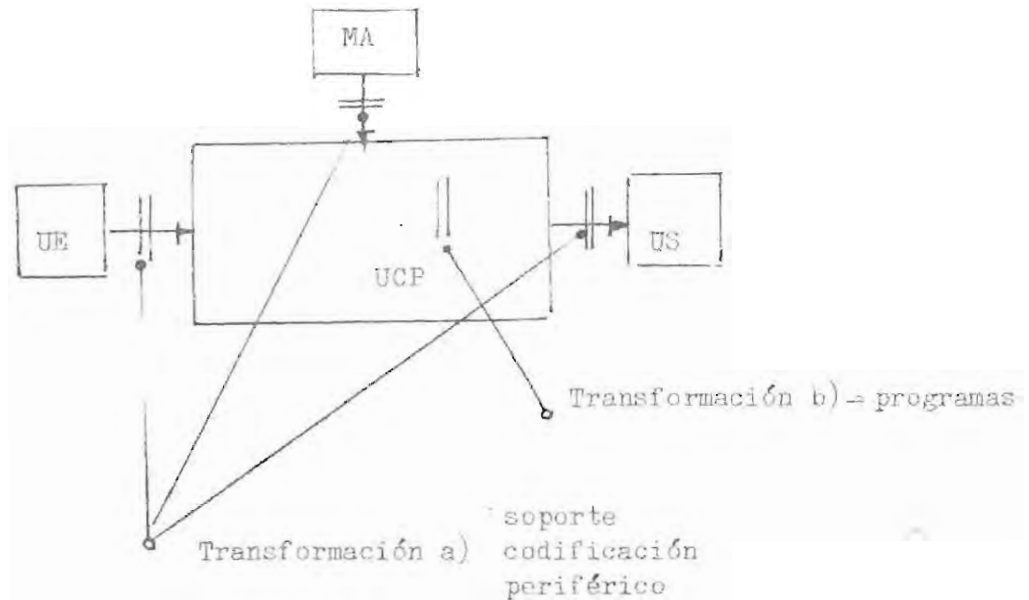
Se ha graficado una Memoria Auxiliar (M.A.) que constituye una extensión, de particulares características de la U.E.

En general, todo proceso comenzará en la U.E., fluirá por la U.C.P. sufriendo transformaciones en los datos y arrojará resultados elaborados a la salida es decir en U.S.

Las transformaciones serán de dos tipos:

- a) en el ingreso y egreso de la información;
- b) en el proceso de esa información entrante.

Ambas se realizarán en el lenguaje entendible por estas máquinas que finalmente es el llamado lenguaje de máquina o binario. Es decir:



La entrada/salida de información exigirá un soporte o medio donde volcar o registrar la información; una codificación conveniente que será función del medio y un elemento físico para leer el soporte llamado periférico.

La ejecución de una tarea con un programa (sucesión lógica de instrucciones) normalmente existente en Memoria Central y cuya lectura de datos se adaptará al soporte/s elegido/s.


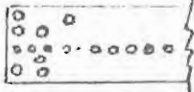



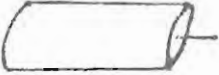
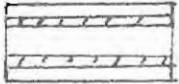
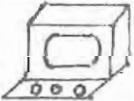
II. Soportes de la información

Se pueden clasificar desde cuatro puntos de vista:

- a) Función:
 - Entrada
 - Salida
 - Entrada/Salida
- b) Reutilización:
 - Reutilizables
 - No Reutilizables
- c) Legibilidad por el hombre:
 - Legibles directamente
 - Legibles por traducción
 - Legibles por máquina
- d) Direccionabilidad:
 - Direccionable
 - No Direccionable

En el siguiente cuadro se muestran los soportes físicos:

Clasificación sintética de soportes físicos

SOPORTE	DESCRIPCION	CLASIFICACION			
		Función	Reutiliza- ble	Legibili- dad	Dirección
	Tarjeta Perforada	E/S	NO	Por Traducción	NO
	Banda de Papel	E/S	NO	Por Traducción	NO
<ul style="list-style-type: none"> . Marcas . Ópticos . Magnéticos 	Documentos	E	NO	Directa	NO
	Impresora	S	NO	Directa	NO
	Cinta Magnética	E/S	SI	Por Máquina	NO
	Disco Magnético	E/S	SI	Por Máquina	SI
	Tambor Magnético	E/S	SI	Por Máquina	SI
	Tarjeta Magnética	E/S	SI	Por Máquina	SI
	Terminal Remota	E/S	SI	Directa	NO

Características para los medios más utilizados:

Veamos como se registra la información en cada uno, aclarando que, sea cual fuere el soporte, la misma se registra caracter por caracter (numérico o alfabético).

1. Tarjeta Perforada

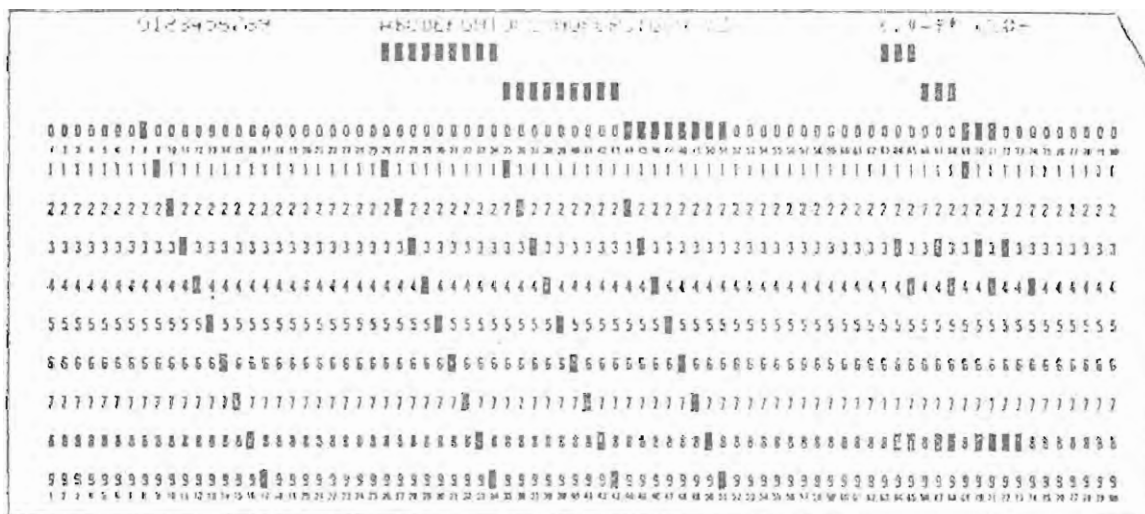
Capacidad: 80 caracteres/tarjeta (1 caracter por columna)

Códigos: Hollerith: números: 1 perforación por columna
letras: 2 perforaciones por columna
símbolos: 3 perforaciones por columna

Powers
Bull

Periférico: Lector de tarjetas perforadas

Velocidad: 200-1.500 tarjetas/minuto



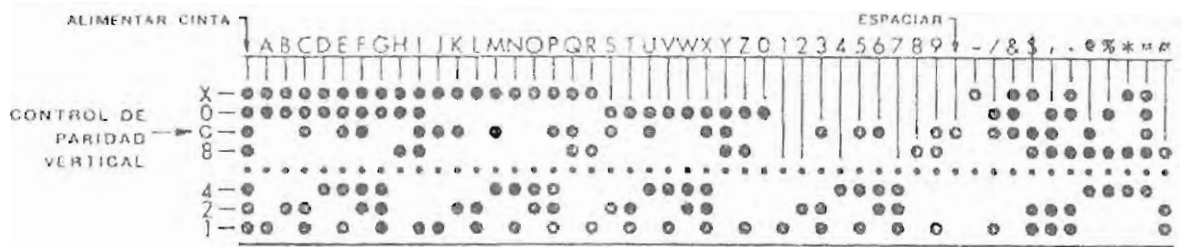
Tarjeta perforada

2. Banda de papel

Capacidad: ilimitada

Codificación: combinación de perforaciones transversales a la dirección del movimiento. Cada caracter puede representarse con una combinación de 5, 6, 7 u 8 perforaciones o bits.

Velocidad lectura: 100-1.000 caracteres/segundo.



Cinta de papel perforada

3. Documentos

De gran aplicación en la gestión bancaria, no se describirán en este texto.

HERNANDO ALONSO LEONARDO

LINEA	CORRE		PRIMER VIA II		VIA II		ALISA		CAMBIO LINEA		
	0	1	0	1	0	1	0	1	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9

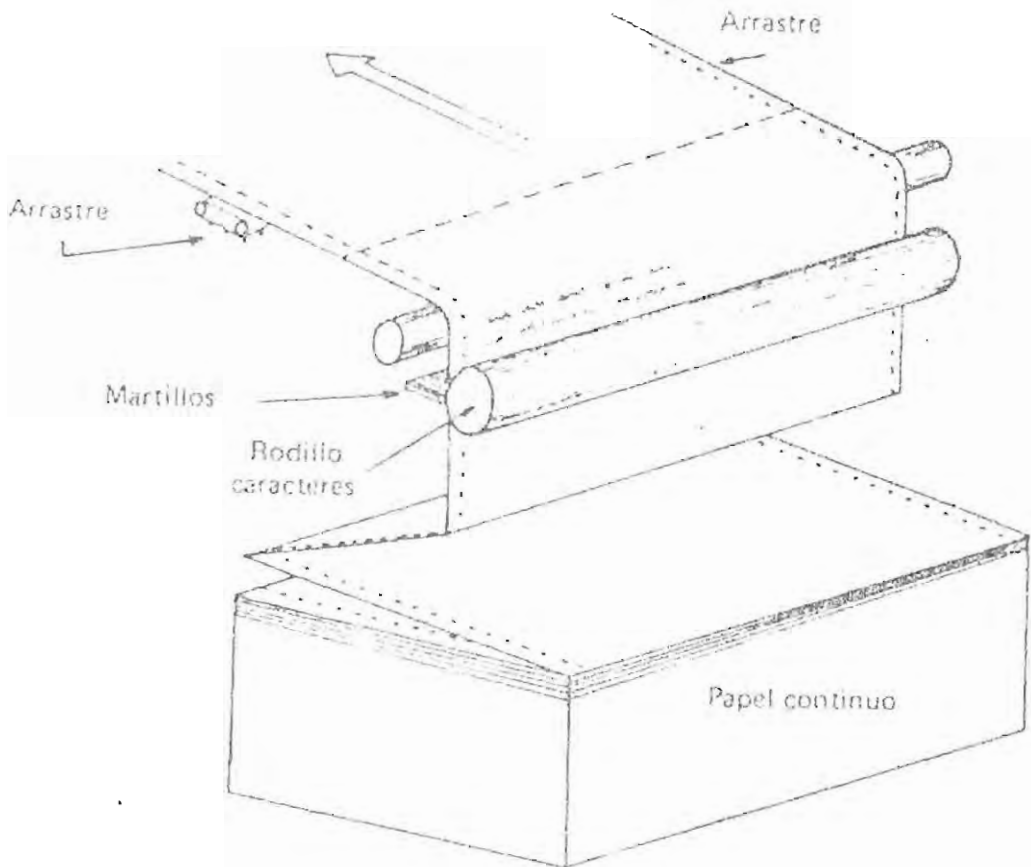
INSPEL 19 DOCUMENTO 11-11A

4. Impresoras

Capacidad: 132/136 caracteres/línea

- Tipo: -Rodillo
- Cadena
- Banda

Velocidad: 1.600 líneas/minuto



Impresora de Rodillo

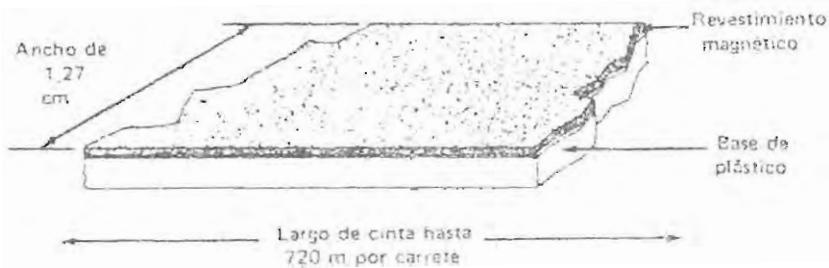
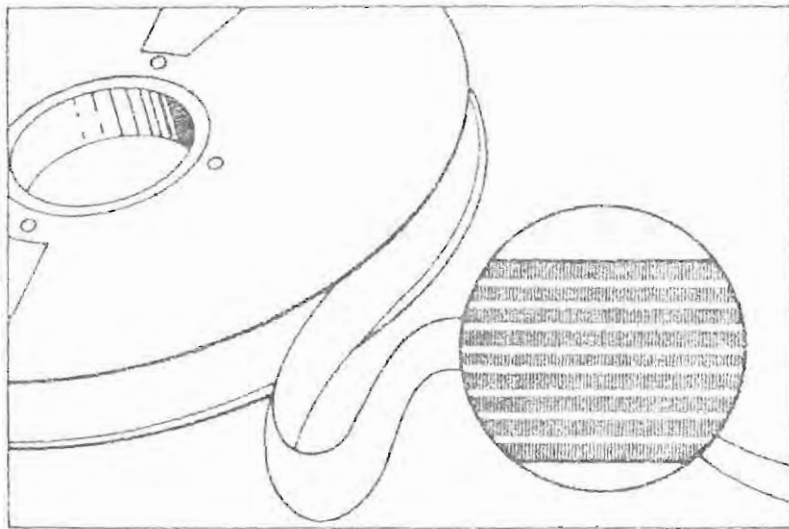
5. Cinta magnética

Capacidad: ilimitada

Densidad: 200-1.600 posiciones (caracteres/pulgada)

Velocidad :10.000 a 120.000 caracteres/segundo





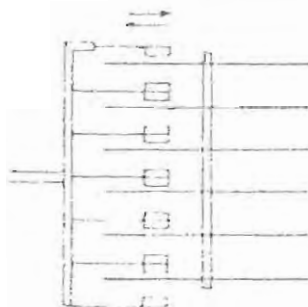
6. Disco magnético

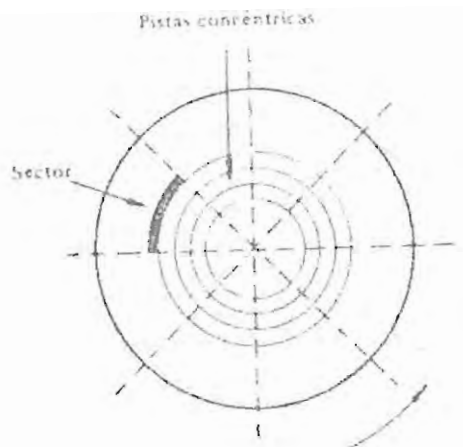
Capacidad: 20 sectores/pista
 100-500 pistas/cara
 10 caras/pack
 aprox. 2.000.000 caracteres/pack

Tiempo de acceso: 30-145 milisegundos

Registros de información: serie o paralelo

Velocidad aproximada: 150.000 caracteres/segundo

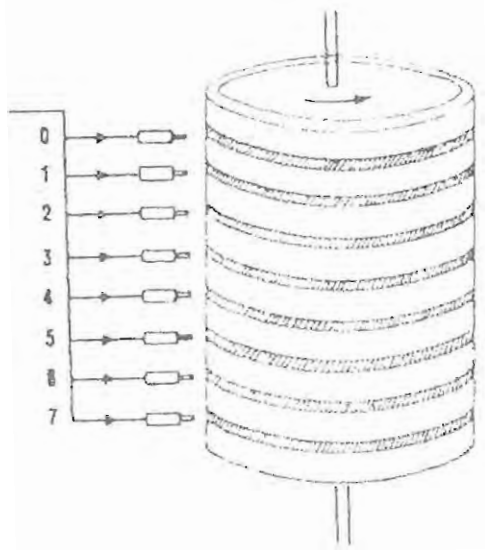




7. Tambor magnético

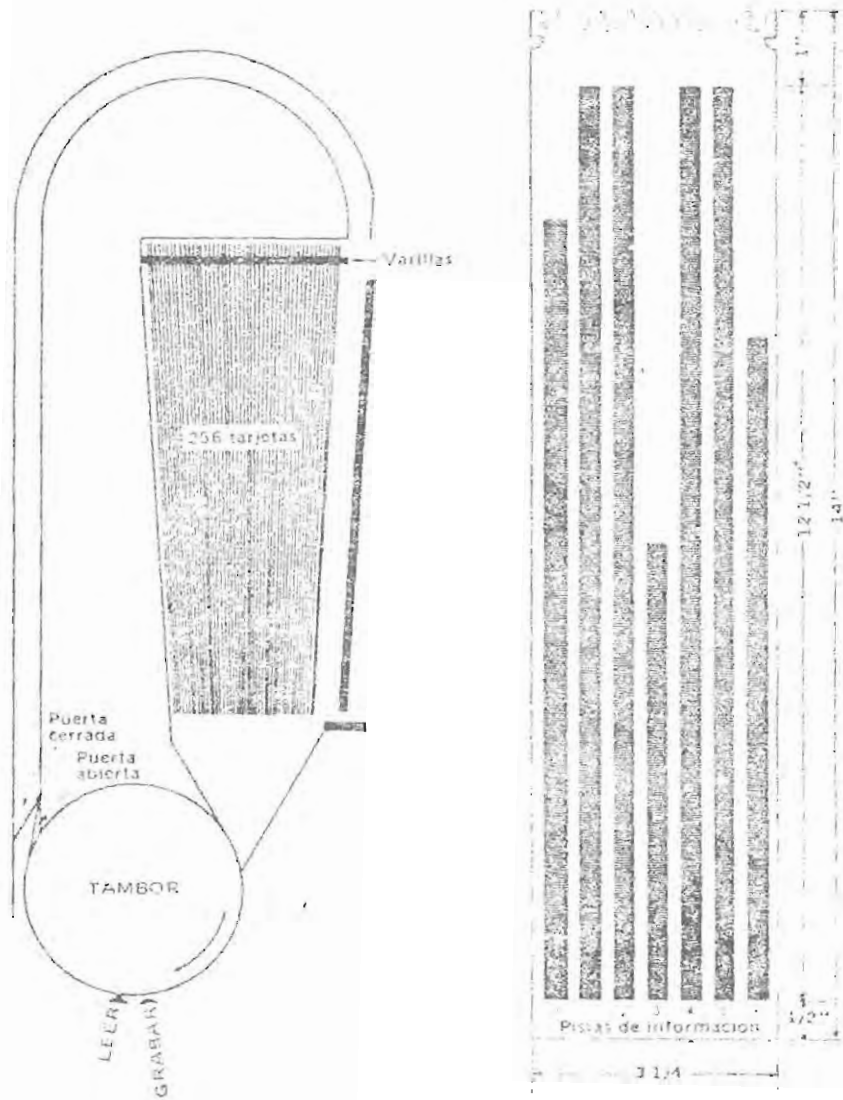
Capacidad: 100.000 a 5.000.000 caracteres
800 pistas

Velocidad transmisión: 1.200.000 caracteres/segundo



8.- Tarjeta magnética

Capacidad: Cartuchos de 256 tarjetas
Cada tarjeta 21.700 caracteres



9. Terminal remota

Capacidad: número de líneas de la pantalla

Apto para los trabajos en tiempo real.

Se puede acompañar con una pequeña impresora para registrar la información de pantalla.



III. Archivos. Organización de archivos. Búsqueda de registros.

Conocidas las características y codificación de los soportes físicos, la información se registrará en ellos estructurada en conceptos lógicos llamados "registro". Este a su vez estará constituido por "campos" que constituyen los items con relación a una aplicación. Podemos decir:

Archivo: Conjunto ordenado de registros

Registro: Conjunto de campos con concepto lógico

Campo: Conjunto de caracteres del item del registro

Por ej.: el archivo de sueldos de una empresa contendrá registros a razón de uno por empleado, a su vez cada registro contendrá campos como nombre y apellido, nº de legajo, cargas sociales, licencias, etc.

Los archivos se estructuran, en general, con registros de longitud constante, es decir de igual cantidad de caracteres.

En cintas magnéticas, a su vez, se agrupan una cantidad constante de registros lógicos en una unidad mayor llamada registro físico. La razón reside en que ésta es la mínima unidad que precisa el computador en memoria central.

La organización de un archivo depende fuertemente del soporte físico y el procesamiento del mismo de su organización.

Se entiende por procesamiento la actualización y la búsqueda de registros. Existen tres tipos fundamentales:

- a) Secuencial
- b) Secuencial con índice (indexado)
- c) Random (o al azar)

La organización del archivo en uno de estos tipos depende fuertemente del medio físico o soporte utilizado.

En la organización secuencial los registros se ordenan uno a continuación de otro siguiendo un orden creciente o decreciente de los datos de un campo particular llamado campo clave. El campo clave es un dato que está en todos los registros y es distinto para cada uno de ellos. Por ej. el nº de padrón de cada alumno inscripto en una Facultad.

En la organización secuencial indexada (que es una variante del tipo anterior) los registros se ordenan en forma secuencial por campo clave y luego se arma un pequeño archivo constituido por registros que solo contienen el campo clave y otro campo que contiene "la dirección" donde se encuentra el registro en el archivo principal. Esto se hace para facilitar la búsqueda de un registro sin recorrer todo el archivo.

En la organización random o al azar los registros se ordenan uno a continuación de otro pero no hay relación alguna entre un registro y el que sigue. Para ubicar un registro particular en este caso se usan métodos.

Desarrollamos aquí la organización factible para cada medio físico como así también el procesamiento del mismo, entendiéndose por procesamiento el de actualización y búsqueda de registros.

a) Tarjetas perforadas

La organización es secuencial conforme a un campo clave y con registros como máximo de 80 caracteres (columnas de una tarjeta).

El proceso de actualización y búsqueda es manual:

El ordenamiento de las tarjetas en forma secuencial se hace en una clasificadora.

b) Cintas magnéticas

La organización es secuencial con registros ordenados en forma creciente o decreciente del campo clave pero su numeración no es consecutiva para permitir la actualización.

El proceso de actualización se hace en computadora, haciendo uso de tres cintas: una con el archivo; otra con las modificaciones y la tercera que será la resultante de la actualización.

La búsqueda o recuperación de la información (búsqueda de un registro) se hace con métodos. Los más comunes son: a) directa, b) uno cada n, c) binaria.

En el método a) se compara uno a uno el identificador de cada registro de la cinta con el identificador (campo clave) buscado.

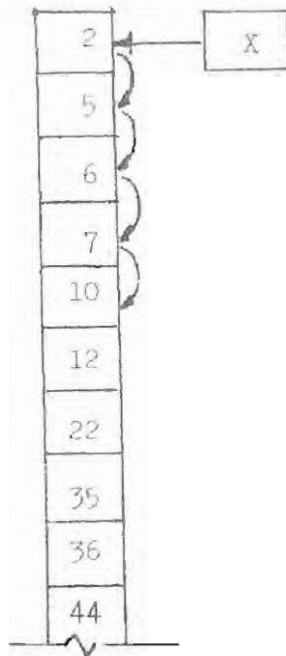
En el método b) se subdivide en n partes el archivo y se compara el identificador buscado con los identificadores cada n registros. Al ubicar el grupo de identificadores del grupo n se recorre todo el tramo comparando uno a uno hasta detectar el registro buscado.

En el método c) se compara por mitades. El identificador del registro buscado se compara con el de la mitad del archivo, el mismo será mayor o menor desechando de esta forma una mitad del archivo. Luego se repite el proceso en la mitad restante y así sucesivamente hasta ubicar el registro buscado.

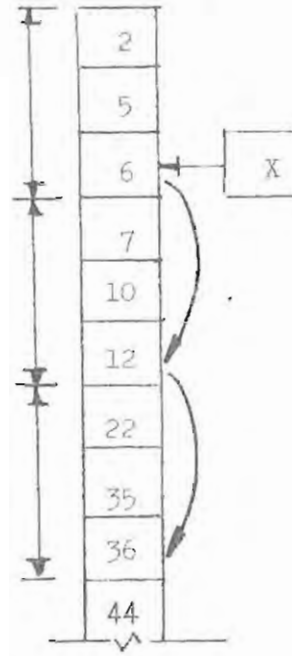
Los dos últimos métodos suponen el archivo principal de la computadora.

Los tres métodos descriptos se ilustran a continuación:

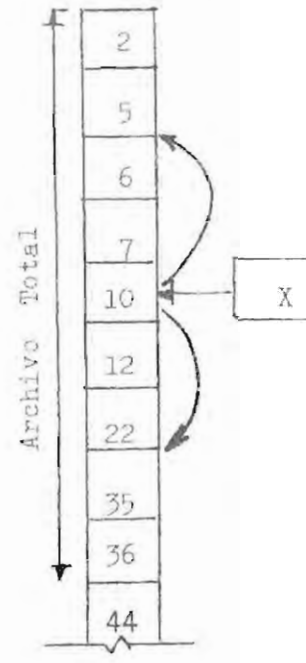
a) Directa



b) Uno de cada n



c) Binaria



En los gráficos, el rectángulo marcado con X corresponde al registro cuyo identificador buscamos en la lista cuyos registros están ordenados.

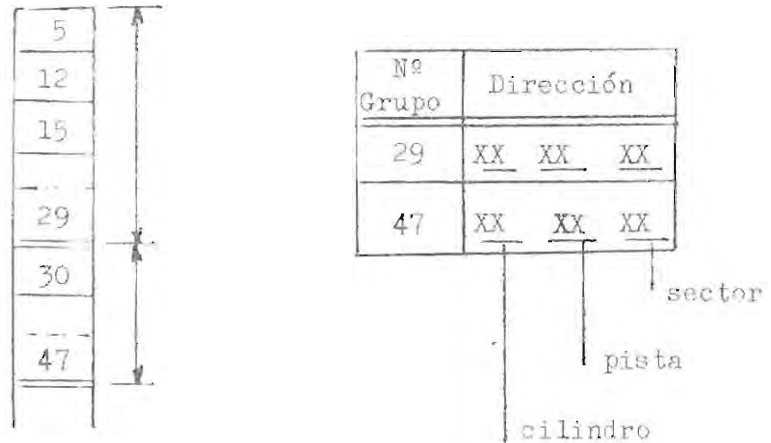
c) Discos magnéticos

La organización puede ser secuencial, secuencial con índice o al azar.

La secuencial es la misma que vimos para cinta.

En la organización secuencial indexada se usa la cara superior para armar un archivo auxiliar que contiene el campo clave de cada grupo de n registros y la dirección donde se encuentra el primer registro del grupo en el paquete de discos. Esta dirección está dada por tres números: el primero indica el cilindro, el segundo la pista y el tercero el sector. (Cilindro se denomina al cilindro imaginario formado por las pistas de igual número en cada cara del disco).

El procesamiento se hace entrando en memoria el archivo índice una vez ubicado el registro buscado, se da la orden al brazo correspondiente, el que accede al registro en forma directa.



En la organización al azar se relaciona el campo clave de cada registro con la dirección del archivo mediante un algoritmo de direccionamiento.

Este tipo de organización es ventajoso respecto del secuencial en que se necesita menor reorganización y posee velocidad de acceso más rápida. Esta última ventaja lo hace apto para el uso con terminales.

La desventaja principal de los archivos random es la creación de sinónimos (2 registros distintos ubicados por el algoritmo en igual lugar físico) y los espacios vacíos.

Los algoritmos usados son varios, podemos nombrar: a) dirección directa, b) tabla, c) Condensación, d) selección de dígitos, e) residuo. Por ej.: en la técnica de dirección directa si el registro tiene una longitud de 1 sector se asignará a cada registro la dirección del sector (cilindro, pista, sector), si ocupara

3 sectores se asignará dirección con la fórmula:

$1000 \times 3 - 3 = 2997$ 1er. registro
 $1001 \times 3 - 3 = 3000$ 2º registro

comenzando su almacenamiento en dirección 1000.

Para finalizar un comentario sobre una nueva técnica hacia la que tiende la organización de archivos apta para el trabajo en tiempo real: la base de datos.

En general, del análisis de una aplicación surgen el diseño de archivos y programas; la explotación de esos archivos por otros programas de la misma u otra aplicación necesita frecuentemente clasificaciones y creación de nuevos archivos. Si se agrega el hecho de que un dato puede ser requerido según distintas vías de acceso (terminales remotas) el sistema es impracticable en tiempo real. El desarrollo de los medios y soportes de almacenamiento con acceso directo hicieron que gradualmente se ejerciera más control coordinado y centralizado sobre los datos. El proceso de integración originó este nuevo concepto en el campo de la informática, el de la base de datos.

Entre los objetivos fundamentales figuran:

- . Permitir que los datos sean estructurados en forma más conveniente a cada aplicación sin redundancia de datos.
- . Permitir que varios programas que se están ejecutando concurrentemente puedan acceder o modificar la base de datos.
- . Proveer y permitir el uso de una variedad de estrategias de búsqueda en toda la base de datos a partir de ella.
- . Proveer la posibilidad de controlar la ubicación física de los datos y facilitar la independencia de los programas respecto de los datos.