

# Taller de Programación

## Clase 5 - Uso de archivos

Mgs. Lic. Marcos Prunello

2017-09-24

- Los programas usan variables para guardar información (datos de entrada, resultados calculados, valores intermedios, etc).
- Pero cuando los programas dejan de correr, el valor almacenado en las variables se pierde.
- En muchas ocasiones, es necesario guardar información de una forma más permanente.

- **Archivo o fichero:** un conjunto de información sobre un mismo tema tratado como una unidad de almacenamiento y organizado de forma estructurada para la búsqueda de un dato individual.
- Generalmente se graba en el disco rígido de la máquina y se usa para recolectar la información en un todo lógicamente cohesivo.
- Los archivos pueden contener instrucciones de programas o información creada o usada por un programa.
- Todos los objetos de datos permanentes que guardamos en nuestra computadora (documentos, juegos, programas ejecutables, código, etc.) son guardados en la forma de archivos.

- La unidad elemental de un archivo: **registro**
- Es una colección de información relativa a una misma entidad.
- Cada registro de un mismo archivo tiene la misma estructura que los demás.
- Los datos individuales sobre dicha entidad ocupan **campos** dentro de los registros.

Ejemplo:

ARCHIVO: Pasajeros

Campo 1: NOMBRE, tipo caracter

Campo 2: NÚMERO DE VUELO, tipo caracter

Campo 3: FECHA DE VUELO, tipo caracter

Campo 4: NÚMERO DE ASIENTO, tipo caracter

Campo 5: CIUDAD ORIGEN, tipo caracter

Campo 6: CIUDAD DESTINO, tipo caracter

Campo 7: PRECIO, tipo numérico

Nombre	NroVuelo	FechaVuelo	NroAsiento	Origen	Destino	Precio
Pamela Suárez	AR6071	12/09/17	17A	Rosario	Córdoba	1532.23
Gonzalo Echarri	AR5423	14/09/17	31B	Rosario	Buenos Aires	1424.10
...	...	...	...	...	...	...

- Identificar un registro dentro del archivo:
  - Eligiendo un campo llamado **clave** (por ejemplo, el nombre del pasajero), que contiene un único valor para cada registro.
  - A través del valor de más de uno de sus campos.

Clasificación según la forma en la que se pueda acceder a cada uno de sus registros:

## Archivo secuencial

- Los registros se encuentran en cierto orden que debe ser respetado para la lectura de los mismos.
- Para leer el registro situado en la posición  $n$ , se debe pasar antes por los  $n - 1$  registros anteriores.
- Los registros se leen uno por uno hasta llegar al final del archivo. La mayoría de los lenguajes de programación disponen de una función lógica que devuelve un valor **VERDADERO** cuando se alcanza el final del archivo.

## Archivo directo

- Está formado por un conjunto de registros que pueden ser recuperados por su posición dentro del archivo sin necesidad de recorrer los anteriores.

## Archivo indexado

- Dispone de una tabla de índices adicional, es decir, una referencia que permite obtener de forma automática la ubicación de la zona del archivo donde se encuentra el registro buscado.
- Permite localizar un registro por medio de su clave, o del valor de algún campo en particular, sin recorrer previamente los registros que lo preceden.



Procedimientos básicos que los programas llevan a cabo sobre los archivos:

- **Creación de un archivo:** consiste en la escritura de los registros que van a conformar el archivo.
- **Apertura de un archivo:** para que un programa pueda operar directamente sobre un archivo, debe realizar es la **apertura** del mismo, que incluye la identificación del archivo a utilizar y el modo (lectura, escritura, etc.).
- **Cierre de un archivo:** cuando un programa no vaya a acceder más a un archivo, es necesario indicarlo a través del **cierre** del mismo.
- **Lectura y escritura en un archivo:** la **lectura** consiste en transferir información del archivo a la memoria principal usada por el programa, mientras que la **escritura** es la transferencia de información guardada en las variables del programa al archivo.

Otras operaciones que se pueden realizar sobre los archivos incluyen acciones de:

- **Actualización:** añadir (dar de alta), modificar o eliminar (dar de baja) algún registro.
- **Clasificación:** reubicar los registros de tal forma que queden ordenados por algún campo determinado.
- **Fusión o mezcla:** combinar dos o más archivos para formar uno nuevo.
- **Partición:** subdividir los registros por el valor que toman en algún campo.

# Pseudo-código

Al escribir los algoritmos en pseudo-código, se puede hacer uso de las siguientes expresiones para representar algunas operaciones a realizar sobre los archivos:

- `ABRIR (nombre del archivo) secuencial`, de entrada/salida: indica la acción de apertura de un archivo con organización secuencial, sobre el cual se va a proceder a realizar lectura o escritura de registros.
- `CERRAR (nombre del archivo)`: para indicar que el archivo no será vuelto a utilizar.
- `LEER (nombre del archivo) campo1, campo2, ...`: indica la lectura del próximo registro cuando se trabaja con un archivo secuencial. La expresión `campo1, campo2, ...` son los nombres de los campos que contienen los registros y constituyen los identificadores de las variables que almacenarán temporalmente los datos del registro que se están leyendo.

- LEER (nombre del archivo, número de registro) campo1, campo2, ...: indica la lectura de un registro en particular cuando se trabaja con un archivo de acceso directo.
- LEER (nombre del archivo, campo1 = valor) campo1, campo2, ...: indica la lectura de cada uno de los registros en los cuales el campo campo1 es igual a valor, para cuando se trabaja con un archivo de acceso indexado.
- Las tres acciones de LEER anteriores tienen su acción análoga de ESCRIBIR, para modificar o añadir registros al archivo.
- FINDE (nombre del archivo): es la función lógica que devuelve el valor FALSO mientras resten más registros por leer en el archivo y el valor VERDADERO cuando se llega al final del archivo y ya no quedan más registros por leer.

- Son un buen soporte para archivos como el del ejemplo y como los que usaremos en el Taller.
- Un **archivo de texto** es un archivo informático que contiene únicamente texto formado por una secuencia ordenada de caracteres.
- El texto almacenado en este tipo de archivo carece de formato tipográfico (negrita, cursiva, colores, subrayado, fuente, etc.).
- Una gran variedad de programas puede leer y editar el contenido de estos archivos.
- Están compuestos por caracteres ordinarios (letras, números y signos de puntuación) y especiales (saltos de línea, tabulaciones).
- Como las computadoras solamente entienden números, cada caracter es codificado internamente con una representación numérica binaria.
- Hay diferentes formatos de codificación de caracteres (como *ASCII*, *ISO-8859-1* o *UTF-8*).