

Organización de los Datos

Unidad 3

Almacenamiento y Estructuras de Archivos

Introducción

Propiedades de los Archivos

- **Persistencia:** Datos escritos en un archivo persisten después de que el programa termina, los datos pueden usarse posteriormente
- **Compartición:** Datos almacenados en archivos pueden ser utilizados por otros programas y usuarios simultáneamente
- **Gran tamaño:** Archivos de datos pueden ser muy grandes, generalmente no caben en memoria principal

Organización de los Archivos

- La organización de los datos en un sistema de cómputo se puede ver como una serie de agrupamientos:
 - Bits
 - Bytes
 - Campos
 - Registros
 - Archivos
- Bases de Datos
 - Atributos
 - Entidades

Campo

- Es la unidad básica de un archivo
- Un campo es una agrupación de caracteres que puede representar una característica o atributo
 - Combinación de caracteres
 - Números
 - Combinación de palabras

Registro

- Un registro es la colección de varios campos que describen a una entidad
- Cuando un registro es almacenado en memoria, es referido como un objeto o estructura, y los campos que lo forman son conocidos como miembros
- Cuando ese objeto es almacenado en un archivo se le conoce solo como registro

Registro

- La información sobre un estudiante puede ser almacenada en un registro que a su vez está formado por campos que contienen cierta información

Nombre	Apellido paterno	Apellido materno	Licenciatura	Matrícula
---------------	-------------------------	-------------------------	---------------------	------------------

Archivo

- Un archivo es la colección de varios registros del mismo tipo

Almacenamiento de un Archivo

- Lo más importante cuando se almacenan datos en un medio de almacenamiento es que la información se podrá recuperar
- Existen dos maneras de representar datos en un medio de almacenamiento
 - Representación textual
 - Representación binaria

Representación Textual

- Se representan los datos con una cadena de caracteres del alfabeto y algunos símbolos especiales
- Es conveniente cuando se representan datos alfa numéricos
- Para datos numéricos se debe considerar el formato del dato (número de posiciones para la fracción, signo presente, etc)
- Una ventaja es que es muy portable ya que no se depende de la manera en que se representa un número en la máquina

Representación Binaria

- Los datos se escriben con el mismo formato en que se encuentran en memoria
- Se debe elegir el tipo de dato más adecuado considerando el rango de valores que éste puede tener

Organización de un Archivo

Formas de Organización

- Las diferentes métodos para organizar campos en un registro son:
 - Hacer que los campos sean de un determinado número de bytes de longitud
 - Indicar la longitud de cada campo
 - Colocar un separador al final de cada campo
 - Utilizar expresiones de la forma dato=valor para identificar un campo y su contenido

Número Fijo de Bytes (*Fixed-Lenght Records*)

- Este método es de los más utilizados en la organización de archivos
- Es un archivo en donde cada registro contiene el mismo número de bytes
- Es sencillo recuperar la información
- Puede presentar problemas con el espacio (espacio desperdiciado o espacio insuficiente)

Nombre	Apellido paterno	Apellido materno	Licenciatura	Matrícula
--------	------------------	------------------	--------------	-----------

Indicadores de Longitud

- Se coloca la longitud al inicio de cada campo
- Este método se utiliza para manejar registros de longitud variable
- Es sencillo recuperar la información
- Se debe considerar la longitud que tiene este indicador para no confundirlo con información

06Nombre

12Apellido
paterno

13Apellido
materno

15Licenciatura

10Matrícula

Uso de Delimitadores

- Se utiliza un carácter especial como delimitador
- El separador se debe distinguir de manera sencilla del resto de los caracteres
- Este carácter no debe interferir en el procesamiento

Nombre |

Paterno |

Materno |

Licenciatura |

Matrícula |

Campos con Descriptores

- El uso de expresiones tipo=valor se conoce como uso de descriptores
- Se utiliza en conjunto con los separadores
- Hay que considerar que el descriptor no sea muy grande
- Un manejo avanzado de los descriptores y llaves se observa en las Bases de Datos

Nom=Nombre

Pat=Paterno

Mat=Materno

Lic=Licenciatura

Mat=Matrícula

Actividad

- Manejo de archivos CSV para consulta por campos

Sistema de Archivos

Uso de Sistemas de Archivos

- Organizar la información de una organización en archivos se conoce como utilizar un Sistema de Archivos
- Aquí existen varios archivos que contienen distinto tipo de información que puede o no estar relacionada entre sí

Desventajas

Desventajas de un uso Tradicional

- El manejar la información con archivos, lo que se denomina un uso tradicional acarrea ciertas desventajas:
 - Redundancia. Almacenamiento de los mismos datos en varios archivos
 - Dificultad en la compartición de archivos
 - Inconsistencia debido a las modificaciones realizadas en un archivo