|  |
| --- |
| cifp Juan de colonia |
| Acceso a Datos |
| UT1: Manejo de ficheros |
|  |
| **Alexis López Briongos Dam2t** |
| **02/11/2023** |

|  |
| --- |
|  |

*Índice*

[FICHERO 2](#_Toc157696761)

[CLASES JAVA 2](#_Toc157696762)

[FLUJO O STREAM 4](#_Toc157696763)

[ACCESO A FICHEROS 5](#_Toc157696764)

[OPERACIONES SOBRE FICHEROS 6](#_Toc157696765)

[OPERACIONES SOBRE REGISTROS DE UN FICHERO 6](#_Toc157696766)

[CLASES PARA LA GESTIÓN DE STREAM DE DATOS 7](#_Toc157696767)

[FICHEROS BINARIOS 8](#_Toc157696768)

[OBJETOS EN FICHEROS BINARIOS 9](#_Toc157696769)

[ACCESO ALEATORIO FICHEROS 9](#_Toc157696770)

[FICHEROS XML 10](#_Toc157696771)

[Acceso a los ficheros XML mediante DOM 11](#_Toc157696772)

[Acceso ficheros con SAX 11](#_Toc157696773)

# FICHERO

¿Qué es un fichero?

* Conjunto de bits (información) almacenados en un dispositivo.

Características:

* No son volátiles.
* Se ubican en directorios o carpetas.
* Dentro de su directorio tienen un nombre (identificador) único.
* Cada fichero suele tener una extensión la cual indica el tipo de datos que contiene.

Compuesto por:

* Registros
  + Campos
* La manera en que se agrupan los datos dentro de un fichero depende del propio diseño del fichero.

# CLASES JAVA

¿Qué es java.io?

* Es la clase que contiene todo lo necesario para tratar ficheros con java

¿Qué es File?

* **File**: es una clase que se utiliza para la manipulación de ficheros.
* **List()**: devuelve un array de strings con los nombres de los ficheros y directorios de un objeto File.
* **ListFiles()** : devuelve un array de files con los ficheros y directorios que existen en ese directorio.
* **getName()**: devuelve el nombre de un fichero o directorio.
* **getPath()**: devuelve la ruta relativa del fichero o directorio que le pasemos.
* **getAbsolutePath()**: devuelve la ruta absoluta del fichero o directorio que le pasemos.
* **exists()**: devuelve un booleano si existe o no un fichero o directorio.
* **canWrite()**: devuelve un booleano si el fichero se puede escribir o no.
* **canRead()**: devuelve un booleano si el fichero se puede leer o no.
* **isFile()**: devuelve un booleano si es fichero o no
* **isDirectory()**: devuelve un booleano si es directorio o no
* **length()**: devuelve un longitud de un fichero en bytes
* **mkdir()**: crea un directorio si no existe y devuelve un booleano si existe o no dicho directorio
* **renameTo(File, “nuevoNombre”)**: renombra el fichero y devuelve booleano si el fichero contiene , o no existe , o no tienes permisos de escritura.
* **delete():** elimina el directorio si está vacío y devuelve un booleano si ha podido hacerlo.
* **createNewFile():** crea un fichero vacío si se puede.
* **getParent():** devuelve el nombre del directorio padre (string) si no devuelve null.

# FLUJO O STREAM

* Para manejar la entrada salida entre una fuente y un destino se utiliza los flujos o streams. Es decir, cualquier programa que necesite obtener información de cualquier fuente necesita abrir un stream y viceversa (enviar).
* La vinculación entre el stream y el dispositivo fuente la realiza el propio sistema de e/s de java (java.io).
* Tipos de streams
  + Streams de bytes (8 bits)
    - Se utilizan para operaciones de e/s de bytes. Es decir, para la lectura y escritura de datos binarios.
    - Todas las clases de streams de bytes descienden de las clases InputStream/OutputStream y cada una de ellas tiene varias subclases dependiendo del dispositivo que se vaya a utilizar.
  + Streams de caracteres (16 bits)
    - Se implementaron porque los streams 8 bits se quedaron cortos.
    - Se utilizan para operaciones de e/s de caracteres y para ello se emplean las clases: 🡺Manejan streams de caracteres UNICODE
      * Reader
      * Writer
* Clases de streams de bytes(8 bits):
  + InputStream: clases que producen entradas de distintas fuentes
    - ByteArrayInputStream permite utilizar un espacio de almacenamiento intermedio (buffer) en memoria.
    - StringBufferInputStream: convierte un string en un InputStream
    - PipedInputStream: almacena en tuberías
    - FilterInputStream: da funcionalidades a otras clases de InputStream
      * LineNumberInputStream
      * DataInputStream
      * BufferedInputStream
      * PushBackInputStream
    - SequenceInputStream: convierte dos o más objetos InputStream en un InputStream único.
    - FileInputStream: realiza un flujo de entrada desde un fichero.
  + OutputStream: clases que producen salidas de distintas fuentes.
    - ByteArrayOutputStream: permite utilizar un espacio de almacenamiento intermedio(buffer) en memoria.
    - PipedOutputStream: se asocia al PipedInputStream y la información que se graba en el PipedOutputStream acaba automáticamente como entrada al PipedInputStream asociado.
    - FilterOutputStream: da funcionalidades a otras clases de OutputStream
      * LineOutputStream
      * BufferedOutputStream
      * PrintStream
    - FileOutputStream: realiza un flujo de salida hacia un fichero.
* Clases de streams de 16 bits
  + Reader
    - BufferedReader (BufferedInputStream)
      * LineNumberReader
    - CharArrayReader (ByteArrayInputStream)
    - InputStreamReader: convierte un InputStream en un Reader
      * FileReader (FileInputStream)
    - FilterReader(FilterInputStream)
      * PushBackReader (PushBackInputStream)
    - PipeReader (PipedInputStream)
    - StringReader (StringBufferedInputStream)
  + Writer
    - BufferedWriter (BufferedOutputStream)
    - CharArrayWriter (ByteArrayOutputStream)
    - OutputStreamWriter
      * FileWriter (FileOutputStream)
    - FilterWriter (FilterOutputStream)
    - PipedWriter (PipedOutputStream)
    - StringWriter (StringOutputStream)
    - PrintWriter (PrintOutputStream)

# ACCESO A FICHEROS

* Existen dos maneras para acceder a ficheros
  + Acceso Secuencial en el cual los datos o registros del fichero se leen y se escriben siempre en orden secuencial (de principio a fin) el nuevo registro ira siempre al final del fichero a parir del último fichero.
    - Se suelen utilizar en procesos por lotes (Batch).
    - Clases para acceder de forma secuencial a los ficheros
      * FileOutputStream
      * FileInputStream
      * FileWriter
      * FileReader
    - Ventajas
      * Son rápidos para acceder a los registros.
      * Aprovechan mejor el espacio.
      * Son sencillos de usar.
    - Inconvenientes
      * No se puede acceder directamente a un registro.
      * Son muy complicados de actualizar.
  + Acceso Aleatorio (ficheros aleatorios) se permite acceder directamente a un dato o registro sin haber pasado antes por los anteriores.
    - Los datos tienen que estar almacenados en registros de tamaño fijo.
    - Clases que se utilizan para acceso aleatorio en ficheros.
      * RandomAccessFile.
    - Los ficheros aleatorios utilizan direcciones relativas, no absolutas.
    - Ventajas
      * Rapidez con la que se accede a un registro.
    - Inconvenientes:
      * Establecer posición de cada registro
      * Se desaprovecha el espacio, puede haber huecos.

# OPERACIONES SOBRE FICHEROS

* Independientemente de cuál sea su forma de acceso, las operaciones que se puedan realizar sobre un fichero son:
  + Creación: de un fichero con un nombre único en el directorio que nos permite acceder a él. Es un proceso que solo se debe hacer una vez.
  + Apertura: es la primera operación que hay que hacer para que un programa pueda operar con el fichero.
  + Lectura: transferir información desde el fichero hacia la memoria (programa).
  + Escritura: transferir información desde la memoria hacia el fichero.
  + Cierre: cuando no se vaya a usar más y suele ser siempre la ultima instrucción del programa.
  + Eliminación

# OPERACIONES SOBRE REGISTROS DE UN FICHERO

* Altas: insertar nuevos registros
  + Si el acceso es secuencial el registro ira al final del fichero.
  + Si el acceso aleatorio el registro irá se aplica la función designada.
* Bajas: eliminar registros existentes:
  + Si el acceso es secuencial el borrado será física.
    - Se emplea algún campo que cambia de valor para indicar que ese registro está borrado lógicamente.
    - Se realizan las con un fichero auxiliar de manera que se leen todos los registros y se copian en otro fichero leyendo todos los registros y obviando el que tiene registro modificado, se elimina el fichero original y se renombra el fichero auxiliar con el nombre del fichero original.
  + Si el acceso es aleatorio el borrado será lógica del fichero.
    - Se emplea algún campo que cambia de valor para indicar que ese registro está borrado lógicamente.
* Modificaciones: cambiar valores de los registros.
  + Si el acceso es secuencial la modificación será lógica.
    - Se emplea algún campo que cambia de valor para indicar que ese registro está borrado lógicamente.
    - Se realizan con un fichero auxiliar de manera que se leen todos los registros y se copian en otro fichero leyendo todos los registros y añadiendo el registro que tiene el campo modificado pero con el valor modificado, se elimina el fichero original y se renombra el fichero auxiliar con el nombre del fichero original.
  + Si el acceso es aleatorio la modificación mediante la clave.
    - Con el campo clave buscar el registro que queremos modificar y cambiar el valor del registro.
* Consultas: accedemos y obtenemos información de los registros:
  + Si el acceso es secuencial consulta será desde el primero hasta el registro a encontrar.
  + Si el acceso es aleatorio se recogerá el valor de los registros por la clave introducida.

# CLASES PARA LA GESTIÓN DE STREAM DE DATOS

* Ficheros de texto manejan caracteres alfanuméricos (ASCII, UTF-8).
* En ellos se emplean las clases FileReader y FileWriter.
* Cada vez que se lee o escribe en un fichero de este tipo, hay que hacerlo en un bloque try catch dado que pueden generar excepciones.
  + En el caso de lectura 🡺 FIleNotFoundException
  + En el caso de escritura 🡺 IOException
* FILEREADER métodos: estos métodos devolverán siempre los caracteres leídos o -1 si se ha llegado a final del fichero o si no se puede leer más (no hay más datos, corruptos).
  + Int Read(): lee un carácter y lo devuelve
  + Int Read(char[] buf): leer hasta buf.length caracteres de una matriz de caracteres buf desde el principio.
  + Int Read(char[]buf, int despl, int n): lee hasta n caracteres de la matriz buf, comenzando por buf[despl] indicando el despl (desplazamiento) .
  + Pasos con los que trabajaremos con los ficheros.
    - Creamos fichero, creamos el flujo de entrada con el FileReader, realizamos operaciones de lectura, cerramos el fichero con el método close().
* FILEWRITER métodos:
  + Void write(int c): escribi
  + Void write(char[]buf)
  + Void write(char[]buf, int despl, int n): escribir n caracteres de la matriz buf comenzando desde int despl.
  + Void Write(String str): escribir una cadena de caracteres
  + Void append(Char c): añadir el carácter al final del fichero
* La clase FileReader no tenia métodos para leen líneas completas pero BufferedReader si con el método readLine();
* BUFFEREDREADER métodos:
  + readLine(): lee líneas completas o null si no hay nada o es fin de fichero
  + read(): lee caracteres o null si no hay nada o es fin de fichero
* BUFFEREDWRITER métodos:
  + Write():
* PRINTWRITER:
  + Print(string): imprime la cadena
  + Println(string) :imprime la cadena y un salto de línea

# FICHEROS BINARIOS

* Ficheros que almacenan secuencias de dígitos binarios que no son legibles directamente.
* Ventajas
  + Ocupan menos espacio que los de caracteres.
  + Para almacenar mucha información son mejores.
* Clases que se utilizan para estos ficheros
  + FileInputStream
  + FileOutputStream
* FILEINPUTSTREAM métodos:
  + Int read(): lee un byte(8bits) y lo devuelve
  + Int Read(byte[] b): leer la matriz b hasta su longitud máxima y los devuelve.
  + Int read[byte[] b, int index, int n): lee n bytes(8bits) de la matriz b, comenzando por index.
* FILEOUTPUTSTREAM métodos:
  + Void write(int b):
  + Void write(byte[] b)
  + Void write(byte [] b, int index,int n) : escribe n bytes, de la matriz b, comenzando desde el index.
* Para leer y escribir tipos de datos primitivos, (int, float,etc) se usan los siguientes métodos
* DataInputStream métodos:
  + Boolean readBoolean();
  + Byte readByte();
  + Int readUnsignedByte();
  + Int readUnsignedShort();
  + Shor readShort();
  + Char readChar();
  + Int readint();
  + Long readLong();
  + Float readFloat();
  + Double readDouble();
  + String readUTF();
* DataOutputStream métodos:
  + void writeBoolean(boolean b);
  + void writeByte(int v);
  + void writeBytes(string s)
  + void writeShort(int v);
  + void writeChar(int v);
  + void writeChars(String s)
  + void writeInt(int v);
  + void writeLong(long v);
  + void writeFloat(float v);
  + void writeDouble(doublé v);
  + void writeUTF(String str);

# OBJETOS EN FICHEROS BINARIOS

* Serializable tiene métodos para leer y guardar objetos en ficheros binarios.
* La serialización de objetos permite tomar cualquier objeto implementado por la interfaz serializable y convertirlo en una secuencia de bits, que pueden ser leídas o guardadas.
* Utilizaremos las clase ObjectOutputStream y ObjectInputStream.
* OBJECTINPUTSTREAM:
  + Object readObject();
    - Pueden lanzar excepciones:
      * IOException: error de entrada/salida
      * ClassNotFoundException
* OBJECTOUTPUTSTREAM:
  + void writeObject(Object obj);
    - Pueden lanzar excepciones:
      * IOException: error de entrada/salida
  + Void writeStreamHeader(): este método sirve para escribir la cabecera (es importante modificar este método creando una clase que extienda ObjectOutputStream para modificarlo y poder sobreescribir un fichero.

# ACCESO ALEATORIO FICHEROS

* RANDOMACCESFILE: esta clase tendrá métodos para acceder a un fichero de acceso aleatorio
  + RandomAccessFile(String nombreFich,String modoAcceso)
    - modoAcceso=“-r” se abre en modo lectura y el fichero debe existir
      * Puede dar exepción IOException
    - modoAcces=”-rw” se abre en modo lectura/escritura y si no existe, se crea
  + RandomAccesFile(File objetofile, String modoAcceso)
    - Pueden lanzar la excepción FileNotFoundException
  + Funcionamiento
    - Se tiene un puntero interno que señala la posición actual dentro del fichero. Se coloca siempre en la posición 0 cuando el fichero se crea y se va moviendo según la cantidad de Bytes escritos o leídos
  + RandomAccessFile métodos:
    - Se puede utilizar los métodos de DataInputSttream y DataOutputStream para leer y escribir.
    - Long getFilePointer(); devuelve la posición actual del puntero
    - Void seek(long posición); coloca el puntero en la posición que le indiquemos.
    - Long lenght; devuelve la longitud de fichero en bytes.
    - Int skipBytes(int despl); desplaza el puntero desde la posición actual un numero de bytes que le indiquemos (despl).

# FICHEROS XML

* XML es un metalenguaje no es un lenguaje de programación que permite estructurar y jerarquizar la información describiendo los contenidos dentro del propio documento.
* Los ficheros XML son ficheros de texto que tienen unas marcas especiales utilizadas para delimitar la estructura del documento, de tal manera que cada etiqueta irá delimitada por los símbolos antes < y > y tendrá un nombre <nombre>. Además podrá contener o no una serie de atributos.
* Se pueden usar para:
  + Proporcionar datos a una base de datos.
  + Almacenar copias de parte del contenido de una base de datos.
  + Ficheros de configuración de algunos programas.
  + Protocolo SOAP(Simple Object Access Protocol), sirve para gestionar servidores en modo remoto.
* Para leer los ficheros XML en java y acceder a su contenido se usan:
  + Procesadores XML (Parsers): son independientes al del lenguaje de programación.
    - DOM (Document Object Model): se almacena toda la estructura del documento en memoria en forma de árbol, es decir, con nodos padre, nodos hijos y nodos finales. Una vez creado el árbol, se recorrerán los diferentes nodos y se verá a que tipo pertenece cada uno. Este procesamiento de la información requiere el uso de mucha memoria y tiempo sobre todo cuando los ficheros XML son grandes.
    - SAX(Simple API for XML): se lee el fichero XML de forma secuencial y se produce una serie de eventos en función de la lectura:
      * Inicio y fin del documento
      * Inicio y fin de la etiqueta
      * Etc.
      * Cada evento producido invoca a un método definido por el programador.
      * Este procesamiento consume muy poca memoria a diferencia del DOM, pero no se tiene una visión global del documento.

# Acceso a los ficheros XML mediante DOM

* Necesitaremos las clases, interfaces contenidas en:
  + org.w3c.dom
  + javax.xml.parsers
* Estas clases ofrecen métodos para cargar documentos XML desde una fuente de datos y contienen 2 clases fundamentales:
  + DocumentBuilderFactory.
  + DocumentBuilder.
* DOM no define ningún mecanismo para generar un fichero XML a partir de un árbol en memoria y para ello se utiliza el paquete:
  + java.xml.transform: Permite especificar una fuente y un resultado.
* Los programas que usen DOM necesitarán las siguientes interfaces:
  + Document: es un objeto que equivale a un documento xml y que permite crear nuevos nodos.
  + Element: es cada elemento del documento XML. Esta interfaz tendrá propiedades y métodos para manipular los elementos y los atributos de estos.
  + Node: cualquier nodo del documento.
  + NodeList: que contiene una lista de los nodos hijos de un nodo dado.
  + Attr: permite acceder a los atributos de un nodo.
  + Text: son los datos de tipo carácter de un elemento.
  + CharacterData: son los datos carácter del documento proporcionando atributos y métodos para manipular los datos de caracteres.
  + DocumentType: proporciona información contenida en la etiqueta <!DOCTYPE>

# Acceso ficheros con SAX

* Recorre de forma secuencial y jerárquica
* Es más complejo que DOM y funciona de la siguiente manera:
  + Al leer un documento XML se producen una serie de eventos que ocasionan la llamada a métodos.
  + Estos son:
    - startDocument(): cuando leemos la etiqueta de inicio del documento.
    - startElement(): cuando se lee la etiqueta de inicio de un elemento y a través de este método también podemos acceder a sus atributos.
    - characters(): cuando se leen los caracteres entre las etiquetas.
    - endElement(): cuando se lee la etiqueta de final del elemento.
    - endDocument(): cuando se llega al final del documento.
* Para tratar los eventos se realizan una serie de pasos:
  + Importar las clases e interfaces de SAX.
  + Crear un objeto procesador de XML. Puede producir SAXException.
  + Indicar al XMLReader que objetos tienen los métodos que tratarán los eventos.
    - Interfaces:
      * ContentHandler: recibe notificaciones de los eventos que ocurren en el documento
        + Métodos que implementan:

setContentHandler()

* + - * DTDHandler: recoge eventos relacionados de DTD.
        + Métodos que implementan:

setDTDHandler()

* + - * ErrorHandler: define métodos para el tratamiento de errores.
        + Métodos que implementan:

setErrorHandler()

* + - * EntityResolver: sus métodos se llaman cada vez que se hace referencia a una entidad.
        + Métodos que implementan:

setEntityResolver()

* + - * DefaultHandler: definir una implementación por defecto para todos los métodos.
        + Métodos que implementan:

setDefaultHandler()