

Instituto Tecnológico de Zacatecas

DUPLICADO DE UN DOMUNENTO PDF EXISTENTE

Manejo básico de archivos en Java

Hasta ahora todos los datos que creábamos en nuestros programas solamente existían durante la ejecución de los mismos. Cuando salíamos del programa, todo lo que habíamos generado se perdía. A veces nos interesaría que la vida de los datos fuera más allá que los programas que los generaron. Es decir, que al salir de un programa los datos generados quedaran guardados en algún lugar que permitiera su recuperación desde el mismo u otros programas. Por tanto, queríamos que dichos datos fueran **persistentes**. En este capítulo veremos el uso básico de archivos en Java para conseguir persistencia de datos. Para ello presentaremos conceptos básicos sobre archivos y algunas de las clases de la biblioteca estándar de Java para su creación y manipulación. Además, el uso de esas bibliotecas nos permitirá introducir algunos conceptos “avanzados” de programación como el uso de excepciones para tratar posibles errores durante la ejecución de un programa, o la manipulación de datos a bajo nivel, para transformar nuestros datos a vectores de bytes.

1. El concepto de archivo

Los programas usan variables para almacenar información: los datos de entrada, los resultados calculados y valores intermedios generados a lo largo del cálculo. Toda esta información es efímera: cuando acaba el programa, todo desaparece. Pero, para muchas aplicaciones, es importante poder almacenar datos de manera permanente.

Cuando se desea guardar información más allá del tiempo de ejecución de un programa lo habitual es organizar esa información en uno o varios ficheros almacenados en algún soporte de almacenamiento persistente. Otras posibilidades como el uso de bases de datos utilizan archivos como soporte para el almacenamiento de la información.

Los archivos desde el bajo nivel

Desde el punto de vista de más bajo nivel, podemos definir un archivo (o fichero) como:

Un conjunto de bits almacenados en un dispositivo, y accesible a través de un camino de acceso (pathname) que lo identifica.

Es decir, un conjunto de 0s y 1s que reside fuera de la memoria del ordenador, ya sea en el disco duro, un pendrive, un CD, entre otros.

Esa versión de bajo nivel, si bien es completamente cierta, desde el punto de vista de la programación de aplicaciones, es demasiado simple.

Por ello definiremos varios criterios para distinguir diferentes subcategorías de archivos. Estos tipos de archivos se diferenciarán desde el punto de vista de la programación ya que uno de ellos proporcionará diferentes funcionalidades (modos) para su manipulación.

El criterio del contenido

Sabemos que es diferente manipular números que Strings, aunque en el fondo ambos acaben siendo bits en la memoria del ordenador. Pero eso, cuando manipulamos archivos, distinguiremos dos clases de archivos dependiendo del tipo de datos que contienen:

- Los archivos de caracteres (o de texto).
- Los archivos de bytes (o binarios).

Un *fichero de texto* es aquél formado exclusivamente por caracteres y que, por tanto, puede crearse y visualizarse directamente en editores. Las operaciones de lectura y escritura trabajarán con caracteres. Por ejemplo, los ficheros con código java son ficheros de texto.

En cambio un *fichero binario* ya no está formado por caracteres sino que los bytes que contiene pueden representar otras cosas como números, imágenes, sonido, etc.

El criterio del modo de acceso

Existen dos modos básicos de acceso a la información contenida en un archivo:

- Secuencial
- Acceso directo

En el *modo secuencial* la información del archivo es una secuencia de bytes (o caracteres) de manera que para acceder al byte (o carácter) i -ésimo se ha de haber accedido anteriormente a los $i-1$ anteriores. Un ejemplo de acceso secuencial lo hemos visto con la clase `StringTokenizer`.

El modo de *acceso directo* nos permite acceder directamente a la información del byte i -ésimo. Un ejemplo muy conocido de acceso directo lo tenemos con los vectores (arrays).

2. Los archivos desde Java

En Java, los distintos tipos de ficheros se diferencian por las clases que usaremos para representarlos y manipularlos. Como las clases que usaremos pertenecen a la biblioteca estándar del lenguaje, su uso es

algo más complejo que las de las clases de la ACM, ya que su diseño se ha realizado pensando en su uso industrial. Las clases que usaremos para el tratamiento de errores están ubicadas en el paquete **java.io** por lo que deben ser importadas. Además, el código que trabaja con archivos debe considerar que muchas cosas pueden ir mal cuando se trabaja con ellos: el archivo puede estar corrupto, alguien ha desmontado el pendrive a medio ejecutar del programa, es un disco duro y esta ha caído, o no tiene más espacio para almacenar información, etc. Es por ello que, aunque de forma breve, deberemos introducir el mecanismo estándar en Java para tratar con los errores que pueden darse en nuestro programas: las excepciones.

Tratamiento de errores: las excepciones

Las excepciones son un mecanismo que permite a los métodos detectar que algo “anómalo” ha sucedido que interrumpe el correcto funcionamiento, de manera que quien los genera puede tratar la situación errónea. Decimos en este caso, que el método **lanza** (throw) una excepción. Cuando esto sucede, en vez de seguir con la ejecución normal de instrucciones, se busca hacia atrás en la secuencia de llamadas¹ si hay alguna que quiera **atraparla** (catch). Si ninguna de las llamadas decide atraparla, el programa acaba su ejecución y se informa al usuario del error que se ha producido la excepción y que nadie ha tratado.

Muchas de las excepciones que existen en Java, por ejemplo, dividir por 0, son **excepciones en tiempo de ejecución** (runtime exceptions) y no obligan a que el programador las trate explícitamente (claro que si el código no las trata y durante la ejecución del programa se producen, el programa finalizará con un “bonito” mensaje de error).

En Java, existe otro tipo de excepciones, las denominadas **excepciones comprobadas** (checked exceptions), que obligan al programador que dentro del código de un método invoca una instrucción que puede lanzarla a

- o bien atrapar dicha excepción (colocando dicha instrucción en un bloque **try-catch**)
- o bien, declarar en la cabecera del método que dicho método puede lanzar esa excepción (usando una declaración **throws**).

¹ El concepto de secuencia de llamadas, de hecho la **pila de llamadas**, fue presentado ya en el tema de recursividad cuando trazábamos la ejecución de un programa recursivo.

El objetivo es hacer que el programador no pueda “olvidarse” de tratar las muchas situaciones anómalas que se puedan producir durante la ejecución de un programa.

Tratamiento simplificado de excepciones

No es tema propio de esta asignatura profundizar en el manejo de excepciones, así que lo que veremos será un **tratamiento muy simplificado** de las mismas, haciendo lo mínimo para que Java funcione correctamente nuestros programas.

Para ello, el código que manipule los ficheros, tendrá la siguiente estructura:

```
1 try {
2   Código que abre y trata el fichero
3 } catch (IOException ex) {
4   Código que trata el error
5 }
```

La idea intuitiva de esta construcción es: intentar ejecutar las instrucciones y, en caso de producirse un error en el tratamiento de los ficheros (se ha lanzado una **IOException**), **atrapa** (catch) ese error y ejecuta el código de corrección. Nosotros simplificaremos el código de corrección y solamente escribiremos un mensaje.

Si en vez de tratar el error nosotros queremos indicar que nuestro método puede lanzar excepciones, en su cabecera pondremos:

```
1 public int methodThatCanThrow(params) throws IOException {
2
3   Código que trata ficheros pero no atrapa IOException
4
5 }
```

Como todos los ejemplos que veremos sobre ficheros utilizan estos mecanismos, no añadiremos aquí ejemplos de su utilización.

3. Lectura de ficheros secuenciales de texto

De cara a presentar la manipulación de ficheros secuenciales de texto, presentaremos un problema y su solución y, sobre la solución, comentaremos las operaciones que hemos usado.

Problema: contar apariciones de diversas letras

El problema consistirá en: dado un fichero de texto, contar el número de veces que aparecen una serie de letras en él.

Como siempre, lo difícil es la estrategia, en este caso:

- Supondremos que tanto el nombre del fichero, como los caracteres a considerar son constantes en el programa.
- Leeremos cada uno de los caracteres hasta llegar al final.
- Si está en los caracteres a considerar, incrementamos el contador asociado a ese carácter.

Como la parte de contar no tiene demasiado que ver con lo de los ficheros, lo mejor es separarlo en otra clase cuyo diseño e implementación quedará como ejercicio.

Con estas ideas, la solución del problema sería:

```

1 public class CountingVocals extends ConsoleProgram {
2
3     private static String FILE_NAME = "input.txt";
4     private static String VOCALS = "aeiou";
5
6     public void run() {
7         try {
8             CharCounter counters = new CharCounter(VOCALS, false);
9             FileReader input = new FileReader(FILE_NAME);
10            int c = input.read();
11            while ( c != -1 ) {
12                counters.countIfTargeted((char) c);
13                c = input.read();
14            }
15            input.close();
16            println(counters.toString());
17        } catch (IOException ex) {
18            println("Something bad has happened :-(");
19        }
20    }
21 }

```

Comentemos las líneas más relevantes:

- **3-4:** Definimos las constantes para el nombre de fichero y para las vocales a contar.
- **7 y 17-19:** Como cualquiera de las instrucciones que manipulan el fichero puede dar un error, encerramos todo el código del run en un bloque try-catch.
- **8:** creamos un contador para los caracteres del String dado (en este caso las vocales). El parámetro booleano indica que no queremos ignorar diferencias entre mayúsculas y minúsculas.
- **9:** creamos una instancia de **FileReader** para leer los caracteres del fichero. En este punto decimos que el fichero está abierto y preparado para que leamos caracteres de él.

- **10:** el método **read()** lee el siguiente carácter en el fichero de entrada. Para poder indicar que se ha llegado al final del fichero, en vez de devolver un carácter, devuelve un entero. Así puede usar el **valor -1 para indicar que no quedan más caracteres por leer.**
- **11-14:** mientras no hemos llegado al final del fichero, hemos de tratar el carácter actual y leer el siguiente.
- **12:** contamos el carácter actual (que sabemos que no es el final del fichero). Toda la parte de saber si es uno de los caracteres a considerar ya la hará la clase **CharCounter**.
- **13:** leemos el siguiente carácter en la entrada.
- **15:** después de haber tratado todo el fichero lo cerramos usando en método **close()**. Esto es especialmente importante cuando escribimos, pero mantener abiertos los ficheros que ya no necesitamos cuesta recursos al sistema.
- **16:** escribimos los contadores.

Sobre nombres de archivos, caminos de acceso y demás

Aunque a simple vista parezca una tontería, una de las cosas que más complica el código que trabaja sobre archivos no es la manipulación de su contenido sino la gestión de su nombre. El motivo es que cada sistema operativo usa convenciones diferentes para referirse a un nombre de fichero.

Por ejemplo, en sistemas tipo Unix tenemos:

`/User/jmgimeno/Prog2/FileExample/src/Main.java`

y en un sistema tipo Windows

`C:\User\jmgimeno\Prog2\FileExample\src\Main.java`

Así que hacer código que funcione independientemente del sistema es, cuando menos, tedioso.

Es por ello que, para simplificar, los nombres de ficheros que usaremos no contendrán camino alguno de acceso, lo que hará que estén ubicados en el directorio raíz del proyecto.

Si queréis aprender más sobre la manipulación de los nombres de fichero en java consultad la documentación de la clase la clase **java.io.File** que es la encargada de manipular nombres de archivo, rutas de acceso e incluso crear y listar directorios².

La declaración del fichero de entrada usando explícitamente la clase **File** sería:

² Y es un ejemplo de mala elección de nombre, ya que lo que trata son o bien los nombres de ficheros (debería llamarse **FileName**) o, desde un punto de vista de bajo nivel, las denominadas entradas del sistema de ficheros del sistema operativo (por lo que **FileEntry** también sería un buen nombre).


```
1 FileReader input = new FileReader(new File(FILE_NAME));
```

Otros métodos interesantes de FileReader

Si los buscáis están definidos en la clase `InputStreamReader` que es extendida por `FileReader`

- **int read(char[] buf, int offset, int length)**
Este método lee como máximo `length` caracteres del archivo y los coloca en el vector `buf` a partir de la posición `offset`. Devuelve el número de caracteres leídos o -1 indicando la finalización del archivo.
- **int read(char[] buf)**
Como la anterior pero usando como `offset` i `buf.length` como `length`.
- **String getEncoding()**
Devuelve el nombre del sistema de codificación usado para convertir los 0s y 1s del fichero en caracteres.

Sobre las codificaciones de caracteres

Un tema que también nos ocuparemos es el de las codificaciones usadas para representar los caracteres y que es otra de las grandes complicaciones existentes al tratar ficheros.

El problema es simple de enunciar:

- existen diversas maneras de asignar a un carácter³ un patrón de bits (que es lo que acaba siendo leído o escrito en un fichero)
- está claro que para que todo funcione correctamente, quién escribe un fichero y quien lo lee han de usar el mismo criterio

En Java existen varias clases para representar estas codificaciones, y versiones de los constructores de ficheros que permiten elegir la codificación a usar.

Nosotros no indicaremos codificación alguna y, si generamos los ficheros en la misma máquina que los consumimos, no deberíamos tener problema alguno.

La versión “avanzada” del mismo problema

En la solución anterior, hemos simplificado el tratamiento de los errores a lo mínimo que hay que hacer para lograr que el programa sea un programa Java correcto.

³ Similares problemas suceden en el caso de otros tipos de datos como `int`, `double`, etc.

Pero que el programa sea correcto no quiere decir que la solución sea perfecta. En este apartado os mostraré cómo sería el tratamiento de excepciones correcto y la forma idiomática en Java de hacer la lectura. Primero el programa:

```

2 public class CountingVocals extends ConsoleProgram {
3
4     private static final String FILE_NAME = "input.txt";
5     private static final String VOCALS = "aeiou",
6
7     public void runAdvanced() {
8         FileReader input = null;
9         try {
10             CharCounter counters = new CharCounter(VOCALS, false);
11             input = new FileReader(FILE_NAME);
12             int c;
13             while ( (c = input.read()) != -1 ) {
14                 counters.countIfTargeted(c);
15             }
16             println(counters);
17         } catch (FileNotFoundException ex) {
18             println("Problems opening " + FILE_NAME);
19         } catch (IOException ex) {
20             println("Problems reading " + FILE_NAME);
21         } finally {
22             try {
23                 input.close();
24             } catch (IOException ex) {
25                 println("Problems closing " + FILE_NAME);
26             }
27         }
28     }
29 }

```

4. Escritura de ficheros secuenciales de texto

Como veremos, los conceptos son similares, tan sólo cambia la clase de fichero (ahora es `FileWriter`) y, en vez de leer, escribimos.

Problema: dada una cadena de texto, escribirla al revés en un fichero

La estrategia en este caso es:

- Pedir una cadena al usuario.
- Recorrerla de atrás hacia delante e ir escribiendo en el fichero los caracteres que vamos encontrando
- Al final, cerrar el fichero.

Es decir:

```

1 public class BackwardsWriter extends ConsoleProgram {
2
3     private static String FILE_NAME = "Backwards.txt";
4
5     public void run() {
6         try {
7             String text = readLine("Enter text: ");
8             FileWriter output = new FileWriter(FILE_NAME);
9             for(int i=text.length()-1; i>=0; i--) {
10                 output.write(text.charAt(i));
11             }
12             output.close();
13         } catch (IOException ex) {
14             println("Something bad happened :-(");
15         }
16     }
17 }

```

Comentemos las líneas más relevantes:

- **6 y 13:** para simplificar usaremos la misma estructura de bloque try-catch que en el caso de la lectura.
- **8:** ahora para manipular el fichero usaremos una instancia de **FileWriter** (ya que escribiremos en él).
- **10:** aquí es dónde escribimos un nuevo carácter en el fichero usando en método **write(int)**. Recordad que siempre que me piden un int puedo usar un char.
- **12:** cierro el fichero (si no lo hacemos pudiera ser que algunos de los caracteres no se acabaran guardando en el fichero).

Otros métodos interesantes de FileWriter

- **new FileWriter(String name, boolean append)**
En caso de que ya existe un archivo de nombre name, si el booleano append es cierto, los nuevos caracteres se añadirán al fichero a partir del final. Si no, se creará el fichero vacío y se empezarán a añadir desde el principio.
- **void write(char[] cbuf, int off, int len)**
Escribe len caracteres del vector cbuf a partir de la posición off en el archivo.
- **void write(char[] cbuf)**
Como la anterior pero usando 0 como off y cbuf.length como len.
- **void write(String str, int off, int len)**
Igual que el anterior, pero en vez de un vector de caracteres tenemos un String.

- **void write(String str)**

Como la anterior pero usando 0 como off y str.length() como len.

5. El concepto de Buffering

El concepto de buffering queda muy bien explicado en el siguiente párrafo extraído del libro Head First Java:

Si no hubiera buffers, sería como comprar sin un carrito: de cada vez llevar los productos uno a uno hasta la caja. Los buffers te dan un lugar en el que dejar temporalmente las cosas hasta que está lleno. Por ello has de hacer menos viajes cuando usas el carrito.

Cualquier operación que implique acceder a memoria externa es cara y costosa, por lo que es interesante intentar reducir al máximo las operaciones de lectura/escritura que se realizan sobre los ficheros, haciendo que cada operación lea o escriba muchos caracteres. Además, eso también permite operar a un ritmo más alto, como el de leer una línea completa y devolverla en un solo paso.

Problema: crear un Howler a partir de un texto

Un Howler, en el universo de Harry Potter, no es más que un correo que chilla. Como chillar en internet es escribir en mayúsculas, lo que vamos a hacer es un programa tal que dado el texto de un mail (separado en líneas), lo “howlerice” y lo convierta en mayúsculas. Como siempre, la solución:

```

1 public class HowlerMaker extends ConsoleProgram {
2
3     private static String MAIL_NAME = "mail.txt";
4     private static String HOWLER_NAME = "howler.txt";
5
6     private String howlerize(String text) {
7         return text.toUpperCase();
8     }
9
10    public void run() {
11        try {
12            BufferedReader input =
13                new BufferedReader(new FileReader(MAIL_NAME));
14            BufferedWriter output =
15                new BufferedWriter(new FileWriter(HOWLER_NAME));
16            String line = input.readLine();
17            while (line != null) {
18                String howledLine = this.howlerize(line);
19                output.write(howledLine, 0, howledLine.length());
20                output.newLine();
21                line = input.readLine();

```



```

22     }
23     input.close();
24     output.close();
25 } catch (IOException ex) {
26     println("MAYbe you know who howled :-(");
27 }
28 }
29 }

```

Comentemos, como siempre, las líneas más relevantes:

- **12,13:** el constructor del `BufferedReader` en vez de recibir el nombre del fichero, recibe una instancia de `FileReader`. La idea es que la clase `BufferedReader` se centra en manejar un buffer de caracteres y cuando ha de leer cosas del fichero usa la instancia de `FileReader` para hacerlo⁴.
- **14,15:** equivalente para `BufferedWriter`.
- **16:** leemos una línea entera en `readLine()` String. En el String que nos devuelve, el marcador de fin de línea está eliminado para indicar que no hay más líneas devueltas.
- **19:** usamos una versión de `write` que nos permite escribir una cadena (indicando la posición del primer carácter y la longitud).
- **20:** para que la salida tenga los mismos saltos de línea, hemos de añadirlo usando el método `newLine` (recordad que `readLine` lo había eliminado de `line`, por lo que al pasarla a mayúsculas en `howledLine` no lo tiene).

El problema de los saltos de línea

Otro de los problemas al manipular ficheros de forma uniforme entre sistemas operativos es que éstos utilizan diversos caracteres para indicar el final de una línea.

Tal y como indica la documentación de `readLine`, un fin de línea puede estar indicado por:

- el carácter line-feed ('\n')
- el carácter carriage-return ('\r')
- el carácter carriage-return seguido inmediatamente de line-feed

El primer caso se usa en sistemas tipo Unix, el segundo en las versiones antiguas de MacOS y el último en sistemas tipo Windows.

El método `newLine` escribe el final de línea usando la convención del sistema operativo de la máquina en el que se está ejecutando.

⁴ Cuando el año que viene estudiéis el tema de la herencia, veréis que la clase `BufferedReader` puede usarse para hacer buffering de caracteres que vienen desde otros tipos de reader como son los que obtienen caracteres via comunicaciones en red.

6. Manipulación de datos a bajo nivel

Hasta ahora nuestros programas han estado manipulando números enteros, números en coma flotante, caracteres y strings y en ningún momento⁵ hemos comentado cómo estos datos están representados internamente (por ejemplo, cuánto valen los bits correspondientes a un determinado número entero).

Conocer esta representación nos servirá tanto a nivel conceptual para entender las diferencias entre números de texto y ficheros binarios, como en la práctica, para leer y escribir datos en formato binario.

Tamaños de los tipos primitivos en Java

Una de las ventajas de Java es que la representación de los datos no depende ni de la arquitectura de la máquina ni de su sistema operativo (es una de las ventajas de usar una máquina virtual). Por ello, lo que diremos funciona en cualquier máquina y sistema operativo.

El estándar de Java define los siguientes tamaños para los tipos de datos primitivos:

Tipo primitivo	Tamaño	Valor mínimo	Valor máximo
byte	8-bits	-128	127
char	16-bits	Unicode 0	Unicode 2 ¹⁶ -1
short	16-bits	-2 ¹⁵ (-32.768)	+2 ¹⁵ -1 (32.767)
int	32-bits	-2 ³¹ (-2.147.483.648)	+2 ³¹ -1 (2.147.483.647)
long	64-bits	-2 ⁶³	+2 ⁶³ -1
float	32-bits	32 bits IEEE-754	
double	64-bits	64 bits IEEE-754	
boolean	indefinido	true OR false	

Operaciones de Java a nivel de bits

Java proporciona operadores para realizar operaciones a nivel de bits para todos los **tipos integrales** (byte, char, short, int, long):

- **~a** (complemento) transforma los 0s en 1s y los 1s en 0s en la representación binaria. Por ejemplo, si el byte *b* contiene 00000001 (0x01), ~*b* será 11111110 (0xFE).
- **a&b** (AND) realiza la operación binaria AND bit a bit.

⁵ Salvo lo poco que hemos comentado respecto la conversión entre caracteres y números enteros.

- **a|b** (OR) realiza la operación binaria OR bit a bit.
 - **a^b** (XOR) realiza la operación binaria XOR (exclusivo) bit a bit
- También tenemos operaciones para hacer desplazamientos:

- **a<<n** (left shift) desplaza el patrón de bits n posiciones hacia la izquierda (rellena con ceros)
- **a>>n** (signed right shift) desplaza el patrón de bits n posiciones hacia la derecha (rellena con el bit de signo)
- **a>>>n** (unsigned right shift) desplaza el patrón de bits n posiciones hacia la derecha (rellena con ceros).

```

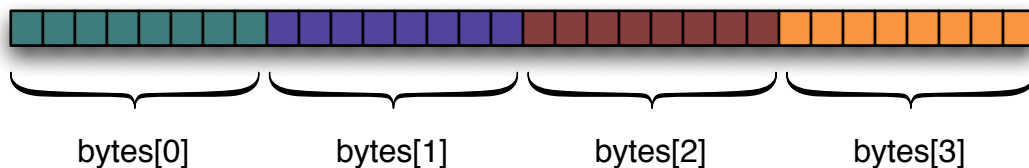
1
2 public class BitOperations extends ConsoleProgram {
3
4     private static String intToBinaryString(int n) {
5         // Integer.toBinaryString(n) adding zeros
6         String binary = "";
7         for (int i = 0; i < 32; i++) {
8             if ( n >= 0 ) {
9                 binary += "0";
10            } else {
11                binary += "1";
12            }
13            n = n << 1;
14        }
15        return binary;
16    }
17
18    private void printIntExpression(String expr, int value) {
19        println(expr + intToBinaryString(value) +
20            " " + String.valueOf(value));
21    }
22
23    public void run() {
24        setFont("Monospaced-40");
25        int a = 84;
26        int b = -29;
27        println("Operations using ints");
28        printIntExpression("    a = ", a);
29        printIntExpression("    b = ", b);
30        printIntExpression("    ~a = ", ~a);
31        printIntExpression("    ~b = ", ~b);
32        printIntExpression("    a&b = ", a&b);
33        printIntExpression("    a|b = ", a|b);
34        printIntExpression("    a^b = ", a^b);

```

⁶ Estos tres operadores tienen mucha letra pequeña en cuanto a su funcionamiento (*Java Language Specification*, §15.19).

[illegible]

Lo que queremos es conseguir una función tal que dado un entero nos lo convierta en un array formado por los cuatro bytes de su representación, es decir:



- **(byte):** convierte un entero a bytes. Como un entero tiene más de un byte, se queda con los ocho bits menos significativos. Es necesario hacer una conversión explícita ya que al convertir de `int` a `byte` podemos perder precisión.
- **b>>8:** desplazar 8 bits a la derecha, para hacer que el siguiente byte ocupe ahora la posición más a la derecha. Lo mismo sucede al desplazar con 16 y 26 para los siguientes bytes.

En código:

```

1 private static byte[] toByteArray(int n) {
2     byte[] bytes = new byte[4];
3     bytes[0] = (byte) (n >> 24);
4     bytes[1] = (byte) (n >> 16);
5     bytes[2] = (byte) (n >> 8);
6     bytes[3] = (byte) (n);
7     return bytes;
8 }

```

¿Cómo recuperar el entero a partir de un array de bytes?

Ahora nos basaremos en estas dos operaciones:

- **b & 0xFF**: que es hacer un AND con un número que solamente tiene 1s en los ocho bytes menos significativos. Así el resultado tiene todo ceros excepto en su parte menos significativa, que tiene los bits como b. Es necesario para las operaciones de desplazamiento, antes de desplazamiento se convierten el byte en int, lo que podría llenar de 1s el byte más significativo (debido a la **extensión del signo**).
- **b << 8**: esta operación desplaza hacia la derecha 8 bits el byte b. Como antes de hacer esa operación Java convierte el byte en int y los bytes tienen signo, es necesario poner todos los bits que no pertenecen a b a cero antes del desplazamiento. Lo mismo se hace con 16, 24 y 32.

```

1 private static int toInteger(byte[] bytes) {
2     int n = bytes[0] << 24;
3     n |= (bytes[1] & 0xFF) << 16;
4     n |= (bytes[2] & 0xFF) << 8;
5     n |= (bytes[3] & 0xFF);
6     return n;
7 }

```

Utilidades de empaquetamiento

Una vez entendido el caso sobre los enteros, vamos a ver la una clase cuyos métodos implementan esta transformación para otros tipos de datos básicos.

Para cada tipo de datos tendremos dos funciones: una para **empaquetar** los valores de ese tipo (packX) y otra para **desempaquetar** los valores del mismo (unpack).

En todos los casos, pasaremos como parámetros:

- un **byte[] buffer** para leer/escribir los datos
- un **int offset**, que indicará la posición inicial a partir de la que leeremos/escribiremos.

Por ejemplo, si realizamos la llamada:

```
PackUtils.packInt(416, buffer, 12);
```

se empaquetarán los 4 bytes correspondientes al entero 416 en las posiciones `buffer[12]`, `buffer[13]`, `buffer[14]` y `buffer[15]`.

Para simplificar, las funciones **no comprueban** si accedemos a posiciones correctas del vector.

Si ahora hiciéramos

```
int n = PackUtils.unpackInt(buffer, 12);
```

el valor de `n` debería de ser 416.

En los siguientes apartados comentaremos la estrategia de empaquetamiento y desempaquetamiento para cada tipo de datos.

Booleanos (boolean)

- **6-10:** Escribimos en `buffer[offset]` un 0 o un 1 dependiendo de si el booleano es cierto o falso.
- **14:** Al leer, devolvemos la comparación de si hay un 0 en la posición `byte[offset]`. En la comparación usamos `byte` para que no haya conversiones de tipo al hacer la comparación (ya que al comparar un `byte` con un `int`, convierte primero el `byte` a `int`). En este caso no hace falta, pero es buena costumbre hacerlo.

Caracteres (char)

- **20-21, 25-26:** Es la misma estrategia comentada con detalle anteriormente para los enteros, pero teniendo en cuenta que un `char` ocupa 2 bytes.

Cadenas (String) de longitud limitada

- **29, 44:** Como queremos poder controlar el final el tamaño, ya que necesitaremos tener todos los registros del tamaño, tendremos que limitar el tamaño máximo de las cadenas. Por ello, solamente guardaremos como máximo **maxLength** caracteres (lo mismo al leer).
- **33-41, 47-55:** La estrategia es empaquetar/desempaquetar los **maxLength** caracteres usando los métodos definidos para los mismos.
- **36-40:** Cuando a medio leer se nos acaba el String, guardamos un cero para marcar el final. Ese cero servirá para, al desempaquetar, parar de leer si la cadena tenía longitud menor que **maxLength**. Fijaos en el uso de la **sentencia break para salir del bucle**.
- **52-54:** Intentamos desempaquetar **maxLength** caracteres, pero si nos encontramos un carácter 0, podemos parar. También usamos un **break** para salir del bucle.

Enteros (int) y enteros largos (long)

- **62-65, 69-72:** La estrategia comentada anteriormente (en vez de varios |= he usado un solo |). Fijaos en la disposición del espacio para una mejor lectura del código.
- **78-85, 89-96:** Lo equivalente, pero para long, que ocupan 8 bytes (64 bits).

Coma flotante con doble precisión (double)

- **102-103, 108-109:** Usamos operaciones predefinidas que nos permiten pasar de double a long (o de long a double) y empaquetamos (o desempaquetamos) usando ese long.

```

1 public class PackUtils {
2
3     public static void packBoolean(boolean b,
4                                     byte[] buffer,
5                                     int offset) {
6         if (b) {
7             buffer[offset] = (byte) 1;
8         } else {
9             buffer[offset] = (byte) 0;
10        }
11    }
12
13    public static boolean unpackBoolean(byte[] buffer,
14                                       int offset) {
15        return buffer[offset] == (byte) 1;
16    }
17
18    public static void packChar(char c,
19                                byte[] buffer,
20                                int offset) {
21        buffer[offset] = (byte) (0xFF & (c >> 8));
22        buffer[offset + 1] = (byte) (0xFF & c);
23    }
24
25    public static char unpackChar(byte[] buffer, int offset) {
26        return (char) ((buffer[offset] << 8) |
27                       (buffer[offset + 1] & 0xFF));
28    }
29
30    public static void packLimitedString(String str,
31                                         int maxLength,
32                                         byte[] buffer,
33                                         int offset) {
34        for (int i = 0; i < maxLength; i++) {
35            if (i < str.length()) {
36                packChar(str.charAt(i), buffer, offset+2*i);
37            }
38        }
39    }
40
41    public static String unpackLimitedString(byte[] buffer,
42                                             int offset,
43                                             int maxLength) {
44        StringBuilder sb = new StringBuilder();
45        for (int i = 0; i < maxLength; i++) {
46            char c = unpackChar(buffer, offset+2*i);
47            sb.append(c);
48        }
49        return sb.toString();
50    }
51
52    public static void packLimitedLong(long l,
53                                       byte[] buffer,
54                                       int offset) {
55        for (int i = 0; i < 8; i++) {
56            buffer[offset+i] = (byte) (l >> (i*8));
57        }
58    }
59
60    public static long unpackLimitedLong(byte[] buffer,
61                                         int offset) {
62        long l = 0;
63        for (int i = 0; i < 8; i++) {
64            l = (l << 8) | (buffer[offset+i] & 0xFF);
65        }
66        return l;
67    }
68
69    public static void packLimitedDouble(double d,
70                                         byte[] buffer,
71                                         int offset) {
72        long l = Double.doubleToLongBits(d);
73        packLimitedLong(l, buffer, offset);
74    }
75
76    public static double unpackLimitedDouble(byte[] buffer,
77                                             int offset) {
78        long l = unpackLimitedLong(buffer, offset);
79        return Double.longBitsToDouble(l);
80    }
81
82    public static void packLimitedFloat(float f,
83                                        byte[] buffer,
84                                        int offset) {
85        long l = Float.floatToRawIntBits(f);
86        packLimitedLong(l, buffer, offset);
87    }
88
89    public static float unpackLimitedFloat(byte[] buffer,
90                                           int offset) {
91        long l = unpackLimitedLong(buffer, offset);
92        return Float.rawIntBitsToFloat(l);
93    }
94
95    public static void packLimitedInt(int i,
96                                     byte[] buffer,
97                                     int offset) {
98        for (int j = 0; j < 4; j++) {
99            buffer[offset+j] = (byte) (i >> (j*8));
100        }
101    }
102
103    public static int unpackLimitedInt(byte[] buffer,
104                                      int offset) {
105        int i = 0;
106        for (int j = 0; j < 4; j++) {
107            i = (i << 8) | (buffer[offset+j] & 0xFF);
108        }
109        return i;
110    }
111
112    public static void packLimitedShort(short s,
113                                       byte[] buffer,
114                                       int offset) {
115        for (int j = 0; j < 2; j++) {
116            buffer[offset+j] = (byte) (s >> (j*8));
117        }
118    }
119
120    public static short unpackLimitedShort(byte[] buffer,
121                                           int offset) {
122        short s = 0;
123        for (int j = 0; j < 2; j++) {
124            s = (s << 8) | (buffer[offset+j] & 0xFF);
125        }
126        return s;
127    }
128
129    public static void packLimitedByte(byte b,
130                                       byte[] buffer,
131                                       int offset) {
132        buffer[offset] = b;
133    }
134
135    public static byte unpackLimitedByte(byte[] buffer,
136                                         int offset) {
137        return buffer[offset];
138    }
139
140    public static void packLimitedChar(char c,
141                                       byte[] buffer,
142                                       int offset) {
143        buffer[offset] = (byte) c;
144    }
145
146    public static char unpackLimitedChar(byte[] buffer,
147                                         int offset) {
148        return (char) buffer[offset];
149    }
150
151    public static void packLimitedByteAndChar(byte b, char c,
152                                              byte[] buffer,
153                                              int offset) {
154        buffer[offset] = (byte) b;
155        buffer[offset+1] = (byte) c;
156    }
157
158    public static byte andChar unpackLimitedByteAndChar(byte[] buffer,
159                                                         int offset) {
160        byte b = buffer[offset];
161        char c = (char) buffer[offset+1];
162        return b and c;
163    }
164
165    public static void packLimitedCharAndByte(char c, byte b,
166                                              byte[] buffer,
167                                              int offset) {
168        buffer[offset] = (byte) c;
169        buffer[offset+1] = (byte) b;
170    }
171
172    public static char andByte unpackLimitedCharAndByte(byte[] buffer,
173                                                         int offset) {
174        char c = (char) buffer[offset];
175        byte b = buffer[offset+1];
176        return c and b;
177    }
178
179    public static void packLimitedShortAndChar(short s, char c,
180                                              byte[] buffer,
181                                              int offset) {
182        buffer[offset] = (byte) (s >> 8);
183        buffer[offset+1] = (byte) s;
184        buffer[offset+2] = (byte) c;
185    }
186
187    public static short andChar unpackLimitedShortAndChar(byte[] buffer,
188                                                           int offset) {
189        short s = 0;
190        s = (s << 8) | (buffer[offset] & 0xFF);
191        s = (s << 8) | (buffer[offset+1] & 0xFF);
192        char c = (char) buffer[offset+2];
193        return s and c;
194    }
195
196    public static void packLimitedCharAndShort(char c, short s,
197                                              byte[] buffer,
198                                              int offset) {
199        buffer[offset] = (byte) c;
200        buffer[offset+1] = (byte) (s >> 8);
201        buffer[offset+2] = (byte) s;
202    }
203
204    public static char andShort unpackLimitedCharAndShort(byte[] buffer,
205                                                           int offset) {
206        char c = (char) buffer[offset];
207        short s = 0;
208        s = (s << 8) | (buffer[offset+1] & 0xFF);
209        s = (s << 8) | (buffer[offset+2] & 0xFF);
210        return c and s;
211    }
212
213    public static void packLimitedIntAndChar(int i, char c,
214                                              byte[] buffer,
215                                              int offset) {
216        buffer[offset] = (byte) (i >> 24);
217        buffer[offset+1] = (byte) (i >> 16);
218        buffer[offset+2] = (byte) (i >> 8);
219        buffer[offset+3] = (byte) i;
220        buffer[offset+4] = (byte) c;
221    }
222
223    public static int andChar unpackLimitedIntAndChar(byte[] buffer,
224                                                         int offset) {
225        int i = 0;
226        i = (i << 8) | (buffer[offset] & 0xFF);
227        i = (i << 8) | (buffer[offset+1] & 0xFF);
228        i = (i << 8) | (buffer[offset+2] & 0xFF);
229        i = (i << 8) | (buffer[offset+3] & 0xFF);
230        char c = (char) buffer[offset+4];
231        return i and c;
232    }
233
234    public static void packLimitedCharAndInt(char c, int i,
235                                              byte[] buffer,
236                                              int offset) {
237        buffer[offset] = (byte) c;
238        buffer[offset+1] = (byte) (i >> 24);
239        buffer[offset+2] = (byte) (i >> 16);
240        buffer[offset+3] = (byte) (i >> 8);
241        buffer[offset+4] = (byte) i;
242    }
243
244    public static char andInt unpackLimitedCharAndInt(byte[] buffer,
245                                                         int offset) {
246        char c = (char) buffer[offset];
247        int i = 0;
248        i = (i << 8) | (buffer[offset+1] & 0xFF);
249        i = (i << 8) | (buffer[offset+2] & 0xFF);
250        i = (i << 8) | (buffer[offset+3] & 0xFF);
251        i = (i << 8) | (buffer[offset+4] & 0xFF);
252        return c and i;
253    }
254
255    public static void packLimitedLongAndChar(long l, char c,
256                                              byte[] buffer,
257                                              int offset) {
258        for (int j = 0; j < 7; j++) {
259            buffer[offset+j] = (byte) (l >> (j*8));
260        }
261        buffer[offset+7] = (byte) c;
262    }
263
264    public static long andChar unpackLimitedLongAndChar(byte[] buffer,
265                                                         int offset) {
266        long l = 0;
267        for (int j = 0; j < 7; j++) {
268            l = (l << 8) | (buffer[offset+j] & 0xFF);
269        }
270        char c = (char) buffer[offset+7];
271        return l and c;
272    }
273
274    public static void packLimitedCharAndLong(char c, long l,
275                                              byte[] buffer,
276                                              int offset) {
277        buffer[offset] = (byte) c;
278        for (int j = 1; j < 8; j++) {
279            buffer[offset+j] = (byte) (l >> ((j-1)*8));
280        }
281    }
282
283    public static char andLong unpackLimitedCharAndLong(byte[] buffer,
284                                                         int offset) {
285        char c = (char) buffer[offset];
286        long l = 0;
287        for (int j = 1; j < 8; j++) {
288            l = (l << 8) | (buffer[offset+j] & 0xFF);
289        }
290        return c and l;
291    }
292
293    public static void packLimitedDoubleAndChar(double d, char c,
294                                              byte[] buffer,
295                                              int offset) {
296        long l = Double.doubleToLongBits(d);
297        packLimitedLongAndChar(l, c, buffer, offset);
298    }
299
300    public static double andChar unpackLimitedDoubleAndChar(byte[] buffer,
301                                                             int offset) {
302        long l = unpackLimitedLongAndChar(buffer, offset);
303        return Double.longBitsToDouble(l);
304    }
305
306    public static void packLimitedFloatAndChar(float f, char c,
307                                              byte[] buffer,
308                                              int offset) {
309        long l = Float.floatToRawIntBits(f);
310        packLimitedLongAndChar(l, c, buffer, offset);
311    }
312
313    public static float andChar unpackLimitedFloatAndChar(byte[] buffer,
314                                                           int offset) {
315        long l = unpackLimitedLongAndChar(buffer, offset);
316        return Float.rawIntBitsToFloat(l);
317    }
318
319    public static void packLimitedIntAndShort(int i, short s,
320                                              byte[] buffer,
321                                              int offset) {
322        buffer[offset] = (byte) (i >> 24);
323        buffer[offset+1] = (byte) (i >> 16);
324        buffer[offset+2] = (byte) (i >> 8);
325        buffer[offset+3] = (byte) i;
326        buffer[offset+4] = (byte) (s >> 8);
327        buffer[offset+5] = (byte) s;
328    }
329
330    public static int andShort unpackLimitedIntAndShort(byte[] buffer,
331                                                         int offset) {
332        int i = 0;
333        i = (i << 8) | (buffer[offset] & 0xFF);
334        i = (i << 8) | (buffer[offset+1] & 0xFF);
335        i = (i << 8) | (buffer[offset+2] & 0xFF);
336        i = (i << 8) | (buffer[offset+3] & 0xFF);
337        short s = 0;
338        s = (s << 8) | (buffer[offset+4] & 0xFF);
339        s = (s << 8) | (buffer[offset+5] & 0xFF);
340        return i and s;
341    }
342
343    public static void packLimitedShortAndInt(short s, int i,
344                                              byte[] buffer,
345                                              int offset) {
346        buffer[offset] = (byte) (s >> 8);
347        buffer[offset+1] = (byte) s;
348        buffer[offset+2] = (byte) (i >> 24);
349        buffer[offset+3] = (byte) (i >> 16);
350        buffer[offset+4] = (byte) (i >> 8);
351        buffer[offset+5] = (byte) i;
352    }
353
354    public static short andInt unpackLimitedShortAndInt(byte[] buffer,
355                                                         int offset) {
356        short s = 0;
357        s = (s << 8) | (buffer[offset] & 0xFF);
358        s = (s << 8) | (buffer[offset+1] & 0xFF);
359        int i = 0;
360        i = (i << 8) | (buffer[offset+2] & 0xFF);
361        i = (i << 8) | (buffer[offset+3] & 0xFF);
362        i = (i << 8) | (buffer[offset+4] & 0xFF);
363        i = (i << 8) | (buffer[offset+5] & 0xFF);
364        return s and i;
365    }
366
367    public static void packLimitedLongAndShort(long l, short s,
368                                              byte[] buffer,
369                                              int offset) {
370        for (int j = 0; j < 6; j++) {
371            buffer[offset+j] = (byte) (l >> (j*8));
372        }
373        buffer[offset+6] = (byte) (s >> 8);
374        buffer[offset+7] = (byte) s;
375    }
376
377    public static long andShort unpackLimitedLongAndShort(byte[] buffer,
378                                                           int offset) {
379        long l = 0;
380        for (int j = 0; j < 6; j++) {
381            l = (l << 8) | (buffer[offset+j] & 0xFF);
382        }
383        short s = 0;
384        s = (s << 8) | (buffer[offset+6] & 0xFF);
385        s = (s << 8) | (buffer[offset+7] & 0xFF);
386        return l and s;
387    }
388
389    public static void packLimitedCharAndLongAndChar(char c, long l, char c2,
390                                              byte[] buffer,
391                                              int offset) {
392        buffer[offset] = (byte) c;
393        for (int j = 1; j < 8; j++) {
394            buffer[offset+j] = (byte) (l >> ((j-1)*8));
395        }
396        buffer[offset+8] = (byte) c2;
397    }
398
399    public static char andLong andChar unpackLimitedCharAndLongAndChar(byte[] buffer,
400                                                                           int offset) {
401        char c = (char) buffer[offset];
402        long l = 0;
403        for (int j = 1; j < 8; j++) {
404            l = (l << 8) | (buffer[offset+j] & 0xFF);
405        }
406        char c2 = (char) buffer[offset+8];
407        return c and l and c2;
408    }
409
410    public static void packLimitedLongAndCharAndChar(long l, char c, char c2,
411                                              byte[] buffer,
412                                              int offset) {
413        for (int j = 0; j < 7; j++) {
414            buffer[offset+j] = (byte) (l >> (j*8));
415        }
416        buffer[offset+7] = (byte) c;
417        buffer[offset+8] = (byte) c2;
418    }
419
420    public static long andChar andChar unpackLimitedLongAndCharAndChar(byte[] buffer,
421                                                                           int offset) {
422        long l = 0;
423        for (int j = 0; j < 7; j++) {
424            l = (l << 8) | (buffer[offset+j] & 0xFF);
425        }
426        char c = (char) buffer[offset+7];
427        char c2 = (char) buffer[offset+8];
428        return l and c and c2;
429    }
430
431    public static void packLimitedIntAndShortAndChar(int i, short s, char c,
432                                              byte[] buffer,
433                                              int offset) {
434        buffer[offset] = (byte) (i >> 24);
435        buffer[offset+1] = (byte) (i >> 16);
436        buffer[offset+2] = (byte) (i >> 8);
437        buffer[offset+3] = (byte) i;
438        buffer[offset+4] = (byte) (s >> 8);
439        buffer[offset+5] = (byte) s;
440        buffer[offset+6] = (byte) c;
441    }
442
443    public static int andShort andChar unpackLimitedIntAndShortAndChar(byte[] buffer,
444                                                                           int offset) {
445        int i = 0;
446        i = (i << 8) | (buffer[offset] & 0xFF);
447        i = (i << 8) | (buffer[offset+1] & 0xFF);
448        i = (i << 8) | (buffer[offset+2] & 0xFF);
449        i = (i << 8) | (buffer[offset+3] & 0xFF);
450        short s = 0;
451        s = (s << 8) | (buffer[offset+4] & 0xFF);
452        s = (s << 8) | (buffer[offset+5] & 0xFF);
453        char c = (char) buffer[offset+6];
454        return i and s and c;
455    }
456
457    public static void packLimitedShortAndIntAndChar(short s, int i, char c,
458                                              byte[] buffer,
459                                              int offset) {
460        buffer[offset] = (byte) (s >> 8);
461        buffer[offset+1] = (byte) s;
462        buffer[offset+2] = (byte) (i >> 24);
463        buffer[offset+3] = (byte) (i >> 16);
464        buffer[offset+4] = (byte) (i >> 8);
465        buffer[offset+5] = (byte) i;
466        buffer[offset+6] = (byte) c;
467    }
468
469    public static short andInt andChar unpackLimitedShortAndIntAndChar(byte[] buffer,
470                                                                           int offset) {
471        short s = 0;
472        s = (s << 8) | (buffer[offset] & 0xFF);
473        s = (s << 8) | (buffer[offset+1] & 0xFF);
474        int i = 0;
475        i = (i << 8) | (buffer[offset+2] & 0xFF);
476        i = (i << 8) | (buffer[offset+3] & 0xFF);
477        i = (i << 8) | (buffer[offset+4] & 0xFF);
478        i = (i << 8) | (buffer[offset+5] & 0xFF);
479        char c = (char) buffer[offset+6];
480        return s and i and c;
481    }
482
483    public static void packLimitedLongAndShortAndChar(long l, short s, char c,
484                                              byte[] buffer,
485                                              int offset) {
486        for (int j = 0; j < 6; j++) {
487            buffer[offset+j] = (byte) (l >> (j*8));
488        }
489        buffer[offset+6] = (byte) (s >> 8);
490        buffer[offset+7] = (byte) s;
491        buffer[offset+8] = (byte) c;
492    }
493
494    public static long andShort andChar unpackLimitedLongAndShortAndChar(byte[] buffer,
495                                                                           int offset) {
496        long l = 0;
497        for (int j = 0; j < 6; j++) {
498            l = (l << 8) | (buffer[offset+j] & 0xFF);
499        }
500        short s = 0;
501        s = (s << 8) | (buffer[offset+6] & 0xFF);
502        s = (s << 8) | (buffer[offset+7] & 0xFF);
503        char c = (char) buffer[offset+8];
504        return l and s and c;
505    }
506
507    public static void packLimitedCharAndLongAndShortAndChar(char c, long l, short s, char c2,
508                                              byte[] buffer,
509                                              int offset) {
510        buffer[offset] = (byte) c;
511        for (int j = 1; j < 7; j++) {
512            buffer[offset+j] = (byte) (l >> ((j-1)*8));
513        }
514        buffer[offset+7] = (byte) (s >> 8);
515        buffer[offset+8] = (byte) s;
516        buffer[offset+9] = (byte) c2;
517    }
518
519    public static char andLong andShort andChar unpackLimitedCharAndLongAndShortAndChar(byte[] buffer,
520                                                                           int offset) {
521        char c = (char) buffer[offset];
522        long l = 0;
523        for (int j = 1; j < 7; j++) {
524            l = (l << 8) | (buffer[offset+j] & 0xFF);
525        }
526        short s = 0;
527        s = (s << 8) | (buffer[offset+7] & 0xFF);
528        s = (s << 8) | (buffer[offset+8] & 0xFF);
529        char c2 = (char) buffer[offset+9];
530        return c and l and s and c2;
531    }
532
533    public static void packLimitedLongAndCharAndShortAndChar(long l, char c, short s, char c2,
534                                              byte[] buffer,
535                                              int offset) {
536        for (int j = 0; j < 6; j++) {
537            buffer[offset+j] = (byte) (l >> (j*8));
538        }
539        buffer[offset+6] = (byte) c;
540        buffer[offset+7] = (byte) (s >> 8);
541        buffer[offset+8] = (byte) s;
542        buffer[offset+9] = (byte) c2;
543    }
544
545    public static long andChar andShort andChar unpackLimitedLongAndCharAndShortAndChar(byte[] buffer,
546                                                                           int offset) {
547        long l = 0;
548        for (int j = 0; j < 6; j++) {
549            l = (l << 8) | (buffer[offset+j] & 0xFF);
550        }
551        char c = (char) buffer[offset+6];
552        short s = 0;
553        s = (s << 8) | (buffer[offset+7] & 0xFF);
554        s = (s << 8) | (buffer[offset+8] & 0xFF);
555        char c2 = (char) buffer[offset+9];
556        return l and c and s and c2;
557    }
558
559    public static void packLimitedIntAndShortAndCharAndChar(int i, short s, char c, char c2,
560                                              byte[] buffer,
561                                              int offset) {
562        buffer[offset] = (byte) (i >> 24);
563        buffer[offset+1] = (byte) (i >> 16);
564        buffer[offset+2] = (byte) (i >> 8);
565        buffer[offset+3] = (byte) i;
566        buffer[offset+4] = (byte) (s >> 8);
567        buffer[offset+5] = (byte) s;
568        buffer[offset+6] = (byte) c;
569        buffer[offset+7] = (byte) c2;
570    }
571
572    public static int andShort andChar andChar unpackLimitedIntAndShortAndCharAndChar(byte[] buffer,
573                                                                           int offset) {
574        int i = 0;
575        i = (i << 8) | (buffer[offset] & 0xFF);
576        i = (i << 8) | (buffer[offset+1] & 0xFF);
577        i = (i << 8) | (buffer[offset+2] & 0xFF);
578        i = (i << 8) | (buffer[offset+3] & 0xFF);
579        short s = 0;
580        s = (s << 8) | (buffer[offset+4] & 0xFF);
581        s = (s << 8) | (buffer[offset+5] & 0xFF);
582        char c = (char) buffer[offset+6];
583        char c2 = (char) buffer[offset+7];
584        return i and s and c and c2;
585    }
586
587    public static void packLimitedShortAndIntAndCharAndChar(short s, int i, char c, char c2,
588                                              byte[] buffer,
589                                              int offset) {
590        buffer[offset] = (byte) (s >> 8);
591        buffer[offset+1] = (byte) s;
592        buffer[offset+2] = (byte) (i >> 24);
593        buffer[offset+3] = (byte) (i >> 16);
594        buffer[offset+4] = (byte) (i >> 8);
595        buffer[offset+5] = (byte) i;
596        buffer[offset+6] = (byte) c;
597        buffer[offset+7] = (byte) c2;
598    }
599
600    public static short andInt andChar andChar unpackLimitedShortAndIntAndCharAndChar(byte[] buffer,
601                                                                           int offset) {
602        short s = 0;
603        s = (s << 8) | (buffer[offset] & 0xFF);
604        s = (s << 8) | (buffer[offset+1] & 0xFF);
605        int i = 0;
606        i = (i << 8) | (buffer[offset+2] & 0xFF);
607        i = (i << 8) | (buffer[offset+3] & 0xFF);
608        i = (i << 8) | (buffer[offset+4] & 0xFF);
609        i = (i << 8) | (buffer[offset+5] & 0xFF);
610        char c = (char) buffer[offset+6];
611        char c2 = (char) buffer[offset+7];
612        return s and i and c and c2;
613    }
614
615    public static void packLimitedLongAndShortAndCharAndChar(long l, short s, char c, char c2,
616                                              byte[] buffer,
617                                              int offset) {
618        for (int j = 0; j < 6; j++) {
619            buffer[offset+j] = (byte) (l >> (j*8));
620        }
621        buffer[offset+6] = (byte) (s >> 8);
622        buffer[offset+7] = (byte) s;
623        buffer[offset+8] = (byte) c;
624        buffer[offset+9] = (byte) c2;
625    }
626
627    public static long andShort andChar andChar unpackLimitedLongAndShortAndCharAndChar(byte[] buffer,
628                                                                           int offset) {
629        long l = 0;
630        for (int j = 0; j < 6; j++) {
631            l = (l << 8) | (buffer[offset+j] & 0xFF);
632        }
633        short s = 0;
634        s = (s << 8) | (buffer[offset+6] & 0xFF);
635        s = (s << 8) | (buffer[offset+7] & 0xFF);
636        char c = (char) buffer[offset+8];
637        char c2 = (char) buffer[offset+9];
638        return l and s and c and c2;
639    }
640
641    public static void packLimitedCharAndLongAndShortAndCharAndChar(char c, long l, short s, char c2, char c3,
642                                              byte[] buffer,
643                                              int offset) {
644        buffer[offset] = (byte) c;
645        for (int j = 1; j < 7; j++) {
646            buffer[offset+j] = (byte) (l >> ((j-1)*8));
647        }
648        buffer[offset+7] = (byte) (s >> 8);
649        buffer[offset+8] = (byte) s;
650        buffer[offset+9] = (byte) c2;
651        buffer[offset+10] = (byte) c3;
652    }
653
654    public static char andLong andShort andChar andChar unpackLimitedCharAndLongAndShortAndCharAndChar(byte[] buffer,
655                                                                           int offset) {
656        char c = (char) buffer[offset];
657        long l = 0;
658        for (int j = 1; j < 7; j++) {
659            l = (l << 8) | (buffer[offset+j] & 0xFF);
660        }
661        short s = 0;
662        s = (s << 8) | (buffer[offset+7] & 0xFF);
663        s = (s << 8) | (buffer[offset+8] & 0xFF);
664        char c2 = (char) buffer[offset+9];
665        char c3 = (char) buffer[offset+10];
666        return c and l and s and c2 and c3;
667    }
668
669    public static void packLimitedLongAndCharAndShortAndCharAndChar(long l, char c, short s, char c2, char c3,
670                                              byte[] buffer,
671                                              int offset) {
672        for (int j = 0; j < 6; j++) {
673            buffer[offset+j] = (byte) (l >> (j*8));
674        }
675        buffer[offset+6] = (byte) c;
676        buffer[offset+7] = (byte) (s >> 8);
677        buffer[offset+8] = (byte) s;
678        buffer[offset+9] = (byte) c2;
679        buffer[offset+10] = (byte) c3;
680    }
681
682    public static long andChar andShort andChar andChar unpackLimitedLongAndCharAndShortAndCharAndChar(byte[] buffer,
683                                                                           int offset) {
684        long l = 0;
685        for (int j = 0; j < 6; j++) {
686            l = (l << 8) | (buffer[offset+j] & 0xFF);
687        }
688        char c = (char) buffer[offset+6];
689        short s = 0;
690        s = (s << 8) | (buffer[offset+7] & 0xFF);
691        s = (s << 8) | (buffer[offset+8] & 0xFF);
692        char c2 = (char) buffer[offset+9];
693        char c3 = (char) buffer[offset+10];
694        return l and c and s and c2 and c3;
695    }
696
697    public static void packLimitedIntAndShortAndCharAndCharAndChar(int i, short s, char c, char c2, char c3,
698                                              byte[] buffer,
699                                              int offset) {
700        buffer[offset] = (byte) (i >> 24);
701        buffer[offset+1] = (byte) (i >> 16);
702        buffer[offset+2] = (byte) (i >> 8);
703        buffer[offset+3] = (byte) i;
704        buffer[offset+4] = (byte) (s >> 8);
705        buffer[offset+5] = (byte) s;
706        buffer[offset+6] = (byte) c;
707        buffer[offset+7] = (byte) c2;
708        buffer[offset+8] = (byte) c3;
709    }
710
711    public static int andShort andChar andChar andChar unpackLimitedIntAndShortAndCharAndCharAndChar(byte[] buffer,
712                                                                           int offset) {
713        int i = 0;
714        i = (i << 8) | (buffer[offset] & 0xFF);
715        i = (i << 8) | (buffer[offset+1] & 0xFF);
716        i = (i << 8) | (buffer[offset+2] & 0xFF);
717        i = (i << 8) | (buffer[offset+3] & 0xFF);
718        short s = 0;
719        s = (s << 8) | (buffer[offset+4] & 0xFF);
720        s = (s << 8) | (buffer[offset+5] & 0xFF);
721        char c = (char) buffer[offset+6];
722        char c2 = (char) buffer[offset+7];
723        char c3 = (char) buffer[offset+8];
724        return i and s and c and c2 and c3;
725    }
726
727    public static void packLimitedShortAndIntAndCharAndCharAndChar(short s, int i, char c, char c2, char c3,
728                                              byte[] buffer,
729                                              int offset) {
730        buffer[offset] = (byte) (s >> 8);
731        buffer[offset+1] = (byte) s;
732        buffer[offset+2] = (byte) (i >> 24);
733        buffer[offset+3] = (byte) (i >> 16);
734        buffer[offset+4] = (byte) (i >> 8);
735        buffer[offset+5] = (byte) i;
736        buffer[offset+6] = (byte) c;
737        buffer[offset+7] = (byte) c2;
738        buffer[offset+8] = (byte) c3;
739    }
740
741    public static short andInt andChar andChar andChar unpackLimitedShortAndIntAndCharAndCharAndChar(byte[] buffer,
742                                                                           int offset) {
743        short s = 0;
744        s = (s << 8) | (buffer[offset] & 0xFF);
745        s = (s << 8) | (buffer[offset+1] & 0xFF);
746        int i = 0;
747        i = (i << 8) | (buffer[offset+2] & 0xFF);
748        i = (i << 8) | (buffer[offset+3] & 0xFF);
749        i = (i << 8) | (buffer[offset+4] & 0xFF);
750        i = (i << 8) | (buffer[offset+5] & 0xFF);
751        char c = (char) buffer[offset+6];
752        char c2 = (char) buffer[offset+7];
753        char c3 = (char) buffer[offset+8];
754        return s and i and c and c2 and c3;
755    }
756
757    public static void packLimitedLongAndShortAndCharAndCharAndChar(long l, short s, char c, char c2, char c3,
758                                              byte[] buffer,
759                                              int offset) {
760        for (int j = 0; j < 6; j++) {
761            buffer[offset+j] = (byte) (l >> (j*8));
762        }
763        buffer[offset+6] = (byte) (s >> 8);
764        buffer[offset+7] = (byte) s;
765        buffer[offset+8] = (byte) c;
766        buffer[offset+9] = (byte) c2;
767        buffer[offset+10] = (byte) c3;
768    }
769
770    public static long andShort andChar andChar andChar unpackLimitedLongAndShortAndCharAndCharAndChar(byte[] buffer,
771                                                                           int offset) {
772        long l = 0;
773        for (int j = 0; j < 6; j++) {
774            l = (l << 8) | (buffer[offset+j] & 0xFF);
775        }
776        short s = 0;
777        s = (s << 8) | (buffer[offset+6] & 0xFF);
778        s = (s << 8) | (buffer[offset+7] & 0xFF);
779        char c = (char) buffer[offset+8];
780        char c2 = (char) buffer[offset+9];
781        char c3 = (char) buffer[offset+10];
782        return l and s and c and c2 and c3;
783    }
784
785    public static void packLimitedCharAndLongAndShortAndCharAndCharAndChar(char c, long l, short s, char c2, char c3, char c4,
786                                              byte[] buffer,
787                                              int offset) {
788        buffer[offset] = (byte) c;
789        for (int j = 1; j < 7; j++) {
790            buffer[offset+j] = (byte) (l >> ((j-1)*8));
791        }
792        buffer[offset+7] = (byte) (s >> 8);
793        buffer[offset+8] = (byte) s;
794        buffer[offset+9] = (byte) c2;
795        buffer[offset+
```

```

36     } else {
37         // We mark with a zero
38         packChar('\0', buffer, offset+2);
39         break;
40     }
41 }
42 }
43
44 public static String unpackChar(int maxLength, byte[] buffer,
45                                 int offset) {
46     String result = "";
47     for (int i = 0; i < maxLength; i++) {
48         char c = unpackChar(buffer, offset+i);
49         if ( c != '\0' ) {
50             result += c;
51         } else {
52             break;
53         }
54     }
55     return result;
56 }
57
58 public static void packInt(int n,
59                             byte[] buffer,
60                             int offset ) {
61     buffer[offset] = (byte) (n >> 24);
62     buffer[offset + 1] = (byte) (n >> 16);
63     buffer[offset + 2] = (byte) (n >> 8);
64     buffer[offset + 3] = (byte) n;
65 }
66
67 public static int unpackInt(byte[] buffer, int offset) {
68     return ((buffer[offset] << 24) |
69            ((buffer[offset + 1] & 0xFF) << 16) |
70            ((buffer[offset + 2] & 0xFF) << 8) |
71            (buffer[offset + 3] & 0xFF));
72 }
73
74 public static void packLong(long n,
75                             byte[] buffer,
76                             int offset) {
77     buffer[offset] = (byte) (n >> 56);
78     buffer[offset + 1] = (byte) (n >> 48);
79     buffer[offset + 2] = (byte) (n >> 40);
80     buffer[offset + 3] = (byte) (n >> 32);
81     buffer[offset + 4] = (byte) (n >> 24);
82     buffer[offset + 5] = (byte) (n >> 16);
83     buffer[offset + 6] = (byte) (n >> 8);
84 }

```

```

85     buffer[offset + 7] = (byte) n ;
86 }
87
88 public static long unpackLong(byte[] buffer, int offset) {
89     return ((long)(buffer[offset] & 0xFF) << 48) |
90            ((long)(buffer[offset + 1] & 0xFF) << 40) |
91            ((long)(buffer[offset + 2] & 0xFF) << 32) |
92            ((long)(buffer[offset + 3] & 0xFF) << 24) |
93            ((long)(buffer[offset + 4] & 0xFF) << 16) |
94            ((long)(buffer[offset + 5] & 0xFF) << 8) |
95            ((long)(buffer[offset + 6] & 0xFF) << 0) |
96            ((long)(buffer[offset + 7] & 0xFF) << 0) ;
97 }
98
99 public static void packDouble(double n,
100                                byte[] buffer,
101                                int offset) {
102     long bits = Double.doubleToLongBits(n);
103     packLong(bits, buffer, offset);
104 }
105
106 public static double unpackDouble(byte[] buffer,
107                                    int offset) {
108     long bits = unpackLong(buffer, offset);
109     return Double.longBitsToDouble(bits);
110 }
111 }

```

7. Archivos binarios de acceso directo

Los archivos de acceso directo están representados por la clase `java.io.RandomAccessFile`, que permite:

- Abrir un archivo en el que se pueda solamente leer (modo “r”) como tanto leer como escribir (modo “rw”).
- Para leer disponemos de las operaciones:
 - **int read(byte[] buff, int off, int len)**
 - **int read(byte[] buff)**
- Para escribir disponemos de las operaciones:
 - **void write(byte[] buff, int off, int len)**
 - **void write(byte[] buff)**
- La operación que caracteriza a este tipo de archivos es:
 - **void seek(long pos)**

que coloca la posición de lectura/escritura en el byte que ocupa la posición pos del archivo.

Así, la siguiente operación de lectura, en vez de usar la posición dejada por la última operación, usará dicha posición como la del

primer byte a leer.

Puede colocarse más allá del final del fichero, pero el tamaño del fichero no aumentará hasta que se haya realizado una operación de escritura.

- Otros métodos relevantes de la clase son:
 - **long length()**, que devuelve el número de bytes ocupados por el archivo.
 - **void setLength(long newLength)**, que define la nueva longitud del archivo como newLength. En caso de ser menor que la longitud actual, el contenido del fichero es truncado. Por ejemplo, setLength(0) hace que las subsiguientes operaciones de escritura se realicen sobre un archivo vacío.
 - **long getFilePointer()**, que devuelve el valor de la posición de lectura/escritura del archivo. Puede ser útil para saber, p.e. estamos intentando leer más allá del último registro del archivo.
- Aunque también provee de operaciones de transformación de tipos básicos a vectores de bytes y viceversa, nosotros usaremos las que hemos definido en la clase PackingUtils, ya que usar las predefinidas obligaría entrar en algunos detalles⁷ que se escapan al contenido del curso.

Uso típico de archivos binarios de acceso directo

Las posibilidades de

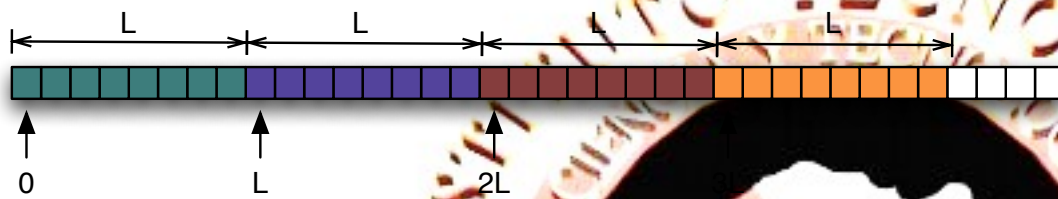
- mezclar operaciones de escritura con operaciones de lectura
- acceder a una posición concreta del archivo

hacen que el uso principal de los archivos de acceso directo sea implementar algo muy parecido a los arrays, pero en memoria secundaria.

El concepto de registro

Si queremos guardar en un archivo de acceso directo los datos correspondientes a las instancias de una clase para poder acceder directamente a cada una de las instancias, deberemos hacer que todas ellas tengan **igual longitud**. De esta manera, si cada instancia tiene longitud L, la instancia i-ésima ocupará L bytes a partir del i*L. Gráficamente:

⁷ Para conocerlos, consultad la documentación de la clase RandomAccessFile y de las interficies asociadas DataInput y DataOutput.



A cada uno de los bloques de bytes que representan los datos de una instancia se le denomina **registro**.

Ejemplo: un archivo de acceso directo de personas

Vamos a mostrar todo lo que hemos comentado sobre archivos de acceso directo sobre un ejemplo concreto: un archivo para representar personas.

La clase que representa los datos

La primera clase que veremos es la que representa una persona. Esta clase contiene:

- campos con información de la persona
- operaciones sobre personas: básicamente getters y el método toString
- operaciones para:
 - dada una persona, obtener un array de bytes
 - dado un array de bytes, construir la persona

Comentarios sobre las líneas destacables:

- **3-6:** declaramos los campos que tendrá cada instancia de Person
- **8:** como los registros han de ser de longitud fija, limitamos la longitud del String name a NAME_LIMIT caracteres. De hecho la limitación afectará solamente cuando leamos/escribamos los datos.
- **9:** en la constante SIZE, calculamos el tamaño que tendrá cada registro asociado a una persona en función de los campos a guardar. Como el tamaño ha de poderse conocer desde fuera, hacemos la constante pública y, **para evitar que se pueda asignar otro valor, la definimos como final.**
- **42-54:** este método, crea un array de bytes de tamaño SIZE y va empaquetando cada uno de los campos a partir de los offsets que le corresponden.
- **56-68:** operación inversa que desempaqueta el array de bytes que recibe como parámetro y crea una instancia con los valores obtenidos. Fijaos en que es un método estático, ya que claramente se refiere a cosas relacionadas con la clase Person, pero no se aplica sobre ninguna instancia (de hecho, es el propio método quién crea una instancia).

```
1 public class Person {
2
3     private long id;
4     private String name;
5     private int age;
6     private boolean married;
7
8     private static final int NAME_LIMIT = 20;
9     public static final int SIZE = 8 + 2 * NAME_LIMIT + 1;
10
11     public Person(long id,
12                  String name,
13                  int age,
14                  boolean married) {
15         this.id = id;
16         this.name = name;
17         this.age = age;
18         this.married = married;
19     }
20
21     public int getAge() {
22         return age;
23     }
24
25     public long getId() {
26         return id;
27     }
28
29     public boolean isMarried() {
30         return married;
31     }
32
33     public String getName() {
34         return name;
35     }
36
37     public String toString() {
38         return "Person{" + "id=" + id + " name=" + name +
39             " age=" + age + " married=" + married + '}';
40     }
41
42     public byte[] toBytes() {
43         byte[] record = new byte[SIZE];
44         int offset = 0;
45         PackUtils.packLong(id, record, offset);
46         offset += 8;
47         PackUtils.packLimitedString(name, NAME_LIMIT,
48                                     record, offset);
49         offset += 2 * NAME_LIMIT;
```



```

50     PackUtils.packInt(age, record, offset);
51     offset += 4;
52     PackUtils.packBoolean(married, record, offset);
53     return record;
54 }
55
56 public static Person fromBytes(byte[] record) {
57     int offset = 0;
58     long id = PackUtils.unpackLong(record, offset);
59     offset += 8;
60     String name = PackUtils.unpackLimitedString(NAME_LIMIT,
61                                                  record,
62                                                  offset);
63     offset += 2 * NAME_LIMIT;
64     int age = PackUtils.unpackInt(record, offset);
65     offset += 4;
66     boolean married = PackUtils.unpackBoolean(record, offset);
67     return new Person(id, name, age, married);
68 }
69
70 }

```

El programa principal

En el programa principal lo que haremos es:

- declarar una referencia al archivo de acceso directo
- crear varias personas
- escribir y leer en diversas posiciones del archivo, indexando por registro.

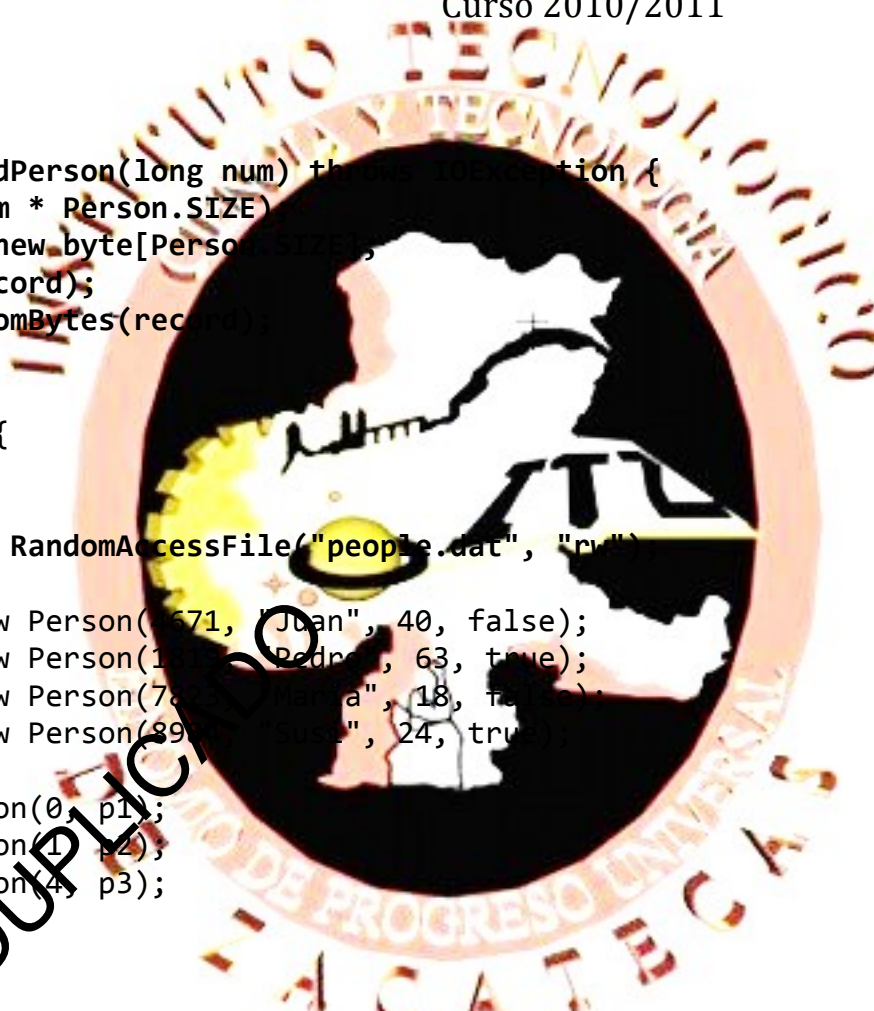
Comentarios de las líneas relevantes:

- **3:** declaramos el archivo de acceso aleatorio como un campo de la clase, así todos los métodos no estáticos de la misma lo podrán utilizar.
- **5-10:** escribe en el archivo el registro con posición num formado con los datos de person.
- **12-17:** lee el registro de la posición num del archivo y crea una instancia de Person con los bytes obtenidos.
- **19-56:** escribimos y leemos en diferentes posiciones del archivo.

```

1 public class Main extends ConsoleProgram {
2
3     private RandomAccessFile raf;
4
5     private void writePerson(long num, Person person)
6                                     throws IOException {
7         this.raf.seek(num * Person.SIZE);
8         byte[] record = person.toBytes();
9         this.raf.write(record);

```



```

10  }
11
12  private Person readPerson(long num) throws IOException {
13      this.raf.seek(num * Person.SIZE);
14      byte[] record = new byte[Person.SIZE];
15      this.raf.read(record);
16      return Person.fromBytes(record);
17  }
18
19  public void run() {
20      try {
21
22          this.raf = new RandomAccessFile("people.dat", "rw");
23
24          Person p1 = new Person(1571, "Juan", 40, false);
25          Person p2 = new Person(1234, "Pedro", 63, true);
26          Person p3 = new Person(7890, "Ana", 18, false);
27          Person p4 = new Person(8901, "Carlos", 24, true);
28
29          this.writePerson(0, p1);
30          this.writePerson(1, p2);
31          this.writePerson(4, p3);
32
33          Person p;
34
35          p = this.readPerson(0);
36          println("p = " + p);
37
38          p = this.readPerson(1);
39          println("p = " + p);
40
41          p = this.readPerson(4);
42          println("p = " + p);
43
44          this.writePerson(3, p4);
45          p = this.readPerson(3);
46          println("p = " + p);
47
48          this.writePerson(1, p1);
49          p = this.readPerson(1);
50          println("p = " + p);
51
52      } catch (IOException e) {
53          println("Algo muy malo ha pasado :-(");
54      }
55  }
56 }

```

La ejecución del programa muestra:

```
p = Person{id=4671 name=Juan age=40 married=false}
p = Person{id=1819 name=Pedro age=60 married=true}
p = Person{id=7823 name=María age=30 married=false}
p = Person{id=8984 name=Susi age=20 married=true}
p = Person{id=4671 name=Juan age=40 married=false}
```

Visualizando el contenido del fichero

Un archivo binario no lo podemos visualizar con un editor de texto. Para ver su contenido, podemos usar la herramienta UNIX `hexdump`, que nos muestra los valores de los bytes del fichero. Si lo aplicamos al fichero generado por el programa anterior, obtenemos:

```
CleoBook >: hexdump -v people.dat
00000000 00 00 00 00 00 00 12 3f 00 4a 00 75 00 61 00 6e
00000100 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000200 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000300 00 00 00 28 00 00 00 00 00 00 00 12 3f 00 4a 00
00000400 75 00 61 00 6e 00 00 00 00 00 00 00 00 00 00 00
00000500 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000600 00 00 00 00 00 00 00 28 00 00 00 00 00 00 00 00
00000700 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000800 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000900 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000a00 00 00 00 00 00 23 18 00 53 00 75 00 73 00 69 00
00000b00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000c00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000d00 00 00 18 01 00 00 00 00 00 00 1e 8f 00 4d 00 61
00000e00 00 72 00 ed 00 61 00 00 00 00 00 00 00 00 00 00
00000f00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00001000 00 00 00 00 00 00 00 12 00
00001009
```

Comentarios:

- En **amarillo** he marcado los 8 bytes correspondientes al id, que es un long. En el caso del primer registro (que empieza en la posición 0), el valor es 0x123f, que podéis comprobar que es lo mismo que 4671.
- En **azul** están los 40 bytes correspondientes a name, dónde cada par de bytes corresponde a un carácter. 0x4a corresponde al carácter 'J', 0x75 a 'u', etc.
- En **rojo** los 4 bytes correspondientes a age. Podéis comprobar que 0x28 es 40.

- En **verde** está el byte correspondiente al booleano married que, valiendo 0x00, es falso.
- En **magenta** he marcado los bytes correspondientes al registro 2, que en ningún momento se ha escrito, por lo que **un registro que no contiene datos es un error**.

8. Ordenación de archivos: el algoritmo MergeSort

Un procedimiento muy habitual a realizar sobre archivos es ordenarlos. Aunque para el caso de los archivos de acceso directo podríamos aplicar los que ya conocemos para ordenar vectores (arrays), no es la mejor forma de hacerlo, ya que las lecturas/escrituras no secuenciales en archivos son muy costosas. El algoritmo MergeSort permite ordenar un archivo, **solamente** realizando lecturas y escrituras secuenciales, por lo que es muy eficiente en el uso de los archivos. En este apartado veremos, en primer lugar, la descripción del algoritmo⁸ y, posteriormente, su implementación para ordenar las líneas de un archivo de texto.

La idea básica del algoritmo

¿Os acordáis de la película de Los Inmortales? En ella se explica que los inmortales, sienten una atracción irresistible de acabar unos con otros y que al final **solamente puede quedar uno**.

La idea detrás del algoritmo MergeSort es muy simple: ir fusionando las partes ordenadas que contiene el fichero, hasta que todo él está en una única parte.

Para ello se basa en dos métodos auxiliares:

- **split**: que dado un fichero de entrada y dos de salida, va colocando las subsecuencias ordenadas del fichero de entrada en cada uno de los de salida.
- **merge**: que dados dos ficheros de entrada y uno de salida, fusiona dos subcadenas, una de cada fichero de entrada, y las guarda en el fichero de salida.

Ambas operaciones se sucederán hasta que el fichero que queremos dividir conste de una única secuencia.

⁸ Una versión que mejora mucho el rendimiento del algoritmo, se plantea en la lista de problemas.

MergeSort sobre un vector

Para visualizar el funcionamiento del algoritmo, lo visualizaremos trabajando sobre la ordenación de un vector. Para hacerlo supongamos que deseamos ordenar el siguiente vector:

15	18	7	9	3	14	6
----	----	---	---	---	----	---

Si marcamos, alternando dos colores, las subsecuencias del vector que ya están ordenadas, tenemos:

15	18	7	9	3	14	6
----	----	---	---	---	----	---

Separando (**split**) por colores en dos vectores auxiliares, obtenemos:

15	18	3	14
7	9	6	

Si fusionamos (**merge**) ordenadamente ambos vectores (los recorremos a la vez y dejando pasar primero al elemento más pequeño de los dos), queda:

7	9	6	15	18	3	14
---	---	---	----	----	---	----

Si repetimos el mismo procedimiento hasta que solamente haya una subsecuencia, obtenemos:

7	9	6	15	18	3	14
7	9	3	14			
6	15	18				
6	7	9	3	14	15	18
6	7	9				
3	14	15	18			
3	6	7	9	14	15	18

Que, como podemos ver, deja el vector ordenado.

Aplicación sobre archivos

Vamos a aplicar la idea del algoritmo para ordenar los números de un archivo de texto. El resultado lo dejaremos en el fichero de texto original.

```

1 public class MergeSort extends ConsoleProgram {
2
3     private static final String INPUT = "input.txt";
4     private static final String AUX1 = "aux1.txt";
5     private static final String AUX2 = "aux2.txt";
6
7     public void run() {
8         try {
9             boolean sorted = split(INPUT, AUX1, AUX2);
10            while (!sorted) {
11                merge(AUX1, AUX2, INPUT);
12                sorted = split(INPUT, AUX1, AUX2);
13            }
14            println("Yataaaa!!!");
15        } catch (IOException ex) {
16            println("Some error has happened");
17        }
18    }
19
20    private boolean split(String input,
21                          String output1,
22                          String output2) throws IOException {
23
24        BufferedReader in = new BufferedReader(
25            new FileReader(input));
26        BufferedWriter out = new BufferedWriter(
27            new FileWriter(output1));
28        BufferedWriter other = new BufferedWriter(
29            new FileWriter(output2));
30
31        boolean sorted = true;
32        String previous = "";
33        String current = in.readLine();
34
35        while (current != null) {
36            if (previous.compareTo(current) > 0) {
37                sorted = false;
38                BufferedWriter tmp = out;
39                out = other;
40                other = tmp;
41            }
42            out.write(current);
43            out.newLine();

```



```
44     previous = current;
45     current = in.readLine();
46 }
47
48 in.close();
49 out.close();
50 other.close();
51
52 return sorted;
53 }
54
55 private void merge(String input1,
56                   String input2,
57                   String output) throws IOException {
58
59     BufferedReader in1 = new BufferedReader(
60         new InputStreamReader(input1));
61     BufferedReader in2 = new BufferedReader(
62         new InputStreamReader(input2));
63     BufferedWriter out = new BufferedWriter(
64         new FileWriter(output));
65
66     String current1 = in1.readLine();
67     String current2 = in2.readLine();
68
69     while (current1 != null && current2 != null) {
70         if (current1.compareTo(current2) <= 0) {
71             out.write(current1);
72             out.newLine();
73             current1 = in1.readLine();
74         } else {
75             out.write(current2);
76             out.newLine();
77             current2 = in2.readLine();
78         }
79     }
80
81     while (current1 != null) {
82         out.write(current1);
83         out.newLine();
84         current1 = in1.readLine();
85     }
86
87     while (current2 != null) {
88         out.write(current2);
89         out.newLine();
90         current2 = in2.readLine();
91     }
92 }
```

```

93     in1.close();
94     in2.close();
95     out.close();
96 }
97 }

```

Algunos comentarios sobre el código:

- **7-18:** el programa principal es el que realiza la ordenación. Inicialmente hace un split y merge, si el fichero no está ordenado hace un merge seguido de un split.
- Fijaros en que el fichero de entrada para split es de salida para merge, y los de salida para split son los de entrada para merge.
- recordad que split, mientras encuentra valores ordenados los va guardando en el mismo fichero, para ello guardamos dos valores:
 - previous, que es el valor que se encuentra guardado
 - current, el valor que se va leyendo actualmente
- **32:** inicializamos previous a la cadena vacía ya que es el menor que cualquier otra cadena, por lo que siempre será menor que lo que leamos forma una subsecuencia ordenada.
- **38-40:** cuando hemos de cambiar de fichero lo que hago es intercambiar las referencias out y other, y siempre escribo en out.
- **31,37:** inicialmente supongo que el fichero de entrada está ordenado, pero si alguna vez detecto un par de elementos consecutivos que están desordenados sé que no lo está.
- **69-79:** si hay elementos en ambos ficheros de entrada, se comparan, y el menor se copia en el de salida.
- **81-85, 87-91:** cuando uno de los ficheros acaba, copiamos los que quedan en el otro en el de salida.

9. Bibliografía

- Eric S.Roberts, The Art & Science of Java, Addison-Wesley (2008).
- [The Java Tutorials](#) (última consulta, 30 de mayo de 2011).
- [The Java Language Specification \(Third edition\)](#) (última consulta, 30 mayo 2011).
- Kathy Sierra & Bert Bates, Head First Java, O'Reilly (2003).