Arquitectura de sistemas

Abelardo Pardo

University of Sydney
School of Electrical and Information Engineering
NSW, 2006, Australia
Autor principal del curso de 2009 a 2012

Iria Estévez Ayres

Damaris Fuentes Lorenzo

Pablo Basanta Val

Pedro J. Muñoz Merino

Hugo A. Parada

Derick Leony

Universidad Carlos III de Madrid
Departamento de Ingeniería Telemática
Avenida Universidad 30, E28911 Leganés (Madrid), España

Capítulo 9. Lectura y escritura de ficheros

Tabla de contenidos

9.1. Introducción

C ofrece un conjunto de funciones para realizar operaciones de entrada y salida (E/S) con las cuales puedes leer y escribir cualquier tipo de fichero. Antes de ver estas funciones, hay que entender algunos conceptos básicos.

9.1.1. Qué es un fichero

En C, un *fichero* se puede referir a un fichero en disco, a un terminal, a una impresora, etc. Dicho de otro modo, un fichero representa un dispositivo concreto con el que puedes intercambiar información. Antes de poder trabajar con un fichero, tienes que abrir ese fichero. Tras terminar el intercambio de información, debes cerrar el fichero que abriste.

9.1.2. Qué es un flujo o *stream*

El conjunto datos que transfieres desde un programa a un fichero, o al revés, se le conoce como *flujo* (*stream* en inglés), y consiste en una serie de bytes (o caracteres). A diferencia de un fichero, que se refiere a un dispositivo concreto de E/S, un flujo es independiente del dispositivo. Es decir, todos los flujos tienen el mismo comportamiento independientemente del dispositivo al que esté asociado. De esta manera puedes realizar operaciones E/S con tan sólo asociar un flujo a un fichero.

Hay dos tipos de flujos. Uno es el *flujo de texto*, consistente en líneas de texto. Una línea es una secuencia de caracteres terminada en el carácter de línea nueva (`'\n'` ó `'\r\n'`). El otro formato es el del *flujo binario*, que consiste en una secuencia de bytes que representan datos internos como números, estructuras o arrays. Se utiliza principalmente para datos que no son de tipo texto, donde no importa la apariencia de esos datos en el fichero (da igual que no se vean "en bonito", como podría verse en un fichero de texto).

9.1.3. E/S mediante buffers

Una porción de memoria que se usa temporalmente para almacenar datos antes de ser enviados a su destino se llama *buffer*. Con ayuda de los buffers, el sistema operativo puede mejorar su eficiencia reduciendo el número de accesos a efectuar en un dispositivo de E/S (es decir, en un fichero).

El acceso a disco o a otros dispositivos de E/S suele ser más lento que a memoria de acceso directo; por eso, si varias operaciones E/S se realizan en un buffer en vez de en el fichero en sí, el funcionamiento de un programa es más eficiente. Por ejemplo, si se envían varias operaciones de escritura a un buffer, las modificaciones de esos datos escritos se quedan en memoria hasta que es hora de guardarlas, que es cuando esos datos se llevan al dispositivo real (a este proceso se le conoce como *flushing*).

Por defecto, todas las operaciones de los flujos E/S en C son con buffer.

9.1.4. Comportamiento del modelo E/S

Ya que la unidad más pequeña que se puede representar en C es un caracter, se puede acceder a un fichero desde cualquiera de sus caracteres (o bytes, que es lo mismo). Se puede leer o escribir cualquier número de caracteres desde un punto móvil, conocido como el *indicador de posición de fichero*. Los caracteres se leen o escriben en secuencia desde ese punto, y el indicador se va moviendo de acuerdo a eso. El indicador se coloca al principio del fichero cuando éste se abre, pero luego se puede mover a cualquier punto explícitamente (no sólo porque se haya leído o escrito algo).

9.2. Funciones básicas

Vamos a centrarnos en cómo abrir y cerrar ficheros y en cómo interpretar los mensajes de error que pueden dar estas dos operaciones.

9.2.1. Punteros a `FILE`

La estructura `FILE` es la estructura que controla los ficheros y se encuentra en la cabecera `stdio.h`. Los flujos utilizan estos punteros a ficheros para realizar las operaciones E/S. La siguiente línea de código declara una variable de tipo puntero a fichero:

```
FILE *ptr_fich;
```

En la estructura `FILE` es donde se encuentra el atributo que representa el indicador de posición de fichero.

9.2.2. Abriendo un fichero

La función `fopen` abre un fichero y lo asocia a un flujo. Necesitas especificar como argumentos el nombre de ese fichero (o la ruta y el nombre) y el modo de apertura.

```
#include <stdio.h>
FILE *fopen(const char *filename, const char *mode);
```

Aquí `filename` es un puntero a `char` que referencia un string con el nombre del fichero. `mode` apunta a otro string que especifica la manera en la que se va a abrir el fichero. La función `fopen` devuelve un puntero de tipo `FILE`. Si ha ocurrido algún error al abrirse, devuelve `NULL`.

El parámetro `mode` es una combinación de los caracteres `r` (*read*, lectura), `w` (*write*, escritura), `b` (binario), `a` (*append*, añadir), y `+` (actualizar). Si usas el caracter `a` y el fichero existe, el contenido de ese fichero se mantiene y los datos nuevos se añaden justo al final, tras el último dato que ya estuviera escrito (el indicador de fichero pues no está situado al inicio del fichero, sino que en este

caso está al final). Si el fichero no existe, lo crea. Con `w` es diferente; este modo siempre borra los datos del fichero si existe (si no, crea uno nuevo). Si al modo le añades el `+`, permite que el fichero se abra para escritura o lectura y puedes modificar cualquier dato que hubiera en él. Si usas `r`, el fichero debe existir; si no, `fopen` dará error y te devolverá `NULL`.

La siguiente lista muestra los posibles modos de abrir un fichero:

- "`r`" abre un fichero de texto existente para lectura.
- "`w`" crea un fichero de texto para escritura.
- "`a`" abre un fichero de texto existente para añadir datos.
- "`r+`" abre un fichero de texto existente para lectura o escritura.
- "`w+`" crea un fichero de texto para lectura o escritura.
- "`a+`" abre o crea un fichero de texto para añadirle datos.
- "`rb`" abre un fichero binario existente para lectura.
- "`wb`" crea un fichero binario para escritura.
- "`ab`" abre un fichero binario existente para añadir datos.
- "`r+b`" abre un fichero binario existente para lectura o escritura.
- "`w+b`" crea un fichero binario para lectura o escritura.
- "`a+b`" abre o crea un fichero binario para añadirle datos.

A veces podrás ver que el modo está escrito como "`rb+`" en vez de "`r+b`", por ejemplo, pero es equivalente.

Responde a las siguientes preguntas para ver si has entendido lo que se explica en este documento:

- ¿Qué hace la siguiente expresión?:

```
fopen("text.bin", r+b);
```

  - Abre un fichero existente para lectura y escritura.
  - Crea un fichero binario para lectura y escritura.
  - Abre un fichero existente para lectura y añadir datos al final.
  - Crea un fichero binario para lectura y añadir datos al final.
  - Ninguna de las anteriores.

9.2.3. Cerrando un fichero

Después de abrir un fichero y leerlo o escribir en él, hay que desenlazarlo del flujo de datos al que fue asociado. Esto se hace con la función `fclose`, cuya sintaxis es la siguiente:

```
#include <stdio.h>
int fclose(FILE *stream);
```

Si `fclose` cierra bien el fichero, devuelve `0`. Si no, la función devuelve `EOF`. Normalmente sólo falla

si se ha borrado el fichero antes de intentar cerrarlo. Acuérdate siempre de cerrar el fichero. Si no lo haces, se pueden perder los datos almacenados en él. Además, si no lo cierras, otros programas pueden tener problemas si lo quieren abrir más tarde.

El siguiente programa abre y cierra un fichero de texto, comprobando los posibles errores que puedan surgir:

```
1  #include <stdio.h>
2
3  enum {EXITO, FALLO};
4  int main(void)
5  {
6    FILE *ptr_fichero;
7    char nombre_fichero[] = "resumen.txt";
8    int resultado = EXITO;
9
10   if ( (ptr_fichero = fopen(nombre_fichero, "r") ) == NULL)
11   {
12     printf("No se ha podido abrir el fichero %s.\n", nombre_fichero);
13     resultado = FALLO;
14   }
15   else
16   {
17     printf("Abierto fichero; listos para cerrarlo.\n");
18     if (fclose(ptr_fichero)!=0)
19     {
20       printf("No se ha podido cerrar el fichero %s.\n", nombre_fichero);
21       resultado = FALLO;
22     }
23   }
24   return resultado;
25 }
```

9.2.4. Abriendo un fichero con un descriptor

Para poder abrir un fichero dado su descriptor de fichero (*file descriptor (fd)* en inglés) se ha de usar la función `fdopen`. Esta función se comporta como la función `fopen`, pero en vez de abrir un fichero dado su nombre, usa el identificador de fichero para abrirlo:

```
#include <stdio.h>
int fdopen(int fildes, const char *mode);
```

Una forma de conseguir un descriptor para un fichero es con la función `mkstemp`:

```
#include <stdio.h>
int mkstemp(char *template);
```

Esta función genera un nombre de fichero temporal único a partir del string `template`. Los últimos seis caracteres de `template` deben ser "XXXXXX", que son reemplazados luego por `mkstemp` con una cadena que hace que el nombre no esté repetido. Ya que será modificada, `template` no debe ser definida como constante. La función `mkstemp` devuelve el descriptor de fichero del fichero temporal creado o `-1` en caso de error.

9.3. Lectura y escritura de ficheros

En C puedes realizar las operaciones de lectura y escritura de varias maneras:

- Leer y escribir carácter a carácter (o lo que es lo mismo, byte a byte), con funciones como `fgetc` y `fputc`.
- Leer y escribir línea a línea, con funciones como `fgets` y `fputs`.
- Leer y escribir un bloque de caracteres (o bytes) cada vez, con `fread` y `fwrite`.

Para este curso nos vamos a centrar en la última, la lectura por bloques, que nos será útil tanto para ficheros de texto como para binarios.

9.3.1. Lectura/Escritura por bloques

En C contamos con las funciones `fread` y `fwrite` para realizar operaciones E/S por bloques.

La sintaxis de `fread` es la siguiente:

```
#include <stdio.h>
size_t fread(void *ptr, size_t size, size_t n, FILE *stream);
```

Aquí `ptr` es un array donde se van a almacenar los datos que se van a leer. `size` indica el tamaño de cada elemento del array. `n` especifica el número de elementos que se van a leer. `stream` es un puntero a un fichero que ha tenido que haber sido abierto previamente. `size_t` es un tipo que está definido también en la cabecera `stdio.h`. La función devuelve el número de elementos que realmente se han leído (que puede ser menor que `n`); si todo ha ido bien o aún no se ha llegado al final del fichero, ese número devuelto debería ser mayor que 0 e igual o menor (en caso de que no sepamos la longitud de lo que vamos a leer) que el tercer argumento, `n`.

Si `fread` no devuelve el resultado esperado, ***habrá que distinguir si es porque llegamos al final del fichero o por un error***; para ello contamos con las funciones `feof` y `ferror` para saber exactamente qué pasó, que veremos a continuación.

La sintaxis de `fwrite` es la siguiente:

```
#include <stdio.h>
size_t fwrite(const void *ptr, size_t size, size_t n, FILE *stream);
```

`ptr` hace referencia a un array con los datos que se van a escribir al fichero abierto apuntado por `stream`. `size` indica el tamaño de cada elemento del array. `n` especifica el número de elementos que se van a escribir. `stream` es un puntero a un fichero que ha tenido que haber sido abierto previamente. La función devuelve el número de elementos que se han escrito; luego si no ha habido ningún error, ese número devuelto debería ser igual que el tercer argumento, `n`. El valor devuelto puede ser menos que este `n` si ha habido algún error.

Como se ha comentado anteriormente, tenemos la función `feof`, que se usa para saber cuándo hemos llegado al final del fichero:

```
#include <stdio.h>
int feof(FILE *stream);
```

Esta función devuelve `0` si aún no se ha llegado al final del fichero. Si se ha llegado, devuelve un entero distinto de cero.

Finalmente, contamos con la función `ferror` para saber si ha ocurrido algún error en la lectura o escritura, que devolverá `0` si no ha ocurrido ninguno y un valor distinto de cero en caso contrario:

```
#include <stdio.h>
int ferror(FILE *stream);
```

El siguiente programa escribe en bloque un array de números y luego los lee de uno en uno para calcular su suma.

```
1  #include <stdio.h>
2  #define TAM 6
3
4  int main(void)
5  {
6    FILE *fichero;
7    int numeros[TAM] = {20, 20, 20, 20, 20, 20};
8    char nombre_fichero[] = "numeros.bin";
9    //Esta variable irá recogiendo los resultados de lectura/escritura
10   size_t resultado;
11
12   /* Abrimos para escritura en bloque*/
13   fichero = fopen(nombre_fichero, "w");
14   if (fichero == NULL)
15   {
16    printf("El fichero no se ha podido abrir para escritura.\n");
17    return -1;
18   }
19   // Se escribe en bloque los elementos del vector.
20   resultado = fwrite(numeros, sizeof(int), TAM, fichero);
21   if (resultado!=TAM)
22   {
23    printf("No se han escrito todos los %d números del array.\n", TAM);
24   }
25
26   if (fclose(fichero)!=0)
27   {
28     printf("No se ha podido cerrar el fichero.\n");
29  return -1;
30   }
31
32   /* Abrimos para lectura */
33   int suma = 0;
34   fichero = fopen(nombre_fichero, "r");
35   if (fichero == NULL)
36   {
37    printf("El fichero no se ha podido abrir para lectura.\n");
38    return -1;
39   }
40   // Se lee número por número.
41   int num;
42   while (!feof(fichero))
43   {
44     resultado = fread(&num, sizeof(int), 1, fichero);
45  if (resultado != 1)
46  {
47   break;
48  }
```

```
49   suma = suma + num;
50    }
51
52    if (ferror(fichero)!=0)
53    {
54     printf("Ha ocurrido algún error en la lectura de números.\n");
55    }
56    else
57    {
58     printf("La suma de los números leídos es: %d\n.",suma);
59    }
60
61    if (fclose(fichero)!=0)
62    {
63      printf("No se ha podido cerrar el fichero.\n");
64   return -1;
65    }
66    return 0;
67  }
68
```

A la hora de escribir el archivo, conocemos de antemano el tamaño del array y se puede escribir todo el bloque. En caso de que hubiera que escribir una serie de números, sin saber exactamente cuántos, se tendría que ir escribiendo uno por uno.

La sentencia `break` de la línea 47 se usa para salir de la lectura, pues ha habido un error o final de fichero y no se necesita efectuar ninguna operación más dentro del bucle. Al salir, se comprueba qué es lo que pasó realmente, para advertir al usuario.

9.4. Acceso aleatorio a ficheros

En la sección anterior hemos visto cómo leer y escribir datos de manera secuencial. En algunos casos, sin embargo, se necesita acceder a una parte específica en mitad de un fichero. Para eso contamos con las funciones, `fseek`, `ftell` y `rewind`, que permiten realizar accesos aleatorios.

La función `fseek` permite desplazar el indicador de posición de fichero al sitio desde donde el cual quieres acceder al fichero. La sintaxis es:

```
#include <stdio.h>
int fseek(FILE *stream, long offset, int whence);
```

`stream` es el puntero asociado con el fichero abierto. `offset` indica el número de bytes que se va a desplazar el indicador desde una posición especificada por `whence`, que puede tener uno de los siguientes valores: `SEEK_SET`, `SEEK_CUR` y `SEEK_END`. Si todo va bien, la función devuelve `0`; si no, devuelve un valor distinto de cero.

Si se utiliza `SEEK_SET`, el desplazamiento se realizará desde el principio del fichero, y el tamaño de `offset` deberá ser mayor o igual que cero. Si se usa `SEEK_END`, el desplazamiento se realizará desde el final del fichero, y el valor de `offset` tendrá que ser negativo. Si se usa `SEEK_CUR`, el desplazamiento se calcula desde la posición actual del indicador de posición de fichero.

Puedes obtener el valor actual del indicador de posición de fichero llamando a `ftell`:

```
#include <stdio.h>
int ftell(FILE *stream);
```

El valor que devuelve `ftell` representa el número de bytes desde el principio del fichero hasta donde nos encontramos en ese momento (sitio indicado por el indicador de posición). Si la función falla, devuelve `-1L` (es decir, el valor largo (`long`) de menos 1).

A veces te interesará reiniciar el indicador de posición de fichero y llevarlo al inicio del mismo. Para esto cuentas con la función `rewind`:

```
#include <stdio.h>
void rewind(FILE *stream);
```

De esta manera, la siguiente línea de código:

```
rewind(ptr_fich);
```

es equivalente a esta otra:

```
fseek(ptr_fich,0L,SEEK_SET);
```

9.5. Manipulación directa de ficheros

Existen otra serie de funciones adicionales para manejar ficheros que puede que te sean útiles.

La función `remove` que, dado el nombre de un fichero, lo borra del sistema de ficheros:

```
int remove(char *filename);
```

Esta función devuelve `0` si el fichero pudo eliminarse correctamente; si no, devuelve otro número.

La función `rename`, que renombra un fichero, teniendo además la posibilidad de poder moverlo entre directorios si en el argumento `newname` ponemos una ruta en vez de sólo un nombre de fichero:

```
int rename(const char *oldname, const char *newname);
```

Esta función devuelve `0` si el fichero pudo renombrarse correctamente; si no, devuelve otro número. Si falla por cualquier razón, el fichero original no se ve afectado.

9.6. Preguntas de autoevaluación

Comprueba con estas preguntas que has entendido cómo se trabaja con ficheros.

1. Hay algo que está mal en este trozo de código, ¿qué es?:

```
1    FILE *fptr;
2    char b = 'b';
3    char *c = &b;
4    if (fptr = fopen("file.txt", "r"))!= NULL)
5    {
6      while(!feof(fptr))
7      {
8          fwrite(c, sizeof(char),1,fptr)
9          printf("Character: %c",*c);
10     }
11     fclose(fptr);
12   }
```

- Pase lo que pase en la apertura del fichero, siempre se tiene que cerrar, así que la llamada a la función `fclose(fptr)` de la línea 11 debería ir fuera de la sentencia `if`, después de la línea 11.
- Las funciones de lectura/escritura (`fread/fwrite`) necesitan como primer parámetro la dirección de la variable donde se va a leer o escribir información, así que la sentencia correcta de la línea 8 es `fwrite(&c, sizeof(char),1,fptr);`
- La función `printf` necesita la dirección de la variable de tipo `char`, no el valor, así que la sentencia correcta de la línea 9 es `printf("Character: %c",c);`
- Se está intentando escribir en un fichero que sólo se ha abierto para lectura.
- El código es correcto, no hay nada mal.

2. Abres el fichero `fptr`, que ocupa 100 bytes. ¿Cuál de loss siguientes pares de sentencias son equivalentes?

- ```
  rewind(fptr);
  ```
  ```
  fseek(fptr,-0L, SEEK_END);
  ```
- ```
  rewind(fptr);
  ```
  ```
  fseek(fptr,100L, SEEK_END);
  ```
- ```
  rewind(fptr);
  ```
  ```
  fseek(fptr,0L, SEEK_CUR);
  ```
- ```
  rewind(fptr);
  ```
  ```
  fseek(fptr,-100L, SEEK_SET);
  ```
- Todas las anteriores son equivalentes.
- Ninguna de las anteriores.

3. Abres el fichero `fptr`. Comienzas a leer byte a byte. Al leer todos los datos y llegar al final, ¿cuál de estas sentencias es correcta?

- El indicador de posición de fichero permanece al comienzo del fichero; sólo se puede mover con las funciones `rewind` ó `fseek`.
- El indicador de posición de fichero está situado al final del fichero; se ha ido moviendo byte a byte, conforme se iba leyendo.
- El indicador de posición de fichero permanece al comienzo del fichero; sólo se va moviendo cuando escribimos en un fichero, no cuando leemos de él.
- Ninguna de las anteriores.

9.7. Bibliografía de apoyo

- L. Joyanes Aguilar, I. Zahonero Martínez, "Programación en C: Metodología, algoritmos y estructura de datos", McGraw Hill, Capítulo 15: Entrada y salida por archivos, páginas .

- J. Badenas Carpio, J.L. Llopis Borrás, O. Coltell Simón, "Curso práctico de programación en C y C++", Capítulo 6: La entrada/salida, páginas 127-142.