

第七章 文件管理

7.1 文件和文件系统

◆ 7.1.1.文件的定义

◆ 文件是计算机系统中信息存放的一种组织形式，目前尚无严格的定义，下面给出两种有代表性的解释：

- （1）文件是具有标识符的相关字符流的集合。
- （2）文件是具有标识符的相关记录（一个有意义的信息单位）的集合。

7.1 文件和文件系统

- ◆ 这两种解释定义了两种文件形式：前者说明文件是由字节组成，这是一种无结构的文件，或称流式文件。
- ◆ 无结构文件由于采用字符流方式，与源程序、目标代码等在形式上是一致的，因此，该方式适用于源程序、目标代码等文件。
- ◆ 目前**UNIX**操作系统，**MS-DOS**系统均采用这种文件形式。

7.1 文件和文件系统

- ◆ 后者说明文件是由记录组成。而记录则是由一组相关信息项组成。
- ◆ 例如每个学生的登记表可视为一个记录，它包括学生姓名，出生年月，性别，籍贯等信息项。所有学生登记表组成一个学生文件。
- ◆ 记录式文件主要用于信息管理。
- ◆ 在现代计算机操作系统中，为方便用户，把设备也作为文件来统一管理，从某种意义上说已拓宽了文件的含义。

7.1.2 文件的分类

- ◆ (1) 以文件的用途分类
- ◆ **系统文件**：由操作系统及其他系统程序和数据组成的文件。这种文件不对用户开放，仅供系统使用，用户只能通过操作系统提供的系统调用来使用它们。
- ◆ **库文件**：是指系统为用户提供的各种标准函数，标准过程和实用程序等。用户只能使用这些文件，而无权对其进行修改。
- ◆ **用户文件**：由用户的信息组成的文件，如源程序文件，数据文件等。这种文件的使用和修改权均属于用户。

7.1.2文件的分类

◆ (2) 按文件的操作保护分类

- **只读文件：** 只允许进行读操作，不能进行写操作的文件。
- **读写文件：** 允许文件主和授权用户对其进行读或写操作的文件。
- **只执行文件：** 该类文件只允许授权的用户调用执行，而不允许其修改或读出文件的内容。

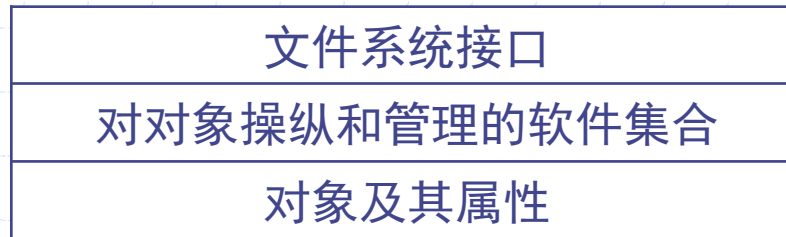
7.1.2文件的分类

◆ (3) 按文件的性质分类

- **普通文件**：指一般的用户文件和系统文件。
- **目录文件**：管理和实现文件系统的文件目录项组成的系统文件，对目录文件可以进行与普通文件一样的各种文件操作。
- **特殊文件**：有的系统把设备作为文件统一管理和使用，并为区别起见，**把设备称为特殊文件**。例如：**unix/linux**

7.1.3 文件系统的层次结构

- ◆ **文件系统**是操作系统中负责存取和管理信息的模块，它用统一的方式管理用户和对系统信息的存储、检索、更新、共享和保护，并为用户提供一整套方便有效的文件使用和操作方法。文件系统的模型如下：



- ◆ 它由管理文件所需的**数据结构**（如文件控制块及存储分配表等）和相应的**管理软件**以及访问文件的一**组操作**组成。

文件系统的功能

- ◆ 1.使用户可执行创建、修改及删除读写文件的命令。
- ◆ 2.使用户在系统控制下共享其他用户的文件，以便用户可共享其它人的工作成果。
- ◆ 3.使用户能以合适的方式构造其他文件
- ◆ 4.使用户能使用在文件间进行数据传输的命令

文件系统的功能

- ◆ 5.为了实现按名存取，需要有一个用户可见的文件逻辑结构，用户按照文件逻辑结构所给定的方式进行信息的存取和加工。这种逻辑结构是独立于物理存储设备的。
- ◆ 6.为了防止意外事故，文件系统具有转储和恢复文件的能力
- ◆ 7.能提供可靠的保护和保密措施

Windows的主流文件系统

- ◆ FAT(File Allocation Table)是“文件分配表”的意思。对我们来说，它的意义在于对硬盘分区的管理。
- ◆ FAT16、FAT32、NTFS是目前最常见的三种文件系统。

其它文件系统

◆ **FAT12:** 是IBM第一台个人电脑中的MS-DOS 1.0使用的文件系统，主要用于软盘。这种系统限制分区的容量最大为16MB——但这根本算不上问题，因为软盘容量从来没有达到16MB。

ISO9660: CD-ROM的文件系统，不过现在已经延伸出很多新的文件系统，对它的一些缺点进行了弥补，如Juliet等。

UDF: 可读写光盘的文件系统。

Mac HFS: 苹果电脑的文件系统，对大容量磁盘有比较好的支持。不过，现在大多数苹果电脑还在使用FAT32文件系统。

7.2 文件的逻辑结构

- ◆ 通常，用户在使用文件时，只关心文件的逻辑结构。
- ◆ 从用户观点观察到的文件组织形式主要有两类：
 - 一类是有结构的文件
 - 另一类是无结构的流式文件
- ◆ 按照文件的组织方式
 - 顺序文件
 - 索引文件
 - 索引顺序文件

文件的两种结构:

(1) 逻辑记录(结构)

- 逻辑记录是文件中按信息在逻辑上的独立含义来划分的信息单位。
- 逻辑记录是对文件进行存取操作的基本单位。

(2) 物理记录(结构)

- 在存储介质上，由连续信息所组成的一个区域称为块，也叫物理记录。

(3) 逻辑记录与物理记录的区别与关系

- 一个是逻辑的概念，一个是物理的概念
- 逻辑记录最终在存放物理记录上

文件的逻辑结构

◆ 1. 记录式文件

➤ 记录式文件是一种**有结构**的文件。

➤ 2. 流式文件

➤ 流式文件是相关的有序字符的集合。是无结构的。

➤ 流式文件是按信息的个数或以特殊字符为界进行存取的。

1. 记录式文件

◆ 每个记录由彼此相关的域构成。记录可按顺序编号为记录1，记录2，...，记录 n 。如果文件中所有记录的长度都相同，则这种文件为定长记录文件。

◆ 例如：学生登记表文件 xsdjb. dbf

姓名	学号	籍贯	通信地址	邮政编码
----	----	----	------	------

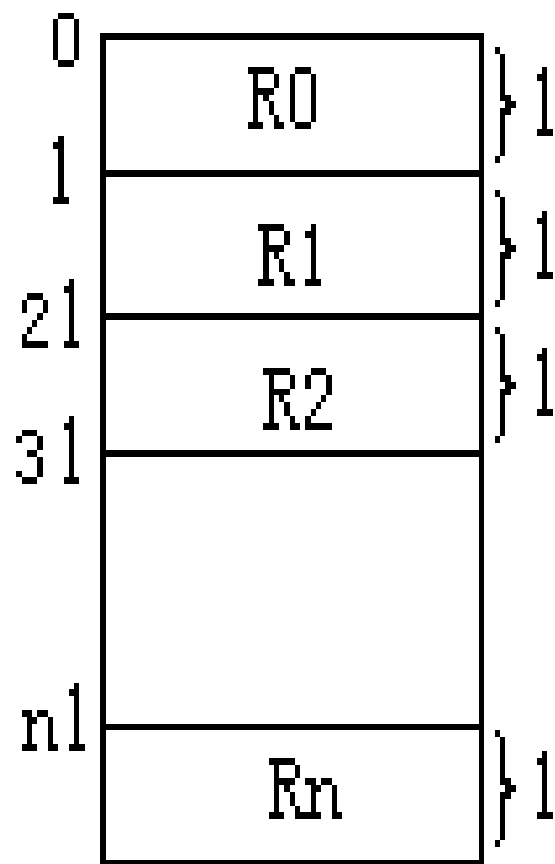
李铭	925678	武昌	武昌关山街125号	430074
----	--------	----	-----------	--------

司马乐	925679	北京	北京海军路88号	100034
-----	--------	----	----------	--------

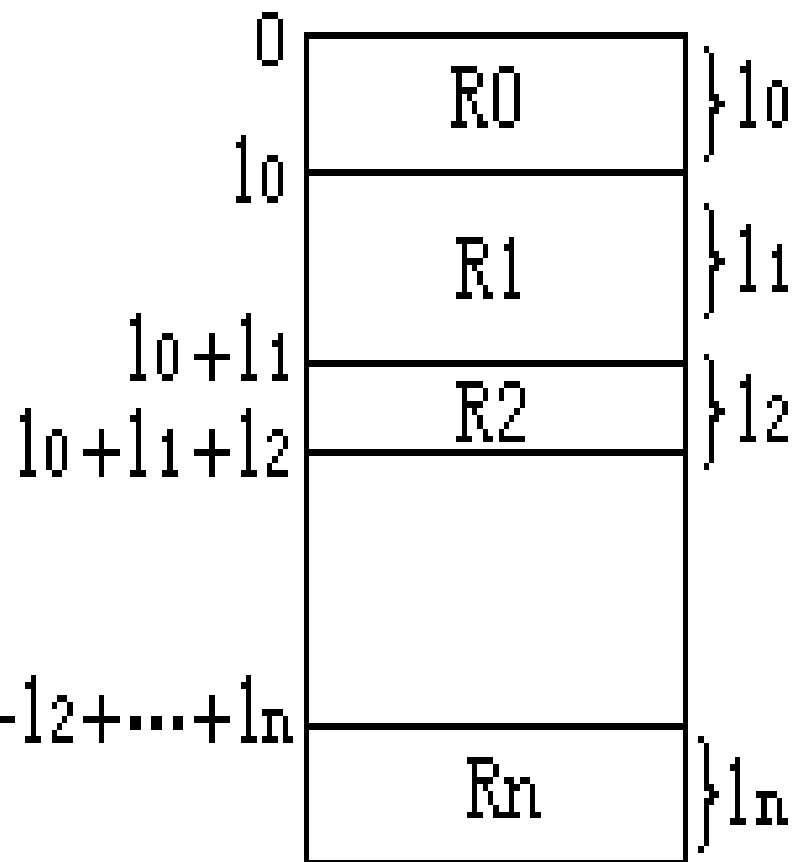
1.记录式文件

- ◆ 记录式文件按照记录长度是否相同，又可分为定长记录文件和不定长记录文件两种。
- ◆ (1) 定长记录：文件中所有记录的长度相等。
- ◆ (2) 变长记录：文件中记录的长度不相等。
- ◆ 定长记录文件的长度 = 记录个数 × 记录长度。
- ◆ 变长记录文件的长度为各记录长度之和。

记录式文件



(a) 定长记录文件



(b) 变长记录文件

2.无结构的文件

- ◆ **无结构文件**是指文件内部不再划分记录，是由一组相关信息组成的有序字符流，即流式文件。其长度直接按字节来计算。
- ◆ **事实上操作系统不知道或不关心文件中存放的内容是什么**，它所见到的都是一个一个的字节。文件中任何信息的含义都由用户级程序解释。

2. 无结构的文件

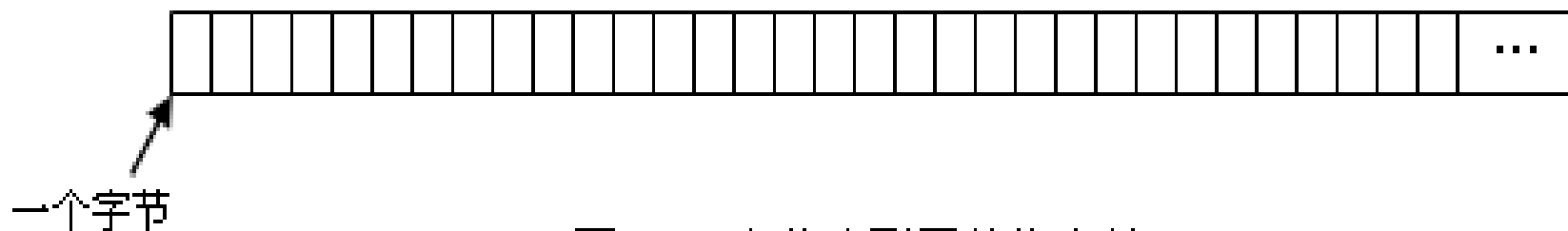


图 6-2 字节序列无结构文件

两种文件的比较

- ◆ 流式文件就象给一张白纸给用户，用户可将他的信息任意地写到纸上，没有任何格式上的限制。
- ◆ 记录式文件就象给一张表格给用户，用户要按表规定的格式填信息。
- ◆ 显然，结构式文件对用户的限制很大，使用起来就不方便，所以记录式文件被淘汰是理所当然的。

7.3 文件目录

- ◆ 通常，在计算机系统中，大量的文件被存储在磁盘上。为了对存储在磁盘上的众多文件进行有效的控制和管理，必须对它们加以妥善组织。
- ◆ 这种组织是**通过文件目录来实现的**，文件目录是一种数据结构，用来标识文件系统中的文件及其物理地址，供检索时使用。

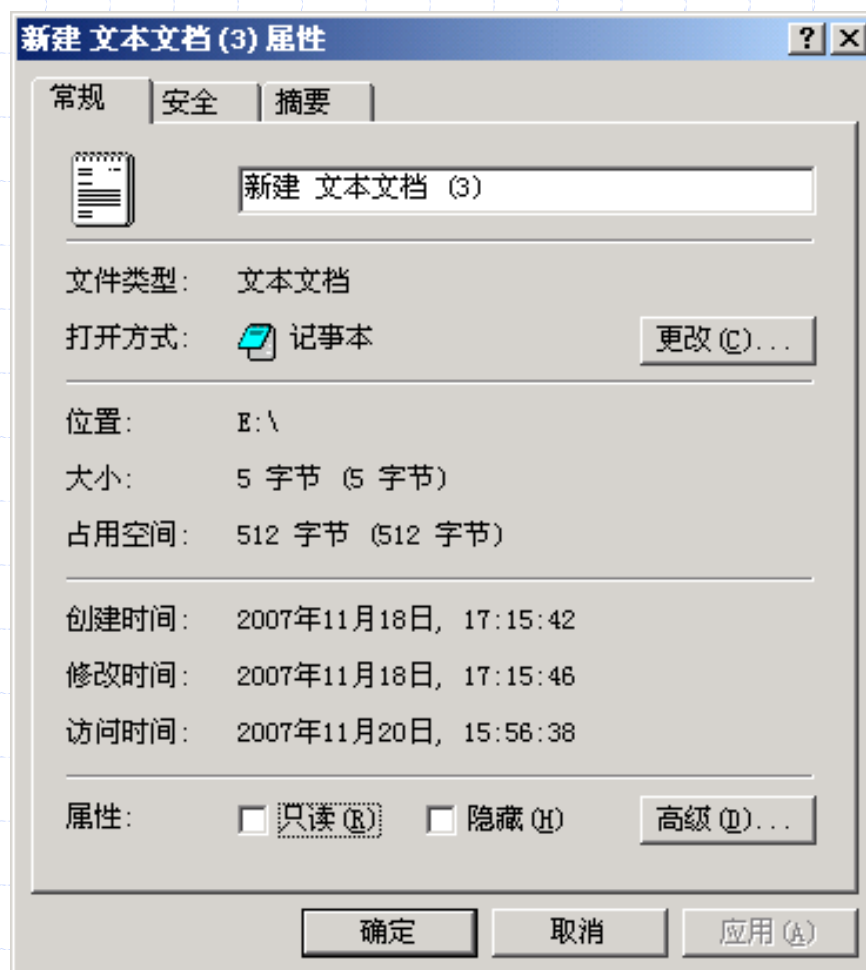
文件目录的组成

- ◆ 文件名：符号文件名，如music、game、file等。
- ◆ 文件类型：指明文件属性是普通文件，还是目录文件或特别文件，是系统文件还是用户文件等。
- ◆ 文件的物理位置：文件在物理设备上的位置，如文件存放在哪台设备的哪些盘块上。
- ◆ 文件的大小：当前文件大小（以字节、字或块为单位）和允许的最大长度。
- ◆ 保护信息：对文件读、写及执行等操作的控制权限标志。
- ◆ 使用计数：表示当前有多少个进程正在使用或打开了该文件。
- ◆ 时间和日期：这个信息反映了文件创建、最后修改、最后使用等情况，可用于对文件实施保护和监控等。

7.3.1 文件控制块和索引结点

- ◆ 而文件目录信息也叫文件控制块（**file control block ,FCB**），它是操作系统为管理文件而设置的数据结构，存放了为管理文件所需的所有有关信息（文件属性）。
- ◆ **文件控制块是文件存在的标志**，它通常由文件属性信息组成。

FCB的外部表现：文件的属性



2.文件目录

- ◆ 为了对众多的文件进行分门别类的管理，提高文件检索的效率，现代操作系统往往将文件的文件控制块集中在一起进行管理。
- ◆ 这种FCB的有序集合就称为文件目录，文件控制块就是其中的目录项（构成文件目录的项目）。
- ◆ 另外，为了实现对文件目录的管理，通常将文件目录以文件的形式保存在外存，这个文件就叫目录文件。

3.索引结点的引入

- ◆ 为了减少系统开销，采用了把文件名与文件描述信息分开的办法，即使文件描述信息单独形成一个称为索引结点的数据结构，简称为i结点。

文件名	索引结点编号
文件名1	
文件名2	
...	

图 6-15 UNIX的文件目录

例题：

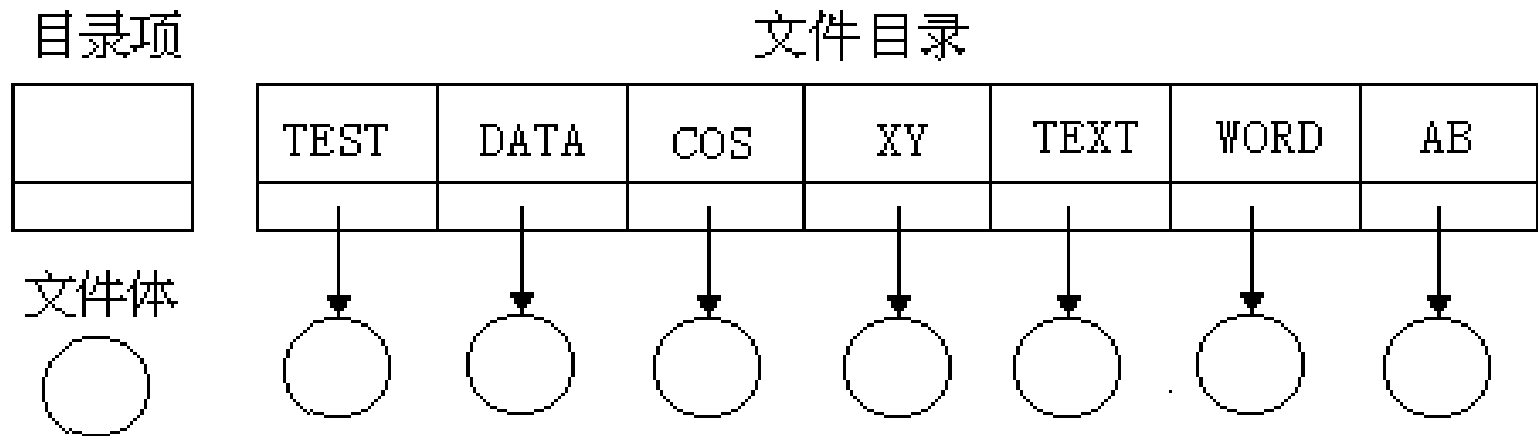
- ◆ 在某个文件系统中，每个盘块为512字节，文件控制块占64个字节，其中文件名占8个字节。如果索引结点编号占2个字节，对一个存放在磁盘上的256个目录项的目录，试比较引入索引结点前后，为找到其中一个文件的FCB，平均其中磁盘的次数。

7.3.2 简单的文件目录

- ◆ 为了方便用户的使用，提高文件系统的效率，也必须对系统内的所有文件目录进行组织。
- ◆ 在现代操作系统中，目录的基本组织方式有：
 - 单级文件目录
 - 两级文件目录
 - 树形结构目录

1 单级文件目录

- ◆ 一级目录是最简单的目录结构。在这种组织方式下，全部文件都登记在同一目录中。

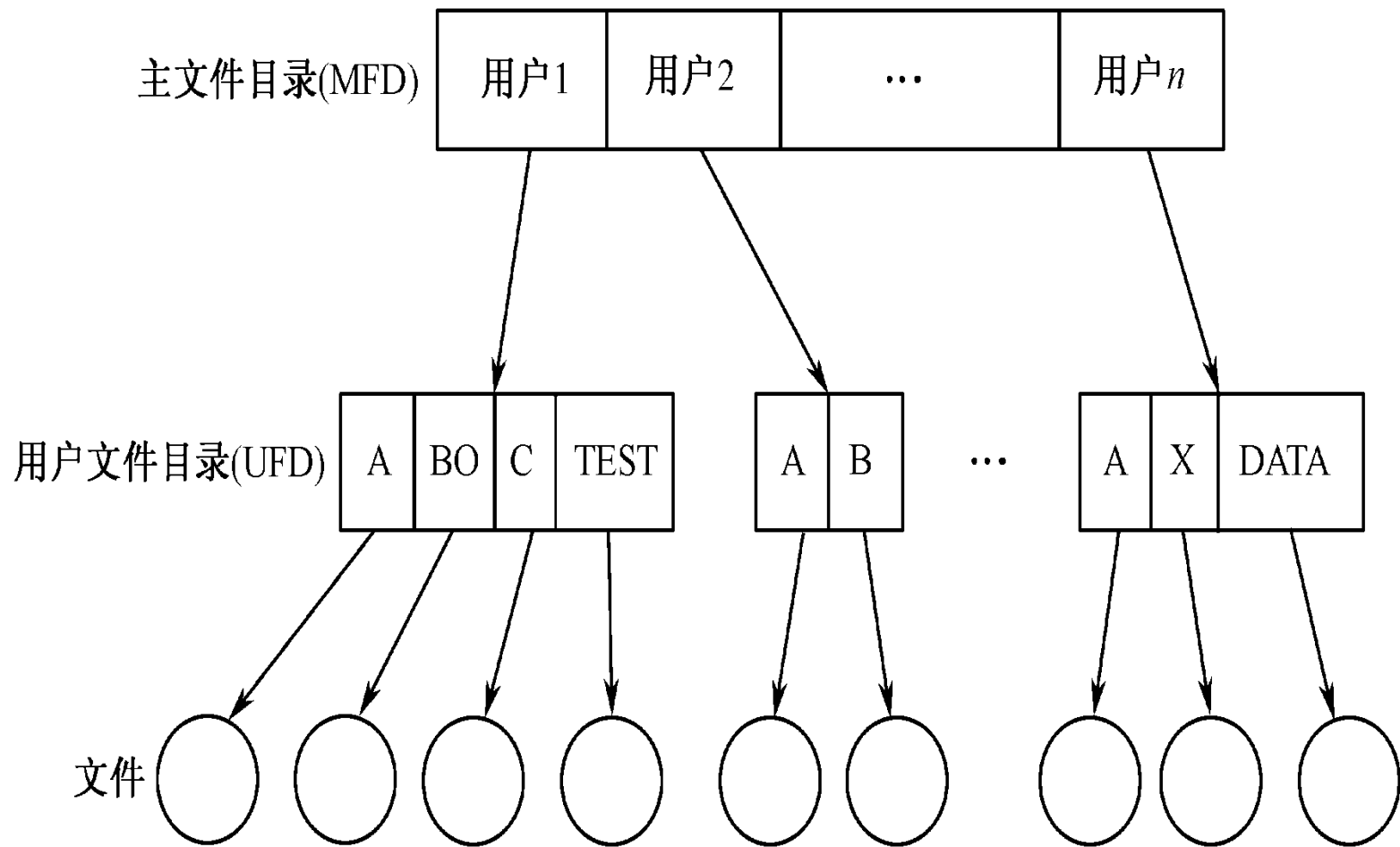


- ◆ 其特点是简单、易于理解和实现，但那也存在以下的缺陷：**查找速度慢、不允许重名和不便于文件的共享**

2 两级文件目录

- 为改变一级目录文件目录命名冲突，并提高对目录文件检索速度而将目录分为两级：
- 一级称为主文件目录，给出用户名，用户子目录所在的物理位置；
- 二级称为用户文件目录，给出该用户所有文件的FCB。
- 文件主目录（MFD）的表目按用户分，每个用户有一个用户文件目录（UFD）

2 两级文件目录



7.2.3 两级目录

- 在二级目录结构中，用户引用特定的文件时，系统只需搜索他自己的UFD，因此，不同用户可拥有具有相同名称的文件，只要每个UFD内的所有文件名称惟一即可。
- 当用户创建文件时，操作系统也只搜索该用户的UFD以确定具有相同名字的文件是否存在。
- 当删除文件时，操作系统只在局部UFD中对其进行搜索，因此，它并不会删除另一个用户的具有相同名称的文件。

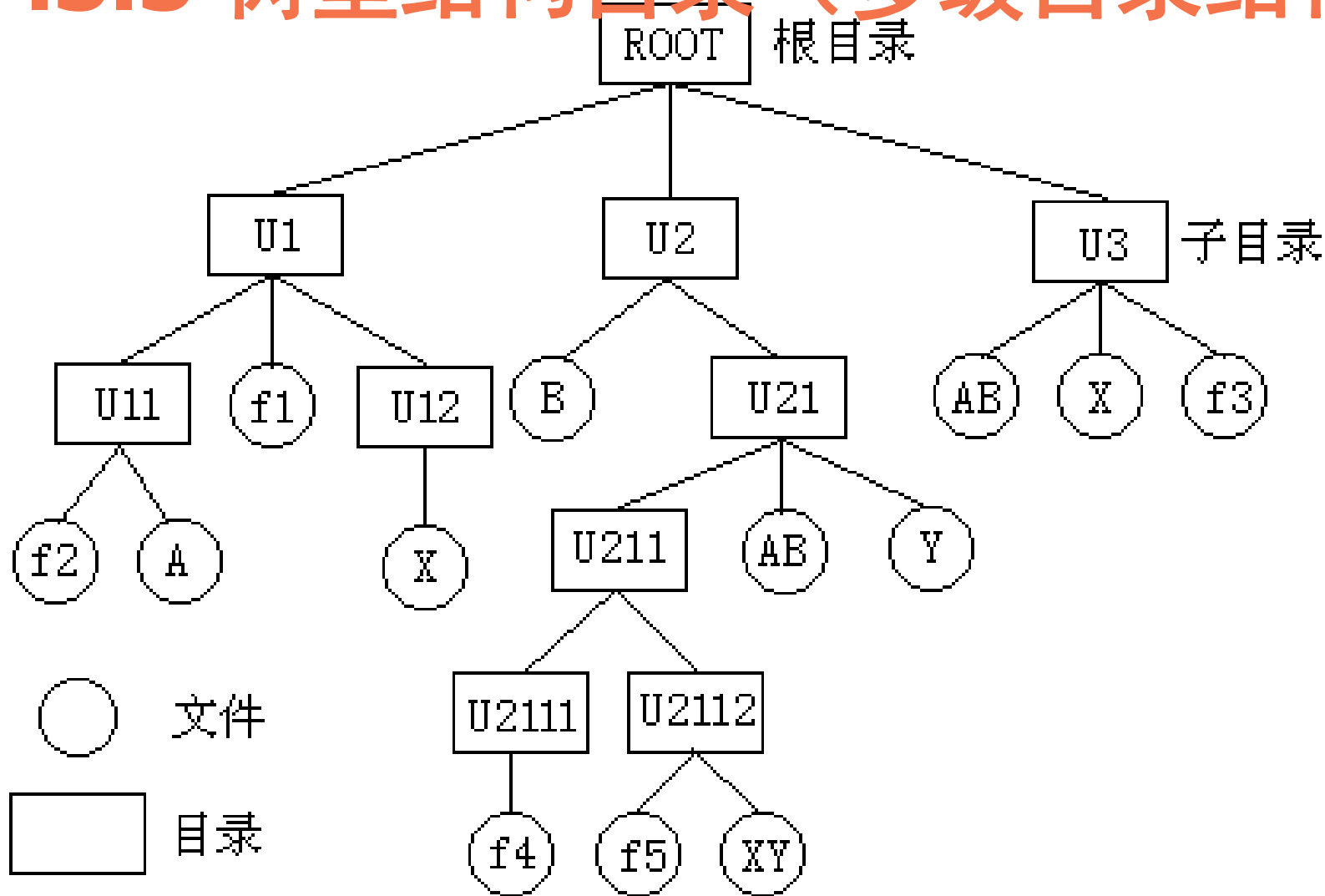
7.2.3 两级目录的特点

- ◆二级目录的优点：解决了名称冲突和文件共享问题，提高了搜索速度，查找时间也降低了。
- ◆但是，它仍有一定的缺陷：缺少灵活性、不能反映现实世界中的多层关系。
- ◆因此就产生了多级目录结构

7.3.3 树型结构目录（多级目录结构）

- ◆ 又称为多级目录结构，它是二级目录结构的扩充。
- ◆ 这种多层次的目录结构如同一棵倒置的树，主目录就是树根，称为**根目录**
- ◆ **每一个树枝结点就是一个子目录**，每一片树叶描述的一个文件。

7.3.3 树型结构目录（多级目录结构）



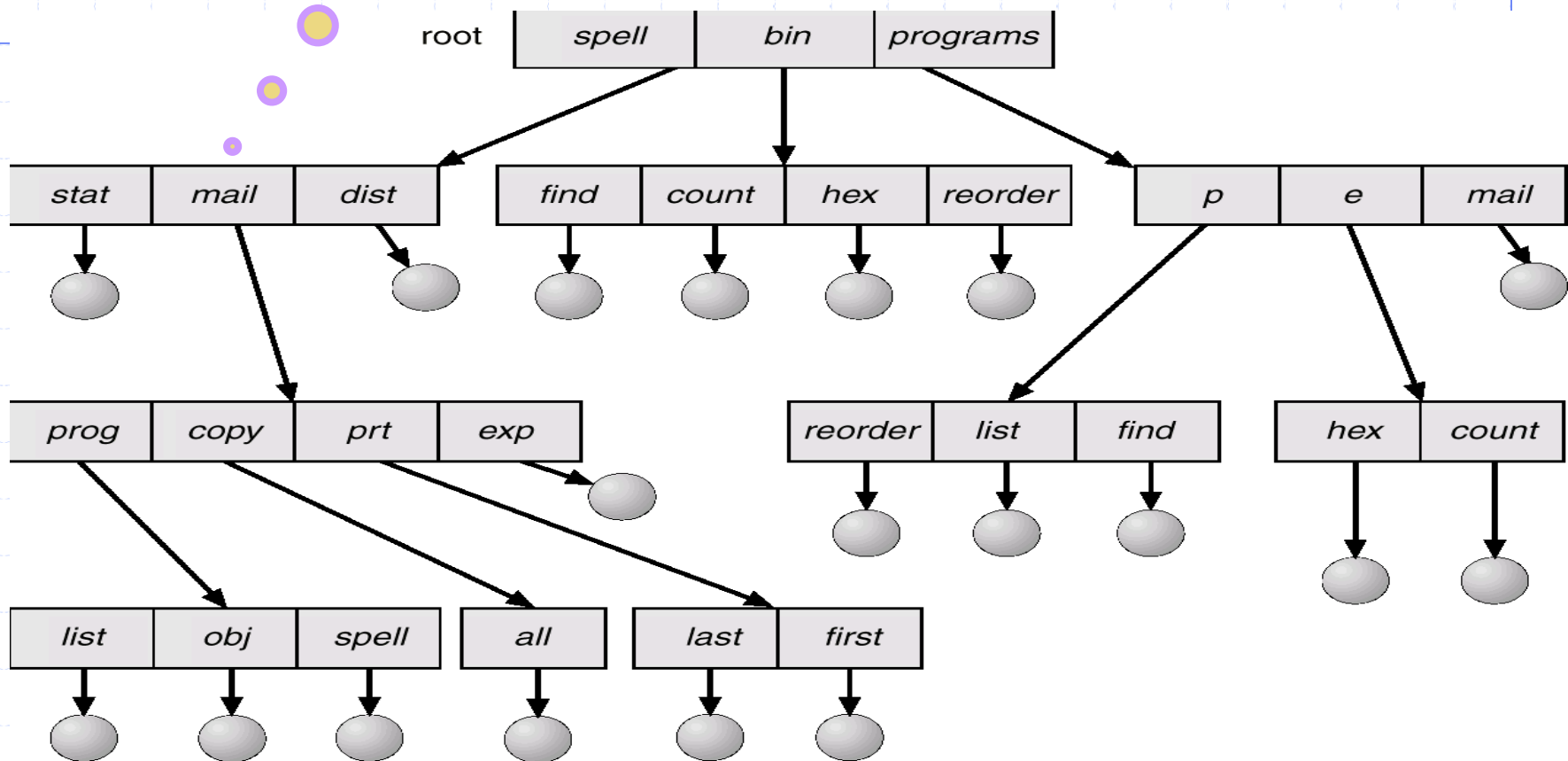
7.3.3 树型结构目录（多级目录结构）

- ◆在树形目录结构中，一个文件的全名将包括从根目录开始到文件为止的通路上遇到的所有子目录路径。
- ◆各子目录名之间用正斜线“/”或反斜线“\”隔开，其中，子目录名组成的部分又称为路径名。

7.3.3 树型结构目录（多级目录结构）

- ◆ 系统内的每个文件都有惟一的路径名。
- ◆ 路径名是从根经过所有子目录再到指定文件的路径。
- ◆ 路径名有两种形式：**绝对路径名**和**相对路径名**。
 - **绝对路径名**从根目录开始并给出路径上的目录名直到指定的文件
 - **相对路径名**从当前目录开始定义一个路径。
- **UNIX/Linux**也使用相对路径名和绝对路径名来标识文件或目录，只不过文件和目录之间采用“/”来分隔，而不是**DOS**的“\”。

当前目录/root/spell/mail
请问**first**的相对路径和绝对路径分别是什么？



树形目录

7.3.3 树型结构目录（多级目录结构）

- ◆ 在上图所示的树形目录中，如果当前目录是 /root/spell/mail,
- ◆ 那么相对路径名 prt/first
- ◆ 和绝对路径名 root/spell/mail/prt/first 指向相同的文件。

7.3.4 目录查询技术

- ◆ 线性检索法

- ◆ Hash方法

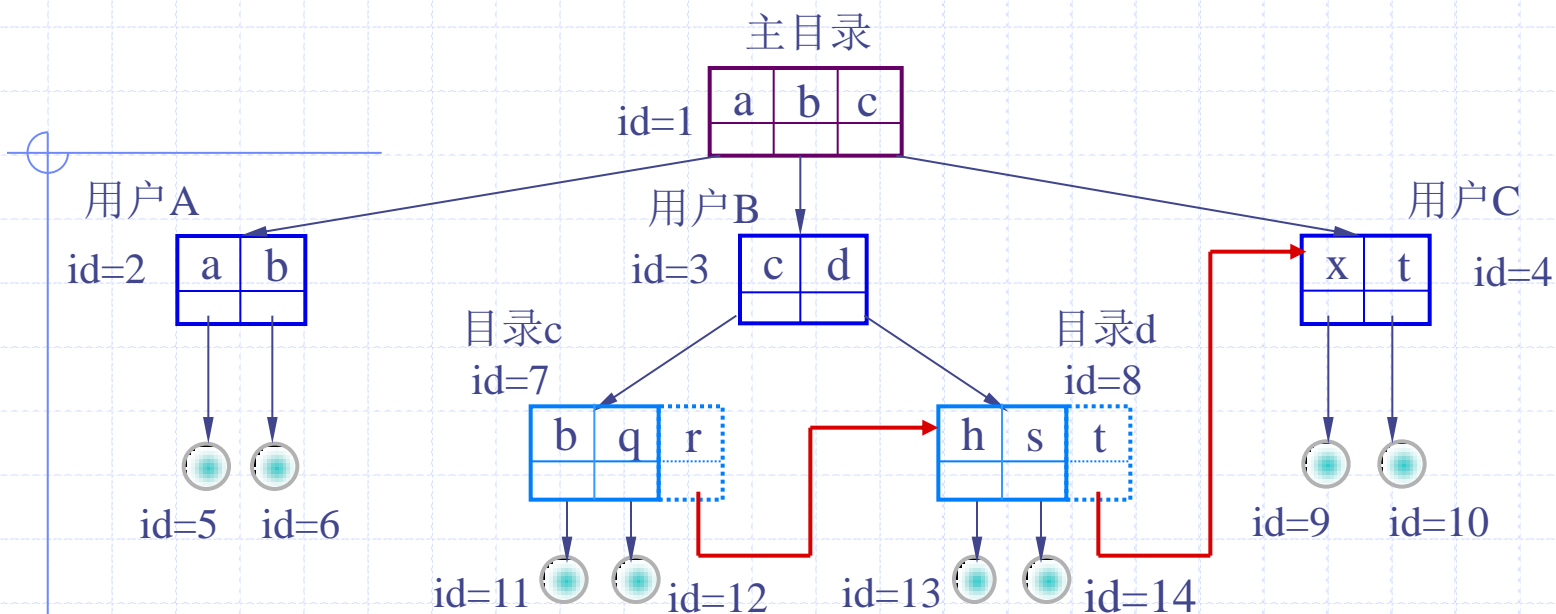
7.4 文件共享

- ◆ 所谓文件共享指系统允许多个用户或进程共享同一份文件。
- ◆ 文件共享不仅是完成共同的任务的多个用户所必需的，更重要的是避免了同一个文件保存多个副本所造成的外存空间的浪费。

7.4 文件共享

◆ 文件共享的方式:

- **采用文件全名访问他人文件:** 直接通过文件目录找到他人文件。
- **基于有向无循环图实现文件共享:** 一个目录项直接用一个指针（或编号）指向另一个目录项达到共享文件的目的——即通过共享同一个文件的索引结点的方式。
- **利用符号链实现文件共享:** 用户A为了共享用户B的Bboot目录下的一个文件f1.c, 可以创建一个LINK类型的新文件x, 新文件x中仅包含被链接文件f1.c的路径名。



例：图中有两个链接，设当前目录id=7:

- 用文件路径名为r，即可访问 id=13（文件h）
- id=9（文件x）——文件路径名为 */d/t

7.5 文件保护

◆ 1. 对用户进行分类

- ◆ 按用户对文件访问权力的差别把用户分成几类，然后对每个文件规定各类用户的存取权限。通常将用户分成三类：

- 文件主
- 文件主的同组用户或合作用户
- 其它用户

7.5 文件保护

◆ 2.对访问权限分类

- ◆ 对文件的访问系统首先要检查访问权限，只允许合法的用户访问。
- ◆ 文件的存取权限一般有以下几种:只读 (R)、只写 (W)、执行 (E)、添加 (A)和修改 (U)等。

1.访问控制矩阵

用户	文件							
	1	2	3	4	5	6	7	8
1	0	0	1	1	0	0	0	0
2	0	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0
4	1	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0

2.用访问控制矩阵实现文件保护

- ◆ 一维代表所有用户，一维代表系统中的所有文件。

存取数 文件名 用户	Wang	lee	Zhang
A.C	RWE	E	RWE		
B.C	RW	R	RWE		
D.C	R	W	WE		
E.C	R	W	RW		