**第7章 文件与外部排序 习题参考答案**

**四、习题**

1. **综述文件有哪几种常用的组织方式，它们各有什么特点。**

**解答：**

1. **顺序文件、索引文件、散列文件**
2. **特点**

**顺序文件又称连续文件**

**优点：处理速度快，存储空间利用率高，可在任何时候媒介上实现**

**缺点：需要对事务预先排序，插入操作要检索整个记录**

**索引文件分为索引顺序文件和索隐非顺序文件**

**优点：可顺序处理，也可随机处理，记录追加、插入方便，查找速度较快**

**缺点：增加索引区和溢出区空间，不利于多码检索**

**散列文件又称直接文件、哈希文件，其优缺点分别为：**

**优点：存取速度快，记录追加、删除容易**

**缺点：存储空间利用率低，且不能进行顺序存储，只能按关键字随机存取，查询方式限于简单查询；在经过多次插入、删除后，可能造成文件结构不合理，需要重新组织文件。**

1. **简单比较文件的多重表和倒排表组织方式各有什么不同。**

**解答：**

1. **多重表文件：记录按主关键字的顺序够成一个串联文件，并建立主关键字的索引；对每一个次关键字项建立次关键字索引。所有具有同一关键字的记录够成一个链表。**

**优点：易于构造，也易于修改。如果不要求保持链表的某种次序，则插入记录容易，此时可将记录插在链表的头指针之后。**

**缺点：删除一个记录较繁琐，需要在每个次关键字的链表中删去该记录。**

1. **倒排文件：用记录的非主属性值（也叫副键）来查找记录而组织的文件**

**优点：检索记录较快**

**缺点：维护困难，在同一索引表中，不同关键字记录数不同，各倒排表的长度不等，同一倒排表中各项长度也不等。**

1. **假设某文件经内部排序得到100个初始归并段，试问:**

**(1)若要使多路归并三趟完成排序，则应取归并的路数至少为多少？**

**(2)假若操作系统要求一个程序同时可用的输入、输出文件的总数不超过13，则按多路归并至少需几趟可完成排序？如果限定这个趟数？则可取的最低路数是多少？**

**解答：**

1. **至少取5路进行归并**
2. **每次取12路进行归并，则至少需2趟才可完成排序，如果限定归并趟数为2，则对于100个初始归并段，取10路归并即可。**
3. 假设一次I/O的物理块大小为150，每次可对750个记录进行内部排序，那么，对含有150000个记录的磁盘文件进行4-路平衡归并排序时，需进行多少次I/O？

解答：

150000个记录存放在150000/150=1000个物理块上，每次可对750/150=5个物理块的记录进行内部排序，可得到1000/5=200个初始归并段，200个初始归并段需要经过log4200=4趟4-路平衡归并排序，所以，总的I/O次数是：1000\*2+1000\*2\*4=10000。

1. 试问输入文件在哪种状态下经由置换-选择排序得到的初始归并段长度最长？其最长的长度是多少？

解答：

记录关键字均非递减排列时，可得到最长的初始归并段，其长度所有记录总数。

1. **试问输入文件在哪种状态下经由置换-选择排序得到的初始归并段长度最短？其最短的长度是多少？**

解答：

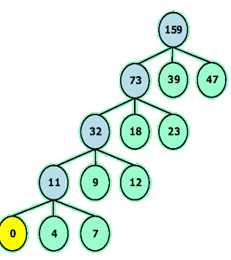
记录关键字均非递增排列时，可得到最短的初始归并段，其长度为1。

1. 已知某文件经过置换-选择排序之后，得到长度分别为47, 9, 39, 18, 4, 12, 23和7的八个初始归并段。试为3-路平衡归并设计一个读写外存次数最少的归并方案，并求出读写外存的次数。

解答：

设归并路数为k，归并段个数为m，则(m-1)MOD(k-1)=1,故需要增加：

k-(m-1)MOD(k-1)-1=1个长度为0的“虚段”。



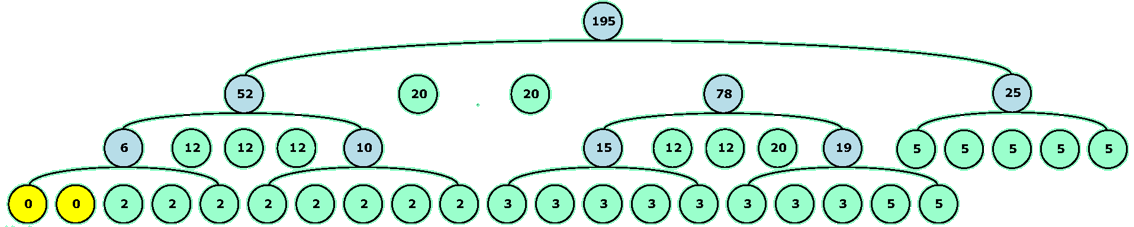
总I/O次数为：[(4+7)\*4+(9+12)\*3+（18+23）\*2+（39+47）\*1]\*2=550

1. **已知有31个长度不等的初始归并段，其中8段长度为2，8段长度为3，7段长度为5，5段长度为12，3段长度为20（单位均为物理块）。请为此设计一个5-路最佳归并方案，并计算总的（归并所需的）读/写外存的次数。**

解答：

设归并路数为k，归并段个数为m，则(m-1)MOD(k-1)=2,故需要增加：

k-(m-1)MOD(k-1)-1=2个长度为0的“虚段”。



总I/O次数为：

[(8\*2+8\*3+5\*2)\*3+(5\*5+5\*12+20)\*2+(2\*20)\*1]\*2=800