

第 8 次书面作业
第九章 外排序

1. FIFO(First in First out)方法和 LRU(Least recently used)方法是用来作为缓冲区替换算法的两种选择。FIFO 每次将最先进入缓冲区的结果替换出去；LRU 每次将距离上次使用时间最长的结果替换出去。假定我们有一个大小为 3 的缓冲区，和一个如下输入的输入序列：
2 3 1 4 0 3 2 4 1 2 4 3
 - (1) 请分别写出 FIFO 方法和 LRU 方法的运行结果（缓冲区的变化），并比较这两个算法，给出 FIFO 方法和 LRU 方法的主要区别。
 - (2) 想要减少访问内存的 miss rate，一个直观的想法是增大缓冲区的大小。请问这样做对于 LRU 和 FIFO 方法是否都是有效的？如果不是请举出反例。
 - (3) 描述一种你所知道的其他替换算法，或者你所设计的一种替换算法。并说明这个算法可能的优势和劣势。
2. 置换选择排序的核心思想是利用堆对数据进行处理。每输出一个值，就从缓冲区中读入下一个数。如果堆的大小是 M ，一个顺串的最小长度就是 M 个记录，至少原来在堆中的那些记录将成为顺串的一部分。最好的情况下，例如输入已经被排序，有可能一次就把整个文件生成成为一个顺串。
 - (1) 现在给出一组输入关键字(17, 2, 20, 40, 10, 19, 8, 13, 11, 25, 21, 7)，假设堆的大小是 5 且起始为空，请写出得到的初始顺串和最后堆的状态。排序时较小的元素在前。
 - (2) 置换选择排序一定要用堆来实现吗？请任意给出一个不同的实现，用上面的输入和设定比较一下两者的差异。
3. 现在有 9 个长度不同的顺串，其长度分别为 17, 40, 16, 55, 25, 11, 6, 21, 41。请用二路归并的方法对其进行归并。
 - (1) 构造最佳归并树。
 - (2) 根据最佳归并树计算每一趟和总的读记录次数。
 - (3) 在多路归并时，显然归并的路数 K 越大速度越快，那么是什么限制了 K 的大小？
4. 胜者树和败者树都是完全二叉树，是树形选择排序的一种变型。每个叶子结点相当于一个选手，每个中间结点相当于一场比赛，每一层相当于一轮比赛。不同的是，胜者树的分支结点记录的是胜者的标号，败者树的分支结点记录的败者的标号。

胜者树与败者树均可在 $\log(n)$ 的时间内找到最值。任何一个叶子结点的值改变后，利用分支结点的信息，还是能够快速找到最值。在 k 路归并排序中经常用到。

举例说明为什么败者树的访问外存次数要比胜者树少，并分析是什么原因造成的。
5. 假设一个记录长为 32 个字节，一个块长 1024 个字节（每个块有 32 个记录），工作内存是 1MB（还有用于 I/O 缓冲区、程序变量等的其他存储空间）。使用置换选择和多路归并，其中归并算法只允许扫描两遍。预计能得到的文件最长为多少？并解释这个结果是怎样计算出来的。