

外部排序(external sorting)和外部搜索(external search)

2019 年 3 月 13 日

1 外部排序 (external sorting)

2 Trie树

Trie树的关键码是可变长的。Trie树的度 ≥ 2 ，他的每层分支靠关键码的一部分决定。比如关键码为“chicken”的结点，在Trie树（根为NULL）的第一层的“c”子树中 \rightarrow “c”子树的“h”子树中 \rightarrow “h”子树中的“i”子树中...

3 外部排序

基于磁盘进行的排序多使用归并排序方法。其基本思路是：1) 将数据分段 (segment)，分别对每段进行排序，获得初始归并段或初始顺串 (*run*)，并将其写回到磁盘上；2) 采用归并树模式，把初始归并段归并。逐渐扩大归并段，同时减少归并段的个数，使之形成一个长段。

减少初始归并段个数 m (增加归并段长度)，或增加归并树路数 k ，即减少归并趟数 S 以减少读写磁盘次数 d ，达到提高排序速度的目的。

3.1 k 路平衡归并

在进行归并时，可以采用两路归并的方式 (将两路初始段，归并成有序的一路数据)，也可以采用 k 路平衡归并 (*k-way balanced merging*)。

参考文献

- [1] 严蔚敏. 数据结构 (C语言版). 北京: 清华大学出版社, 2007.
- [2] 邓俊辉. 数据结构 (C++语言版) (第三版). 北京: 清华大学出版社, 2013.
- [3] 李春葆. 数据结构考研指导. 北京: 清华大学出版社, 2002.
- [4] 殷人昆. 数据结构: 用面向对象方法与C++描述. 北京: 清华大学出版社, 1999.