

排序(sorting)

2019 年 3 月 13 日

1 排序 (sorting)

排序，一般是针对数据表 (*datalist*) 和排序码 (*key*) 进行排序。

如果相等的两值，在排序之后交换位置，则称这种排序方法是不稳定的；如果没有交换位置，则称这种排序方法是稳定的。

如果排序时，所有数据均可以加载到内存中，则称这时使用的排序方法为的内部排序；如果不能将所有数据加载到内存中，数据需要在内存和外存间移动，则称这时使用的排序方法为的外部排序。

排序可以通过顺序表或链表实现。在顺序表中，通过多次比较和移动可以将元素放置到合适的位置，使数组有序。这种排序称为静态排序，因为其结构没有发生改变。如果使用链表，可以通过改变链表指针链接关系实现排序。这种排序称为动态排序，因为其结构发生了改变。

衡量排序算法，需考虑时间开销和存储空间。

- 排序算法的时间开销主要受数据比较次数和数据移动次数影响。对于直插排序、起泡排序、选择排序，其时间开销为 $\Theta(n^2)$ ，对于如快速排序、归并排序和堆排序算法，其时间开销为 $\Theta(n \log_2 n)$ 。
- 排序算法也需额外的内存空间：1) 除自己所占空间外，不需额外空间；2) 需要额外的指针（下标）的存储空间；3) 需要额外的空间存储数据副本或是中间结果。

对于插入排序、快速排序、选择排序、归并排序，主要是基于数组实现。

2 插入排序 (insert sort)

其基本思想是：1) 首先保证前缀有序；2) 对于表中第2个元素，将其插入到自己前方的合适位置上；3) 重复上述过程，直至整个表有序。

采用插入排序思想的排序算法，又可以分为：

- 直接插入法，通过将当前元素和前缀元素依次比较，确定合适位置。
- 折半插入法，确定合适位置时，采用折半搜索（二分搜索）的方法。
- 希尔排序，选定间隔gap，挑选间隔为gap的多个元素组成子序列，对子序列进行直插排序；不断缩小gap，并重复上述过程，直至gap=1。

2.1 直接插入排序 (insert sort)

2.2 折半插入排序 (binary insert sort)

2.3 希尔排序 (shell sort)

3 快速排序 (quick sort)

4 选择排序 (select sort)

选择排序的基本思想是：对于某个位置，在其后缀元素中挑选出最小元素，放入到当前位置。

- 直接选择法，通过对后缀元素的依次比较，确定最小元素。
- 堆排序，借助堆，挑选出最小元素。
- 锦标赛排序 (tournament sort)

4.1 直接选择排序 (select sort)

4.2 堆排序 (heap sort)

5 归并排序 (merge sort)

归并排序主要思想为：1) 将两个等长的待排的有序子序列，归并成一个有序的长序列；2) 上述操作通过不断递归，直至变成对两个元素进行归并；3) 具体操作过程是：两个元素归并→两个长度为2 的子序列进行归并→两个长度为4的子序列进行归并...

6 基于链表的排序

7 分配排序 (sort by distribution)

与其他排序方法不同，其他排序方法均需要对关键码进行比较，而分配排序方法采用的是“分配”和“采集”的方法。

7.1 桶式排序 (bucket sort)

7.2 基数排序 (radix sort)

7.3 LSD基数排序

8 排序算法的分析与比较

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.