**位图索引的实现**

**设计思想**

假设需要建索引的属性为A，则位图索引就是对A的值域里每个值v建一个01数组，当第i行的A属性为v时，v对应的01数组的第i位为1，否则为0

举个例子：

对于一个表：

行号 姓名 性别 年龄

1 liameitimie 男 20

2 shaojifu 男 20

3 wydl 女 20

性别索引

男： 110

女： 001

年龄索引

20：111

而对于一些确定选值的选择运算，直接用对应选值的01数组求与，得到的01数组中为1的位置就是满足条件的行

可以发现位图索引对于一些询问可以得到满足条件的行的行号，而使用顺序文件存取就能通过行号快速地获取对应行的信息

**实现方案**

考虑使用动态开点的线段树来维护每个位向量，若一个节点存在，则该节点对应区间存在1，例如：

对于01数组： 0 1 0 1 0 0 0 1

线段树为这样：

O

/ \

O O

/ \ \

O O O

\ \ \

O O O

0 1 0 1 0 0 0 1

此处使用的线段树值域是二的幂且可以动态变化的，当添加一个行后，发现线段树的值域不够，则直接将值域乘二即可，在树的结构上的变化就是新建了一个节点作为根，原树作为新节点的左儿子

01数组求与的过程即是选出满足条件的线段树，从上往下dfs，当发现某个树当前节点为空时即可返回，可以发现这么做的时间复杂度跟满足条件的行的个数有关

设满足条件的行个数为m，总行数为n

若m<<n，则复杂度为O(mlogn)

若m与n规模相近，则复杂度为O(n)

对于查找，删除，修改操作，与b+树类似，代价均为O(logn)

**数据结构**

动态开点线段树，普通线段树的节点数与值域相关，动态开点线段树与点数相关，达到节省空间的目的

**文件结构**

线段树节点存储在磁盘中，用散列索引维护

原文件使用顺序存储，通过行号找到对应块来访问

**实现细节**

线段树可类比b+树，通过增加一个节点的指针数来减少树高，达到减少磁盘io数的目的