

《数据库系统概念》课程大作业中期报告

计65 赵鋈峰 2016011373

任务分工

单人小组

项目进度

- 使用了斯坦福课程CS346Database System Implementation提供的文件系统模块、其他模块的接口和测试¹
- 记录管理模块
 - 新建文件、打开文件、关闭文件、删除文件
 - 定长记录管理
 - 插入、删除记录
 - 使用Bitmap进行加速
 - 获取属性值满足一定条件的记录
 - 扫描全部记录
 - 支持变长字符串的存储
 - 专门存放在另一个文件中
 - 没有基于现有的文件系统，对文件指针进行暴力的移动，时间上不是很有优势，但关键在于可变长的概念对于存储空间的节省
 - 在定长记录中表示为offset+length的形式
- 索引管理模块
 - 新建索引、打开索引、关闭索引、删除索引
 - B+树

- 参考了开源的基于内存的B+树实现² 完整的节点插入、删除
- 结构的改变：不区分簇集索引和非簇集索引，所有索引存到B+树中的都是（key，RID）对（也就是value就是RID），但是每个节点同时还记录value，并且在比较大小时RID也作为第二键值。因为RID的唯一性，因此所有的（key，RID）对也是唯一的。因此不要求建索引的列的唯一性。
- 获取属性值满足一定条件的记录
 - 利用B+树进行快速检索
 - 支持=、!=、<、<=、>、>=
- 添加了大量的模块测试，包括边界和规模的corner case

存在困难

- 变长字符串的支持
 - 存储、检索、表示方式，由于和一般的存放位置不同，所以支持还存在着诸多没有考虑的问题
 - 可变长字符串列做索引
 - 目前的实现方式难以支持（连每个节点能放的key的个数都无法计算，如果用offset+length的存储格式可能过于麻烦而且对于磁盘的读写次数过高）
- 联合索引
 - 难以支持，每一列都需要不同的比较器，而且列数还不确定
- Bitmap实现
 - 我目前的实现方法是直接用的位数组，检索复杂度为线性，后续可以考虑用助教提供的文件系统中附带的bitmap实现

课程建议

- 课程内容应当尽量和作业内容相关
 - 我们大作业结构与MySQL重叠部分较多，但课上却一直在讲SQL Server的内容
- 对大作业的讲解过少
 - 对比斯坦福课程CS346对于大作业内容的介绍，我们的过于精简，难以下手
 - 如果能像CS346那样接口和各个模块的测试就最好了
 - 每个模块可以多花一节课来讲解、讨论，只有寥寥的几段文字连需求都很难搞清楚

1. - <https://github.com/huzecong/rebaseDB/tree/initial>↩

2. - <https://github.com/parachvte/B-Plus-Tree>↩