**实验5 查询优化与查询执行**

**一.实验完成情况**

1. SELECT语句的优化

(1) 基于启发式规则逻辑优化

(2) 基于代价的物理优化

a. 选择、连接、排序等操作符的操作算法选择

**二.实现功能**

1. 查询优化方法。

(1) 基于启发式规则逻辑优化

<1>投影下推

我们在查询分析时，查询数据字典，得到每一个表在后期操作所全部需要 的属性（包括条件where语句中出现的属性和select的属性），对每个表都在 这些属性上做出投影操作。这样可以降低表的大小。

<2>选择下推

在投影操作之前，将选择操作下推到每个表，使得表的大小显著降低，在 对选择后的数据进行连接等后续操作。

(2)连接、排序等操作符的操作算法选择

a.连接操作算法选择

<1> 如果两个表的数据量都不大，则选择nest loop连接方式

<2> 如果有表走索引，就选择index join 连接方式

<3> 如果数据量很大，就选择hash join 连接方式

<4> 如果数据量不是特别大，并且数据量差不多，就走merge sort join

b. 排序算法选择

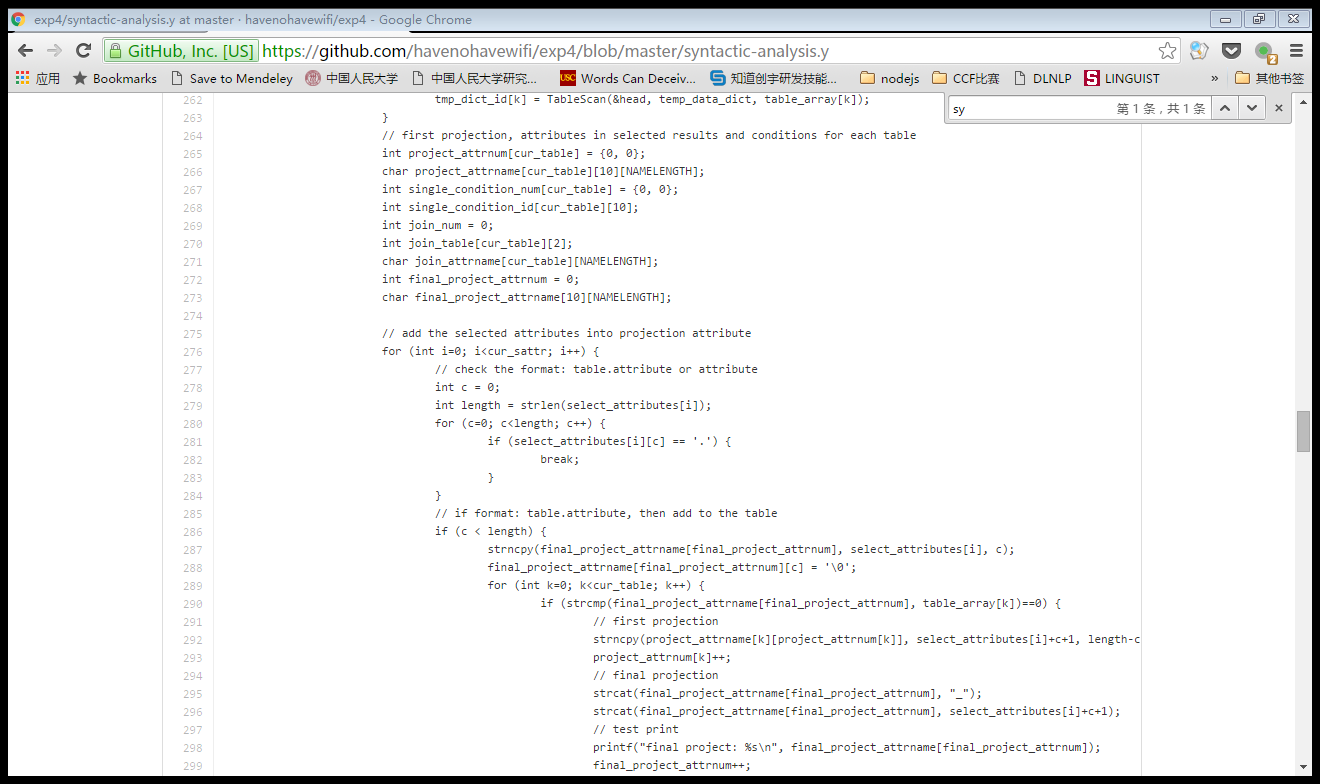
快速排序

2. 数据字典

保存在属性上是否建立索引，是否排序的信息

**三 分模块介绍**

1. 基于启发式规则逻辑优化

project\_attrnum[]数组：保存了对于每个表在整个操作中使用的全部属性个数

project\_attrname[][NAMELENGTH]数组：保存了对于每个表在整个操作中使用的全部属性名

single\_condition\_num[]数组：保存每个表单独的where选择条件个数

single\_condition\_id[]数组：保存每个表单独的where选择条件的id（可在全局选择条件变量中查询）

join\_\*：与连接条件相关，包括需要连接的次数、连接的表及其对应的连接属性

final\_project\_attrnum：用户定义的选择操作输出的属性

final\_project\_attrname[]数组：保存了连接后的表需要输出的属性

* 执行流程

查询分析时，根据数据字典和对SQL语句的分析得到上述数据结构。然后依次执行：对各个表所需属性的投影🡪对各个表单独的选择操作🡪多个表的连接操作🡪在输出的属性上进行投影。

2. 连接选择算法

int choosejoin(struct dbSysHead \*head, relation \*temp\_datadic,int old\_relation\_1, int old\_relation\_2, const char \*name);

参数意义：

head: 头文件

temp\_datadic: 临时数据字典数组

old\_relation\_1: 第一个表在临时数据字典数组的位置

old\_relation\_2: 第二个表在临时数据字典数组的位置

name: 表连接的属性名字

函数调用：

indexjoin,hashjoin,mergesortjoin,nestloop

;

3. 排序算法选择（梦溪）

struct sortNode

{

int value;

char \*record;

};//结构体，用来对每一条记录排序，value为所要排序的属性值

具体的排序函数见tempSort()函数

int tempSort(struct dbSysHead \*head, relation \*temp\_datadic, int old\_relation, char \*attr)

功能：对buffer中的表记录进行快速排序

参数：head数据库头，temp\_datadic临时数据字典数组，old\_relation临时表的位置，

attr属性名

返回值：1排序成功，-1排序失败

4.数据字典

添加相关操作，用来判断某个属性上是否排序，是否建立索引等，用于数据库优化

void changeIndexedByName(char\* name, bool s);

功能：修改Indexed值

参数：name属性名，s是否建立索引（true/false）

返回值：空

void changeOrderedByName(char\* name, bool s);

功能：修改Ordered值

参数：name属性名，s是否排序（true/false）

返回值：空

bool getIndexedByName(char\* name);

功能：得到Indexed值

参数：name属性名

返回值：true已建立索引，false未建立索引

bool getOrderedByName(char\* name);

功能：修改Ordered值

参数：name属性名

返回值：true已排序，false未排序