

实验题目	查询执行器实现			实验日期	2022/5/1
班级	1903102	学号	1190200122	姓名	袁野

# CS33503 数据库系统实验

## 实验检查记录

实验结果的正确性(60%)		表达能力(10%)	
实验过程的规范性(10%)		实验报告(20%)	
加分(5%)		总成绩(100%)	

## 实验报告

### 一、实验目的（介绍实验目的）

1. 掌握各种关系代数操作的实现算法，特别是连接操作的实现算法。
2. 在实验 3 完成的缓冲区管理器的基础上，使用 C++面向对象程序设计方法实现查询执行器。

### 二、实验环境（介绍实验使用的硬件设备、软件系统、开发工具等）

- 1、硬件设备：  
i7-9750H CPU@2.60GHz 2.59GHz; 1.8GHz; 16G RAM;
- 2、软件系统  
Ubuntu 20.04.4 LTS (GNU/Linux 5.10.102.1-microsoft-standard-WSL2 x86\_64)
- 3、开发工具  
G++

### 三、实验过程（介绍实验过程、设计方案、实现方法、实验结果等）

#### 3.1 算法设计

在本次实验中，我们需要使用基于块的嵌套循环连接算法来实现对两个关系的自然连接。

基于块的嵌套循环连接算法就是一种朴素的暴力枚举算法。该算法会将两个关系中较小的关系（外关系）成批的读入内存中，之后将较大的关系（内关系）依次读入内存中。枚举两个块中的所有元组，之后尝试进行自然连接。如果连接成功，就将其插入到结果关系中。

实验题目	查询执行器实现			实验日期	2022/5/1
班级	1903102	学号	1190200122	姓名	袁野

#### Algorithm 1: 基于块的嵌套循环连接算法

**Input:** 两个待连接的关系  $S$  和  $R$ , 以及可用的缓冲块数目  $M$

**Output:** 两个关系自然连接的结果

```

1 if  $B(S) > B(R)$  then
2   | 交换  $S$  和  $R$ 
3 end
4 for 外关系  $S$  的每  $M-1$  块 do
5   | 将这  $M-1$  块读入缓存块
6   | 利用内存查找结构来组织  $M-1$  块中的元组
7   for 内关系  $R$  中的每一个元组  $P$  do
8     | 将  $P$  读入缓冲区
9     | for  $P$  中的每一个元组  $r$  do
10      |   | for 内存查找结构中能与  $r$  进行自然连接的元组  $s$  do
11        |   |   | 连接  $r$  和  $s$ , 将结果写入缓冲区
12      |   |   end
13    |   end
14  end
15 end
16 return

```

## 3.2 代码实现

我们只需要编辑 `executor.cpp` 文件, 实现 `NestedLoopJoinOperator` 类的 `NestedLoopJoinOperator::execute` 方法就可以完成本次实验。

### 3.2.1 getTableSize ()

该函数的目的是计算关系块的数目, 输入是一个文件, 然后遍历文件中的所有页, 然后统计数目并返回, 用于在 `NestedLoopJoinOperator::execute` 中判断关系的大小, 从而判断内关系和外关系。

### 3.2.2 didit ()

该函数用来对某一个关系进行解析, 由于元组是经过序列化的, 所有的值均是以二进制字节码的方式存储的。因此为了便于后续的连接, 我们通过关系的模式来将一个二进制码解析为一个二进制列表来便于访问。该函数仿照 `TableScanner` 的 `print` 方法, 按照三种属性值来实现对二进制码的分割, 将对应的部分存储在一个线性表中。这样就可以通过下标实现  $O(1)$  的查找。

### 3.2.3 NestedLoopJoinOperator::execute ()

这部分的代码就是用 C++ 语言实现了上面给出的伪代码。

首先用 `getTableSize` 函数比较两个关系块的大小, 判断内关系外关系, 然后用 `didit` 函数将关系的二进制码转换为二进制序列, 然后枚举结果关系模式, 来进行连接。枚举每一个属性, 并用 `hasAttr` 方法判断  $S$  或  $R$  中是否包含该属性, 如果  $S$  和  $R$  都包含, 那么判断两个元组这个属性的值是否相等, 如果相等则将其接在结果元组的二进制码中, 然后进行后面

实验题目	查询执行器实现			实验日期	2022/5/1
班级	1903102	学号	1190200122	姓名	袁野

的判断，否则我们认为连接失败，终止这两个元组的连接。如果之后 S 或 R 中的一个关系中包含这个属性，那么直接将对应值连接在结果元组的二进制码中即可。如果二者都不包含这个属性，那么报错。

如果连接成功，那么利用 HeapFileManager 的 insertTuple 方法将结果元组插入堆中。

#### 四、实验结论（总结实验发现及结论）

在本次实验中，我加深了对于基于块的嵌套循环连接算法的理解，初步了解了二进制序列的存储元组的具体方式。通过一些代码了解了获取关系中的属性、元组的内部表示方法、元组在文件中的存储方式的具体细节。