数据库实验三 BenSQL 完整性数据库实现

钟辉 陈岑
PB07210003 PB07210013
中国科学技术大学 2010.6.8

程序背景

本段讲述与程序相关的一些背景知识。它们有助于理解 BenSQL 系统。

实验内容

本程序是数据库系统课程实验的一部分。数据库系统课程共有三个。在前两实验中已经完成
SQL 语言的基础训练，掌握了数据库用法。本实验将要求用 C ++ 实现数据库完整性限制
完整性限制语法可以类似 SQL DDL 中的要求。完整性功能可以选择（如主键完整，外键完
整，空值，引用完整，自定义完整等等）。

需要实现的语句

1. 建表及 Load 表记录的语句
2. 显示已建表数据结构和数据的语句
3. 完整性定义语句
4. 更新语句（insert、delete、或 update）

设计思想

我们希望建立一个功能强大的数据库系统。它不仅能数据查询，而且具有一定的编程能力
这样一些数据应用领域就不必依托赖其它语言。数据库语言的发展也不必受限于其它语言
这是很伟大一个理想，本实验不可能完全实现它。我们希望我们的实验能朝这个方向努力
作一些基础工作。
为了达到上述目标，我们需要模块我们的数据库系统以方便后续扩展。另外，为了支持后续
编程功能的开发，我们需要一个强大的语句解析器，以方便 SQL 语言的扩充。
内存管理是编程语言的核心部分，我们需要保证数据库系统内存管理的高效性。
文件是数据库语言的核心内容。我们希望数据库能很完美地处理文件问题。

BenSQL 简介

BenSQL 特点

支持复杂表达式计算
（独立的表达式计算模块鼎力支持）
高效的数据读取
（卓越的内存管理，文件管理技术保证）
强大的语句解析能力
（昱姐的编译技术驱动护航)
弱数据类型
（同上）

BenSQL 新理念

全能数据库
不仅仅数据查询（好 Ben 哦！）
编程能力
表达式缓存（使用时才计算，Pi）
数据处理
（加密，压缩，转换等，与编程语言无关）

使用方法

BenSQL.exe
或者运行时导入输入 BenSQL 语句
BenSQL.exe　source.ben

逻辑结构

BenSQL 由七个模块组成，关系如下。

- 输入/输出

BenSQL 与用户的交互接口
字符输入输出重定向支持

- 解释器

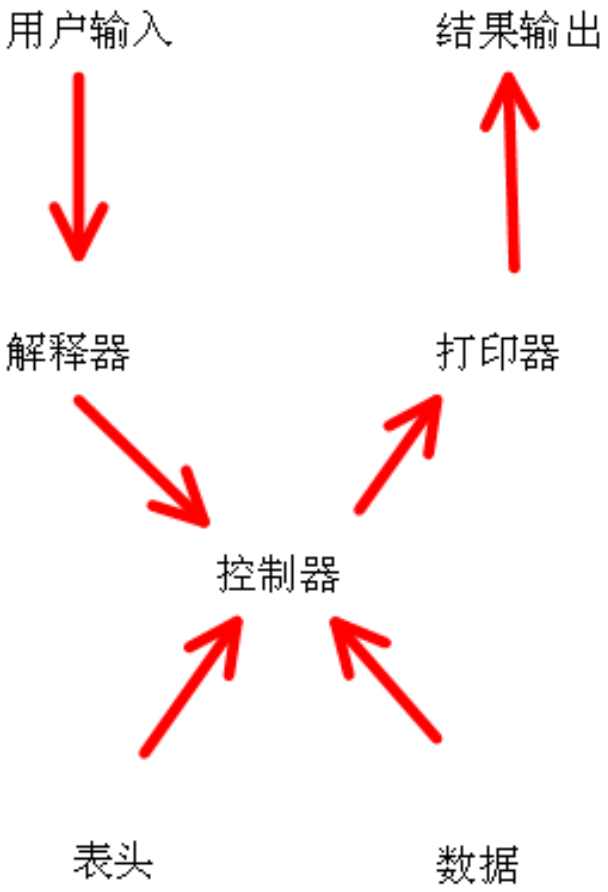
对输入的 BenSQL 命令进行词法、语法分析。
排除语法错误
把格式化后的语法树传给控制器。

- 控制器

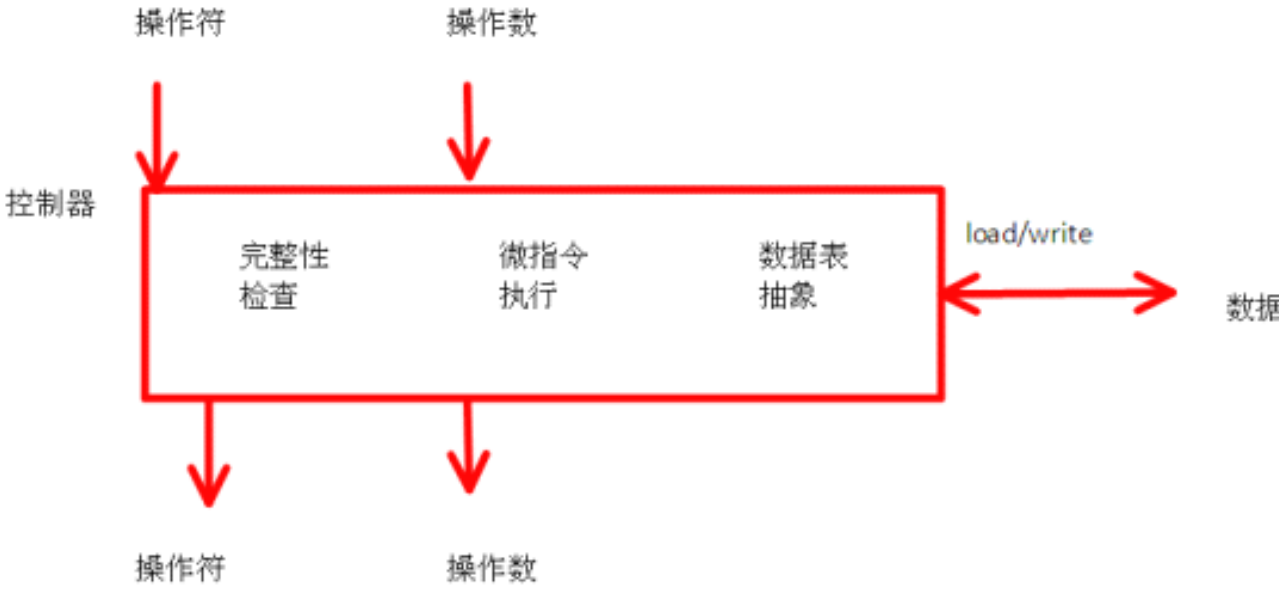
解析指令，分解成能被表操作的微指令。
完整性检查。
控制微指令执行

- 数据访问

文件读写
内存管理
抽象成数据表



BenSQL 的核心部件的接口如下面所示



控制器，解释器和打印器构成 BenSQL 的一个内核。BenSQL 系统核与核之间独立运行，这样能实现嵌套查询，复杂指令。

指令格式

指令集

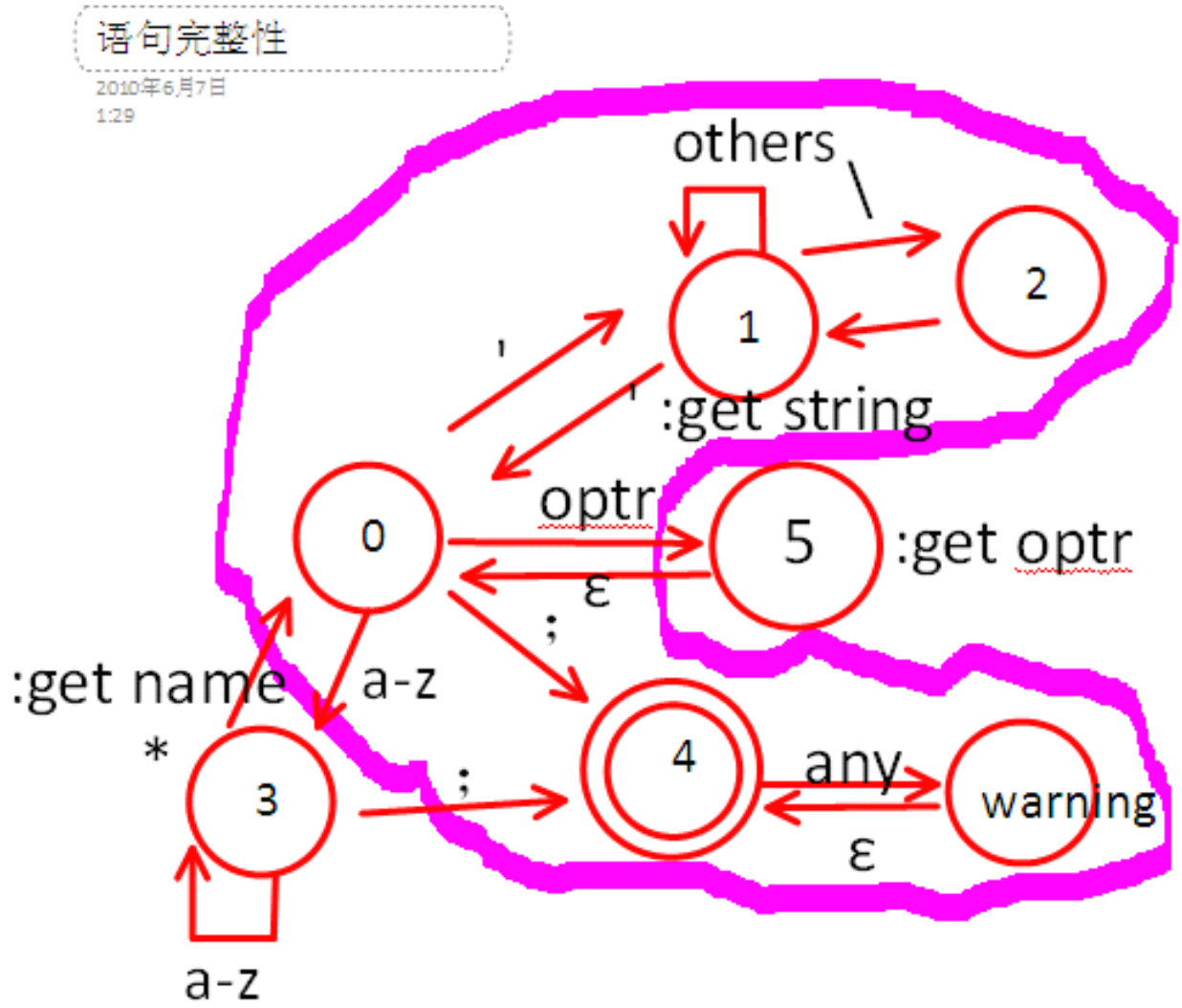
操作数	操作符	操作数(table略去)
0	SELECT	condition, cols
1	INSERT	value
2	UPDATE	condition, action
3	DELETE	condition
4	CREATE	tableInfo
5	DROP	
6	SHOW	
7	LOAD	
8	Quit	
9	Exit	

限制性修饰符

关键字	含义	备注
notnull	非空	
key	关键字	可多重
integer	正整数	

解释器

解释器负责把输入字符流解析成控制器能接受的字符标记（Token）或串（String）



如上图中，解释器可以识别出"name", "optr", "string".
比如 String："this is a string"
比如 Token：this_is_a_token
解释器返回给控制器的是一个结构体
(type, table, arg)
type 给出语句命令的类型；
table 给出语句操作的表；
arg 给出语句操作需要的参数。

表

表有三个域

Name	表的名字
Buffer	内存空间
Information	表信息（列数，列信息）

表可以有两种操作，原子操作和复合操作。

原子操作

Rename	重命名表
Read	从文件读表
Write	写表到文件
Show	打印表结构
Insert	插入元组到表
Select	从表中选择元组
Delete	从表中删除元组
Update	更新表

复合操作

Create	创建表
Load	导入表
Drop	删除表

复合操作是通过一组原子操作来实现。

比如 Create 是通过在元数据表 BenSQL 中 Insert 表名及引用信息来完成的。

文件

每个表对应一个文件，文件头存储表信息，接着是列信息，最后是表数据。

表信息（TableInfo）存储的数据有

数据类型	名称	含义
int	colWidth	行数
int	colNum	列数
ColInfo*	info	列信息

列信息（ColInfo）存储的数据有

数据类型	名称	含义
char*	name	列名
int	length	列长度
int	index	列相对行的位置

表数据（char buf[]）存储的数据有

行 1
行 2
。。。