

CTDPS 查询语言文本与实现

陈启泉 吴扬扬 郑跃斌

[计算机科学(电脑)系]

摘 要

本文主要介绍 CTDPS 的查询语言及实现技术。语言文本的设计基本上是结构英语查询语言 SEQUEL 2 的子集, 查询表达能力较强而又非过程化, 易为广大用户掌握。实现时除考虑各种情况的嵌套查询块、多表的复合映射和连接外, 为提高查找速度还采取了一系列的优化措施。

CTDPS 是在微机“北极星”上设计并实现的汉字事务数据处理系统。本系统兼具有信息管理和数据处理功能。系统的设计技术基本上应用了关系数据库管理系统 systemR 的设计思想。允许用户定义任意个基本表, 各基本表间的信息联系通过元组的连接运算实现。用户在一个查询语句中可涉及多个基本表。系统提供使用交互式 and 查询语言两种方式, 查询语句可独立使用, 也可嵌入到宿主语言 MBASIC 中。有关本系统的总体设计及其功能介绍将在《计算机应用与软件》上发表。本文主要介绍 CTDPS 查询语言的实现方法。

一、CTDPS 查询语言文本

CTDPS 查询语言文本可用 BNF 表示如下:

〈数据操纵〉 ::= 〈信息检索〉 | 〈更新元组〉 | 〈删除元组〉 | 〈数据处理〉
〈信息检索〉 ::= 〈文件名〉 = [L] 〈查询块〉 | [L] 〈查询块〉
〈查询块〉 ::= select 〈检索目标〉 From 〈基本表名列〉 [where 〈条件〉]
〈检索目标〉 ::= 〈目标列〉 | * | count (*)
〈目标列〉 ::= 〈检索项目〉 | 〈检索域段〉 | 〈检索项目〉, 〈目标列〉 | 〈检索域段〉,
 〈目标列〉
〈检索项目〉 ::= 〈检索域〉 | 基本表名 > . 〈检索域〉 | 〈库函数〉 (〈检索域〉)
〈检索域〉 ::= 〈域名〉 | 〈域号〉
〈检索域段〉 ::= 〈域号〉 — 〈域号〉 | 〈域名〉 — 〈域名〉
〈库函数〉 ::= AVG | SUM | MAX | MIN

本文 1985 年 1 月 15 日收到。

<基本表名列>::=<基本表名>,<基本表名列>|<基本表名>

<条件>::=<简单条件>|<复合条件>|<极值条件>

<简单条件>::=[<基本表名>]•<检索域><关系运算符><字值>|<基本表名1>
•<检索域>

=<基本表名2>•<检索域>|<检索域>IN<集合>

<集合>::=《字值列》|<查询块>

<字值列>::=<字值>|<字值>,<字值列>

<字值>::=<数值字值>|{<非数值字值>}

<关系运算符>::=AND|OR

<极值条件>::=MAX(<检索域>)|MIN(<检索域>)

<更新元组>::=update<更新域列>From<基本表名>[where<条件>]

<更新域列>::=<检索域>

=<字值>|<检索域>

=<字值>,<更新域列>

<删除元组>::=Delete From<文件名>[where<条件>]

<数据处理>::=Execute<数据处理定义>From<基本表名>[where<条件>]

例如,对于学校教师实体,定义了这样三个基本表:

人员基本表、部门基本表和工资基本表。

人员基本表

职工号	姓 名	住 址	性 别	出生年月	职 称	...	工作部门号
70239	张英	一号楼	男	1944年3月	讲师	...	020
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

部门基本表

部门号	部门名称	所属系别	人 数	工资总额	...
020	软件教研室	电脑	18
⋮	⋮	⋮	⋮	...	⋮

工资基本表

职工号	工资级别	基本工资	扣 除 金 额				补 贴			实发工资
			房租	水费	电费	其它	付食	粮油	其它	
70239	高教10	80	1.20	1.50	1.40	0.40	10	3	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

我们可进行如下操作:

例1 找出电脑系所有人员的姓名、职称

Select 姓名, 职称

From 人员

Where 工作部门号 IN

Select 部门号

From 部门

Where 所属系别 = { 电脑 }

例2 找出所有人员所属的系别

Select 人员. 姓名, 部门. 所属系别

From 人员, 部门

Where 人员. 工作部门号 = 部门. 部门号

例3 求 0010 号部门有多少讲师

Select count(*)

From 人员

Where (工作部门号 = { 0010号 }) AND (职称 = { 讲师 })

例4 求工资总额最多的部门名称

Select 部门名称

From 部门

Where MAX(工资总额)

例5 求电脑系职工的平均工资並在行印机上打印输出

Lselect AVG(基本工资)

From 工资

Where 职工号 IN

Select 职工号

From 人员

Where 工作部门号 IN

Select 部门号

From 部门

Where 所属系别 = { 电脑 }

例6 把职工号等于 70239 的基本工资修改为 83 元

update 基本工资 = 83

From 工资

Where 职工号 = { 70239 }

为了使输入语句嵌套层次分明, 系统允许用户使用换行符, TAB 键移动光标或连续多个空格作分隔符。

二、语言解释程序的实现方法

解释程序的结构可用框图表示(图1)。

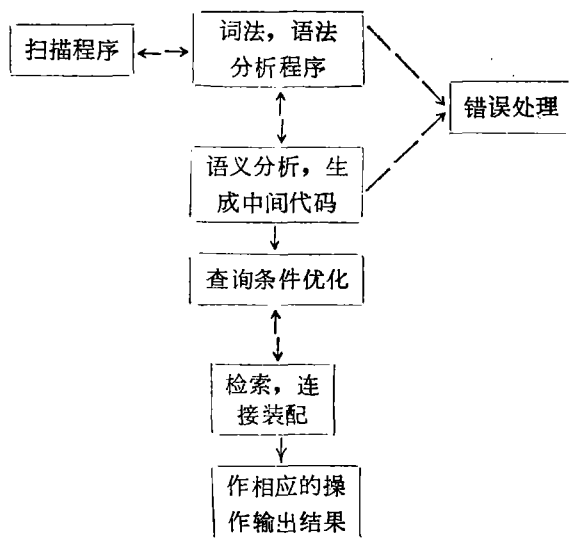


图1

1. 词法分析和语法分析(略)

2. 语义分析、生成中间结果

为了保证检索处理的正确执行，通过词法和语法分析后，系统还必须对查询语句的语义进行分析检查。

如对上面对描述的三个表，查询语句：

```

Select  人员. 职工号, 部门. 部门号, 工资. 实发工资
From    人员, 部门, 工资
Where   ( 人员. 职工号 = 工资. 职工号 ) AND ( 部门. 部门号 = { 0020 } )
  
```

从语法、词法上分析是没错的，符合文法，但这个语句不能实现，因为基本表部门无法与人

员、工资这两个基本表连接，无法投影映射。所以系统必须对连接条件等进行分析检查。

为了便于检索和优化，解释程序把通过词法、语法、语义检查被认为正确的语句转换成一个二维的代码表：

其中：①操作信息包括操作类型，多表或单表标志，有无条件标志等。

②条件i有关信息有：条件类型，条件域类型，有无索引标志，逻辑链，压缩存储标志等。

操作信息
条件 1
条件 2
⋮

3. 优化

本系统采用的是以条件表达式变换为主的优化方法，主要原则为：

①同时进行一串运算，对同一基本表的多个条件，尽可能同时处理，减少扫描次数。

②对每一嵌套层均采用先选择，次连接，最后投影的原则。

③尽可能利用索引，如果限制条件表达式中每个布尔合取因子都至少有一个条件域有索引就使用索引。若有多个条件域有索引，尽量选择数值型的索引进行查找。顺序查找时，所有条件同时处理。

④对于多表连接，先处理限制条件多且有索引的基本表；对具有相同基本表的多个连接条件(若有多条的话)同时进行处理。

为了实现上述优化,系统在检索前对中间代码进行一些预处理,在不改变条件表达式结果值的前提下,对条件排列次序作调整。

4. 检索、处理

①如果目标列只涉及一个表(整个查询语句可能涉及多个表,因为不同嵌套层的基本表可能不同)。我们采用由下而上逐层处理的方法。从最底层开始每处理完一层,都把结果值作为一个中间文件保存,然后把该层转换成一个简单条件,作为最靠近它的外层的条件之一进行处理。

目标列为单表的情况可用粗框表示(图2)。

一个嵌套层的所有条件处理,如图3所示。

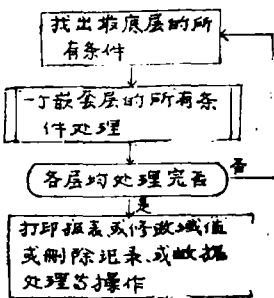


图 2

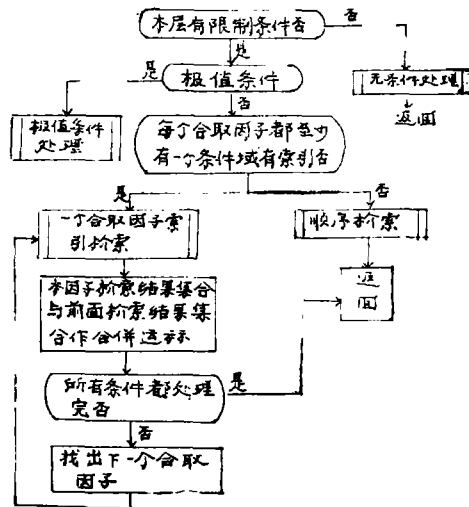


图 3

②多个基本表复合映射:

由于内存空间的限制,我们只能一个表一个表地进行处理,把每个基本表满足条件的所有记录找出来,放在中间结果文件,然后再进行连接、装配。每处理完一个表,系统都把该表与其他表的连接条件转换成对应基本表的一个限制条件。这样能保证下一个表检索出来的记录一定是连接装配时所需要的,排除一切无用的中间结果,节省空间,提高速度。

多表连接处理可用粗框表示(图4)。

本系统采用高级语言编程,内存空间小受到很大限制,整个解释程序分成三块,运行时各模块得调进调出,因而速度受到一定的影响。

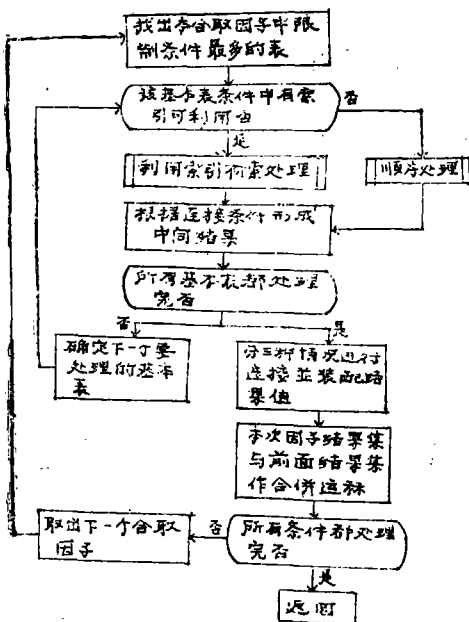


图 4

参 考 文 献

- [1] M. MAstrahan and D. DChamberlin, Implementation of a structured English Query Language, IBM Research Division, San Jose, (1975).
- [2] E. F Codd, A Relational Model of Data for Large shared Data Banks, IBM Research Laboratory, San Jose, California, (1970).

CTDPS Query Language and Implementation

Chen Qiquan Wu Yangyang Zheng Yuebin

Abstract

CTDPS is Chinese character transaction data processing system that had been designed and implemented on micro-computer "North Star". It has both facilities of information management and data processing. The overall system design and facilities will be given on «Computer applications and software». This paper introduces CTDPS query language and implementation techniques. The language text is basically a subset of Structured Query English Language SEQUEL2. Its ability of expressing query is superior and non-procedure, easily for users to master. In implementation, besides considering the nest query blocks in many cases, the complex mapping and join of multi-tables, it also takes a series of optimization measures for increasing search rate.