ISSN 1673-9418 CODEN JKYTA8
Journal of Frontiers of Computer Science and Technology
1673-9418/2008/02(01)-0097-07

E-mail:fcst@public2.bta.net.cn
http://www.ceaj.org
Tel:+86-10-51616056

关系代数派生算子语义表达式间等价性证明*

杨 波[1,2,3+],薛锦云[1]

YANG Bo[1,2,3+], XUE Jinyun[1]

1. 江西省高性能计算技术重点实验室(江西师范大学),南昌 330022
2. 中国科学院 软件研究所,北京 100080
3. 江西财经大学 信息管理学院,南昌 330013
1. Jiangxi Key Lab of High Performance Computing Technology, Nanchang 330022, China
2. Institute of Software, Chinese Academy of Sciences, Beijing 100080, China
3. School of Information, Jiangxi University of Finance and Economy, Nanchang 330013, China
+ Corresponding author: E-mail: jxncyangbo2002@163.com

YANG Bo, XUE Jinyun. Proof of equivalence between semantic expressions of derived relational algebra operators. Journal of Frontiers of Computer Science and Technology, 2008,2(1):97-103.

Abstract: The derived relational algebra operators are widely used in the relational database query languages. There are two kinds of representations for their semantics. One is the expression described by original relational operators, the other is the expression described by one order logic. However, the strict equivalence proof of the two different semantic expressions is not given in related documents. The present study manages to prove the equivalence of the two different kinds of semantic expressions through a series of formal deductions step by step according to some related characteristics of relational algebra. The beginning of these deductions is the semantic expression described by original relational operators; the end is the semantic expression described by one order logic. The foundation of correctness proof of relational algebra expressions can be laid by the present deduction method.

Key words: derived relational algebra operators; original relational algebra operators; one order logic; relational division operator; formal deductions

摘 要: 关系代数的派生算子在关系数据库查询语言中得到了广泛应用。它们的语义有两种常见的表示方式,一种是基于原始算子的表达式,一种是基于一阶逻辑的表达式。但有关的文献资料都没有给出这两种表达式

* the National Natural Science Foundation of China under Grant No.60573080 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.2003CCA02800 (国家重点基础研究发展规划(973)前期研究专项).
Received 2007-08, Accepted 2007-10.

等价性的严格证明。文章尝试通过一系列等价变换,证明派生算子语义的这两种表达式间的等价性。从派生算子(主要是除算子)语义的原始算子表达式出发,根据关系代数表达式的特点,通过一步步的等价变换,得到派生算子语义的一阶逻辑表达式。所使用的变换方法能为关系代数表达式的正确性证明打下基础。

关键词:关系代数的派生算子;原始算子;一阶逻辑;除算子;等价变换

文献标识码:A **中图分类号:**TP311

1 引言

关系代数是研究关系及其运算的代数理论[1]。关系代数在关系数据库的设计理论、查询语言的设计[2]与查询优化[3]等领域有重要应用。

关系代数的算子可分为原始算子和派生算子。原始算子有五个,分别是:并(∪),差(-),笛卡尔积(×),选择(σ),投影(π)等;派生算子有四个,分别是:关系的交(∩)、θ连接($\rhd_F\lhd$)、自然连接(⋈)、除(÷)等,它们的定义基于原始算子。这些派生算子表达能力强,在查询语言中得到了广泛应用[1]。

原始算子语义的表达式只有一种,它们是用一阶逻辑式给出的[4]。但派生算子语义的表达式除了一阶逻辑表达式之外,还有一种表达式,这种表达式是用基本算子组成的关系代数表达式来表示派生算子的语义。有关的文献资料中都没有给出派生算子语义的这两种表达式之间等价性的形式化证明,至多只是用自然语言非形式化地解释了一下二者为什么等价。

本文从原始算子的语义表达式出发,结合关系代数的相关特点,基于一阶逻辑演算,经过一步步等价变换得到全部派生算子(重点在于除算子)语义的一阶逻辑表达式,从而形式化地证明了关系代数派生算子语义的两种不同表达式之间的等价性。本文所使用的变换方法能为关系代数表达式的正确性证明打下基础。

在本文接下来的叙述中, 首先介绍相关工作,然后列出原始算子语义的表达式,这些语义表达式[5]可以作为等价变换的出发点;然后分析了关系代数表达式的相关特点[5],根据这些特点,给出变换过程中应遵循的规则和引理;由交、θ连接、自然连接、除算子四个派生算子语义的原始算子表达式变换得到了它们语义的一阶逻辑表达式;从而严格证明了这两种不同表达式的等价性。最后进行了一些讨论。

2 相关工作

关系代数除法是关系代数的一个较复杂的派生算子。多数有关数据库原理的文献资料都会通过组合原始算子,给出除法语义的原始算子表达式。如下式[5]:

$$A\div B=\prod\{AD\}(A)-\prod\{AD\}((\prod\{AD\}(A)\times \prod\{A.a-AD\}(B))-A)$$

其中,$A.a$,$B.a$ 分别表示关系 A 和 B 的属性集合,$AD=A.a-B.a$ 表示关系 A 和关系 B 属性集的差。

从这个很难看出除算子在做一件什么事情。因此有关文献通常会用一阶逻辑给出其语义的另一种表达式。例如 E.F.Codd 给出了除算子的一阶逻辑表达式,如下式[6]:

$$R[A\div B]S=\{r[\bar{A}]: r\in R\wedge S[B]\subseteq g_R(r[\bar{A}])\}$$

其中$\bar{A}$表示属性集 A 和属性集 B 的差;$g_R(r[\bar{A}])$表示 $r[\bar{A}]$在 R 中的象集,其形式定义为:$g_R(x)=\{y|<x,y>\in R\}$[6]。Codd 在文献[6]中没有说明除算子语义的两种不同表达式为什么等价。Silberschatz 在文献[5]中非形式化地说明了除算子语义的两种不同表达式为什么等价,但这种说明是文字叙述,不是严格的形式化变换。

除文献[4-6]外,其他相关资料中也未见如本文所述的派生算子语义表达式间的等价性证明方法。例如,文献[7]研究了判断不同关系代数表达式运算结果是否相等的方法。其中提到要完全判断不同关系代数表达式运算结果是否相等是个 NP 难问题,因此仅

提出了判断关系代数表达式中某些重要的子集是否相等的方法，并且这里关系代数表达式仅限于使用σ、π、×三个算子，以便降低问题的复杂度。这个方法首先分别用矩阵（是 Zloof 提出的"Query-by-Example"语言的一种形式）来表示关系代数表达式中的重要子集，然后通过一个算法判断它们的结果是否相等。没有涉及关系代数的派生算子，也没有涉及派生算子语义表达式间的等价问题。文献[8]给出了算法用于判断关系代数中连接运算的结果是否达到预期的要求。没有涉及派生算子语义表达式间的等价问题。文献[9，10]系统介绍了关系数据库理论，没有提到如本文所述的派生算子语义表达式间的等价性证明方法。Ullman 在文献[11]中叙述了关系代数表达式如何转换成关系逻辑表达式，直接给出了∩、∪、-、×、σ、π算子对应的关系逻辑表达式，没有给出派生算子（尤其是除算子）语义表达式间的等价性证明。

本文目的是尝试通过一系列等价变换，建立关系代数除算子以及其他派生算子语义表达式间严格的逻辑联系，证明它们的等价性。原始算子和派生算子的语义共同形成关系代数算子的语义模型。有了这个语义模型，结合本文提出的变换规则引理和方法，就可以进一步讨论关系代数表达式的正确性证明。

3 原始算子的形式化语义[5]

关系代数五个原始算子：∪，-，×，σ，π 语义表达式在文献[5]中已给出，因为它们是后面等价变换的出发点，为了阐述方便，在此一一列出。

（1）∪：$A\cup B=\{t|t\in A\vee t\in B\}$

（2）-：$A-B=\{t|t\in A\wedge t\notin B\}$

（3）×：$A\times B=\{t|(t=<t1,t2>)\wedge t1\in A\wedge t2\in B\}$

[$<t1,t2>$表示 $t1$ 和 $t2$ 的并置]

（4）σ：$\sigma_{\{F\}}(\mathrm{A})=\{t|t\in A\wedge F(t)\}$

[F 是条件表达式谓词]

（5）∏：$\prod_{\{E\}}(A)=\{t[E]|t\in A\}$ [E 是投影的属性组，$t[E]$表示元组 t 在属性组 E 上的投影]

4 变换规则和引理

读者在后面将看到，派生算子语义的不同表达式等价性证明过程实际上就是原始算子语义的嵌套组合变换过程，这种变换要在一定的规则和引理下进行。首先分析了关系代数表达式的相关特点，然后根据这些特点，提出变换过程中应遵循的规则和引理。

4.1 关系代数表达式的特点

关系代数算子的运算对象是关系，运算结果是新的关系，这个新的关系又可作为算子的运算对象，再运算得到新的关系。这个过程可以一直嵌套进行，这一特点是后面进行嵌套组合变换的依据。关系代数五种原始算子按照这种方式与关系相互组合，形成各种各样的关系代数表达式，所有的关系代数表达式都是这样形成的。

派生算子同样是由原始算子与关系经过嵌套组合运算得到的。所以派生算子的语义可由原始算子的语义以及组合方式确定。派生算子语义的一阶逻辑表达式可以由原始算子语义的表达式经过嵌套组合的方式形式化地变换出来。

原始算子的语义表示式只有一种，是用一阶谓词逻辑来描述的，描述分为两部分。"|"左边的是结果部分，可以是一个元组变量或元组变量的投影，这个变量（以后称为结果变量）代表一个元组，作为运算结果，通常用 t 或 t 在某些属性上的投影表示。"|"右边是一个一阶谓词逻辑表达式，形式化地描述了结果变量应该满足的条件。

本文根据这些特点，结合一阶逻辑的推演，提出变换过程中的一些规则和引理，将派生算子语义的两种表达式等价性证明转化为一阶逻辑的推演。

4.2 变换规则

下面是本文提出的两条规则，它们是变换的基础。

规则 1 设 $E1=\{t|F1(t)\}$，$E2=\{t|F2(t)\}$；其中 $E1$、$E2$ 是关系代数表达式，$F1(t)$、$F2(t)$是谓词逻辑表达式，作用是用来确定 t 的取值范围，若 $F1(\mathrm{t})$和 $F2(t)$确定的 t 的取值范围相等，则 $E1=E2$。

规则 2 若 $E1=\{t|F1(t)\}$，且 $F1(t)$中不含 t'（t'是任意不同于 t 的变量名），则 $E1=\{t'|F1(t')\}$。

4.3 引理

本文在上述规则的基础上提出并证明若干引理，这些引理将用于第5、6章的证明过程中。

引理1 设 $E1=\{t|F1(t)\}$，$E2=\{t|F2(t)\}$；其中 $E1$、$E2$ 是关系代数表达式，$F1(t)$、$F2(t)$ 是谓词逻辑表达式，若 $F1(t)\equiv F2(t)$，则 $E1=E2$。

证明 当 $F1(t)\equiv F2(t)$ 时 $F1(t)$ 和 $F2(t)$ 确定的 t 的取值范围相等。由规则1可知，$E1=E2$。 □

引理2 若 $E1=\{t|(t=f(x,y,\cdots))\wedge F1(f(x,y,\cdots))\}$，则 $E1=\{t|F1(t)\}$。其中 $E1$ 是关系代数表达式，$F1$ 是一个谓词，$t=f(x,y,\cdots)$ 是一个满射，$f(x,y,\cdots)$ 中不含 t。

证明 令 $F2(t)$ 为：$(t=f(x,y,\cdots))\wedge F1(f(x,y,\cdots))$，则 $E1=\{t|F2(t)\}$。

因为 $t=f(x,y,\cdots)$，所以 $F2(t)\equiv(t=f(x,y,\cdots))\wedge F1(t)$。

因为 $t=f(x,y,\cdots)$ 是一个满射，所以 $(t=f(x,y,\cdots))$ 在 $F2(t)$ 中对于 t 的取值范围没有影响；所以 $F1(t)$ 和 $F2(t)$ 确定的 t 的取值范围相等；由规则1可知，$E1=\{t|F1(t)\}$。 □

引理3 $E1$、$E2$ 是关系代数表达式，$F1(t)$、$F2(t)$ 是谓词逻辑表达式，t 是 $E1$ 的结果变量，$E2$ 的结果变量中不包含投影动作。

若 $E1=\{t|F1(t)\wedge t\in E2\}$，$E2=\{t|F2(t)\}$，则 $E1=\{t|F1(t)\wedge F2(t)\}$；

若 $E1=\{t|F1(t)\wedge t\notin E2\}$，$E2=\{t|F2(t)\}$，则 $E1=\{t|F1(t)\wedge\neg F2(t)\}$。

证明 由 $E2=\{t|F2(t)\}$ 可知 $t\in E2\equiv F2(t)$，$t\notin E2\equiv\neg F2(t)$，所以 $F1(t)\wedge t\in E2\equiv F1(t)\wedge F2(t)$；$F1(t)\wedge t\notin E2\equiv F1(t)\wedge\neg F2(t)$。

由引理1可知：

若 $E1=\{t|F1(t)\wedge t\in E2\}$，$E2=\{t|F2(t)\}$，则 $E1=\{t|F1(t)\wedge F2(t)\}$；

若 $E1=\{t|F1(t)\wedge t\notin E2\}$，$E2=\{t|F2(t)\}$，则 $E1=\{t|F1(t)\wedge\neg F2(t)\}$。 □

引理4 $E1$、$E2$ 是关系代数表达式，$F1(t)$、$F2(t)$ 是谓词逻辑表达式，t 是 $E1$ 的结果变量；$E1$、$E2$ 在组合之前不包含 t'，t' 是任意不同于 t 的变量名，在使用引理之前 t' 不出现在 $E1$、$E2$ 中，$E2$ 的结果变量中含投影动作。

若 $E1=\{t|F1(t)\wedge t\in E2\}$，$E2=\{t[E]|F2(t)\}$，则 $E1=\{t|F1(t)\wedge(t=t'[E])\wedge F2(t')\}$；

若 $E1=\{t|F1(t)\wedge t\notin E2\}$，$E2=\{t[E]|F2(t)\}$，则 $E1=\{t|F1(t)\wedge(t=t'[E])\wedge\neg F2(t')\}$。

证明 由 $E2=\{t[E]|F2(t)\}$ 可知，$t\in E2\equiv(t=t'[E])\wedge F2(t')$，$t\notin E2\equiv(t=t'[E])\wedge\neg F2(t')$，所以，$F1(t)\wedge t\in E2\equiv F1(t)\wedge(t=t'[E])\wedge F2(t')$；$F1(t)\wedge t\notin E2\equiv F1(t)\wedge(t=t'[E])\wedge\neg F2(t')$。

由引理1可知：

若 $E1=\{t|F1(t)\wedge t\in E2\}$，$E2=\{t[E]|F2(t)\}$，则 $E1=\{t|F1(t)\wedge(t=t'[E])\wedge F2(t')\}$；

若 $E1=\{t|F1(t)\wedge t\notin E2\}$，$E2=\{t[E]|F2(t)\}$，则 $E1=\{t|F1(t)\wedge(t=t'[E])\wedge\neg F2(t')\}$。 □

通过上述变换规则和引理，可以由派生算子语义的原始算子表达式变换得到一阶逻辑表达式，从而严格地证明这两种表达式等价。

注 在下面第5、6章的变换过程中，基本上每一步都有注释，这些注释放在中括号“[ ]”中，用来说明位于“[ ]”之前(不是之后)的式子是依据什么式子或规则引理得到的。

5 交、θ连接、自然连接算子语义表达式间等价性证明

5.1 交算子

交算子语义的原始算子表达式：

$A\cap B=A-(A-B)$

交算子语义的一阶逻辑表达式：

$A\cap B=\{t|t\in A\wedge t\in B\}$

下面证明这两个表达式等价。

$A-(A-B)=\{t|t\in A\wedge t\notin(A-B)\}=$

[−算子的语义]

$\{t|t\in A\wedge\neg(t\in A\wedge t\notin B)\}=$

[−算子的语义和引理3]

$\{t|t\in A\wedge(t\notin A\vee t\in B)\}=\{t|t\in A\wedge t\in B\}$　[引理1]

5.2　θ连接算子

连接算子语义的原始算子表达式:

$A\rhd_F\lhd B=\sigma\{F\}(A\times B)$

连接算子语义的一阶逻辑表达式:

$A\rhd_F\lhd B=\{t|(t=<t1,t2>)\wedge t1\in A\wedge t2\in B\wedge F(t)\}$

下面证明这两个表达式等价。

$\sigma\{F\}(A\times B)=\{t|t\in(A\times B)\wedge F(t)\}=$

[σ算子的形式化语义]

$\{t|(t=<t1,t2>)\wedge t1\in A\wedge t2\in B\wedge F(t)\}$

[×算子的形式化语义和引理3]

5.3　自然连接算子

设 A 和 B 的公共属性是 $ab1,\cdots,abk$;$A.a$,$B.a$ 分别表示关系 A 和 B 的属性集合,则

自然连接算子语义的原始算子表达式如下:

$A\rhd\lhd B=\prod\{A.a\cup B.a\}(\sigma\{A.ab1=B.ab1\wedge\cdots\wedge A.abk=B.abk\}(A\times B))$

自然连接算子语义的一阶逻辑表达式如下:

$A\rhd\lhd B=\{t[A.a\cup B.a]|(t=<t1,t2>)\wedge t1\in A\wedge t2\in B\wedge(t[A.ab1]=t[B.ab1]\wedge\cdots\wedge t[A.abk]=t[B.abk])\}$

下面证明这两个表达式等价。

$\prod\{A.a\cup B.a\}(\sigma\{A.ab1=B.ab1\wedge\cdots\wedge A.abk=B.abk\}(A\times B))=$

$\{t[A.a\cup B.a]|t\in(\sigma\{A.ab1=B.ab1\wedge\cdots\wedge A.abk=B.abk\}(A\times B))\}=$　[∏算子的语义]

$\{t[A.a\cup B.a]|t\in(A\times B)\wedge(t[A.ab1]=t[B.ab1]\wedge\cdots\wedge t[A.abk]=t[B.abk])\}=$　[σ算子的语义和引理3]

$\{t[A.a\cup B.a]|(t=<t1,t2>)\wedge t1\in A\wedge t2\in B\wedge(t[A.ab1]=t[B.ab1]\wedge\cdots\wedge t[A.abk]=t[B.abk])\}$

[×算子的语义和引理3]

6　除算子语义表达式间等价性证明

除算子是关系代数中比较复杂的派生算子,其语义的原始算子表达式如下:

$A\div B=\prod\{AD\}(A)-\prod\{AD\}((\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A)$

[$A.a$,$B.a$ 分别表示 A 和 B 的属性集合,$AD=A.a-B.a$]

除算子语义的一阶逻辑表达式可以是:

$A\div B=\{t[AD]|(t\in A)\wedge((t'\in B)\rightarrow(<t[AD],t'[A.a-AD]>\in A))\}$

下面证明这两个表达式等价。

$\because\ \prod\{AD\}(A)=\{t[AD]|t\in A\}$　(1)

[∏算子的语义]

$\prod\{A.a-AD\}(B)=\{t[A.a-AD]|t\in B\}$

[∏算子的语义]

$\therefore\ \prod\{AD\}(A)\times\prod\{A.a-AD\}(B)=\{t|(t=<t1,t2>)\wedge(t1\in\prod\{AD\}(A))\wedge(t2\in\prod\{A.a-AD\}(B))\}]=$

[×算子的语义]

$\{t|(t=<t1,t2>)\wedge(t1=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(t\notin A)\}$　[引理3]

$\therefore\ (\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A=$

$\{t|(t\in(\prod\{AD\}(A)\times\prod\{A.a-AD\}(B)))\wedge t\notin A\}=$

[−算子的语义]

$\{t|(t=<t1,t2>)\wedge(t1=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(t\notin A)\}$　[引理3]

$\therefore\ \prod\{AD\}((\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A)=$

$\{t[AD]|t\in((\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A)\}=$

[∏算子的语义]

$\{t[AD]|(t=<t1,t2>)\wedge(t1=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(t\notin A)\}=$　[引理3]

$\{t1|(t=<t1,t2>)\wedge(t1=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(t\notin A)\}=$

[因为$(t=<t1,t2>)\Rightarrow(t1=t[AD])$,所以用 $t1$ 代换 $t[AD]$]

$\{t1|(t1=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(<t1,t2>\notin A)\}=$

[根据引理2,$t=<t1,t2>$是满射,因此用$<t1,t2>$代换 t]

$\{t|(t=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(<t,t2>\notin A)\}$　(2)

[根据规则2,将 $t1$ 换名成 t]

$\therefore\ A\div B=\prod\{AD\}(A)-\prod\{AD\}((\prod\{AD\}(A)\times\prod\{A.a-$

$AD\}(B))-A)=$ [÷算子语义的原始算子表示式]

$\{t|(t\in\prod\{AD\}(A))\wedge(t\notin\prod\{AD\}((\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A))\}=$ [–算子的语义]

$\{t|(t=t'[AD]\wedge t'\in A)\wedge(t\notin\prod\{AD\}((\prod\{AD\}(A)\times\prod\{A.a-AD\}(B))-A))\}=$ [式(1)和引理4]

$\{t|(t=t'[AD]\wedge t'\in A)\wedge\neg((t=t1'[AD])\wedge(t1'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(<t,t2>\notin A))\}=$

[式(2)和引理4]

$\{t|(t=t'[AD]\wedge t'\in A)\wedge\neg((t=t'[AD])\wedge(t'\in A)\wedge(t2=t2'[A.a-AD])\wedge(t2'\in B)\wedge(<t,t2>\notin A))\}=$

[t'和$t1'$都是为了确定结果变量t属于关系A，而且投影属性相同，所起的作用一样，因此可以用t'代换$t1'$]

$\{t|(t=t'[AD]\wedge t'\in A)\wedge\neg((t=t'[AD])\wedge(t'\in A)\wedge(t2'\in B)\wedge(<t,t2'[A.a-AD]>\notin A))\}=$

[根据引理2，用$t2'[A.a-AD]$代换$t2$]

$\{t'[AD]|(t'\in A)\wedge\neg((t'[AD]=t'[AD])\wedge(t'\in A)\wedge(t2'\in B)\wedge(<t'[AD],t2'[A.a-AD]>\notin A))\}=$

[根据引理2，用$t'[AD]$代换t]

$\{t'[AD]|(t'\in A)\wedge\neg((t'\in A)\wedge(t2'\in B)\wedge(<t'[AD],t2'[A.a-AD]>\notin A))\}=$ [引理1]

$\{t'[AD]|((t'\in A)\wedge\neg(t'\in A))\vee((t'\in A)\wedge\neg((t2'\in B)\wedge(<t'[AD],t2'[A.a-AD]>\notin A)))\}=$ [引理1]

$\{t'[AD]|(t'\in A)\wedge\neg((t2'\in B)\wedge(<t'[AD],t2'[A.a-AD]>\in A))\}=$ [引理1]

$\{t'[AD]|(t'\in A)\wedge(\neg(t2'\in B)\vee(<t'[AD],t2'[A.a-AD]>\in A))\}=$ [引理1]

$\{t'[AD]|(t'\in A)\wedge((t2'\in B)\rightarrow(<t'[AD],t2'[A.a-AD]>\in A))\}=$ [引理1]

$\{t[AD]|(t\in A)\wedge((t'\in B)\rightarrow(<t[AD],t'[A.a-AD]>\in A))\}$

[根据规则2，先将t'换名成t，然后将$t2'$换名成t']

经过以上的等价变换，得到用一阶谓词逻辑描述的“÷”算子的语义表达式为：

$$A\div B=\{t[AD]|(t\in A)\wedge((t'\in B)\rightarrow(<t[AD],t'[A.a-AD]>\in A))\} \quad (3)$$

[$A.a$,$B.a$分别表示A和B的属性集合，$AD=A.a-B.a$]

由式(3)可以较方便地看出除算子在做一件什么事情。对除算子这个派生概念的把握应包含严格证明和直觉理解两个方面；严格证明的载体是数学语言，直觉理解的载体是自然语言。上述嵌套组合法的形式等价变换过程属于严格证明，根据对式(3)的观察可以得到对除算子这个概念直觉理解，用自然语言叙述如下：

关系A除以关系B，结果是一些元组在属性组AD上投影的集合，这些元组t首先属于关系A，并且符合这样的条件：任意属于关系B的元组t'，它在属性组$A.a-AD$上的投影，与t在属性组AD上的投影并置所形成的新元组$<t[AD],t'[A.a-AD]>$必定属于关系A。

7 总结和讨论

本文通过等价变换形式化地证明了关系代数派生算子语义表达式之间的等价性。下面进行一些相关讨论。

(1)相对于其他派生算子，除算子的变换过程要复杂些，但仔细地看还是不难理解的，因为在除算子形式化语义的变换过程中使用的方法和其他派生算子是一样的，都是采用嵌套组合的办法进行形式变换。

(2)原始算子和派生算子的语义共同形成关系代数算子的语义模型。有了这个语义模型，就可以进一步讨论关系代数表达式的正确性证明。

(3)文章第4章所提出的变换规则和引理在变换过程中扮演着重要的角色。这些变换规则和引理是根据关系代数表达式的特点总结出来的，这些特点详见第4.1节。

(4)本文变换过程中所使用的方法和规则引理不仅可用于派生算子语义的不同表达式之间等价性证明，也可以用于各种关系代数表达式向一阶逻辑语义表达式的变换。基于本文所给出的关系代数算子的语义表达式，使用本文在第4章中所提出的变换规则和引理，可以得到任意复杂的关系代数表达式语义的一

阶逻辑表达式,由此可判断关系代数表达式之间是否等价,或者进行关系代数表达式的正确性证明。这些都将在后续工作中继续探讨。

References:

[1] Zhang Xiaoxiang, Xu Jiafu. Encyclopedia of computer science and technology[M]. 2nd ed. Beijing: Tsinghua University Press, 2005.

[2] Li Gaohe, Shi Jun. XML query based on relational algebra[J]. Computer Engineering and Design, 2004,25(8): 1415-1418.

[3] Yang Weiming. Query optimization and application of relational database[J]. Today Panorama of Modern Sciences, 2006(9):2.

[4] Ralston A, Reilly E D, Hemmendinger D. Encyclopedia of Computer Science[M]. 4th ed. [S.l.]: Nature Publishing Group, 2000.

[5] Silberschatz A. Database system concepts[M]. 5th ed. Beijing: Higher Education Press, 2005.

[6] Codd E. Relational completeness of data base sublanguages[M]//Data Base Systems. [S.l.]: Prentice Hall, 1972.

[7] Ullman J D, Aho A V, Sagiv Y. Equivalences among relational expressions[J]. SIAM J Computing, 1979,8: 218-246.

[8] Ullman J D, Aho A V, Beeri C. The theory of joins in relational databases[J]. ACM Trans on Database Systems, 1979,4:297-314.

[9] Kanellakis P. Elements of relational database theory[M]//Van Leeuwen J. Handbook of Theoretical Computer Science. [S.l.]: Elsevier, 1991.

[10] Abiteboul S, Hu H R, Vianu V. Foundations of databases[M]. [S.l.]:Addison Wesley, 1995.

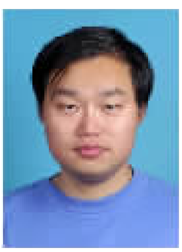
[11] Ullman J D. A first course in database systems[M]. 2nd ed. Beijing: China Machine Press, 2006.

附中文参考文献:

[1] 张效祥,徐家福.计算机科学技术百科全书[M].2 版.北京:清华大学出版社,2005.

[2] 李高和,石军.基于关系代数的 XML 数据查询[J].计算机工程与设计,2004,25(8):1415-1418.

[3] 杨蔚鸣.关系数据库的查询优化以及在现实中的应用[J].今日科苑,2006(9):2.



杨波(1975-),男,江西南昌人,博士生,讲师,主要研究领域为:软件形式化和自动化。

YANG Bo was born in 1975. He is a Ph.D. candidate and lecturer. His research interests include software formalization and automatization.

薛锦云(1947-),男,江苏海门人,教授,博士生导师,1970 年毕业于南京大学数学系,1985 年 12 月-1988 年 4 月在美国康奈尔大学计算机系进修访问,1995 年 6 月-1996 年 5 月在美国圣塔克拉大学任客座研究员,主要研究领域为:可信软件开发方法、工具和环境。

XUE Jinyun was born in 1947. He is a professor and doctoral supervisor. He was graduated in Department of Mathematics, Nanjing University in 1970. He ever visited Department of Computer Science in Cornell from December 1985 to April 1988 and was a research follow in Santa Clara University. His main research interests include the method, tools and environment for developing trustworthy software.