

用逻辑方法提高数据库技术研究水平

徐洁磐 (南京大学计算机系)

摘 要

数据库技术的发展需要有合适的数学工具与方法的支撑以提高其研究水平。目前,数理逻辑的一些分支正被用于数据库研究,其中有证明论的方法、模型论的方法以及函数的方法,这些方法的应用从观念上、思想上及方法上对数据库的研究进行了革新,并且从统一表示形式、扩大功能、简化实现方法等方面为数据库发展作出了贡献。在这些方法基础上目前已研制成功一些试验性数据库系统,同时,商品化系统也相继问世。

数据库技术自60年代问世以来,经历了70年代与80年代的发展,已成为计算机领域的一个重要分支,目前正被广泛地应用于各个领域。

数据库应用的扩大促进了数据库技术的发展,而数据库技术发展需要有合适的研究方法工具,利用数学特别是利用数理逻辑作为研究工具与方法以提高数据库技术的研究水平,已成为目前数据库研究中的一个重要研究动向,同时这种研究又推动了数据库应用的发展,预计今后还会产生更大的效果。

对数据库的这种研究方法一般是对数据库建立一种数学模型,使之对数据库的研究变成对某种数学(模型)的研究。由于数学的形式化研究方法完整系统的理论以及各种数学的研究特色,使得数据库的研究更方便、容易。

目前用数理逻辑的一些分支对数据库系统建立数学模型的研究包括:数据库的证明论模型;数据库的模型论模型;数据库的函数模型。

一 数据库的证明论模型

数据库系统可以用数理逻辑中的证明论描述,也就是说可以将数据库系统看成是一个证明系统。证明系统具体说来,由如下几

部分组成:

(1) 合式公式的构成:这里给出证明论中基本公式的构成法则。一般情况下,它由一阶谓词演算的合式公式构成。

(2) 公理:公理由一组公式组成,它一般分两种:事实性公理与规则性公理。事实性公理给出基本事实,而规则性公理则给出演绎的基本规则。

(3) 证明(过程)与定理:证明(过程)是一个公式序列,序列中的每个公式均由事实性公理或由序列中的前面公式经规则性公理的演绎而得,而证明(过程)的公式序列中最后一个公式叫定理。

自70年代初 Codd 等人研究数据库中的关系演算后将一阶谓词演算的书写方法引入到关系数据库系统中,在70年代末与80年代初 Gallaire 及 Minker^{[1]、[2]} 等对关系数据库系统与逻辑间关系作了深入研究,并将逻辑中的证明论方法带入到数据库研究中,此后逐渐形成了数据库的证明论模型,其基本思想如下:

(1) 可以用合式公式表示关系数据库系统的各种形式要求。

(2) 关系数据库系统中的关系数据(即数据元组)是公理中的事实性公理。

(3) 关系数据库系统中的窗口是公理中的规则

性公理。

(4) 关系数据库系统中的基本假设是公理。

(5) 数据查询是定理。

(6) 数据模式、完整性约束条件、安全性要求以及数据间的函数依赖关系等均是定理。

这样,关系数据库系统即可由证明系统表示,这种表示方法对数据库研究和实现产生很大影响,它们主要表现在:

(1) 传统的数据库系统的每个功能均由一个功能模块构成,功能越多,模块越多。这样一个大型数据库系统就是由很多个功能模块组成的一个庞大的软件。而在证明论模型下,整个数据库系统是一个证明系统。这个证明系统的唯一功能是由公理经证明而得到定理。这样,证明论模型下的数据库系统只需要一个功能模块就能实现整个系统。从而使得整个数据库系统的实现变得简单而方便。数据库系统的庞大存储空间也可以大为减少。

(2) 证明论模型目前在计算机领域中有很多软件工具可用,如PROLOG语言即是基于消解原理的演绎性证明系统,因此可以利用PROLOG语言为工具很方便地书写出数据库系统来。

(3) 基于证明论模型的数据库系统具有证明功能,它不仅可以证明数据库系统自身的定理,也可以扩充数据库系统功能,用以证明与数据库系统有关的另一些定理。这些都是传统数据库系统所做不到的。基于证明论模型的这种具有扩充功能的数据库系统即是目前处于发展时期的演绎数据库系统。因此基于证明论的数据库系统还具有比传统数据库系统更大的功能。

总之,在数据库研究中引入证明论模型使数据库系统的结构变得简单,存储空间大为减少,实现也较为容易而功能有所扩大。

二 数据库的模型论模型

模型论是数理逻辑的一个分支,模型论目前在形式语义学、数据库理论中得到广泛应用。

模型理论由一个三元组定义:

$$(L, \Sigma, M)$$

其中L是这个理论所使用的语言,它是由谓词、函数及常量所构成的一阶谓词演算的合式公式,这种合式公式在模型理论中称为语句或句子。

Σ 是一个语句集,在模型理论中叫规则集。

M称为结构,它给出L中谓词、函数及常量的一个解释。

如果结构M的解释使 Σ 中的每个语句为真,则称此结构M为 Σ 的一个模型,并记为:

$$M \models \Sigma$$

在模型论中,我们一般所研究的理论都是满足M为 Σ 的模型的那些理论。

在80年代初期及中期, Jacobs^[4]、Reither^[5]等先后将模型论引入数据库研究中并在该领域中得以发展。用模型论描述数据库系统的主要思想如下:

(1) 可以用模型理论中的语言L书写数据库系统的各种形式要求。

(2) 数据库系统中的数据模式、完整性约束条件、安全性要求、数据间的函数依赖关系等都是模型理论中的规则集。

(3) 数据库系统中的数据、函数是模型理论中的一个解释(即是它的一个结构)。

(4) 在数据库系统中,数据必须满足数据模式、完整性约束条件、安全性要求及函数依赖关系,在模型理论中则可表示为:

$$M \models \Sigma$$

这样,一个静态的数据库即可以用一个三元组模型理论表示。

更进一步,我们还有:

(5) 对数据库的查询我们可以用模型理论表示,查询要求是规则集,而查询结果则是一个解释,并且满足 $M \models \Sigma$ 。

(6) 对数据库的删除与插入我们可以用模型理论表示,删除与插入要求是规则集,而删除与插入的结果则是一个解释,并且满足 $M \models \Sigma$ 。

由于修改可分解成删除与插入两种操作,因此修改也可以用模型理论表示。

(7) 对数据库的窗口我们可以用模型理论表示,窗口表示式是规则集,而窗口则是一个解释,并且满足 $M \models \Sigma$ 。

这样,一个动态数据库系统也可以用模型理论表示。

由上可以看出,一个基于模型论的数据库系统是一个完整的数据库系统,这种表示

方法对数据库的研究和实现产生重大影响，它们主要表现在：

(1) 在模型理论中，静态与动态数据库系统的功能实现都表现为结构 M 与规则集 E 间的模型匹配，这样，使得庞大的数据库系统软件的多种功能实现变成仅仅是一种单一的模型匹配功能的实现。这样就使数据库系统的实现变得较为简单，所占空间也较少。

(2) 可以用一种模型论语言实现完整的数据库系统。

(3) 模型论语言除了能实现数据库系统功能外，还能实现数据库系统其它的一些有关模型匹配功能，如可以更多地增加数据库语义规则以及实现虚关系等功能。因此，基于模型论的数据库系统具有比传统数据库系统更大的功能。

与证明论模型一样，引入模型论模型，使数据库的研究与实现变得更为简单、方便，实现更为容易，空间占用量更少，还具有更多的功能。

三 数据库的函数模型

数据库的函数模型是一种构造性模型，它是用构造性方法表示数据库系统，从这点看，这种模型是一种很有特色的模型。

数据库的函数模型是建立在下面四个层次上的数学模型：

(1) 集合：集合是函数理论的基础。集合有并、交、差等运算。有属于、包含、相等关系。

(2) 函数：函数建立在集合之上。它有多种形式：单值函数与多值函数、一元函数与多元函数、显函数与隐函数。

(3) 算子：算子建立在函数之上，它是一种由旧函数构造新函数的方法。算子共有六个，它们是反函数算子、选置算子、复合算子、递归算子、插入算子及删除算子。

(4) 变换：变换给出了由函数构造集合的方法。变换有两种，一种叫 λ -变换，用它可以将显函数构造集合。另一种叫 μ -变换，用它可以将隐函数构造集合。

在函数理论中有一些基本函数或原始的

数，由它们通过六种算子可以构造出很多其它函数，再通过二种变换可以构造出很多集合来。

对数据库的函数模型研究始于70年代末期，Sibley、Kerschery 等人^[1]首先将函数概念引入数据库中，到80年代相继出现了一些基于函数模型的数据库系统，其中较为著名的有 DEPLEX^[2]、MORSE、JFO 等。目前，对数据库的函数模型的研究与实现已取得了很显著的成果。

可以用函数理论表示数据库系统，其要点如下：

(1) 数据模式及数据可以用基本函数表示。

(2) 数据的完整性约束条件、安全性要求、数据间的函数依赖关系等可以用一组隐函数表示，而对它们的检验则是对隐函数作 μ -变换，如结果为空集则表示数据满足各类约束条件，如结果不为空集，则表示数据不满足约束条件，且这个非空集的元素即为不满足约束条件的数据。

(3) 数据的查询要求可由基本函数经反函数、复合函数、选置及递归等算子构造而成的函数表示，而查询结果则是对此函数作 λ -变换的结果。

(4) 窗口表示式是函数，而窗口所表示的数据则是对函数的 λ -变换的结果。

(5) 数据的删除、插入与修改可由基本函数利用删除算子与插入算子以及其它算子构成的新函数表示，而删除、插入与修改后的结果则是对函数作 λ -变换的结果。

这样，数据库系统就可以用一种函数理论构造而成。

数据库的函数模型表示具有很大的优越性，它还与前面两种模型有明显的不同，其主要表现在以下几个方面：

(1) 数据库的函数模型是一种构造性的数学模型，它有一些基本构件，具体说来就是六个算子、两个变换，而整个数据库系统即可由这八个构件构造而成。这种构造思想对数据库系统的硬件与软件的实现有很大的指导意义。在数据库机的研究中，可以用八种基本电路板构成一个完整的数据库机。在数据库的软件实现中可以用八个基本模块完成数据库系统的构造。

(2) 由于 LISP 语言是一种函数式程序设计语言，因此，可以用 LISP 语言很方便地构造八种基

本模块,从而实现数据库系统的构成。

(3) 用函数构造的方式还可以完成数据库系统中的其它构造,从而使得基于函数模型的数据库系统具有比一般传统数据库系统更多的功能。

* * *

引入了多种逻辑方法后使数据库的研究变得丰富多彩。不同方法从不同角度以不同特色观察、分析与研究数据库,使得数据库研究出现了新气象,从观念、思想、方法上对数据库进行了革新,它们从统一表示形式、扩大功能、简化实现方法等方面为数据库的发展作出了贡献。

目前,基于这三种数学模型的数据库实验性系统与商品化系统都相继问世,对它们的研究还在继续。同时,一些新的数据库的数学模型也陆续有所报导,如代数模型、图论模型及范畴论模型等,预计这些数学模型也将对数据库的发展产生重大影响。

参 考 文 献

[1] H. Gallaire, J. Minker, "Logic

And Database" New York Plenum Press, 1978.

[2] H. Gallaire, J. Minker, "Advance In Database Theory" Vol. 1 New York Plenum Press, 1981.

[3] R. Reither, "Towards A Logic Reconstruction of Relational Database Theory On Conceptual Modelling" Springer-Verlag New York Inc. 1984.

[4] B. E. Jacobs, "Database Logic" ACM 1982-29, No.2.

[5] F. H. Sibley, L. Kershbery, "Data Architecture And Data Model Consideration" Proc. AFIPS Nat. Computer conf. Dallas, Tex, June 1977

[6] D. shipman, "The Functional Data Model And The Data Language DAPLEX" ACM TODS 6 No. 1, 1981.



知识工程进展1990

由中国第三届全国知识工程研讨会程序委员会按学报级杂志审稿程序组织评审、编辑,中国地质大学出版社出版的中国第三届全国知识工程研讨会论文选集《知识工程进展1990》将于今年八月中旬公开发行人。这本文集是《知识工程进展》的第二卷,共选入3篇特约论文和27篇征文,它们比较集中地反映了中国知识工程学界在机器学习,专家系统,智能控制,自然语言处理等各方面的研究进展和学术动态,是从事知识工程和人工智能研究、教学和发展工作的工程技术人员、教师及学生的重要参考资料。《知识工程进展1990》印刷精美、质量较高,每本25元(含邮资),需要者可汇款至430074武汉市武昌中国地质大学计算机科学系宋力同志收,待第三届全国知识工程研讨会召开(9月1日)后即发书。

智能软件学组供稿