

一个用于异构的数据库集成的 多 context 逻辑^{*}

刘海燕 陈火旺 刘凤岐

(国防科技大学计算机系 长沙 410073)

摘要 Internet 的发展要求能同时使用来自不同知识源的信息. 本文定义了一个两层的多 context 逻辑结构 DBCL. 第 1 层的每个 context 都是一个传统的多类逻辑理论, 对应于一个数据库; 第 2 层的 context 是个特殊的多类逻辑理论, 可以声明第 1 层中各 context 间的各种联系, DBCL 为异构的数据库的集成提供了一个形式化描述. 本文还为 DBCL 定义了一类可靠且完备的语义解释.

关键词 数据库, 集成, context, 多类逻辑, 异构的.

中图法分类号 TP18

最近几年, 计算机领域最热门的话题当属 internet 的构建、开发和使用. 不仅计算机界人士, 凡人网的用户都是它的受益者. internet 蕴含着巨量的数据信息, 涉及到各门学科和行业, 用户可以不受地理位置的限制, 通过一定的机制来调用远方的资源. 所以 internet 的出现缩短了人与人之间的通讯距离, 扩大了每个人所面对的世界.

然而与任何其它新技术一样, internet 也需要各项配套技术的支持. internet 是个开放的网络系统, 网上提供了许多信息数据库, 它们以不同的方式、含义存储在网络的各结点上, 用户也可以创建自己的数据库. 这一方面增强了系统的灵活性和开放性, 另一方面, 由于各数据库可以拥有不同的模式和设计, 同一符号在不同的数据库中可能代表不同的含义, 所以在同一个应用中很难同时使用来自不同数据库的信息. 从而如何集成使用来自异构的数据库的信息在近些年再次成为人们研究的热点.

为了集成不同的数据库, 首先必须分析各模式所反映的各种实体的性质, 分析不同模式间各类对象、属性的联系; 然后把表示现实世界中同一个实体的相同实例的对象作为一个对象, 解决不同模式间可能存在的语义不相容性和数据不协调现象; 最后得到一个全局的数据库模式或一个全局的数据库.^[1,2] 与以往的集成方式相比, 我们的方法更强调语义不相容性的解决, 它能显式地表示不同数据库间的语义联系和逻辑联系.

* 本文研究得到国家 863 高科技项目基金资助. 作者刘海燕, 女, 1970 年生, 博士生, 主要研究领域为计算机软件 and 人工智能. 陈火旺, 1936 年生, 教授, 博士生导师, 主要研究领域为计算机软件, 人工智能和形式化理论等. 刘凤岐, 1938 年生, 教授, 主要研究领域为人工智能, 专家系统等.

本文通讯联系人: 刘海燕, 长沙 410073, 国防科技大学计算机系

本文 1996-11-13 收到修改稿

由于不同的数据库可能由不同的设计者设计,所以数据库之间可能存在下列语义不相容现象^[1]:

命名冲突(Naming Conflicts):由于每个模式的设计者都用自己的字母表为对象和属性命名,所以可能出现同名异义词(Homonyms)和同义词(Synonyms),而且同名异义词和同义词只能通过数据库之外的描述指出。

类型冲突(Type Conflicts):当相同的概念在不同的模式中以不同的编码结构表示时,会出现类型冲突。

关键字冲突(Key Conflicts),即在不同的模式中同一概念有不同的关键字。

数据遗失(Missing Data),表现为在不同的模式下为同一概念定义了不同的属性。

抽象层次的不同(Levels of Abstraction),即在不同的 DB 中,关于同一个实体的信息以不同的详细程度存储着。例如在一个 DB 中 labor-cost 和 material-cost 单独地存储,而在另一个 DB 中,他们作为一个整体 total-cost 存储。

度量冲突(Scaling Conflicts):一个实体的同一属性在不同的 DB 中以不同的度量存储。如一个部件的重量在一个 DB 中按公斤存储,在另一个 DB 中按克存储。

当把多个 DB 集成时,需要弄清不同 DB 模式中各类实体间的联系,而上述的语义不相容现象使得这项工作难以顺利完成。本文定义了一个两层的多 context 逻辑结构 DBCL,它使用推广了的多类逻辑形式化地描述 DB 的集成。

1 DB 到多类理论的转换

本文仍沿用数据库领域的一些概念;数据库设计者用实体描述各类客观事物和各种联系。实体可分为对象和属性两类,属性反映的是对象的性质,如张三的年龄,张三是对象,年龄是属性。实体又可分为总体和个体两级,如学生是总体,而张三是一个个体。数据模型是实体模型的数据化,描述对象的数据称为记录,而描述属性的数据称为数据项,总体对应的是数据的型,个体对应的是值。

数据库与逻辑有着密切的联系,在物理上数据库表现为存储的数据的集合,在逻辑上每个数据库都可以看作是一个领域公理的集合。在文献[3]中,Guha 描述了一种把 DB 转化成逻辑理论的方法。在文献[4]中,Buvac 指出 context 可以用于集成那些本来并未期望能共同使用的数据库。本文基本上借鉴了他们的思想,并针对 DB 间的语义不相容现象给出一个形式化的公理系统。

本节主要定义数据库到多类逻辑理论的一种转换。设 DB_i 是第 i 个数据库,则与 DB_i 对应的多类逻辑系统 c_i 定义为:

为 DB_i 的每个记录类型和每个数据项类型都在 c_i 中定义一个类型;

为每个数据项类型定义一个二元谓词,它的第 1 个参数的类型为相应的记录类型,而第 2 个参数的类型为相应的数据项类型;

为每个记录选用一个该记录类型的新的个体常元,为每个数据项类型的每个可能取值都选用一个该项类型的常元。

那么 DB_i 的每个记录的每个数据项都对应着 c_i 的一个原子公式。例如,在描述学生的 DB 中,对象类型为 student,设描述张三的记录所对应的常元为 a ,属性年龄的类型是整型,

它所对应的谓词为 age;属性姓名的类型是字符串型,所对应的谓词为 name.那么描述张三的记录对应于公式 name(a,“ZhangSan”), age(a,30),…….

关键字有唯一确定一个记录的作用,关键字所对应的谓词称为关键谓词,设 P_1, \dots, P_n 为关键谓词,那么有 $\forall x, y, v_1, \dots, v_n (P_1(x, v_1) \wedge \dots \wedge P_n(x, v_n) \wedge P_1(y, v_1) \wedge \dots \wedge P_n(y, v_n) \rightarrow (x=y))$.

上述转换与数据库所选用的模型无关,无论是关系型数据库、实体关系型数据库,还是面向对象型数据库,通过相似的转换都可以转化成多类逻辑的一个理论,所以,本文定义的逻辑系统可用于异构的数据库的集成.

DB 到多类逻辑理论的转换是本文工作的第 1 步,只有转化成相应的逻辑公式后,才能用逻辑的方法形式化 DB 的集成.但是这种转化只是概念上的,并不真正地把每个数据项都改写成逻辑公式,只需在集成时把它看作公式即可.

2 数据库 Context 逻辑 DBCL

语法:DBCL 是一个两层的多 context 逻辑结构.第 1 层的 context 称为数据库 context (简称为 DB context),每个 DB context 都是一个传统的含等词的多类一阶逻辑,本文用 $c_i (1 \leq i \leq n)$ 标识 DB context.第 2 层的 context 是个特殊的多类逻辑理论,本文用 c_0 标识(见图 1).

设 $c_i (0 \leq i \leq n)$ 的字母表为 Σ_i ,设 Σ_0 包含了许多个体类型,每种类型都有可数多个个体常元. $\Sigma_i (1 \leq i \leq n)$ 的类型是 Σ_0 的类型的子集,而每种类型的个体常元与 Σ_0 中同型的个体常元都相同.此外,每个 $\Sigma_i (0 \leq i \leq n)$ 都有自己的谓词符号,为简便起见,我们假设 Σ_i 不含等词.

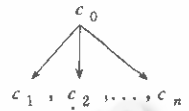


图1

设 VAR 是各种类型的个体变元的集合.我们用 VAR_{Σ_i} 表示 VAR 中所有类型在 Σ_i 内的元素的集合. $c_i (1 \leq i \leq n)$ 的项和公式的定义与建立在 Σ_i 和 VAR_{Σ_i} 上的经典多类一阶逻辑(含等词)的项和公式的定义相同. c_0 的项的定义与建立在 Σ_0 和 VAR 上的多类逻辑的项的定义相同.

定义 2.1. c_0 的公式归纳定义为

- 1) 若 P 为 Σ_0 中的一个 m 元谓词, t_1, t_2, \dots, t_m 为 c_0 的 m 个项,且 t_i 与 P 的第 i 个参数的类型相同,则 $P(t_1, t_2, \dots, t_m)$ 为 c_0 的公式;
 - 2) 若 t_1, t_2 是类型为 T_i 的项,则 $t_1 = t_2$ 为 c_0 的公式(本文中,一般不明确指出等词的类型,读者可根据该等词的参数类型推断);
 - 3) 若 t_1 是类型 T_1 的一个项, t_2 是类型 T_2 的一个项,则 $coReferT(T_1, t_1, T_2, t_2)$ 为 c_0 的一个公式,它的直观含义是 t_1 和 t_2 的指称物相同;
 - 4) 若 A 为 $c_i (1 \leq i \leq n)$ 的公式,则 $ist(c_i, A)$ 为 c_0 的公式;
 - 5) 若 A, B 为 c_0 的公式, $x \in VAR$,则 $\sim A, A \rightarrow B, \forall x A$ 都是 c_0 的公式;
- $\wedge, \vee, \Leftrightarrow$ 和 \exists 都是通常的简写形式.

DBCL 的公理模式有 $(0 \leq i \leq n, 1 \leq j \leq n)$:

$A_1: \vdash c_i \emptyset, \emptyset$ 为 c_i 的公式,且为带等词的多类一阶逻辑的一个永真式的一个实例;

- $A_2: \vdash_{c_0} t_1 = t_2 \Leftrightarrow coReferT(T, t_1, T, t_2)$, 其中 t_1 和 t_2 的类型均为 T ;
- $A_3: \vdash_{c_0} corReferT(T_1, t_1, T_2, t_2) \rightarrow coReferT(T_2, t_2, T_1, t_1)$;
- $A_4: \vdash_{c_0} corReferT(T_1, t_1, T_2, t_2) \wedge coReferT(T_2, t_2, T_3, t_3) \rightarrow coReferT(T_1, t_1, T_3, t_3)$;
- $A_5: \vdash_{c_0} ist(c_j, A \rightarrow B) \rightarrow (ist(c_j, A) \rightarrow ist(c_j, B))$;
- $A_6: \vdash_{c_0} ist(c_j, \forall x A) \Leftrightarrow \forall x ist(c_j, A) \quad x \in VAR_{\Sigma_j}$;
- $A_7: \vdash_{c_0} ist(c_j, \sim A) \rightarrow \sim ist(c_j, A)$;
- $A_8: \vdash_{c_0} t_1 = t_2 \Leftrightarrow ist(c_j, t_1 = t_2)$; t_1 和 t_2 是 c_j 的项.

DBCL 的推理规则有 ($0 \leq i \leq n, 1 \leq j \leq n$):

- $R_1(GMP)$: 若 $\vdash_{c_i} A \rightarrow B$ 且 $\vdash_{c_i} A$ 则 $\vdash_{c_i} B$;
- $R_2(GGEN)$: 若 $\vdash_{c_i} A$ 则 $\vdash_{c_i} \forall x A \quad x \in VAR_{\Sigma_i}$;
- $R_3(Enter)$: 若 $\vdash_{c_0} ist(c_j, A)$ 则 $\vdash_{c_j} A$;
- $R_4(Exit)$: 若 $\vdash_{c_j} A$ 则 $\vdash_{c_0} ist(c_j, A)$;

定义 2.2. 设 A 为 c_i ($0 \leq i \leq n$) 的一个公式, Γ 是 c_i 的一个公式集, 则

- 1) $\vdash_{c_i} A$ 当且仅当 存在有穷序列 $\vdash_{k_1} A_1, \vdash_{k_2} A_2, \dots, \vdash_{k_m} A_m$,

其中 $c_i = k_m, A_m = A$ 且对每个 $1 \leq j \leq m$ 下列条件之一满足:

- i) $\vdash_{k_j} A_j$ 是 DBCL 的公理;
 - ii) $\vdash_{k_j} A_j$ 由前面的元素及 DBCL 的某个推理规则而得.
- 2) $\Gamma \vdash_{c_i} A$ 当且仅当 存在 $B_1, B_2, \dots, B_m \in \Gamma$ 使得 $\vdash_{c_i} ((B_1 \wedge B_2 \wedge \dots \wedge B_m) \rightarrow A)$ 成立.

由公理 A_2, A_3 和 A_4 知, $coReferT$ 是等词“=”的推广, 用 $coReferT$ 可以建立不同数据库中数据项间的同义词关系, 如 $coReferT(T_i, "M", T'_j, "Male")$.

设 P_1 和 P_2 分别为 c_i 和 c_j ($1 \leq i, j \leq n$) 的 m 元谓词, P_1 的参数类型依次为 T_1, T_2, \dots, T_m , P_2 的参数类型依次为 T'_1, T'_2, \dots, T'_m . 若对 T_i ($2 \leq i \leq m$) 型的每个个体常元 a , 都存在 T'_i 型的个体常元 b , 满足 $coReferT(T_i, a, T'_i, b)$, 反之亦然. 则 $coReferP(c_i, P_1, c_j, P_2)$ 是公式: $\forall x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_m (coReferT(T_1, x_1, T'_1, y_1) \wedge coReferT(T_2, x_2, T'_2, y_2) \wedge \dots \wedge coReferT(T_m, x_m, T'_m, y_m) \rightarrow (ist(c_i, P_1(x_1, x_2, \dots, x_m)) \Leftrightarrow ist(c_j, P_2(y_1, y_2, \dots, y_m))))$ 的简写形式.

由 $coReferP$ 的定义, 可得出 DBCL 的如下定理 ($1 \leq i, j, k \leq n$):

- $D_1: \vdash_{c_0} coReferP(c_i, P, c_i, P)$;
- $D_2: \vdash_{c_0} coReferP(c_i, P_1, c_j, P_2) \rightarrow coReferP(c_j, P_2, c_i, P_1)$;
- $D_3: \vdash_{c_0} coReferP(c_i, P_1, c_j, P_2) \wedge coReferP(c_j, P_2, c_k, P_3) \rightarrow coReferP(c_i, P_1, c_k, P_3)$.

所以 $coReferP$ 是各 DB context 的谓词间的一个等价关系, 它反映了数据库中属性类型的同义词性质, 如 $coReferP(c_i, favoritecolor, c_j, colorpreferred)$.

语义: 每个 context 都对应于一个标准的多类解释的集合, $ist(c_i, A)$ 为真当且仅当 A 在 c_i 所对应的所有多类解释下都为真.

定义 2.3. 设 c_0 的个体类型为 T_1, T_2, \dots , 若 M 满足下列条件, 则称 M 为 DBCL 的一个解释:

对任意的 c_i ($0 \leq i \leq n$), $M(c_i)$ 是 Σ_i 的一个标准的多类解释的非空集合, 设 $M(c_0) \subseteq P$

$\{(\cup_1^\infty D_i, I) \mid (\cup_1^\infty D_i, I) \text{ 是 } \Sigma_0 \text{ 的一个标准的多类解释}\}$, 则

1) 在所有的多类解释中, 同型的论域相同. 即对任意的 $u \in M(c_i), v \in M(c_j)$, 设 D_{uk} 是 u 中类型 T_k 所对应的论域, D_{vk} 是 v 中类型 T_k 所对应的论域, 则 $D_{uk} = D_{vk}$. 所以我们用 D_i 标识所有多类解释中与类型 T_i 对应的域;

2) $\cup_1^\infty D_i$ 上存在一个等价关系 E ;

3) 对任意的 $u \in M(c_i), v \in M(c_j)$, 若 a 既是 c_i 的常元又是 c_j 的常元, 则 $I_u(a) = I_v(a)$.

定义 2.4. M 下的赋值 $S: VAR \rightarrow \cup_1^\infty D_i$ 对任意变元赋予相应类型的值.

项的值 $I_u(t)[S]$ 的定义与多类一阶逻辑中项 t 在解释 I_u 和赋值 S 下的值的定义相同.

可以证明: 项的值与 context 无关, 即对任意的项 t , 任意的 context c_i 和 c_j 及任意的 $u \in M(c_i), v \in M(c_j)$, 若 t 既是 c_i 的项又是 c_j 的项, 则 $I_u(t)[S] = I_v(t)[S]$. 所以在下面的讨论中, 当计算项的值时, 略去 I 的下标, 统一写成 $I(t)[S]$.

定义 2.5. 设 A 为 c_i 的公式, $u \in M(c_i)$, 则 $M, u \models_{c_i} A[S]$ 归纳定义为:

1) $M, u \models_{c_i} P(t_1, t_2, \dots, t_n)[S]$ iff $\langle I(t_1)[S], \dots, I(t_n)[S] \rangle \in I_u(P)$;

2) $M, u \models_{c_i} t_1 = t_2[S]$ iff t_1 和 t_2 类型相同且 $\langle I(t_1)[S], I(t_2)[S] \rangle \in E$;

3) $M, u \models_{c_0} coReferT(T_i, t_1, T_j, t_2)$ iff $\langle I(t_1)[S], I(t_2)[S] \rangle \in E$;

4) $M, u \models_{c_i} \sim B[S]$ iff $M, u \models_{c_i} B[S]$ 不成立;

5) $M, u \models_{c_i} (A \rightarrow B)[S]$ iff $M, u \not\models_{c_i} A[S]$ 或 $M, u \models_{c_i} B[S]$;

6) $M, u \models_{c_i} \forall x A[S]$ 设 x 的类型为 T_j , iff 对任意的 $d \in D_j$, 都有 $M, u \models_{c_i} A[S(x/d)]$;

7) $M, u \models_{c_0} ist(c_j, A)[S]$ iff 对任意的 $v \in M(c_j), M, v \models_{c_j} A[S]$ 都成立.

定义 2.6. 设 A 是 c_i 的一个公式, Γ 是 c_i 的一个公式集, M 为 DBCL 的一个解释, S 是 M 下的一个赋值, $u \in M(c_i)$, 则

1) $\models_{c_i} A$ iff 对任意的 M, S, u , 都有 $M, u \models_{c_i} A[S]$;

2) $\langle M, u, S \rangle$ 在 c_i 中满足 Γ iff 对 Γ 中每个公式 A 都有 $M, u \models_{c_i} A[S]$;

3) $\Gamma \models_{c_i} A$ iff 对任意 M, S, u , 若 $\langle M, u, S \rangle$ 在 c_i 中满足 Γ , 则 $M, u \models_{c_i} A[S]$;

定理 2.1. (完备性) $\Gamma \vdash_{c_i} A$ 当且仅当 $\Gamma \models_{c_i} A$.

\Rightarrow 方向是可靠性定理, 容易证明, DBCL 的每条公理都可靠, 每条推理规则都保真;

在 \Leftarrow 方向的证明中, 我们仅证明若 $\Gamma \models_{c_0} A$ 则 $\Gamma \vdash_{c_0} A$

首先证明: 若 Γ 在 c_0 中协调, 则 Γ 在 c_0 可满足:

与一阶逻辑的 Lindenbaum^[5]引理相似, 可以把 Γ 扩充成对于 c_0 来说既完全又协调的公式集 Δ , 对任意的 $1 \leq i \leq n$, 取 $\Gamma_\alpha = \{A \mid A \text{ 为 } c_i \text{ 的公式且 } \Delta \vdash_{c_0} ist(c_i, A)\}$, 则 Γ_α 协调. Γ_α 定义了从 c_0 的观点看 context c_i 所描述的事物状态. 记 $\Gamma_{c_0} = \Delta$. 我们将用所有的 Γ_α 构造解释 M . 取论域 D_i 为括充后的语言中所有类型为 T_i 的项. Γ_α 在 c_i 中的每个极大协调扩张 Γ' 都对应于一个 $u \in M(c_i)$, 满足 $I_u(a) = a, \langle t_1, \dots, t_m \rangle \in I_u(P)$ 当且仅当 $P(t_1, t_2, \dots, t_m) \in \Gamma'$, 取 $S(x) = x$. 设 $V \in M(c_0)$ 可以证明: 对任意的 $A \in \Gamma, M \models_{c_0} A[S]$ 都成立.

用反证法可证明, 若 $\Gamma \models_{c_0} A$, 则 $\tau \vdash_{c_0} A$.

3 领域公理在不同 context 中声明的 DBCL

在定义 2.2 中, $\Gamma \vdash_{c_j} A$ 要求 Γ 和 A 同属于 context c_j . 然而, 在实际应用中, 每个数据库

都存储自己的数据,对应到 DBCL 系统,即领域公理在不同的 context 中声明.例如 DB_1 和 DB_2 都是描述学生的数据库, DB_1 中有“*The sex of ZhangSan is 'M'*”,而且 *ZhangSan* 与张三同一个人, DB_1 中的 *sex* 即 DB_2 中的性别, DB_1 中的 ‘*M*’ 即 DB_2 中的男, DB_1 中的 ‘*F*’ 即 DB_2 中的女,那么“张三是男性”应该是集成后的数据库中的一条信息.

用 DBCL 的语言表示为:

$\vdash_{c_1} sex(ZhangSan, "M") \vdash_{c_0} coReferP(c_1, sex, c_2, 性别), \vdash_{c_0} ZhangSan = 张三,$
 $\vdash_{c_0} coReferT(T_{sex}, "M", T_{性别}, "男"), \vdash_{c_0} coReferT(T_{sex}, "F", T_{性别}, "女").$

设 Γ 是由如上形式的领域公理组成的一个集合,下面定义在这种情况下 DBCL 的推理.

定义 3.1. $FOR(\Gamma)$ 是满足下列条件的最小集合:

- 1) 若 $\vdash_{c_i} A \in \Gamma (1 \leq i \leq n)$, 则 $ist(c_i, A) \in FOR(\Gamma)$;
- 2) 若 $\vdash_{c_0} A \in \Gamma$, 则 $A \in FOR(\Gamma)$.

定义 3.2. $\vdash_{c_i} A$ 是 Γ 的一个 DBCL 结论 iff

- 1) 若 $1 \leq i \leq n$, 则 $\vdash_{c_i} A$ 是 Γ 的一个 DBCL 结论 iff $FOR(\Gamma) \vdash_{c_0} ist(c_i, A)$;
- 2) $\vdash_{c_0} A$ 是 Γ 的一个 DBCL 结论 iff $FOR(\Gamma) \vdash_{c_0} A$.

上例中, $\Gamma = \{ \vdash_{c_1} sex(ZhangSan, "M"), \vdash_{c_0} coReferP(c_1, sex, c_2, 性别), \vdash_{c_0} ZhangSan = 张三, \vdash_{c_0} coReferT(T_{sex}, "M", T_{性别}, "男"), \vdash_{c_0} coReferT(T_{sex}, "F", T_{性别}, "女") \}$

由定义 3.1 和 3.2 知, $\vdash_{c_2} 性别(张三, "男")$ 是 Γ 的一个 DBCL 结论.

4 DBCL 对 DB 的语义不相容现象的表示

本节讨论 DBCL 如何表示引言中指出的 DB 间的各种语义不相容现象.

由于每个 DB 都对应着一个 context, context 的类型名、常量名、谓词名分别对应于 DB 的对象类型和属性类型、类型的值空间、属性名,而且由于每个 context 都有自己的解释集合,所以 DBCL 中的任意名称都缺省地体现了同名异义词的性质.同义词用等词、*coReferT* 和 *coReferP* 表示.

由于 DBCL 是多类逻辑的扩充,它用逻辑描述各类对象的各种属性,所以类型冲突、数据遗失等语义不相容现象在 DBCL 中都表现为不同 DB context 的内容不同,不影响集成后的数据库的整体作用.

用 c_0 的公式可以表示 DB 间的关键字冲突,如 DB_1 和 DB_2 都是描述学生情况的数据库, DB_1 用“#SS”(即学号)作为关键字, DB_2 用“name”作为关键字,它们间的关系为:

$ist(c_1, \forall x, y, z \#SS(x, z) (\#SS(y, z) \rightarrow x=y),$
 $ist(c_2, \forall x, y, z \text{ name}(x, z) (\text{name}(y, z) \rightarrow x=y),$
 $ist(c_1, \#SS(a, 0001)), ist(c_2, \text{name}(a, "ZhangSan"))).$

抽象层次的不同和度量的不同也可用 c_0 的公式描述,如文章开始时的例子可表示为:

$\forall x, y, z (ist(c_i, \text{labor-cost}(x, y)) \wedge ist(c_i, \text{material-cost}(x, z)) \rightarrow ist(c_j, \text{total-cost}(x, y + z)));$
 $\forall x, y (ist(c_i, \text{weight}(x, y)) \rightarrow ist(c_j, \text{weight}(x, 1000 * y))).$

5 结 论

internet 的发展要求能综合使用来自不同信息源的信息. 本文设计了一个两层的多 context 逻辑结构 DBCL, 用于形式化地描述异构的数据库的集成, 进一步的工作将着重研究数据库间数据不协调性的处理.

参考文献

- 1 Reddy M P *et al.* A methodology for integration of heterogeneous databases. *IEEE Trans. on Knowledge and Data Engineering*, 1994, **6(6)**:920~933.
- 2 Spaccapietra S. View integration: a step forward in solving structural conflicts. *IEEE Trans. on Knowledge and Data Engineering*, 1994, **6(2)**:258~274.
- 3 Guha R V. Context: a formalization and some application. MCC Technical Report, Number ACT-CYC-423-91, MCC, Austin Texas, 1991.
- 4 Buvac S. Quantificational logic of context. In: the Proceedings of AAAI'96, 1996. 600~607.
- 5 Bell J L, Machover M. A course in mathematical logic. North-Holland Publishing Company, 1977.

A MULTICONTEXT LOGIC USED FOR INTEGRATING HETEROGENEOUS DATABASES

LIU Haiyan CHEN Huowang LIU Fengqi

(Department of Computer Science National University of Defense Technology Changsha 410073)

Abstract To use information from different sources together is a desirable property for the users, especially in these internet days. A two-level multicontext architecture DBCL is defined in this paper which can be used to integrate heterogeneous databases. Each context in its first level is a classical many-sorted theory, which corresponds to a database. The context in the second level is a generalized many-sorted theory, where metatheories about the various relations among databases can be stated. A kind of semantics, with respect to which DBCL is sound as well as complete, is also provided in this paper.

Key words Database, integration, context, many-sorted logic, heterogeneous.

Class number TP18