文章编号:1001 - 2486(2004)04 - 0015 - 07

异构数据源集成系统中能力对象集成模型及代数*

唐九阳,张维明,宋峻峰,修保新,肖卫东
(国防科技大学人文与管理学院,湖南 长沙 410073)

摘 要:异构数据源集成系统中,参与集成各数据源的查询处理能力有自身特殊的限制,导致中介器支持的查询类型变得复杂。提出一种便于异构数据源集成的公共能力对象数据模型——COIM 对象模型,在此基础上,进而定义数据源能力对象代数,给出基于局部数据源的能力对象构造集成系统中介器能力对象的方法。由于可以通过计算得到中介器能力对象,用户能够提前预知中介器支持的查询类型,从而直接构造可以执行的查询,减少查询提交的盲目性;同时中介器本身也可像数据源一样参与其它中介器的集成,集成方式变得更为灵活。

关键词:异构数据源集成;能力描述;能力对象模型;能力对象代数
中图分类号:TP311 文献标识码:A

A Data Model and Algebra for Capability Object Integration in Heterogeneous Data Integration Systems

TANG Jiu-yang, ZHANG Wei-ming, Song Jun-feng, XIU Bao-xin, XIAO Wei-dong
(College of Humanities and Management, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: In heterogeneous data sources integration systems, there can be many special limited access patterns to access interface of data sources, i.e., binding patterns impose a restriction that one or more input attributes must have a binding, and also restrict the output attributes that are projected. Constrained by the expressiveness of the access interface, users often have difficulty in articulating a precise query over the data sources. In this paper, we propose a data model for data sources' capability object integration (COIM) to capture a rich variety of query-processing capabilities and present algebra to compute the set of mediator-supported queries based on the capability limitations of the sources they integrate. By computing mediator query capabilities and representing them in the same way as those of data sources, we enable mediators to be used by other mediators as well as other applications, and we also make it easier for end users to know in advance which mediator queries are feasible.

Key words: heterogeneous data sources integration; capability description; capability object model; capability object algebra

异构数据源集成系统集成的数据来源广泛,包括数据库系统、文件系统、WWW 等。这些异构的数据源即使通过封装器包装成关系型数据库,出于以下两个原因往往只能提供有限的查询能力:从系统安全保密或运行性能等方面考虑,系统仅向外界提供有限信息;数据源是文件系统或者原有的访问界面只允许用户进行有限的操作[1,2]。通常,数据源需要将自身内容以及支持的查询类型通知中介器,以方便中介器构造查询计划。另一方面,中介器应当向用户展示自身支持的查询类型,从而引导用户提交可行的查询。

为解决基于受限能力的异构数据集成问题,国外一些机构进行了相应研究。TSIMMIS[3]借助关系查询描述语言(RQDL)构造表达连接查询能力的查询模板;Garlic[4]和 DISCO[5]系统对目标查询表达式没有限制,容许数据源描述连接和非连接的查询能力;Versatile[6]系统采用能力向量描述各数据源的基本查询能力;Information Manifold[7]借助基于五元组的能力记录刻画数据源的查询能力;文献[8]使用约束模式(binding pattern)来描述数据源的有限查询处理能力,约束模式 $< b, f >$ 规定了访问数据源时必须提供哪些属性的值。上述方法局限于描述数据源自身的查询能力,都没有考虑刻画中介器的查询能力。

* 收稿日期:2004 - 03 - 15
基金项目:国家自然科学基金资助项目(60172012);湖南省自然科学基金资助项目(03JJY3110)
作者简介:唐九阳(1978—),男,博士生。

本文提出一种便于异构数据源集成的能力对象数据模型——能力对象集成模型(简称 COIM 模型),并以 COIM 对象代数作为构建新 COIM 对象的数学基础。COIM 对象代数的操作对象是 COIM 对象,它提供对象并,对象连接,对象选择,对象投影 4 种操作。根据这些提供的操作,可以基于局部数据源的 COIM 对象构造中介器的 COIM 对象。由于可以通过计算得到中介器的能力,用户能够提前预知中介器支持的查询类型,从而直接构造可行的查询,减少查询提交的盲目性;同时中介器本身也可像数据源一样参与其它中介器的集成,集成方式变得更为灵活。

1 COIM 对象模型

关系数据库系统采用关系模型描述自身的结构,支持诸如连接的复杂操作,文件系统仅容许对其内容的简单扫描,文本检索系统则支持利用布尔表达式的简单查询,而 WWW 由非结构化数据构成,往往只允许通过查询表格进行查询,不接受任意类型的 SQL 语句查询,用户只有指定相应属性输入值,系统才能做出响应输出值[9]。

为简化对问题的讨论,本文假定数据源对外提供关系模式集合,如果数据源采用了其它的数据模型,可以通过封装器(Wrapper)包装成关系模式。因此,针对每个数据源提交的查询相当于指定关系模式的某些属性值,而执行查询后得到的结果是关系模式相应投影属性集的元组集合。

例 1 设数据源的输出模式为 $R(X,Y,Z)$。R 中的元组集合为 $\{(x_1,y_1,z_1),(x_1,y_2,z_1),(x_2,y_2,z_2)\}$。那么针对 $R(X,Y,z_1)$ 的查询结果为 $\{(x_1,y_1,z_1),(x_1,y_2,z_1)\}$,而针对 $R(X,y_1,Z)$ 的查询结果为 $\{(x_1,y_1,z_1)\}$。

下面先引入相应的形式化定义。

定义 1 关系的描述称为关系模式,通常简记为

$$R(A_1,A_2,\cdots,A_n)$$

其中 R 为关系名,$A_1,A_2,\cdots,A_n$ 为属性名。

由于中介器也可以看成是虚拟的数据源,因此提出公共能力对象模型如下:

定义 2 能力对象模型 COIM 是一个三元组(id,R,cp),其中:

id 是数据源的标识符,COIM 对象模型的每个对象都含有标识符。

R 是通用数据源(包含局部数据源和中介器)的模式。本文假定各局部数据源属性所来自的域之间不存在语义冲突,同时假定局部数据源模式和中介模式之间也不存在语法冲突,这些同样可以通过封装器对语法和语义冲突进行消解[10~12]。那么,局部数据源模式与中介模式中同名属性的差别在于前者属性所来自的域是后者属性所来自的域的子集。

cp 是一个称作能力描述的二元组(ior,cd),其中 ior 是称作输入输出关系的二元组$(Input,Output)$,$Input$ 和 $Output$ 是 R 中属性名集合的子集。$Input$ 是输入约束属性集,$Output$ 则反映满足输入约束的情况下待投影的属性集。输入输出关系描述了数据源的受限查询能力。cd 是描述数据源内容的二元组(att_i,dom_i),att_i 是 R 中的属性,dom_i 是中介模式中该属性所来自的域 D_i 的子集。如果说一个元组 t 满足 cd 当且仅当 cd 中的每个元组(att_i,dom_i),都有 $t.att_i\in dom_i$。对于 $Input$ 中的每个属性 att_j,cd 中必然存在元组(att_j,dom_j)。

注:一个数据源可以包含多个能力描述,其中每个能力描述都对应一个不同的 COIM 对象,但这些 COIM 对象的标识符(id)相同。下文提出的 COIM 对象代数应用于标识符不同的 COIM 对象。

2 COIM 对象代数

COIM 对象模型从便于异构数据源能力对象的集成出发,提供一系列 COIM 对象操作的定义,它们分别是对象并、对象连接、对象选择和对象投影,这些操作的总和称为 COIM 对象代数。COIM 对象代数满足封闭特性,即 COIM 对象运算的结果仍是 COIM 对象。

由于中介器的处理较复杂,可以支持简单和高级的应用,诸如对数据源返回的结果进行过滤、选择等后处理,通过在不同的操作对象间传递属性值连接返回结果。因此,以下定义的 COIM 对象操作分情

况进行讨论。

2.1 对象并

COIM 对象 O_1 和 O_2 的并记作 $O_1 \cup O_2$。

设 $O_1=(id_1,R_1,cp_1)$,$O_2=(id_2,R_2,cp_2)$,$R_1=R_2$(即两个关系都有 n 个属性,且相应的属性取自同一个域),$cp_1=(ior_1,cd_1)$,$ior_1=(Input_1,Output_1)$,$cd_1=\{(att1_i,dom1_{att1i})\mid att1_i\in Att_1\}$,$cp_2=(ior_2,cd_2)$,$ior_2=(Input_2,Output_2)$,$cd_2=\{(att2_j,dom2_{att2j})\mid att2_j\in Att_2\}$,$Att_1$ 和 Att_2 分别是 cd_1 和 cd_2 的属性名集合。

为了叙述上的方便,先定义几个记号:

定义 3 设 COIM 对象 O 中 cd 的属性名集合为 Att,那么 cd 在其属性名子集合 A 上的投影 $cd\mid_A=\{(att_k,dom_{attk})\mid att_k\in A\subseteq Att\}$。

定义 4 设 R 是 COIM 对象 O 中的关系模式,$ATT(R)$表示取 R 的属性名集合。

定义 5 cd_1 和 cd_2 的复合运算$\oplus$定义为

$cd_1\oplus cd_2=cd_1\mid_{Att1-Att2}\oplus cd_2\mid_{Att2-Att1}=\{(att_k,dom_{attk})\mid\in(att_k,dom_{attk})\in cd_1\mid_{Att1-Att2})\cup((att_k,dom_{attk})\in cd_2\mid_{Att2-Att1})\}$。

定义 6 cd_1 和 cd_2 的复合运算 Θ 定义为

$cd_1\Theta cd_2=cd_1\mid_{Att1\cap Att2}\Theta cd_2\mid_{Att1\cap Att2}=\{(att_k,dom_{attk})\mid att_k\in Att_1\cap Att_2,dom_{attk}=dom1_{attk}\cup dom2_{attk}\}$。

定义 7 cd_1 和 cd_2 的复合运算$\odot$定义为

$cd_1\odot cd_2=cd_1\mid_{Att1\cap Att2}\odot cd_2\mid_{Att1\cap Att2}=\{(att_k,dom_{attk})\mid att_k\in Att_1\cap Att_2,dom_{attk}=dom1_{attk}\cap dom2_{attk}\}$。

对象 O_1 与对象 O_2 的并由属于 O_1 或属于 O_2 的元组组成,记 $O_1\cup O_2=(id,R,cp)$。O_1 和 O_2 对象并的结果是一个新的 COIM 对象。各操作结果 COIM 对象的标识符 id 不具有实际意义,都由系统自动生成,以后不再赘述。下面仅定义 COIM 对象的模式集 R 和能力集 cp。

$ATT(R)=ATT(R_1)=ATT(R_2)$,

$cp=(ior,cd)$,

$ior=(Input,Output)$,$Input=\{in_i\mid in_i\in(Input_1\cup Input_2)\}$,$Output=\{out_i\mid out_i\in(Output_1\cup Output_2)\}$,

$cd=(cd_1\oplus cd_2)\cup(cd_1\Theta cd_2)$。

例 2 设中介能力对象 $O(id,R,cp)$由三个数据源能力对象 $O_1(id_1,R_1,cp_1)$,$O_2(id_2,R_2,cp_2)$和 $O_3(id_3,R_3,cp_3)$的并得到,三个数据源能力对象的描述如图 1。

在图 1 中,对于 O_1 来说,用户只有提供 x 属性的值,数据源才能做出响应,返回(x,y)元组集合。

因此,cp_1 和 cp_2 的并 $cp_t=((\{x,y\},\{x,y,z\}),\{(x,\{0,1,2\}),(y,\{True,False\}),(z,\{Jan,\cdots,June\})\})$,而 cp_t 和 cp_3 的并 $cp=((\{x,y,z\},\{x,y,z\}),\{(x,\{0,1,2,4,5\}),(y,\{True,False\}),(z,\{Jan,\cdots,Dec\})\})$。$ATT(R)=\{x,y,z\}$。则中介器能力对象的描述如图 2 所示。

O_1	$(id_1, R_1(x,y,z),$ $\quad (\{x\}, \{x,y\}),$ cp_1 $\quad \{(x, \{0,1,2\})\}$ ior_1 cd_1 $)$
O_2	$(id_2, R_2(x,y,z),$ $\quad (\{y\}, \{y,z\}),$ $\quad \{(y, \{True, False\}), (z, \{Jan, \cdots, June\})\}$ $)$
O_3	$(id_3, R_3(x,y,z),$ $\quad (\{z\}, \{x,y,z\}),$ $\quad \{(x, \{4,5\}), (z, \{Jan, \cdots, Dec\})\}$ $)$

图1 数据源能力对象描述

Fig.1 Descriptions for capability objects of data sources

O	$(id, R(x,y,z),$ $\quad (\{x,y,z\}, \{x,y,z\}),$ $\quad \{(x, \{0,1,2,4,5\}), (y, \{True, False\}), (z, \{Jan, \cdots, Dec\})\}$ $)$

图2 中介器能力对象描述

Fig.2 Description for capability object of mediator

如果中介器支持后处理等高级应用,对象并操作的定义不变。

2.2 对象连接

COIM **对象** O_1 **和** O_2 **的连接记作** $O_1 \bowtie O_2$**。它是从两个对象的关系的笛卡尔积中选取属性间满足一定条件的元组。本文仅讨论自然连接,其他连接的定义可以类似得到。**

设 O_1 **和** O_2 **的定义如** 2.1 **节,记** $O_1 \bowtie O_2 = (id, R, cp)$**。**

不考虑属性值在数据源间的传递,则 R **和** cp **的定义如下:**

$ATT(R) = ATT(R_1) \cup ATT(R_2)$,

$cp = (ior, cd)$,

$ior = (Input, Output)$, $Input = \{in_i \mid in_i \in (Input_1 \cup Input_2)\}$, $Output = \{out_i \mid out_i \in (Output_1 \cup Output_2)\}$,

$cd = (cd_1 \oplus cd_2) \cup (cd_1 \odot cd_2)$。

例3 设中介对象 $O(id, R, cp)$ **由三个数据源能力对象** $O_1(id_1, R_1, cp_1)$, $O_2(id_2, R_2, cp_2)$ **和** $O_3(id_3, R_3, cp_3)$ **的连接得到。即** $O = O_1 \bowtie O_2 \bowtie O_3$**。其中:**

$ATT(R_1) = \{x,y,z\}$, $cp_1 = ((\{x\}, \{y,z\}), \{(x, \{0,1,2\}), (z, \{Jan, \cdots, Dec\})\})$, $ATT(R_2) = \{z,u,v\}$, $cp_2 = ((\{z\}, \{z,u\}), \{(z, \{Jan, \cdots, June\})\})$, $ATT(R_3) = \{v,w\}$, $cp_3 = ((\{v\}, \{w\}), \{(v, \{True, False\})\})$。

那么,$ATT(R) = \{x,y,z,u,v,w\}$, $cp = ((\{x,z,v\}, \{y,z,u,w\}), \{(x, \{0,1,2\}), (z, \{Jan, \cdots, June\}), (v, \{True, False\})\})$。

如果考虑属性值在数据源间的传递,则 R **和** cp **的定义为:**

$ATT(R) = ATT(R_1) \cup ATT(R_2)$,

$cp = (ior, cd)$,

$ior = (Input, Output)$, $Input = \{in_i \mid in_i \in ((Input_1 \cup Input_2) - Output_1)\}$, $Output = \{out_i \mid out_i \in (Output_1 \cup Output_2)\}$,

$cd=(cd_1\oplus cd_2)\cup(cd_1\odot cd_2)$。

例4 设中介对象 $O(id,R,cp)$ 由三个数据源能力对象 $O_1(id_1,R_1,cp_1)$,$O_2(id_2,R_2,cp_2)$ 和 $O_3(id_3,R_3,cp_3)$ 的连接得到,其中三个对象的定义如例3。那么,$ATT(R)=\{x,y,z,u,v,w\}$,$cp=((\{x,v\},\{y,z,u,w\}),\{(x,\{0,1,2\}),(z,\{Jan,\cdots,June\}),(v,\{True,False\})\})$。

与例2相比,由于中介器可以通过在操作对象 O_1 和 O_2 之间传递绑定属性值 z 来执行连接操作,因此 cp 中的 $Input$ 集合中略去了绑定属性 z。

2.3 对象选择

对象选择是按照一定条件 f,在给定 COIM 对象 $O_1(id_1,R_1,cp_1)$ 中进行选择,用公式表示为

$$\delta[f](O_1)=(id,R,cp)$$

这里,f 表示选择条件,f 的形式如下:$att\ \theta_c$,$\theta\in\{<,>,\leqslant,\geqslant,=\}$,$att\in Att1$,$c\in const$。

设 O_1 的定义如2.1节。不考虑中介器的后处理,则 R 和 cp 的定义如下:

$ATT(R)=ATT(R_1)$,

$cp=(ior,cd)$,

$ior=(Input,Output)$,$Input=\{in_i\mid in_i\in Input_1\}$,$Output=\{out_i\mid out_i\in Output_1\}$,

$cd=(cd_1-cd_1\mid_{att})\cup\{(att,dom_{att}\cap f)\}$。

例5 设中介对象 $O(id,R,cp)$ 由数据源能力对象 $O_1(id_1,R_1,cp_1)$ 通过选择条件($x\leqslant 3$)得到。$ATT(R_1)=\{x,y,z\}$,$cp_1=((\{x,y\},\{y,z\}),\{(x,\{0,1,2,3,4,5\}),(y,\{True,False\})\})$。

那么,$ATT(R)=\{x,y,z\}$,$cp=((\{x,y\},\{y,z\}),\{(x,\{0,1,2,3\}),(y,\{True,False\})\})$。

如存在后处理,并且选择条件形如 $att=c$,由于输入绑定属性可以根据选择条件直接推导得到,则绑定属性集可以略去选择条件中的属性。因此,R 和 cp 的定义如下:

$ATT(R)=ATT(R_1)$,

$cp=(ior,cd)$,

$ior=(Input,Output)$,$Input=\{in_i\mid in_i\in(Input_1-att)\}$,$Output=\{out_i\mid out_i\in Output_1\}$,

$cd=(cd_1-cd_1\mid att)\cup\{(att,dom_{att}\cap f)\}$。

例6 设中介对象 $O(id,R,cp)$ 由数据源能力对象 $O_1(id_1,R_1,cp_1)$ 通过选择条件($x=3$)得到。O_1 的定义如例5。

那么,$ATT(R)=\{x,y,z\}$,$cp=((\{y\},\{y,z\}),\{(x,\{3\}),(y,\{True,False\})\})$。

2.4 对象投影

对象投影是在给定 COIM 对象 $O_1(id_1,R_1,cp_1)$ 中选取子对象。

给定 COIM 对象 $O_1(id_1,R_1,cp_1)$ 以及相关的属性集 $Att=\{att_1,\cdots,att_k\}$。COIM 对象在 Att 上的投影可写作

$$\Pi[att_1,\cdots,att_k](O_1)=(id,R,cp)$$

如果 O_1 的输入属性集 $Input_1$ 不属于 Att,即 $Input_1\not\subset Att$,那么中介器没有相应的能力模板(能力对象)。如果 $Input_1\subseteq Att$,则 O_1 进行投影操作相当于将能力对象的 R、$Output_1$ 和 cd_1 中未被投影的属性过滤掉。

$ATT(R)=Att$,

$cp=(ior,cd)$,

$ior=(Input,Output)$,$Input=\{in_i\mid in_i\in Input_1\}$,$Output=\{out_i\mid out_i\in(Output_1\cap Att)\}$,

$cd=cd_1\mid_{Att}$。

例7 中介对象 $O(id,R,cp)$ 由数据源能力对象 $O_1(id_1,R_1,cp_1)$ 投影属性集 $\{x,y,z\}$ 得到。$ATT(R_1)=\{x,y,z,u\}$,$cp_1=((\{x,y\},\{z,u\}),\{(x,\{0,1,2\}),(u,\{True,False\})\})$。

那么,$ATT(R)=\{x,y,z\}$,$cp=((\{x,y\},\{z\}),\{(x,\{0,1,2\})\})$。

3 应用

假定三个待集成的数据源通过封装器包装后的模式 R_1、R_2 和 R_3 如下所示：

$R_1(title, ISBN, publisher, publication\_date)$,

$R_2(author, title, subject, abstract)$,

$R_3(author, title, subject)$。

其中数据源1是提供华南地区1992到2004年出版图书类信息的网站，该网站对外提供两个访问接口：用户通过指定书籍名可以得到该书籍的ISBN号、出版社和出版时间，用户也可通过指定出版社和出版时间来浏览满足条件的所有书籍；数据源2的关系数据库中存储了计算机类图书的信息，外部用户提交基于书籍名的查询，系统返回作者和书籍摘要元组；数据源3是计算机、通信和经济管理类图书的信息提供网站，用户点击选择书籍的类别，就得到相关的所有书籍和作者。根据以上描述，3个数据源对应的能力对象如图3所示，其中数据源1包含 O_1 和 O_2 两个能力对象。

O_1	$(id_1, R_1(title, ISBN, publisher, publication\_date),$ $(\{title\}, \{ISBN, publisher, publication\_date\}),$ ior_1 $\{(title, String), (ISBN, String), (publisher, SOUTHCHINAPUBLISHER), (publication\_date,$ $\{1992, \cdots, 2004\})\}$ cd_1 $)$
O_2	$(id_2, R_1(title, ISBN, publisher, publication\_date),$ $(\{publisher, publication\_date\}, \{title, ISBN\}),$ ior_2 $\{(title, String), (ISBN, String), (publisher, SOUTHCHINAPUBLISHER), (publication\_date,$ $\{1992, \cdots, 2004\})\}$ cd_2 $)$
O_3	$(id_3, R_2(author, title, subject, abstract),$ $(\{title\}, \{author, abstract\}),$ ior_3 $\{(author, String), (title, String), (subject, \{computer\}), (abstract, String)\}$ cd_3 $)$
O_4	$(id_4, R_3(author, title, subject),$ $(\{subject\}, \{author, title\}),$ ior_4 $\{(author, String), (title, String), (subject, \{computer, communication, economical\ management\})\}$ cd_4 $)$

图3 图书类数据源能力对象描述

Fig.3 Descriptions for capability objects of book data sources

可以根据现有的3个应用系统构建一个集成的网上图书查询系统，集成模式 R 为

$$R(author, title, subject, ISBN, publisher, publication\_date)$$

应用COIM对象代数，将 O_3 投影后的对象与 O_4 并，然后与 O_1 或 O_2 进行连接，得到中介器的能力对象 O_5 和 O_6，其中

$$O_5 = (\prod[author, title, subject](O_3) \cup O_4) \bowtie O_1$$

$$O_6 = (\prod[author, title, subject](O_3) \cup O_4) \bowtie O_2$$

假设中介器支持后处理，属性值可以在数据源间进行传递，得到中介器的能力对象 O_5 和 O_6 如图4。

O_5	$(id_5, R(author, title, subject, ISBN, publisher, publication\_date),$ $(\{title, subject\}, \{author, ISBN, publisher, publication\_date\}),$ $\boldsymbol{ior_5}$ $\{(author, String), (title, String), (subject, \{computer, communication, economical\ management\}), (ISBN, String), (publisher, SOUTHCHINAPUBLISHER), (publication\_date, \{1992, \cdots, 2004\})\}$ $\boldsymbol{cd_5}$ $)$
O_6	$(id_6, R(author, title, subject, ISBN, publisher, publication\_date),$ $(\{title, subject, publisher, publication\_date\}, \{author, title, ISBN\}),$ $\boldsymbol{ior_6}$ $\{(author, String), (title, String), (subject, \{computer, communication, economical\ management\}), (ISBN, String), (publisher, SOUTHCHINAPUBLISHER), (publication\_date, Y\ \{1992, \cdots, 2004\})\}$ $\boldsymbol{cd_6}$ $)$

图 4 图书集成系统的中介器能力对象描述

Fig.4 Descriptions for capability objects of mediator in book integration system

对于通过计算得到的中介器能力对象 O_5,用户提交查询时必须同时提供 *title* 和 *subject* 属性值,系统才能做出响应;而针对 O_6 提交的查询,则需提供 *title*,*subject*,*publisher* 和 *publication_date* 属性值,而且出版社的域值必须限定在华南地区,出版时间的域值限定于 1992 年到 2004 年。显然,用户由于提前预知中介器支持的查询类型,从而能够直接构造可以执行的查询,减少查询提交的盲目性。

4 结束语

异构数据源集成系统中,有必要描述中介器的能力,以便中介器自身可以像局部数据源一样灵活地参与集成。本文提出一种便于异构数据源集成的 COIM 模型。COIM 模型基于输入输出属性集以及数据源内容的描述,因而能自然地刻画数据源和中介模式的能力。COIM 对象代数提供对象并、连接、选择、投影 4 种操作,是 COIM 能力对象集成的数学基础。基于局部数据源的能力对象,应用 COIM 对象代数中提供的操作,可以计算得到中介器的能力,从而引导用户预先了解中介器支持的查询类型,减少查询提交的盲目性。

参 考 文 献:

[1] Roth M T, Schwarz P M. Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources[A]. VLDB 1997:266 – 275.

[2] Li C, Chang E. Query Planning with Limited Source Capabilities[A]. In ICDE, 2000,401 – 412.

[3] Li C, Yerneni, Vassalos V, Papakonstantinou Y, Garcia-Molina H, Ullman J. Capability based Mediation in Tsimmis[A]. In Proceedings of ACM SIGMOD Conference, 1998.

[4] Haas L, Kossman D, Wimmers E, Yang J. Optimizing Queries across Diverse Data Sources[A]. In Proceedings of VLDB Conference, 1997.

[5] Kapitskaia O, Tomasic A, Valduriez P. Scaling Heterogeneous Databases and the Design of Disco[R]. INRIA Technical Report, 1997.

[6] 王宁,王能斌.异构数据源集成系统查询分解和优化的实现[J].软件学报,2000,11(2):222 – 228.

[7] Levy A, Rajaraman A, Ordille J. Querying Heterogeneous Information Sources using Source Description[A]. In Proceedings of International Conference on Very Large Data Bases, 1996.

[8] Florescu D, Levy A, Manolescu I, Suciu D. Query Optimization in the Presence of Limited Access Patterns[A]. In Proceedings of the ACM SIGMOD Conference, 1999.

[9] Yerneni R, Li C, Ullman J D, Molina H G. Optimizing Large Join Queries in Mediation Systems[A]. In ICDT, 1999:348 – 364.

[10] Batini C, Lenzerini M, Navathe S B. A Comparative Analysis of Methodologies for Database Schema Integration[J]. ACM Computing Surveys, December 1986,18(4):323 – 364.

[11] Kim W, Choi I, Gala S, Scheevel M. On Resolving Schematic Heterogeneity in Multidatabase Systems[J]. Distributed and Parallel Databases, 1993,1(3).

[12] Krishnamurthy R, Litwin W, Kent W. Language Features for Interoperability of Databases with Schematic Discrepancies[A]. In Proceedings of the ACM SIGMOD Conference, 1991.