XML 强闭包依赖的研究*)

殷丽凤[1]　郝忠孝[1,2]
(哈尔滨理工大学计算机科学与技术学院　哈尔滨 150080)[1]
(哈尔滨工业大学计算机科学与技术学院　哈尔滨 150001)[2]

摘　要　XML 闭包依赖是基本的完整性约束。当 XML 文档出现不完全信息时,XML 闭包依赖同样对于阻止 XML 数据的更新异常、查询优化以及索引设计具有重要的意义。提出了在不完全信息环境下基于一致路径集合的 XML 强函数依赖和 XML 强闭包依赖的定义,研究了 XML 强闭包依赖的判定定理。在提出 XML 强闭包依赖推理规则的基础上,对其有效性和完备性进行了证明。最后分析了 XML 强闭包依赖和 XML 强函数依赖的关系。

关键词　不完全信息,XML 强函数依赖,XML 强闭包依赖,推理规则

Research on XML Strong Inclusion Dependency

YIN Li-feng[1]　HAO Zhong-xiao[1,2]
(School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)[1]
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)[2]

Abstract　XML inclusion dependency is essential integrity constraints. When XML documents come forth incomplete information, XML inclusion dependency has also important meaning for updating anomaly prevention, query optimization and index design in XML document. XML strong functional dependency' definition and XML strong inclusion dependency' definition based on consistent path set under the incomplete information circumstances were formalized, judging theorem for XML strong inclusion dependency was studied. Based on formalizing inference rules of XML strong inclusion dependency, the soundness and completeness of inference rules were proved. At last the relation of XML strong functional dependency and XML strong inclusion dependency was analysed.

Keywords　Incomplete information, XML strong functional dependency, XML strong inclusion dependency, Inference rule

1　引言

XML 文档已经成为 Web 应用之间的信息交换标准。但 XML 在提供了从不同种类数据源表示数据的一种方便语法的同时,却没有提供完善的完整性约束方面的语义信息。在数据库领域,完整性约束对于查询优化、阻止更新异常等方面都具有很重要的意义。为此,有必要对 XML 数据的完整性约束进行深入研究。目前国外许多作者已经对 XML 完整性约束进行了探讨,文献[1]研究了 XML 文档的键约束,文献[2]研究了 XML 文档的路径约束,文献[3]研究了 XML 文档的函数依赖约束,文献[4]研究了 XML 文档的多值依赖约束,文献[5-7]研究了 XML 文档的闭包约束。

由于客观世界存在大量不完全信息,所以研究不完全信息环境下数据库的完整性约束具有重要意义。在关系数据库中,文献[8]研究了基于存在型不完全信息对不完全信息环境下的函数依赖语义约束,文献[9]对不完全信息环境下的闭包依赖进行了研究,但是这些理论还没有延伸到 XML 数据库理论环境,当 XML 文档出现不完全信息时,XML 完整性约束的研究还很少。为了研究不完全信息环境下 XML 外键以及范式问题,本文提出 XML 强闭包依赖理论。

本文在一致路径集合的基础上提出了不完全信息环境下 XML 强函数依赖和 XML 强闭包依赖的定义,给出了 XML 强闭包依赖的判定定理。为了解决逻辑蕴涵问题,提出了 XML 强闭包依赖推理规则,并对其有效性和完备性进行了证明。为了研究范式,最后分析了 XML 强闭包依赖和 XML 强函数依赖的关系。

2　基本知识

为了便于形式化描述 XML 强闭包依赖,本文引进了路径和路径节点集[10]等概念。本文中,$s \triangleright p$ 表示路径节点集 s 通过路径 p,$X \subseteq Y$ 表示 X 是 Y 的子树;φ 表示不完全信息中的存在型不完全信息[8],简记为不完全信息。

定义 1[10]　不完全的 XML 文档树是一个六元组 $T=(V, lab, ele, att, val, v_r)$。其中,$V$ 表示 T 的节点集合;lab 表示从 V 到 $E \cup A \cup \{S\} \cup \varphi$ 的函数,满足 $\forall v \in V$。若 v 为分枝节点,则 $lab(v) \in E$;若 v 为叶子节点,则 $lab(v) \in \{A \mid S \mid \varphi\}$。其中 E 表示元素名字的有限集合,$\mid$ 表示或,A 表示属性名字的有限集合,S 表示文本字符串,φ 表示不完全信息。ele 表示从节点 V 到 V 中一系列节点的部分映射,满足 $\forall v \in V$。若存在 $ele(v)$,则 $lab(v) \in E$。att 表示从 $V \times A$ 到 V 的部分函数,满足 $\forall v \in V, l \in A$。若 $att(v, l)=v_1$,则 $lab(v) \in E$ 且 $lab(v_1)=l$。val 表示节点的函数值,$\forall v \in V$。若 $lab(v) \in E$,

*)该项目得到黑龙江省自然科学基金资助(F200702)。**殷丽凤**　博士研究生,主要研究领域为 XML 数据库不完全信息理论和主动 XML 数据库理论;**郝忠孝**　博导,教授,主要研究领域为数据库系统与理论。

则 $val(v)=v$；若 $lab(v)=S$ 或 $lab(v)\in A$，则 $val(v)$ 为字符串；若 $lab(v)=\varphi$，则 $val(v)=\varphi$；v_r 为 T 的根节点。若 T 不出现 φ，则称为完全的 XML 文档树。

定义 2[10] 设 P 为路径的集合，$\forall p\in P$，若路径 p_1 是路径 p 的路径前缀，则 $p_1\in P$，称 P 为一致路径集合。最后标识是叶子节点标识的路径，称为全路径。

定义 3 XML 数据库模式为一致路径集合中的全路径集合的集合，记作 $S(P_1,\cdots,P_n)$，其中 P_i 为一致路径集合中的全路径集合，$i\in[1,n]$，简记为 S。

定义 4 设 P 为一致路径集合中的全路径集合，T 为不完全的 XML 文档树。若 $s\in T$，$p\in P$，$s\rhd p$，函数 $N(p)$ 返回通过路径 p 的路径节点集 s 的最后节点；若 $\forall s\in T$，则 $\exists p\in P$ 且 $s\rhd p$，称 T 符合 P，记作 $T\triangle P$。

定义 5 设 XML 的数据库模式为 $S(P_1,\cdots,P_n)$，符合 P_i 的完全 XML 文档树 T_i 所构成的森林 $(T_1,\cdots,T_n)$ 为完全 XML 数据库，记作 $FD(i\in[1,n])$，也称 FD 符合 S，记作 $FD\geqslant S$。若 $T_1,\cdots,T_n$ 中有一个为不完全的 XML 文档树，则称为不完全的 XML 数据库，记作 ND，也称 ND 符合 S，记作 $ND\geqslant S$。

定义 6 设 P 和 T 的意义同定义 4，若 $\forall p\in P$，则 $\exists s\in T$，$s\rhd p$ 且 s 唯一；反之，若 $\forall s\in T$，则 $\exists p\in P$，$s\rhd p$ 且 p 唯一，称 T 与 P 为一对一关系，记作 $T\leftrightarrow P$。

定义 7 设 P 和 T 的意义同定义 4 且 $T\triangle P$，T 的可能世界的集合记为 $POSS(T)$，$POSS(T)=\{T'\mid T'\triangle P$，存在映射 $f: T\rightarrow T'$，满足 $\forall s\in T$，$\exists p\in P$，$s\rhd p$，若 $val(N(p))=\varphi$，则 $f(val(N(p)))\neq\varphi\}$。

定义 8 设 P 和 T 的意义同定义 4 且 $T\triangle P$，$T_1\subseteq T$，$T_2\subseteq T$，$T_1\leftrightarrow P$，$T_2\leftrightarrow P$。对于 $p\in P$，T_1 中的 $val(N(p))$ 的值记为 $val(N_1(p))$，T_2 中的 $val(N(p))$ 的值记为 $val(N_2(p))$，满足下面的条件之一，则称 $val(N_1(p))$ 和 $val(N_2(p))$ 等价，记作 $val(N_1(p))\neq val(N_2(p))$。

(1)若 $val(N_1(p))$ 和 $val(N_2(p))$ 都为完全信息，则 $val(N_1(p))=val(N_2(p))$。

(2)若 $val(N_1(p))$ 和 $val(N_2(p))$ 都为不完全信息，它们的语义信息相同，进行不完全信息代换时，$val(N_1(p))$ 和 $val(N_2(p))$ 代换为同一值。

定义 9 设 P,T,T_1,T_2 的意义同定义 7。对于 $p\in P$，满足下面的条件之一，则称 $val(N_1(p))$ 和 $val(N_2(p))$ 相容，记作 $val(N_1(p))\doteqdot val(N_2(p))$；否则称 $val(N_1(p))$ 和 $val(N_2(p))$ 不相容，记作 $val(N_1(p))\not\doteqdot val(N_2(p))$。

(1) $val(N_1(p))\doteq val(N_2(p))$。

(2)若 $val(N_1(p))$ 为完全信息，$val(N_2(p))$ 为不完全信息，则至少存在一个 $T'\in POSS(T)$，使 $val(N_1(p))=val(N_2(p))$ 成立。

(3)若 $val(N_1(p))$ 和 $val(N_2(p))$ 都为不完全信息，且它们的限定的代换范围的交集不是空的，即至少有一种不完全信息代换过程，使 $val(N_1(p))=val(N_2(p))$。

3 定义、定理

定义 10 设 $ND\geqslant S$，$P_i\in S$，$T_i\triangle P_i$，$X\subseteq P_i$，$X=\{x_1,\cdots,x_n\}$，$Y\subseteq P_i$，$Y=\{y_1,\cdots,y_m\}$。通过 P_i 的 XML 强函数依赖（记作 XSFD）表示形式为 $P_i: X\xrightarrow{s}Y$，只要 $\forall T_1,T_2\subseteq T_i$ 且 $T_1\leftrightarrow X\cup Y$，$T_2\leftrightarrow X\cup Y$。若 $val(N_1(x_a))\doteqdot val(N_2(x_a))$ $(a\in[1,n])$，则 $val(N_1(y_b))\doteq val(N_2(y_b))(b\in[1,m])$，则此 XSFD 在 T_i 上成立，记作 $T_i\models P_i: X\xrightarrow{s}Y$。

定义 11 设 $FD\geqslant S$，$P_i,P_j\in S$，$T_i\triangle P_i$，$T_j\triangle P_j$，$X\subseteq P_i$，$X=\{x_1,\cdots,x_n\}$，$Y\subseteq P_j$，$Y=\{y_1,\cdots,y_n\}$，$x=x_1\cap\cdots\cap x_n$，$y=y_1\cap\cdots\cap y_n$，$i,j\in[1,n]$。若 $i\neq j$，则 $x_i\neq x_j$ 且 $y_i\neq y_j$。XML 闭包依赖（记作 XIND）的表示形式为 $P_i[X]\subseteq P_j[Y]$，只要 $\forall T_X\subseteq T_i$，$T_X\leftrightarrow X$，则 $\exists T_Y\subseteq T_j$，$T_Y\leftrightarrow Y$ 且满足 $val(N_X(x_a))=val(N_Y(y_a))(a\in[1,n])$，则此 XIND 在 FD 上成立，记作 $FD\models P_i[X]\subseteq P_j[Y]$。

定义 12 设 $ND\geqslant S$，其它条件同定义 11。XML 强闭包依赖（记作 XSIND）表示形式为 $P_i[X]\overset{s}{\subseteq}P_j[Y]$，只要 $\forall T_X\subseteq T_i$，$T_X\leftrightarrow X$，则 $\exists T_Y\subseteq T_j$，$T_Y\leftrightarrow Y$ 且满足 $val(N_X(x_a))\doteq val(N_Y(y_a))(a\in[1,n])$，则此 XSIND 在 ND 上成立，记作 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$。

定理 1 若 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 成立，$\forall ND'\in POSS(ND)$，则 $ND'\models P_i[X]\subseteq P_j[Y]$。

证明：由 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 得：$\forall T_X\subseteq T_i$，$T_X\leftrightarrow X$，则 $\exists T_Y\subseteq T_j$，$T_Y\leftrightarrow Y$ 且满足 $val(N_X(x_a))\doteq val(N_Y(y_a))(a\in[1,n])$。根据等价的定义，$\forall ND'\in POSS(ND)$ 都满足 $val(N_X(x_a))=val(N_Y(y_a))(a\in[1,n])$，所以 $ND'\models P_i[X]\subseteq P_j[Y]$ 成立。证毕。

4 XSIND 的推理规则

为了解决 XML 逻辑蕴含的判定问题，需要给出一个推理规则集。推理规则集首先必须是正确的，即推导出的 XML 强闭包依赖确实是成立的；其次推理规则集应该是完备的，即可以推导出所有成立的 XML 强闭包依赖。下面给出 XML 强闭包依赖的推理规则集 R：

(1) 自反规则。$ND\models P_i[X]\overset{s}{\subseteq}P_i[X]$。其中 $X=\{x_1,\cdots,x_n\}$。

(2) 排列投影规则。若 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 成立，则 $ND\models P_i[X']\overset{s}{\subseteq}P_j[Y']$ 也成立。其中 $X=\{x_1,\cdots,x_n\}$，$Y=\{y_1,\cdots,y_n\}$，$X'=\{x_{i_1},\cdots,x_{i_m}\}$，$Y'=\{y_{i_1},\cdots,y_{i_m}\}$，$[i_1,\ i_m]\in[1,n]$ $(i_m\leqslant n)$。

(3) 传递规则。若 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 和 $ND\models P_j[Y]\overset{s}{\subseteq}P_k[Z]$ 成立，则 $ND\models P_i[X]\overset{s}{\subseteq}P_k[Z]$ 也成立。其中 $X=\{x_1,\cdots,x_n\}$，$Y=\{y_1,\cdots,y_n\}$，$Z=\{z_1,\cdots,z_n\}$。

定理 2 XSIND 的推理规则集(1)－(3)是正确的。

证明：(1) 设 $T_i\triangle P_i$，$T_{X1}\subseteq T_i$，$T_{X1}\leftrightarrow X$，$T_{X2}\subseteq T_i$，$T_{X2}\leftrightarrow X$，由 $T_{X1}\subseteq T_i$、$T_{X2}\subseteq T_i$ 成立得 T_{X2} 和 T_{X1} 可以为同一个文档子树。所以 $\forall T_{X1}\subseteq T_i$，$T_{X1}\leftrightarrow X$，则 $\exists T_{X2}\subseteq T_i$、$T_{X2}\leftrightarrow X$ 且满足 $val(N_{X1}(x_a))\doteq val(N_{X2}(x_a))(a\in[1,n])$，所以 $ND\models P_i[X]\overset{s}{\subseteq}P_i[X]$ 成立。

(2) 由 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 得：$\forall T_X\subseteq T_i$、$T_X\leftrightarrow X$，则 $\exists T_Y\subseteq T_j$、$T_Y\leftrightarrow Y$ 且满足 $val(N_X(x_a))\doteq val(N_Y(y_a))(a\in[1,n])$。又由 $[i_1,i_m]\in[1,n](i_m\leqslant n)$，$X'=\{x_{i_1},\cdots,x_{i_m}\}$，$Y'=\{y_{i_1'},\cdots,y_{i_m'}\}$，则 $val(N_X(x_b))\doteq val(N_Y(y_b))(b\in[i_1,i_m])$ 成立，即 $ND\models P_i[X']\overset{s}{\subseteq}P_j[Y']$ 成立。

(3) 由 $ND\models P_i[X]\overset{s}{\subseteq}P_j[Y]$ 得：$\forall T_X\subseteq T_i$，$T_X\leftrightarrow X$，则

$\exists T_Y \subseteq T_j, T_Y \leftrightarrow Y$ 且满足 $val(N_X(x_a)) \doteq val(N_Y(y_a))(a \in [1,n])$。又由 $ND \models P_j[Y] \overset{s}{\subseteq} P_k[Z]$ 得：$\forall T_Y \subseteq T_j$、$T_Y \leftrightarrow Y$，则 $\exists T_Z \subseteq T_k, T_Z \leftrightarrow Z$ 且满足 $val(N_Y(y_b)) \doteq val(N_Z(z_b))(b \in [1,n])$。所以 $\forall T_X \subseteq T_i, T_X \leftrightarrow X$，则 $\exists T_Z \subseteq T_k$、$T_Z \leftrightarrow Z$ 且满足 $val(N_X(x_c)) \doteq val(N_Z(z_c))(c \in [1,n])$，即 $ND \models P_i[X] \overset{s}{\subseteq} P_k[Z]$ 成立。证毕。

在证明 XSIND 的推理规则集完备性之前，首先给出逻辑蕴含、推导和追逐规则的定义：

定义 13 设 I 表示通过 S 的 XSIND 的集合，若 $ND \models I$，则 $ND \models P_i[X] \overset{s}{\subseteq} P_j[Y]$，则称 I 逻辑蕴含 $P_i[X] \overset{s}{\subseteq} P_j[Y]$，记作 $I \mapsto P_i[X] \overset{s}{\subseteq} P_j[Y]$。

定义 14 给定 XSIND 的推理规则集 R 和 XSIND 集合 I，若 XSIND：$P_i[X] \overset{s}{\subseteq} P_j[Y]$ 能够由 I 根据 R 推出，称 I 能够推导出 $P_i[X] \overset{s}{\subseteq} P_j[Y]$，记作 $I \vdash P_i[X] \overset{s}{\subseteq} P_j[Y]$。

定义 15(追逐规则)设 $ND \models P_i[c_1, \cdots, c_k] \overset{s}{\subseteq} P_j[d_1, \cdots, d_k] \in I, T_i \in ND, T_i \triangle P_i, \exists T_{i'} \subseteq T_i, T_{i'} \leftrightarrow P_i$。设 $T_j \in ND$，$T_j \triangle P_j, \exists T_{j'} \subseteq T_j, T_{j'} \leftrightarrow P_j$ 满足 $val(N_{j'}(d_u)) \doteq val(N_{i'}(c_u)), val(N_{j'}(q_v)) = \varphi(u \in [1,k], v \in [1,m])$，其中 $P_j = \{d_1, \cdots, d_k, q_1, \cdots, q_m\}$。若 $T_{j'}$ 不在 T_j 中，则插入 $T_{j'}$ 到 T_j 中。

推理规则集的完备性保证可以推出所有被逻辑蕴含的 XSIND，即令 $\sigma = P_a[a_1, \cdots, a_k] \overset{s}{\subseteq} P_b[y_1, \cdots, y_k]$，若 $I \mapsto \sigma$，则 $I \vdash \sigma$。下面给出相关的引理。

引理 1 设 $P_a \in S$、$T_a \triangle P_a$、$T_{a'} \subseteq T_a$、$T_{a'} \leftrightarrow P_a$ 满足 $val(N_{a'}(a_i)) = i, val(N_{a'}(p_b)) = \varphi(i \in [1,k], b \in [1,n])$，其中 φ 的取值范围为 0，$P_a = \{a_1, \cdots, a_k, p_1, \cdots, p_n\}(A = \{a_1, \cdots, a_k\})$。初始化 ND 为 $T_{a'}$，其它文档树为空，如图 1 所示。设 $P_j \in S, T_j \triangle P_j, T_{j'} \subseteq T_j, T_{j'} \leftrightarrow P_j$ 满足 $val(N_{j'}(e_u)) = i_u \geq 1$ $(E = \{e, \cdots, e_k\}, E \subseteq (P_j, u \in [1,k])$，若应用追逐规则 $T_{j'} \in ND$，则 $I \vdash P_a[a_{i_1}, \cdots, a_{i_k}] \overset{s}{\subseteq} P_j[e_1, \cdots, e_k]$ 成立。

证明：(归纳法)。应用追逐规则对插入 ND 中的文档子树进行归纳。

(1) 初始化 ND。若 $T_{j'}$ 是文档子树 $T_{a'}$，初始化时 $T_{j'}$ 被插入到 T_j 中，则 T_j 和 T_a 包含的文档子树相同，如图 2 所示。通过自反规则，$I \vdash P_a[a_{i_1} \cdots a_{i_k}] \overset{s}{\subseteq} P_a[a_{i_1} \cdots a_{i_k}]$，所以引理 1 成立。

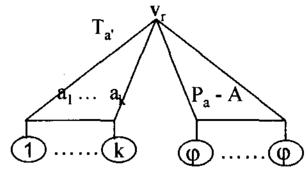

图 1 初始化 $ND = T_a{'}$

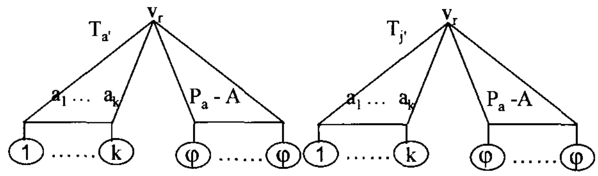

图 2 应用追逐规则初始化 $ND = \{T_{a'}, T_{j'}\}$

(2) 根据 I，应用追逐规则插入所有的文档子树构成的森林为 ND。由 $T_{j'} \in ND$，根据 XSIND 的追逐规则一定存在 $P_m[f_1, \cdots, f_k] \overset{s}{\subseteq} P_j[e_1, \cdots, e_k] \in I$ 满足 $P_m \in S, T_m \triangle P_m, T_{m'} \subseteq T_m, T_{m'} \leftrightarrow P_m, T_{m'} \in ND, val(N_{m'}(f_u)) = val(N_{j'}(e_u)) = i_u$ $(u \in [1,k], F = \{f_1, \cdots, f_k\})$，否则 $T_{j'}$ 不可能在 ND 中。对 $\{a_1, \cdots, a_k\}$ 重新排列投影为 $\{a_{i_1}, \cdots, a_{i_k}\}([i_1, i_k] \in [1,k], i_k \leq k)$，则 $val(N_{a'}(a_{i_u})) = i_u(u \in [1,k])$，再根据 XSIND 的追逐规则和传递规则一定存在 $P_a[a_{i_1}, \cdots, a_{i_k}] \overset{s}{\subseteq} P_m[f_1, \cdots, f_k] \in I$，否则 $T_{m'}$ 不能应用追逐规则加入 ND 中，图 3 显示了 $T_{j'}$, $T_{m'}$ 和 $T_{a'}$ 三者之间的对应关系。由 $P_a[a_{i_i} \cdots a_{i_k}]a_{i_1}, \cdots, a_{i_k} P_m[f_1, \cdots, f_k] \in I$、$P_m[f_1, \cdots, f_k] \overset{s}{\subseteq} P_j[e_1, \cdots, e_k] \in I$，再通过 $XSIND$ 的传递规则，$I \vdash P_a[a_{i_i} \cdots, a_{i_k}] \overset{s}{\subseteq} P_j[e_1, \cdots, e_k]$。证毕。

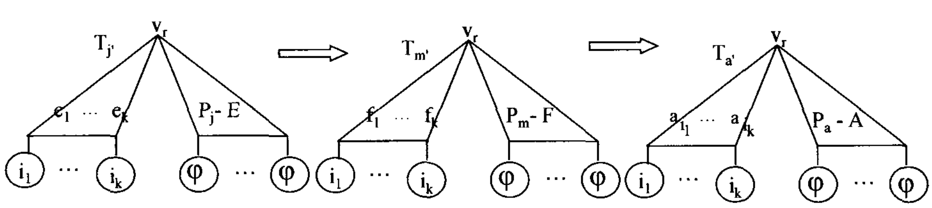

图 3 文档树 $T_{j'}$、$T_{m'}$ 和 $T_{a'}$ 三者之间的对应关系

定理 3 XSIND 的推理规则集是完备的(即 $I \rightarrow \sigma$，则 $I \vdash \sigma$)。

证明：设 XML 数据库模式为 $S(P_1, \cdots, P_n)$，XSIND 的集合为 $I, \sigma = P_a[a_1, \cdots, a_k] \overset{s}{\subseteq} P_b[y_1, \cdots, y_k]$。首先应用追逐规则插入 XML 文档子树，创建符合 S 的不完全的 XML 数据库 $ND = \{T_1, \cdots, T_n\}$。然后证明 $I \vdash \sigma$。

初始化 ND，设 ND 的初始值为符合 P_a 的 XML 文档树 $T_a = \{T_{a'}\}$($T_{a'}$ 的意义同图 1)，ND 的其它所有的 XML 文档树为空，即 $ND = \{T_{a'}\}$。然后根据 I 应用追逐规则插入 XML 文档子树到 ND 中，直到没有 XML 文档子树插入到 ND 中为止。由此，最后的 ND 满足 I，否则能够应用追逐规则插入新的 XML 文档子树。由 $I \mapsto \sigma$，则 $ND \models \sigma$，即 $ND \models P_a[a_1, \cdots, a_k] \overset{s}{\subseteq} P_b[y_1, \cdots, y_k]$。又由 $T_{a'} \in ND$，根据 $P_a[a_1, \cdots, a_k] \overset{s}{\subseteq} P_b[y_1, \cdots, y_k]$ 和追逐规则，则一定存在 $T_{b'} \in ND$，$T_{b'}$ 满足 $T_{b'} \subseteq T_b$、$val(N_{b'}(y_i)) = i$，其中 $T_b \triangle P_b$、$i \in [1,k]$。通过引理 1，$I \vdash P_a[a_1, \cdots, a_k] \overset{s}{\subseteq} P_b[y_1, \cdots, y_k]$，即 $I \vdash \sigma$ 成立，所以结论成立。证毕。

5 XSFD 和 XSIND 的关系

一般情况下，给定 XSFD 和 XSIND，能够推导出 XSFD，给定 XSFD 和 XSIND，也能推导出 XSIND。为了研究范式，下面给出 XSFD 和 XSIND 之间关系的定理。

定理 4 设 XML 数据库模式为 $S(P_1, \cdots, P_n)$，不完全的 XML 数据库为 ND，则有：

(1)若 $ND\vdash P_i[XY]\overset{s}{\subseteq}P_j[WZ]$和 $ND\vdash P_j:W\xrightarrow{s}Z$ 成立,则 $ND\vdash_{P_i}:X\xrightarrow{s}Y$ 也成立。其中 $X=x_1,\cdots,x_m,W=w_1,\cdots,w_m,Y=y_1,\cdots,y_n,Z=z_1,\cdots,z_n$。

(2)若 $ND\vdash P_i[XY]\overset{s}{\subseteq}P_j[WU]$、$ND\vdash P_i[XZ]\overset{s}{\subseteq}P_j[WV]$、$ND\vdash_{P_j}:W\xrightarrow{s}U$ 成立,则 $ND\vdash P_i[XYZ]\overset{s}{\subseteq}P_j[WUV]$也成立。其中 $X=x_1,\cdots,x_m,W=w_1,\cdots,w_m,Y=y_1,\cdots,y_k,U=u_1,\cdots,u_k,Z=z_1,\cdots,z_n,V=v_1,\cdots,v_n$。

(3)设$\Sigma=\{P_i[XY]\overset{s}{\subseteq}P_j[WU],P_i[XZ]\overset{s}{\subseteq}P_j[WU],P_j:W\xrightarrow{s}U\}$。若 $T_i\triangle P_i,T_j\triangle P_j$ 满足Σ,$T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i$,则 $val(N_{i1}(y_b))\doteq val(N_{i1}(z_b))(b\in[1,n])$。其中 $X=x_1,\cdots,x_m,W=w_1,\cdots,w_m,Y=y_1,\cdots,y_n,U=u_1,\cdots,u_n,Z=z_1,\cdots,z_n$。

证明:(1)设 $T_i\triangle P_i,T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i,T_{i2}\subseteq T_i,T_{i2}\leftrightarrow P_i$ 满足 $val(N_{i1}(x_a))\div val(N_{i2}(x_a))(a\in[1,m])$。由 $ND\vdash_{P_i}[XY]\overset{s}{\subseteq}P_j[WZ]$,则存在 $T_j\triangle P_j,T_{j1}\subseteq T_j,T_{j1}\leftrightarrow P_j,T_{j2}\subseteq T_j,T_{j2}\leftrightarrow P_j$ 满足 $val(N_{j1}(w_a))\doteq val(N_{i1}(x_a)),val(N_{j1}(z_b))\doteq val(N_{i1}(y_b))$和 $val(N_{j2}(w_a))\doteq val(N_{i2}(x_a)),val(N_{j2}(z_b))\doteq val(N_{i2}(y_b))$成立$(b\in[1,n])$。由 $val(N_{i1}(x_a))\div val(N_{i2}(x_a))$得 $val(N_{j2}(w_a))\div val(N_{j1}(w_a))$,又由 $ND\vdash P_j:W\xrightarrow{s}Z$ 得 $val(N_{j2}(z_b))\doteq val(N_{j1}(z_b))$,所以 $val(N_{i2}(y_b))\doteq val(N_{i1}(y_b))$。即若 $T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i,T_{i2}\subseteq T_i,T_{i2}\leftrightarrow P_i$ 满足 $val(N_{i1}(x_a))\div val(N_{i2}(x_a))$,则 $val(N_{i1}(y_b))\doteq val(N_{i2}(y_b))$,所以 $ND\vdash P_i:X\xrightarrow{s}Y$ 成立。

(2)设 $T_i,T_j\in ND,T_i\triangle P_i,T_j\triangle P_j$ 满足 $P_i[XY]\overset{s}{\subseteq}P_j[WU],P_i[XZ]\overset{s}{\subseteq}P_j[WV],P_j:W\xrightarrow{s}U$。设 $T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i$,则有 $T_{j1}\subseteq T_j,T_{j1}\leftrightarrow P_j,T_{j2}\subseteq T_j,T_{j2}\leftrightarrow P_j$ 满足 $val(N_{j1}(w_a))\doteq val(N_{i1}(x_a)),val(N_{j1}(u_b))\doteq val(N_{i1}(y_b))$且 $val(N_{j2}(w_a))\doteq val(N_{i1}(x_a)),val(N_{j2}(v_c))\doteq val(N_{i1}(z_c))(a\in[1,m],b\in[1,k],c\in[1,n])$。所以 $val(N_{j1}(w_a))\doteq val(N_{j2}(w_a))$,则 $val(N_{j1}(w_a))\div val(N_{j2}(w_a))$,由 $P_j:W\xrightarrow{s}U$,则 $val(N_{j1}(u_b))\doteq val(N_{j2}(u_b))$成立。又由 $val(N_{j1}(u_b))\doteq val(N_{i1}(y_b))$,则 $val(N_{j2}(u_b))\doteq val(N_{i1}(y_b))$,所以对于每个 $T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i$,则存在 $T_{j2}\subseteq T_j,T_{j2}\leftrightarrow P_j$ 且满足 $val(N_{j2}(w_au_bv_c))\doteq val(N_{i1}(x_ay_bz_c))$,所以 $ND\vdash P_i[XYZ]\overset{s}{\subseteq}P_j[WUV]$成立。

(3) 设 $T_i,T_j\in ND,T_i\triangle P_i,T_j\triangle P_j$ 满足 $P_i[XY]\overset{s}{\subseteq}P_j[WU],P_i[XZ]\overset{s}{\subseteq}P_j[WU],P_j:W\xrightarrow{s}U$。设 $T_{i1}\subseteq T_i,T_{i1}\leftrightarrow P_i$,则存在 $T_{j1}\subseteq T_j,T_{j1}\leftrightarrow P_j$ 和 $T_{j2}\subseteq T_j,T_{j2}\leftrightarrow P_j$ 满足 $val(N_{j1}(w_a))\doteq val(N_{i1}(x_a)),val(N_{j1}(u_b))\doteq val(N_{i1}(y_b))$和 $val(N_{j2}(w_a))\doteq val(N_{i1}(x_a)),val(N_{j2}(u_b))\doteq val(N_{i1}(z_b))(a\in[1,m],b\in[1,n])$成立。所以 $val(N_{j1}(w_a))\doteq val(N_{j2}(w_a))$,则 $val(N_{j1}(w_a))\div val(N_{j2}(w_a))$,又由 $P_j:W\xrightarrow{s}U$,则 $val(N_{j1}(u_b))\doteq val(N_{j2}(u_b))$成立,所以 $val(N_{i1}(y_b))\doteq val(N_{i1}(z_b))$也成立。证毕。

结束语 本文研究了在不完全信息环境下的 XML 强闭包依赖理论。通过对 XML 强闭包依赖的研究,对 XML 文档中的不确定性数据进行了约束,为研究不完全信息环境下的 XML 数据库的外键奠定了基础。在以后的工作中,将进一步对不完全信息环境下 XML 函数依赖和闭包依赖的范式进行研究。

参 考 文 献

[1] Buneman P, Davidson S, Fan W, et al. Keys for xml. Computer Networks, 2002, 39(5): 473-487

[2] Buneman P, Fan W, Weinstein S. Path constraints on structured and semistructured data//Proc. ACM PODS Conference. 1998: 129-138

[3] Vincent M W, Liu Jixue. Functional Dependencies for XML, AP-Web, 2003: 22-34

[4] Vincent M W , Liu Jixue . Multivalued Dependencies in XML , BNCOD, 2003: 4-18

[5] Fan W, Libkin L. On XML integrity constraints in the presence of DTDs. Journal of the ACM, 2002, 49(3): 368-406

[6] Fan W, Simeon J. Integrity constraints for xml. Journal of Computer and System Sciences, 2003, 66(1): 254-291

[7] Vincent M W, Schrefl M, Liu Jixue, et al. Generalized Inclusion Dependencies in XML// APWeb 2004, LNCS 3007: 224-33

[8] 郝忠孝. 空值环境下数据库导论. 北京:机械工业出版社, 1996

[9] Levene M, Loizu G. Null Inclusion Dependencies in Relational Databases. Inf. Comput, 1997(136): 67-108

[10] Vincent M W , Liu Jixue , Liu Chengfei . Strong functional dependencies and their application to normal forms in XML. ACM Trans. Database Syst, 2004, 29(3): 445-462

(上接第 44 页)

得到归一化向量:

$$\boldsymbol{B}'=(b_1',b_2',\cdots,b_m') \tag{5}$$

结束语 度量是信息安全管理的一个重要环节。信息安全度量是收集安全证据的过程,它包括度量模型、度量指标、度量基线和度量方法等关键因素。本文系统分析了信息安全度量的概念、原理和方法,给出了一个基于基线的信息安全度量模型,并研究了信息安全模糊综合度量方法。信息安全度量是一个复杂的过程,它在今后信息安全保障体系建设中的地位和作用日益重要。我们下一步的工作将侧重于对信息安全度量的共性技术和工具的研究,以及在信息系统的实践。

参 考 文 献

[1] 吕欣. 我国信息安全现状和趋势. 国家信息中心:中国信息安全年鉴, 2007: 54-68

[2] Zhang K . A theory for system security // Computer Security Foundations Workshop, 1997. Proceedings. 1997: 148-155

[3] Maconachy W V, Schou C D, Ragsdale D, et al. A Model for Information Assurance: An Integrated Approach//Proceedings of the 2001 IEEE Workshop on Information Assurance and Security. United States Military Academy, 2001: 306-310

[4] Lü Xin. Information Security Assurance Evaluation for Network Information Systems// CIS2006: Computational Intelligence and Security. LNAI. , Vol 4456. Springer, 2007: 869-877

[5] Fowler K, Schmalzel J. Why do we care about measurement?. Instrumentation & Measurement Magzaine, IEEE, 2004, 7(1): 38-46

[6] 秦寿康, 等. 综合评价原理. 北京:电子工业出版社, 2003

[7] British Standards Institute. Code of practice for information security management, BS 7799, London, 1999

[8] ISO/IEC 13335-5: 2001. Information technology—Guidelines for the management of IT Security —Part5: Management guidance on network security, 2001

[9] ISO/IEC15408. Information Technology-Security Techniques-Evaluation Criteria for IT Security, 2005