Tạp chí Tin học và Điều khiển học, T. 15, S. 1 (1999), 62–65

SOME RESULTS ABOUT THE THIRD NORMAL FORM FOR RELATION SCHEME

VU DUC THI

Abstract. The third normal form (3NF) which was introduced by E. F. Codd is an important normal form for relation schemes in the relational database. It is known [6] that a set of minimal key's of a relation scheme is a Sperner system (some times it is called an antichain) and for an arbitrary Sperner system there exists a relation scheme the set of minimal keys of which is exactly this Sperner system. This paper gives new necessary and sufficient conditions for an arbitrary relation scheme is in 3NF and its set of minimal keys is a given Sperner system.

1. INTRODUCTION

Now we start with some necessary definitions, and in the next sections we formulate our results.

Definition 1.1. Let $R = \{h_1, \ldots, h_n\}$ be a relation over U, and $A, B \subseteq U$. Then we say that B functionally depends on A in R (denoted $A\,f/R\,B$) iff

$$(\forall h_i,\ h_j \in R)\,(\forall a \in A)\,(h_i(a) = h_j(a)) \Rightarrow (\forall b \in B)\,(h_i(b) = h_j(b)).$$

Let $F_R = \{(A, B)\} : A, B \subseteq U,\ A\,f/R\,B\}$. F_R is called the full family of functional dependencies of R. Where we write (A, B) or $A \to B$ for $A\,f/R\,B$ when R, f are clear from the next context.

Definition 1.2. A functional dependency over U is a statement of the form $A \to B$, where $A, B \subseteq U$. The FD $A \to B$ holds in a relation R if $A\,f/R\,B$. We also say that R satisfies the FD $A \to B$.

Definition 1.3. Let U be a finite set, and denote $P(U)$ its power set. Let $Y \subseteq P(U) \times P(U)$. We say that Y is an f-family over U iff for all $A, B, C, D \subseteq U$

(1) $(A, A) \in Y$,

(2) $(A, B) \in Y,\ (B, C) \in Y \Rightarrow (A, C) \in Y$,

(3) $(A, B) \in Y,\ A \subseteq C,\ D \subseteq B \Rightarrow (C, D) \in Y$,

(4) $(A, B) \in Y,\ (C, D) \in Y \Rightarrow (A \cup C,\ B \cup D) \in Y$.

Clearly, F_R is an f-family over U.

It is known [1] that if Y is an arbitrary f-family, then there is a relation R over U such that $F_R = Y$.

Definition 1.4. A relation scheme S is a pair $\langle U, F\rangle$. Where U is a set of attributes, and F is a set of FDs over U. Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 1.3.

Clearly, in [1] if $S = \langle U, F\rangle$ is a relation scheme, then there is a relation R over U such that $F_R = F^+$. Such a relation is called an Armstrong relation of S.

Definition 1.5. Let R be a relation, $S = \langle U, F\rangle$ be a relation scheme, Y be an f-family over U, and $A \subseteq U$. Then A is a key of R (respectively a key of S, a key of Y) if $A\,f/R\,U$ ($A \to U \in F^+$, $(A, U) \in Y$). A is a minimal key of R (respectively S, Y) if A is a key of R (S, Y) and any proper subset of A is not a key of R (S, Y). Denote K_R (respectively K_S, K_Y) the set of all minimal keys of R (respectively S, Y),

Clearly, K_R, K_S, K_Y are Sperner systems over U.

Definition 1.6. Let K be a Sperner system over U. We define the set of antikeys of K, denote by K^{-1}, as follows:

$$K^{-1} = \{A \subset U : (B \in K) \Rightarrow (B \not\subseteq K) \text{ and } (A \subset C) \Rightarrow (\exists B \in K)(B \subseteq C)\}.$$

It is easy to see that K^{-1} is also a Sperner system over U.

It is known [4] that if K is an arbitrary Sperner system playing the role of the set of minimal keys (respectively antikeys), then this Sperner system is not empty (respectively does't contain U). We also regard the comparison of two attributes to be the elementary step of algorithms. Thus, if we assume that subsets of U are represented as sorted list of attributes, then a Boolean operation on two subsets of requires at most $|U|$ elementary steps.

Definition 1.7. Let $I \subseteq P(U)$, $U \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(U)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $U \in M^+$ but not in M, since it is the intersection of the empty collection of sets.

Denote $N = \{A \in I : A \neq \cap\{A' \in I : A \subset A'\}\}$.
In [6] it is proved that N is the unique minimal generator of I. Thus, for any generator N' of I we obtain $N \subseteq N'$.

Definition 1.8. Let R be a relation over U, and E_R the equality set of R, i.e.

$$E_R = \{E_{ij} : 1 \leq i < j \leq |R|\},$$

where $E_{ij} = \{a \in U : h_i(a) = h_j(a)\}$.
Let $T_R = \{A \in P(U) : \exists E_{ij} = A,\ \nexists E_{pq} : A \subset E_{pq}\}$. Then T_R is called the maximal equality system of R.

Definition 1.9. Let R be a relation, and K a Sperner system over U. We say that R represents K iff $K_R = K$.

The following theorem is known in [8].

Theorem 1.10. *Let K be a relation, and K a Sperner system over U. R presents K iff $K^{-1} = T_R$, where T_R is the maximal equality system of R.*

Let $s = \langle U, F\rangle$ be a relation scheme over U. From s we construct $Z(s) = \{X^+ : X \subseteq U\}$, and compute the minimal generator N_s of $Z(s)$.

We put $T_s = \{A : A \in N_s,\ \nexists B \in N_s : A \subset B\}$.

In [8] we presented the following result.

Proposition 1.11. *Let $s = \langle U, F\rangle$ be a relation scheme over U. Then*

$$K_s^{-1} = T_s.$$

Definition 1.12. Let $s = \langle R, F\rangle$ be a relation scheme over R. We say that an attribute a is prime if it belong to a minimal key of s, and nonprime otherwise.
$s = \langle R, F\rangle$ is in the third normal form (3NF) if $A \to \{a\} \notin F^+$ for $A^+ \neq R$, $a \notin A$, a is nonprime.

If a relation scheme is changed to a relation we have the definition of 3NF for relation.

2. RESULTS

In this section we show the following result. It is a new necessary and sufficient condition for an arbitrary relation scheme is in 3NF and its set of minimal keys is a given Sperner system.

First we denote some following concepts.

Let K be a Sperner over U.

Denote

$$T(K^{-1}) = \{A : \exists B \in K^{-1} : A \subseteq B\},$$
$$K_n = \{a \in U : \nexists A \in K : a \in A\}.$$

K_n is called the set of nonprime attributes of K.
We have the following theorem

Theorem 2.1. *Let $s = \langle U, F\rangle$ be a relation scheme and K is a Sperner system over U. ${K_n}^{-1} = \{B - a : a \in K_n, B \in K^{-1}\}$, where K_n is the set of nonprime attributes of K. Then s is in 3NF and $K_s = K$ if and only if*

$$\{U\} \cup K^{-1} \cup K_n^{-1} \subseteq Z(s) \subseteq \{U\} \cup T(K^{-1}). \qquad (*)$$

Proof. Assume that s is in 3NF and $K = K_s$. By Proposition 1.11 and from definitions of $Z(s)$, $T(K^{-1})$ we obtain the right-hand side of $(*)$. If $K_n = \emptyset$ then the left-hand side of $(*)$ is obvious. Assume that $K_n \neq \emptyset$. According to Proposition 1.11, $K_s = K$ and by definition of $Z(s)$ we have $\{R\} \cup K^{-1} \subseteq Z(s)$. According to definition of 3NF, s is in 3NF then for every $B \in K^{-1}$, $a \in K_n : B - a \in Z(s)$. Consequently, we obtain $\{R\} \cup K^{-1} \subseteq Z(s)$.

Conversely, assume that we have $(*)$. By Proposition 1.11 and according to definitions of $Z(s)$, $T(K^{-1})$, K^{-1} we obtain $K_s = K$. If $K_n = \emptyset$ then s is in 3NF. Assume that $K_n \neq \emptyset$. If s isn't in 3NF then there exists a set A and $a \in K_n : a \notin A$ such that $A^+ \subseteq B$. From $a \in A^+$ we obtain $a \in B$. By $a \notin A$ we obtain $A \subseteq B - a$. Consequently, we have $B - a \subset (B-a)^+$. Thus, there exists a $C \in K_n^{-1}$ such that $C \notin Z(s)$. This conflicts with the fact that $K_n^{-1} \subseteq Z(s)$. The proof is complete.

Clearly, the right-hand side and the left-hand side of $(*)$ don't depend on s.

Based on definition of $Z(s)$ and according to Proposition 1.11 and Theorem 2.1 the following corollary is clear.

Corollary 2.2. *Let $s = \langle U, F\rangle$ be a relation scheme. Denote F_n the set of all nonprime attributes of s. Then s is in 3NF iff $\forall B \in K_s^{-1}$, $a \in F_n : (B-a)^+ = B - a$.*

REFERENCES

[1] Armstrong W. W., Dependency structures of database relationships, *Information Processing*, Holland publ. Co. **74** (1974) 580-583.

[2] Beeri C, Bernstein P. A., Computational problems related to the design of normal form relational schemas, *ACM Trans on Database Syst.* **4** (1) (1979) 30-59.

[3] Beeri C., Dowd M., Fagin R., Staman R., On the structure of Armstrong relations for functional dependencies, *J. ACM* **31** (1) (1984) 30-46.

[4] Demetrovics J., Logical and structural investigation of relational datamodel, *MTA - SZTAKI Tanulmanyok, Budapest* **114** (1980) 1-97.

[5] Demetrovics J., Thi V. D., Some results about functional dependencies, *Acta Cybernetica* **8** (3) (1988) 273-278.

[6] Demetrovics J., Thi V. D., Relations and minimal keys, *Acta Cybernetica* **8** (3) (1988) 279-285.

[7] Demetrovics J., Thi V. D., On keys in the relational datamodel, *Inform. Process. Cybern. EIK* **24** (10) (1988) 515-519.

[8] Demetrovics J., Thi V. D., Algorithm for generating Armstrong relations and inferring functional dependencies in the relational datamodel, *Computers and Mathematics with Applications, Great Britain* **26** (4) (1993) 43-45.

[9] Demetrovics J., Thi V. D., Some problems concerning keys for relation schemes and relationals in the relational datamodel, *Information Processing Letters, North Holland* **46** (4) (1993) 179-183.

[10] Demetrovics J., Thi V. D., Some computational problems related to the functional dependency in the relational datamodel, *Acta Scientiarum Mathematicarum* **57** (1-4) (1993) 627-628.

[11] Demetrovics J., Thi V. D., Armstrong relation, functional dependencies and strong dependencies, *Comput. and AI* (submitted for publication).

[12] Garey M. R., Johnson S. S., *Computers and Intractability: A Guide to Theory of NP - Completeness*, Bell Laboratories, W. H. Freeman and Company, San Francisco, 1979.

[13] Gottlob G., Libkin L., Investigations on Armstrong relations dependency inference, and excluded functional dependencies, *Acta Cybernetica Hungary* **IX** (4) (1990) 385-402.

[14] Lucchesi C. L., Osborn S. L., Candidate keys for relations, *J. Comput. Syst. Sci.* **17** (2) (1978) 270-279.

[15] Thi V. D., Investigation on combinatorial characterizations related to functional dependency in the relational datamodel (Ph. D. dissertation), *MTA - SZTAKI Tanulmanyok, Budapest* **191** (1986) 1-157.

[16] Thi V. D., Minimal keys and antikeys, *Acta Cybernetica* **7** (4) (1986) 361-371.

[17] Demetrovics J., Thi V. D., Some results about normal forms for functional dependency in the relational datamodel, *Discrete Applied Mathematics* **69** (1996) 61-74.

Nhận bài ngày 23 - 3 - 1998

Viện Công nghệ thông tin, Trung tâm KHTN và CNQG.