Tạp chí Tin học và Điều khiển học, T. 16, S. 2 (2000), 1–4

SOME OBSERVATIONS ON THE SECOND NORMAL FORM

LUONG CAO SON, VU DUC THI

Abstract. For functional dependency, the second normal form (2NF) which was introduced by E. F. Codd has been widely investigated both theoretically and practically. In this paper, we give a new necessary and sufficient condition for an arbitrary relation scheme is in 2NF and its set of minimal keys is a given Sperner system.

1. INTRODUCTION

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

Definition 1. Let $r = \{h_1, \ldots, h_n\}$ be a relation over R, and $A, B \subseteq R$. Then we say that B functionally depends on A in r (denoted $A \xrightarrow[r]{f} B$) 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 $F_r = \{(A, B) : A, B \subseteq R,\ A \xrightarrow[r]{f} B\}$. F_r is called the full family of functional dependencies of r. Where we write (A, B) or $A \to B$ for $A \xrightarrow[r]{f} B$ when r, f are clear from the next context.

Definition 2. A functional dependency (FD) over R is a statement of the form $A \to B$, where $A, B \subseteq R$. The FD $A \to B$ holds in a relation r if $A \xrightarrow[r]{f} B$. We also say that r satisfies the FD $A \to B$.

Definition 3. Let R be a finite set, and denotes $P(R)$ its power set. Let $Y \subseteq P(R) \times P(R)$. We say that Y is an f-family over R iff for all $A, B, C, D \subseteq R$

(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 R.

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

Definition 4. A relation scheme s is a pair $\langle R, F\rangle$, where R is a set of attributes, and F is a set of FDs over R. Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 3.

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

Definition 5. Let r be a relation, $s = \langle R, F\rangle$ be a relation scheme, Y be an f-family over R, and $A \subseteq R$. Then A is a key of r (a key of s, a key of Y) if $A \xrightarrow[r]{f} B$ ($A \to R \in F^+$, $(A, R) \in Y$). A is a minimal key of $r(s, Y)$ if A is a key of $r(s, Y)$ and any proper subset of A is not a key of $r(s, Y)$. Denote $K_r, (K_s, K_Y)$ the set of all minimal keys of $r(s, Y)$.

Clearly, K_r, K_s, K_Y are Sperner systems over R.

Definition 6. Let K be a Sperner system over R. We define the set of antikeys of K, denote by K^{-1}, as follows:

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

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

It is known [4] that if K is an arbitrary Sperner system plays the role of the set of minimal keys (antikeys), then this Sperner system is not empty (does't contain R).

Definitions 7. Let $I \subseteq P(R)$, $R \in I$, and $A, B \in I \Rightarrow A \cap B \in I$. Let $M \subseteq P(R)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $R \in M^+$ 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 8. Let r be a relation over R, and E_r the equality set of r, i.e. $E_r = \{E_{ij} : 1 \leq i < j \leq |R|\}$, where $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$. Let $T_R = \{A \in P(R) : \exists E_{ij} = A, \text{ no } \exists E_{pq} : A \subset E_{pq}\}$. Then T_R is called the maximal equality system of r.

Definition 9. Let r be a relation, and K a Sperner system over R. We say that r represents K if $K_r = K$.

The following theorem is known in [10]

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

From Theorem 1 we obtain the following corollary.

Corollary 1. *Let $s = \langle R, F\rangle$ be a relation scheme and r a relation over R. We say that r represents s if $K_r = K_s$. Then r represents s iff $K_s^{-1} = T_r$, where T_r is the maximal equality system of r.*

In [9] we proved the following theorem.

Theorem 2. *Let $r = \{h_1, ..., h_m\}$ be a relation, and F and f-family over R. Then $F_r = F$ iff for every $A \in P(R)$.*

$$H_F(A) = \begin{cases} \bigcap\limits_{A \subseteq E_g} E_g & \text{if } \exists E_g : A \subseteq E_g, \\ R & \text{otherwise}, \end{cases}$$

where $H_F(A) = \{A \in R : (A, \{a\}) \in F\}$ and E_r is the equality set of r.

Definition 10. Let $s = \langle R, F\rangle$ be a relation scheme over R. We say that an attributet a is prime if belong to a minimal key of s, and nonprime otherwise. Then $s = \langle R, F\rangle$ is in 2NF if $K' \to \{a\} \notin F^+$ for each $K \in K_s$, $K' \subset K$, and a is nonprime.

If a relation scheme is changed to a relation we have the definition of 2NF for relation.

Definition 11. [5] Let P be a set of all f-families over R. An ordering over P is defined as follows: For $F, F' \in P$ let $F \leq F'$ iff for all $A \subseteq R$, $H_{F'}(A) \subseteq H_F(A)$, where $H_F(A) = \{a \in R : (A, \{a\}) \in F\}$.

Theorem 3. [9] *Let K be a Sperner system over R. Let:*

$$L(A) = \begin{cases} \bigcap\limits_{A \subseteq B} B & \text{if } \exists B : A \subseteq B, \\ R & \text{otherwise}, \end{cases}$$

and $F = \{(C, D) : D \subseteq L(C)\}$.

Then F *is an f-family over* R, $H_F = L$, *and* $K_F = K$. *If* F' *is an arbitrary f-family over* R *such* $K_{F'} = K$ *then* $F \leq F'$ *holds.*

2. RESULTS

Now we present a new necessary and sufficient condition for an arbitrary relation scheme is in 2NF and its set of minimal keys is a given Sperner system.

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

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

It is know [1] that for a given relation scheme s there is a relation r such that r is an Armstrong relation of s. On the other hand, by Corollary 1 and Theorem 2 the following proposition is clear.

Proposition 1. *Let* $s = \langle R, F \rangle$ *be a relation scheme over* R. *Then:*

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

Let K be a Sperner system over R. Denote $T(K^{-1}) = \{A : \exists B \in K^{-1},\ A \subseteq B\}$ and $K_n = \{a \in R : \text{no } \exists A \in K,\ a \in A\}$. K_n is called the set of nonprime attributes of K.

Theorem 4. *Let* $s = \langle R, S \rangle$ *be a relation scheme and* K *a Sperner system over* R. *Denote* $M(K) = \{A - a : a \in A,\ A \in K\}$. *Set* $I^* = \{B : B = C^+,\ C \in M(K)\}$ *and* $I_s = \{B - a :, a \in K_n,\ B \in I^*\}$, *where* K_n *is the set of nonprime attributes of* K.

Then s *is in 2NF and* $K_s = K$ *if and only if*

$$\{R\} \cup K^{-1} \cup I_s \subseteq \{R\} \cup T(K^{-1}). \qquad (*)$$

Proof. Assume that s is in 2NF and $K = K_s$. From definitions of $Z(s)$, $T(K^{-1})$ and by Proposion 1 we obtain $Z(s) \subseteq \{R\} \cup T(K^{-1})$. It is easy to see that if $K_n = \emptyset$ then $\{R\} \cup K^{-1} \cup I_s \subseteq Z(s)$. Assume that $K_n \neq \emptyset$. According to Proposition 1, $K_s = K$ and by definition of $Z(s)$ we have $\{R\} \cup K^{-1} \subseteq Z(s)$. If there is a $B \in I^*$ and a $K_n : B - a \notin Z(s)$. Thus, $B - a \subset (B - a)^+$ holds. From definition of I^* there is a $C \in M(K) : C^+ = B$. Clearly, $a \in (B - a)^+$ holds. According to definition of closure we obtain $(B - a)^+ = C^+ = B$. Hence, $a \in C^+$ holds. Thus, $C \to \{a\} \in F^+$, $a \notin C$ hold. This conflicts with the fact that s is in 2NF. Consequently, we obtain $\{R\} \cup K^{-1} \cup I_s \subseteq Z(s)$.

Now, assume that we have $(*)$. By definitions of $Z(s)$, K^{-1}, $T(K^{-1})$, and by Proposition 1 we obtain $K_s = K$. If $K_n = \emptyset$ then s is in 2FN. Assume that $K_n \neq \emptyset$. Suppose that there is a $D \subset A$. $A \in K_s(1)$ and $a \in K_n$, $a \notin D$, but $D \to \{a\}F^+$. By (1) and according to constructions of $M(K)$ and I^* there is a $C \in M(K) : D \subset C$. I_s is obvious that $a \notin C$.

Clearly, $D^+ \subseteq C^+$ and $a \subseteq C^+$. Set $B = C^+$. It can be seen that $C \subseteq B - a$. Consequently, $B - a \subset C^+ = (B - a)^1$. This conflicts with the fact that $I_s \subseteq Z(s)$. Thus, s is in 2NF. The proof is complete. □

REFERENCES

[1] Armstrong W. W., *Dependency Structures of Database Relationships*, Information Processing 74, Holland Publ. Co., 1974, p. 580-583.

[2] Beeri C., Bernstein P. A., Computational problems related to the design of normal form relational schemes, *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] Bernstein P. A., Synthesizing third normal form relations from functional dependencies, *ACM Trans. on Database Synt.* **1** (1976) 277-298.

[5] Burosch G., Demetrovics J., Katona G. O., The poset of closures as a model of changing databases, *Order* **4** (1987) 127-142.

[6] Demetrovics J., On the equivalence of candidate keys with Sperner system, *Acta Cybernetica* **4** (1979) 247-252.

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

[8] Demetrovics J., Hencsey G., Libkin L. O., Muchik I. B., Normal form relation scheme: a new characterization, Manuscript.

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

[10] Demetrovics J., Thi V. D., Relations and minimal keys, *Acta Cybernetica Hungary* **8** (3) (1998) 279-285.

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

[12] Demetrovics J., Thi V. D. Normal form and minimal keys in the relational datamodel, *Acta Cybernetica* **11** (3) (1994) 205-215.

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

[14] Gottlob G., Libkin L., Investigations on Armstrong relations, dependency inference, and exclused functional dependencies, *Acta Cybernical Hungary* **7** (4) (1990) 385-402.

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

[16] Maier D., Minimum cover in the relational database model, *JACM* **27** (4) (1980) 664-674.

[17] Mannila H., Raika K. J., Design by example: an application of Armstrong relations, *J. Comput. Syst. Scien.* **33** (1986) 125-141.

[18] Mannila H., Raiha K. J., Practical algorithms for finding prime attributes and testing normal forms, *Processings of the eighth ACM SIGACT-SIGART Syposium on Principles of Database Systems Philadelphia*, PA (ACM, New York, 1989), p. 128-133.

[19] Osborn S. L., Testing for existence of a covering Boyce–Codd normal form, *Infor. Proc. Lett.* **8** (1) (1979) 11-14.

[20] Thi V. D., *Investigation on Combinatorial Characterzation Related to Functional Dependency in the Relational Datamodel* (in Hungarian), MTA-SZTAKI Tanulmanyok, 191 (1986) 1-157, Ph.S. Dissertation.

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

[22] Thi V. D., On antikeys in the relational datamodel (in Hungarian), *Alkalmazott Matematikai Lapok* **12** (1986) 111-124.

[23] Tsou D. M., Fischer P. C., Decomposition of a relation scheme into Boyce-Codd normal form, *SIGACT NEWS* **14** (1982) 23-29.

Received December 14, 1998

Vu Duc Thi - Institute of Information Technology.
Luong Cao Son - Informatic Centre of Office of Government.