Tạp chí Tin học và Điều khiển học, T. 15, S. 4 (1999), 26–30

SOME RESULTS RELATED TO THE MINIMAL ARMSTRONG RELATION 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. The Armstrong relation is an essential concept in investigating the relational datamodel.

In this paper we present some new estimations for the size of minimal Armstrong relations for 3NF relation schemes.

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

Definition 1.2. A functional dependency (FD) over U is a statement of the form $A \rightarrow B$, where $A, B \subseteq U$. The FD $A \rightarrow B$ holds in a relation R if $A \xrightarrow[R]{f} B$. We also say that R satisfies the FD $A \rightarrow B$.

Definition 1.3. Let U be a finite set, and denotes $P(U)$ 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 over U, $S = \langle U, F\rangle$ be a relation scheme, Y be an f-family over U, and $A \subseteq U$. Then A is a key of R (a key of S, a key of Y) if $A\, f/R\, U$ ($A \rightarrow U \in F^+$, $(A, U) \in Y$). 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 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:

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

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

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 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.

Definitions 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 $T_R = \{A \in P(U) : \exists E_{ij} = A$, no $\exists E_{pq} : A \subset E_{pq}\}$. 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 if $K_R = K$.

The following theorem is known in [8].

Theorem 1.10. *Let K be a relation, and K a Sperner system over U. We say that 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 $T_s = \{A : A \in N_s, \text{no} \exists B \in N_s : A \subset B\}$.

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.

We say that $s = \langle R, F\rangle$ is in the third normal form (3NF) if $A \rightarrow \{a\} \notin F^+$ 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.

Definition 1.13. [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 $H_F(A) = \{a \in R : (A, \{a\}) \in F\}$.

2. RESULTS

The size of minimal Armstrong relations was investigated in some papers (see [2,7,11,15]). Now we present some new bounds for the size of minimal Armstrong relations for relation schemes in 3NF.

Definition 2.1 (Minimal Armstrong relation). Let F be an f-family over U. Let

$$M(F) = \min\{m : |R| = m,\ F_R = F\}.$$

Denote $H_F(A) = \{a \in U : (A, \{a\}) \in F\}$, and $Z(F) = \{A : H_F(A) = A\}$.

Note that from this definition for a relation scheme $s = \langle U, F\rangle$ we have $Z(s) = Z(F^+)$.

Proposition 2.2 [4]. *Let F be an f-family over U. Then*

$$(|N(F)|)^{1/2} < M(F) \leq |N(F)| + 1,$$

where $N(F)$ is the minimal generator of $Z(F)$.

Now we give some following concepts.

Let K be a Sperner system over U. Denote

$$T(K^{-1}) = \{A : \exists B \in K^{-1},\ A \subseteq B\},$$

$$K_n = \{a \in U : \text{no} \exists A \in K,\ a \in A\}.$$

K_n is called the set of nonprime attributes of K.

Then we have the following result [18]

Theorem 2.3. *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

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

Lemma 2.4. *Let K be a Sperner system over U. Denote*

$$K_n = \{a \in U : no \exists A \in K : a \in A\}$$

and

$$K_n^{-1} = \{B - a : a \in K_n,\ B \in K^{-1}\}.$$

Then (1) $C, D \in K_n^{-1} : C \neq D$ and K_n^{-1} is Sperner system over U,

(2) each element of K_n^{-1} isn't the intersection of elements of K^{-1}.

Proof. It is obvious that if $K_n = \emptyset$ then we have (1) and (2). Assume that $K_n \neq \emptyset$. It is known [11] that K_n is the intersection of all elements of K^{-1}.

Suppose that there are $C, D \in K_n^{-1} : C \not\subseteq D$. Consequently, there are $a, b \in K_n$, $A, B \in K^{-1}$: $C = A - a$, $D = B$. Hence, $A - a \subseteq B - b$ holds. By (*) we have $A \subseteq (B - b) \cup a = B - b$. From this $A \subset B$ holds. This conflicts with the fact that K^{-1} is a Sperner system. Thus, we have (1).

Because K_n is the intersection of all elements of K^{-1} we have (2). The proof is complete.

Given a relation scheme $s = \langle U, F\rangle$, we say that a functional dependency $A \to B \in F$ is redundant either $A \subseteq B$ or there is $C \to D \in F$ such that $C \subsetneq A$.

Remark 2.5. Let K be a Sperner system over R. Denote K^{-1} the set of all antikeys of K. From K, K^{-1}, K_n^{-1} we construct the following relation scheme $s_1 = \langle U, F_1\rangle$.

For each $A \subset R$, if there is $B \in K^{-1} \cup K_n^{-1}$ such that $A \subseteq B$ then we set $C = \cap\{B \in K^{-1} \cup K_n^{-1} : A \subseteq B\}$. Set $A \to C$. Denote T the set of all such functional dependencies. Set $F_1 = \{E \to U : E \in K\} \cup (T - Q)$ where $Q = \{X \to Y \in T : X \to Y$ is a redundant functional dependency$\}$. From Theorem 1.10, Theorem 2.3, Lemma 2.4 and definition of Sperner system we obtain $K_{s_1} = K$ and s_1 is in 3NF. Clearly, by Lemma 2.4 $K^{-1} \cup K_n^{-1}$ is the minimal generator of $Z(s_1)$.

From K we construct $s_2 = \langle R, F_2 \rangle$ as follows: $F_2 = \{E \to R : E \in K\}$. It can be seen that $K = K_{s_2}$, s_2 is in the 3NF and $Z(s_2) = \{R\} \cup T(K^{-1})$.

From Definition 1.14, Theorem 2.3, Lemma 2.4 and Remark 2.5 we have the following

Corollary 2.6. *Let K be a Sperner system over R. Denote by V the set of all 3NF relation schemes over R, the minimal keys of which are exactly the elements of K.*
Then s_1 and s_2 which constructed in Remark 2.6 are the unique minimal and maximal elements of the partially ordered set V for the ordering defined Definition 1.14.

Based on these results we obtain the following main theorem.

Theorem 2.7. *Let $s = \langle U, F \rangle$ be a 3NF relation scheme, K a Sperner system over R. Denote*

$$K_n^{-1} = \{B - a : a \in K_n,\ B \in K^{-1}\},$$

where K_n is the set of nonprime attributes of K.

Set $K_n^* = K^{-1} \cup K_n^{-1}$,

$I_n^* = \{C : C \in T(K^{-1}) - K_n^*,\ C \neq \cap\{D : C \subset D,\ D \in K_n^*\}\}$.

Then if $K_s = K$ then

$$|N(F^+)| \leq |K_n{}^*| + |I_n{}^*|.$$

Proof. According to definition of minimal generator for an f-family over R and by Theorem 2.3, Remark 2.5 and Corollary 2.6 we obtain $|K_n{}^*|$. By Lemma 2.4 $K_n{}^*$ is the minimal generator of $Z(s_1)$ in Remark 2.5. According to Theorem 2.3 and construction of $I_n{}^*$ we have $|N(F^+)| \leq |K_n{}^*| + |I_n{}^*|$. The proof is complete.

From Theorem 2.7 and Proposition 2.2 we have following.

Theorem 2.8. *Let K be a Sperner system, $s = \langle R, F \rangle$ a 3NF relation scheme over R. Let*

$$M(F^*) = \min\{m : |r| = m,\ F_r = F^+,\ s = \langle R, F \rangle \text{ a relation scheme}\}.$$

Denote

$$K_n^{-1} = \{B - a : a \in K_n,\ B \in K^{-1}\},$$

where K_n is the set of nonprime attributes of K.

Set $K_n{}^* = K^{-1} \cup K_n^{-1}$,

$I_n{}^* = \{C : C \in T(K^{-1}) - K_n{}^*,\ C \neq \cap\{D : C \subset D,\ D \in K_n{}^*\}\}$.

Then if $K_s = K$ then

$$2|K_n{}^*|^{1/2} < M(F^*) \leq |K_n{}^*| + |I_n{}^*| + 1.$$

REFERENCES

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

[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] 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) (1998) 279-285.

[7] Demetrovics J., Thi V. D., On keys in the relation 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 relation 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 D. 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 relation 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. Scien.* **17** (2) (1978) 270-279.

[15] Thi V. D., Investigation on Combinatorial Characterization Related to Functional Dependency in 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.

[18] V. D. Thi, Some results about the third normal form for relation scheme, *J. Computer Science and Cybernetic*, Hanoi Vietnam (Submitted 1999).

Received May 25, 1998

Institute of Information Technology