

## SOME RESULTS ABOUT RELATIONS IN THE RELATIONAL DATAMODEL

VU DUC THI

**Abstract.** We introduce the concepts of minimal family of a relation. First, we show the algorithm finding a minimal family of a given relation. After that, we prove that the time complexity of finding a minimal family of a given relation is exponential in the number of attributes.

**Tóm tắt.** Trong bài này chúng tôi trình bày họ tối thiểu của một quan hệ. Chúng tôi đưa ra thuật toán tìm họ tối thiểu của một quan hệ cho trước, sau đó chứng minh độ phức tạp của việc tìm họ tối thiểu của một quan hệ cho trước tuân theo luật số mũ đối với số lượng các thuộc tính.

### 1. INTRODUCTION

The relational datamodel which was introduced by E.F. Codd is one of the most powerful database models. The basic concept of this model is a relation. It is a table every row of which corresponds to a record and every column to an attribute. Because the structure of this model is clear, simple and mathematical instruments can be applied in it, it becomes the theoretical basis of database models. Semantic constrains between sets of attributes play very important roles in logical and structural investigations of the relational datamodel both in practice and design theory. Important among these constraints is functional dependency.

This paper gives some results about computational problems related to relations.

Let us give some necessary definitions and results that are used in next section. The concepts given in this section can be found in [1, 2, 4, 6, 7, 8, 16].

Let  $R = \{a_1, \dots, a_n\}$  be a nonempty finite set of attributes. A functional dependency (FD) is a statement of the form  $A \rightarrow B$ , where  $A, B \in R$ . The FD  $A \rightarrow B$  holds in a relation  $r = \{h_1, \dots, h_m\}$  over  $R$  if  $\forall h_i, h_i \in r$  we have  $h_i(a) = h_j(a)$  for all  $a \in A$  implies  $h_i(b) = h_j(b)$  for all  $b \in B$ . We also say that  $r$  satisfies the FD  $A \rightarrow B$ .

Let  $F_r$  be a family of all FDs that hold in  $r$ . Then  $F = F_r$  satisfies

- (1)  $A \rightarrow A \in F$ ,
- (2)  $(A \rightarrow B \in F, B \rightarrow C \in F) \implies (A \rightarrow C \in F)$ ,
- (3)  $(A \rightarrow B \in F, A \subseteq C, D \subseteq B) \implies (C \rightarrow D \in F)$ ,
- (4)  $(A \rightarrow B \in F, C \rightarrow D \in F) \implies (A \cup C \rightarrow B \cup D \in F)$ .

A family of FDs satisfying (1) - (4) is called an  $f$ -family (sometimes it is called the full family) over  $R$ .

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

Given a family  $F$  of FDs, there exists a unique minimal  $f$ -family  $F^+$  that contains  $F$ . It can be seen that  $F^+$  contains all FDs which can be derived from  $F$  by the rules (1) - (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  $R$ . Denote  $A^+ = \{a : A \rightarrow \{a\} \in F^+\}$ .  $A^+$  is called the closure of  $A$  over  $s$ . It is clear that  $A \rightarrow B \in F^+$  iff  $B \subseteq A^+$ .

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

Let  $R$  be a nonempty finite set of attributes and  $P(R)$  its power set. The mapping  $H : P(R) \rightarrow P(R)$  is called a closure operation over  $R$  if for  $A, B \in P(R)$ , the following conditions are satisfied:

- (1)  $A \subseteq H(A)$ ,
- (2)  $A \subseteq B$  implies  $H(A) \subseteq H(B)$ ,
- (3)  $H(H(A)) = H(A)$ .

Let  $s = \langle R, F \rangle$  be a relation scheme. Set  $H_s(A) = \{a : A \rightarrow \{a\} \in F^+\}$ , we can see that  $H_s$  is a closure operation over  $R$ .

Let  $r$  be a relation,  $s = \langle R, F \rangle$  be a relation scheme. Then  $A$  is a key of  $r$  (a key of  $s$ ) if  $A \rightarrow R \in F_r$  ( $A \rightarrow R \in F^+$ ).  $A$  is a minimal key of  $r(s)$  if  $A$  is a key of  $r(s)$  and any proper subset of  $A$  is not a key of  $r(s)$ .

Denote  $K_r$  ( $K_s$ ) the set of all minimal keys of  $r$  ( $s$ ).

Clearly,  $K_r, K_s$  are Sperner systems over  $R$ , i.e.  $A, B \in K_r$  implies  $A \not\subseteq B$ .

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

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

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

It is known [5] that if  $K$  is an arbitrary Sperner system over  $R$ , then there is a relation scheme  $s$  such that  $K_s = K$ .

In this paper we always assume that if a Sperner system plays the role of the set of minimal keys (antikeys), then this Sperner system is not empty (doesn't contain  $R$ ). We consider the comparison of two attributes as an elementary step of algorithms. Thus, if we assume that subsets of  $R$  are represented as sorted lists of attributes, then a Boolean operation on two subsets of  $R$  requires at most  $|R|$  elementary steps.

Let  $L \subseteq P(R)$ .  $L$  is called a meet-irreducible family over  $R$  (sometimes it is called a family of members which are not intersections of two other members) if  $\forall A, B, C \in L$ , then  $A = B \cap C$  implies  $A = B$  or  $A = C$ .

Let  $I \subseteq P(R)$ ,  $R \in I$ , and  $A, B \in I \implies A \cap B \in I$ .  $I$  is called a meet-semilattice over  $R$ . Let  $M \subseteq P(R)$ . Denote  $M^+ = \{\cap M' : M' \subseteq M\}$ . We say that  $M$  is a generator of  $I$  if  $M^+ = I$ . Note that  $R \in M^+$  but not in  $M$ , by convention it is the intersection of the empty collection of sets.

Denote  $N = \{A \in I : A \neq \cap \{A' \in I : A \subseteq A'\}\}$ .

In [5] it is proved that  $N$  is the unique minimal generator of  $I$ .

It can be seen that  $N$  is a family of members which are not intersections of two other members.

Let  $H$  be a closure operation over  $R$ . Denote  $Z(H) = \{A : H(A) = A\}$  and  $N(H) = \{A \in Z(H) : A \neq \cap \{A' \in Z(H) : A \subseteq A'\}\}$ .  $Z(H)$  is called the family of closed set  $s$  of  $H$ . We say that  $N(H)$  is the minimal generator of  $H$ .

It is shown [5] that if  $L$  is a meet-irreducible family then  $L$  is the minimal generator of some closure operation over  $R$ . It is known [1] that there is an one-to-one correspondence between these families and  $f$ -families.

Let  $r$  be a relation. Denote  $A_r^+ = \{a : A \rightarrow \{a\} \in F_r\}$ . Then  $r$  is a Boyce-Codd normal form (BCNF) relation if  $\forall A \subseteq R : A_r^+ = A$  or  $A_r^+ = R$ .

Let  $r$  be a relation over  $R$ . Denote  $E_r = \{E_{ij} : 1 \leq i < j \leq |r|\}$ , where  $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$ . Then  $E_r$  is called the equality set of  $r$ .

Let  $T_r = \{A \in P(R) : \exists E_{ij} = A, \nexists E_{pq} : A \subseteq E_{pq}\}$ . We say that  $T_r$  is the maximal equality system of  $r$ .

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 [7].

**Theorem 1.1.** *Let  $K$  be a non-empty Sperner system and  $r$  a relation over  $R$ . Then  $r$  represents  $K$  iff  $K^{-1} = T_r$ , where  $T_r$  is the maximal equality system of  $r$ .*

Let  $s = \langle R, F \rangle$  be a relation scheme over  $R$ ,  $K_s$  is a set of all minimal keys of  $s$ . Denote by  $K_s^{-1}$  the set of all antikeys of  $s$ . From Theorem 1.1 we obtain the following corollary.

**Corollary 1.2.** *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 [6] we proved the following theorem.

**Theorem 1.3.** Let  $r = \{h_1, \dots, h_m\}$  be a relation, and  $F$  an  $f$ -family over  $R$ . Then  $F_r = F$  iff for every  $A \subseteq R$

$$H_F(A) = \begin{cases} \bigcap_{A \subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij} \in E_r : A \subseteq E_{ij}, \\ R & \text{otherwise,} \end{cases}$$

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

**Theorem 1.4.** [3] Let  $K = \{K_1, \dots, K_m\}$  be a Sperner system over  $R$ . Set  $s = \langle R, F \rangle$  with  $F = \{K_1 \rightarrow R, \dots, K_m \rightarrow R\}$ . Then  $K_s = K$ .

## 2. RESULTS

In this section we introduce the concepts of minimal family of a relation. We show that the time complexity of finding a minimal family of a given relation is exponential in the number of attributes.

Now we introduce the following concept.

**Definition 2.1.** Let  $r$  be a relation over  $R$  and  $F_r$  a family of all FDs that hold in  $r$ . Put  $A_r^+ = \{a : A \rightarrow \{a\} \in F_r\}$ . Set  $Z_r = \{A : A = A_r^+\}$ . Then  $M(Z_r)$  is called the minimal family of  $r$ .

We construct a following exponential time algorithm finding a minimal family of a given relation.

### Algorithm 2.2.

Input: a relation  $r = \{h_1, \dots, h_m\}$  over  $R$ .

Output: a minimal family of  $r$ .

Step 1: Find the equality set  $E_r = \{E_{ij} : 1 \leq i < j \leq m\}$ .

Step 2: Find the minimal generator  $N$ , where  $N = \{A \in E_r : A \neq \bigcap \{B \in E_r : A \subset B\}\}$ .

Denote elements of  $N$  by  $A_1, \dots, A_t$ .

Step 3: For every  $B \subseteq R$  if there is an  $A_i$  ( $1 \leq i \leq t$ ) such that  $B \subseteq A_i$ , then compute  $C = \bigcap_{B \subseteq A_i} A_i$  and set  $B \rightarrow C$ . In the converse case set  $B \rightarrow R$ .

Denote by  $T$  the set of all such functional dependencies.

Step 4: Set  $F = T - Q$ , where  $Q = \{X \rightarrow Y \in T : X \rightarrow Y \text{ is a redundant functional dependency}\}$ .

Step 5: Put  $M(Z_r) = \{(B, C) : B \rightarrow C \in F\}$ .

According to Theorem 1.3 and definition of  $M(Z_r)$ , Algorithm 2.2 finds a minimal family of  $r$ .

It can be seen that the time complexity of Algorithm 2.2 is exponential in the number of attributes.

**Proposition 2.3.** Given a BCNF relation  $r$  over  $R$ . The time complexity of finding a minimal family of  $r$  is exponential in the number of elements of  $R$ .

*Proof.* From a given BCNF relation  $r$  we use Algorithm 2.2 to construct the minimal family of  $r$ . By definition of BCNF, we obtain

$$M(Z_r) = \{(B, C) : B \rightarrow C \in F_r\} = \{(B, R) : B \in K_r\}.$$

Let us take a partition  $R = X_1 \cup \dots \cup X_m \cup W$ , where  $|R| = n$ ,  $m = \lceil n/3 \rceil$ , and  $|X_i| = 3$  ( $1 \leq i \leq m$ ).

Set  $M = (K^{-1})^{-1}$ , i.e.  $K^{-1}$  is a set of minimal keys of  $M$ , we have:

$$M = \{C : |C| = n - 3, C \cap X_i = \emptyset \text{ for some } i\} \text{ if } |W| = 0,$$

$M = \{C : |C| = n - 3, C \cap X_i = \emptyset \text{ for some } i (1 \leq i \leq m - 1) \text{ or } |C| = n - 4, C \cap (X_m \cup W) = \emptyset\}$  if  $|W| = 1$ .

$$M = \{C : |C| = n - 3, C \cap X_i = \emptyset \text{ for some } i (1 \leq i \leq m) \text{ or } |C| = n - 2, C \cap W = \emptyset\} \text{ if } |W| = 2.$$

It is clear that  $3^{\lceil n/4 \rceil} < |K^{-1}|$ ,  $|M| \leq m + 1$ .

Denote elements of  $M$  by  $C_1, \dots, C_t$ . Construct a relation  $r = \{h_0, h_1, \dots, h_t\}$  as follows:

For all  $a \in R$   $h_0(a) = 0$ , for  $i = 1, \dots, t$

$$h_i(a) = \begin{cases} 0 & \text{if } a \in C_i, \\ i & \text{otherwise.} \end{cases}$$

Clearly,  $|r| < |R|$  holds. According to Theorem 1.1  $M$  is the set of antikeys of  $r$  and  $K^{-1}$  is the set of minimal keys of  $r$ . From definition of BCNF, we can see that  $M(Z_r) = \{(B, R) : B \in K^{-1}\}$ .

Thus, we can construct a relation  $r$  in which the number of rows of  $r$  is less than  $|R|$ , but the number of elements of  $M(Z_r)$  is exponential in the number of attributes. The proposition is proved.

Because class of BCNF relations is a special subfamily of the family of relations over  $R$ , the next corollary is obvious.

**Corollary 2.4.** *The time complexity of finding a minimal family of a given relation  $r$  is exponential in the number of attributes.*

## 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] Bekessy A., Demetrovics J., Contribution to the theory of database relations, *Discrete Math.* **27** (1979) 1–10.
- [5] Demetrovics J., Logical and structural investigation of relational datamodel, *MTA - SZTAKI Tanulmányok, Budapest* **114** (1980) 1–97 (Hungarian).
- [6] Demetrovics J., Thi V. D., Some results about functional dependencies, *Acta Cybernetica* **8** (3) (1988) 273–278.
- [7] Demetrovics J., Thi V. D., Relations and minimal keys, *Acta Cybernetica* **8** (3) (1988) 279–285.
- [8] Demetrovics J., Thi V. D., On keys in the relational datamodel, *Inform. Process. Cybern. EIK* **24** (10) (1988) 515–519.
- [9] Demetrovics J., Thi V. D., On algorithm for generating Armstrong relations and inferring functional dependencies in the relational datamodel, *Computers and Mathematics with Applications* **26** (4) (1993) 43–55.
- [10] Demetrovics J., Thi V. D., Armstrong relation, functional dependencies and strong dependencies, *Comput. and AI* **14** (3) (1995) 279–298.
- [11] Mannila H., Raiha K. J., Design by example: an application of Armstrong relations, *J. Comput. Syst. Scien.* **33** (1986) 126–141.
- [12] Osborn S. L., Testing for existence of a covering Boyce–Codd normal form, *Infor. Proc. Lett.* **8** (1) (1979) 11–14.
- [13] Thi V. D. Investigations on combinatorial characterizations related to functional dependencies in the relational datamodel, *MTA-SZTAKI Tanulmányok, Budapest* **191** (1986) 1–157, Ph.D. Dissertation (Hungarian).
- [14] Thi V. D. Minimal keys and antikeys, *Acta Cybernetica* **7** (4) (1986) 361–371.
- [15] Thi V. D. On the antikeys in the relational datamodel, *Alkalmazott Matematikai Lapok* **12** (1986) 111–124 (Hungarian).
- [16] Thi V. D., Logical dependencies and irredundant relations, *Computers and Artificial Intelligence* **7** (2) (1988) 165–184.
- [17] Yu C. T., Johnson D. T., On the complexity of finding the set of candidate keys for a given set of functional dependencies, *IPL* **5** (4) (1976) 100–101.

*Received May 2, 2000*

*Revised January 4, 2001*