

SOME OBSERVATIONS ON THE RELATION SCHEMES IN THE RELATIONAL DATAMODEL

VU DUC THI

Abstract. In this paper, we introduce the new concept of maximal family of a relation scheme. The time complexity of finding this family is presented in this paper.

Tóm tắt. Trong bài này, chúng tôi trình bày họ cực đại của một sơ đồ quan hệ.

1. DEFINITIONS AND PRELIMINARY RESULTS

The relational datamodel which was introduced by E. F. Codd is one of the most powerful database models. This paper gives some results about computational problems related to relation schemes. 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].

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 $A, B \subseteq R$. The FD $A \rightarrow B$ holds in a relation $r = \{h_1, \dots, h_m\}$ over R if $\forall h_i, h_j \in r$ 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:

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

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

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 $N(H) = \{A \in Z(H) : A \neq \cap\{A' \in Z(H) : A \subset A'\}\}$. $Z(H)$ 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 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 $T_r = \{A \in P(R) : \exists E_{ij} = A, \nexists E_{pq} : A \subset E_{pq}\}$. 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, 10].

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 $A \subseteq R$*

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

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. MAXIMAL FAMILY OF A RELATION SCHEME

In this section we introduce the new concept of maximal family of a relation scheme. We show that the time complexity of finding a maximal family of a given relation scheme is exponential in the number of attributes.

Now we prove that the time complexity of finding a set of antikeys for relation scheme is exponential in the number of attributes. We show that finding a maximal family of a relation scheme can be polynomially transformed to this problem.

Definition 2.1. Let $s = \langle R, F \rangle$ be a relation scheme. Set $H_s(A) = A^+$ for all $A \subseteq R$. Put $Z(s) = \{A : A = A^+\}$. Denote by N_s the minimal generator of $Z(s)$. Set $M(s) = \{(A, \{a\}) : a \notin A, A \in Z(s) \text{ and } B \in Z(s), a \notin B, A \subseteq B \text{ imply } A = B\}$. Then we say that $M(s)$ is a maximal family of s .

Put $T_a = \{A, \{a\}\} \in M(s)$ and $L(T_a) = \{A : (A, \{a\}) \in T_a\}$.

Let $s = \langle R, F \rangle$ be a relation scheme over R . From s we construct $Z(s)$ 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 known [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 1.2 and Theorem 1.3 the following proposition is clear.

Proposition 2.2. Let $s = \langle R, F \rangle$ be a relation scheme over R . Then

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

It is shown [7] that the problem of finding all antikeys of a relation is solved by polynomial time algorithm. For a relation scheme we have the following theorem.

Theorem 2.3. The time complexity of finding a set of all antikeys of a given relation scheme is exponential in the number of attributes.

Proof. We have to prove that:

- (1) There is an algorithm which finds a set of all antikeys of a given relation scheme in exponential time in the number of attributes.
- (2) There exists a relation scheme $s = \langle R, F \rangle$ such that the number of elements of K_s^{-1} is exponential in the number of attributes (in our example $|K_s^{-1}|$ is exponential not only in the number of attributes, but also in the number of elements of F).

For (1), we construct a following algorithm:

Let $s = \langle R, F \rangle$ be a relation scheme over R .

Step 1: For every $A \subseteq R$ compute A^+ , and set $Z(s) = \{A^+ : A \subseteq R\}$.

Step 2: Construct the minimal generator N_s of $Z(s)$.

Step 3: Compute the set T_s from N_s .

According to Proposition 2.2 we have $T_s = K_s$.

Clearly, the time complexity of this algorithm is exponential in $|R|$.

As to (2): Let us take a partition $R = X_1 \cup \dots \cup X_m \cup W$, where $m = \lfloor n/3 \rfloor$, $|R| = n$ and $|X_i| = 3$ ($1 \leq i \leq m$).

Set

$$K = \{B : |B| = 2, B \subseteq X_i \text{ for some } i\} \text{ if } |W| = 0,$$

$$K = \{B : |B| = 2, B \subseteq X_i \text{ for some } i : 1 \leq i \leq m-1 \text{ or } B \subseteq X_m \cup W\} \text{ if } |W| = 1,$$

$$K = \{B : |B| = 2, B \subseteq X_i \text{ for some } i : 1 \leq i \leq m \text{ or } B = W\} \text{ if } |W| = 2.$$

It is easy to see that

$$K^{-1} = \{A : |A \cap X_i| = 1, \forall i\} \text{ if } |W| = 0,$$

$$K^{-1} = \{A : |A \cap X_i| = 1 (1 \leq i \leq m-1) \text{ and } |A \cap (X_m \cup W)| = 1\} \text{ if } |W| = 1,$$

$$K^{-1} = \{A : |A \cap X_i| = 1 (1 \leq i \leq m) \text{ and } |A \cap W| = 1\} \text{ if } |W| = 2.$$

Let $f : N \rightarrow N$ (N is the set of natural numbers) be the function defined as follows:

$$f(n) = \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \pmod{3}, \\ \frac{4}{3} \cdot 3^{\lfloor n/3 \rfloor} & \text{if } n \equiv 1 \pmod{3}, \\ 2 \cdot 3^{\lfloor n/3 \rfloor} & \text{if } n \equiv 2 \pmod{3}. \end{cases}$$

It can be seen that $f(n) = |K^{-1}|$ and $3^{\lfloor n/4 \rfloor} < f(n)$.

It is clear that $n-1 \leq |K| \leq n+2$, $3^{\lfloor n/4 \rfloor} < |K^{-1}|$.

Thus, if denote the elements of K by K_1, \dots, K_t , then we set $s = \langle R, F \rangle$, where $F = \{K_1 \rightarrow R, \dots, K_t \rightarrow R\}$. By Theorem 1.4 K^{-1} is the set of all antikeys of s . Consequently, for an arbitrary set of attributes we can always construct a relation scheme $s = \langle R, F \rangle$ such that $|F| < |R| + 2$, but the number of antikeys of s is exponential not only in the number of attributes, but also in the number of elements of F . The theorem is proved.

According to Proposition 2.2 we show that finding a maximal family $M(s)$ can be polynomially transformed to problem of finding all antikeys of given relation scheme.

Algorithm 2.4.

Input: Let $s = \langle R, F \rangle$ be relation scheme.

Output: K_s^{-1} .

Step 1: For each $a \in R$ we construct T_a .

Step 2: Set

$$N_s = \bigcup_{a \in R} L(T_a).$$

Step 3: Put

$$K_s^{-1} = \{A \in N_s : \exists B \in N_s : A \subset B\}.$$

Clearly, the steps 2 and 3 of this algorithm require polynomial time in the number of attributes. On the other hand, according to Theorem 2.3 we have the following.

Corollary 2.5. *Let $s = \langle R, F \rangle$ be a relation scheme. Then the time complexity of finding the family $M(s)$ 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] Thi V. D., Demetrovics J., Some results about normal forms for functional dependency in the relational datamodel, *J. Discrete Applied Mathematics, North Holland* **69** (1996) 61–74.
- [18] Thi V. D., Demetrovics J., Describing Candidate Keys by hypergraphs, *J. Computers and Artificial Intelligence* **18** (2) (1999) 191–207.
- [19] Thi V. D., Demetrovics J., Some computational problems related to Boyce–Codd normal form, *Anales Univer. Sci. Budapest, Sect. Comp.* No. 19 (2000) 119–132.
- [20] 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 16, 2000