

ON THE CONNECTIONS BETWEEN RELATIONS, RELATION SCHEMES AND KEYS

VU DUC THI

Abstract. The relation, relation scheme, keys and antikeys are essential concepts in the relational datamodel. In this paper, we present some computational connections between relations, relation schemes and sets of minimal keys.

Some another combinatorial results connecting them also are given.

1. INTRODUCTION

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

Definition 1.1. Let $R = \{h_1, \dots, h_n\}$ be a relation over U , and $A, B \subseteq U$. Then we say that B functionally depends on A in R (denoted $A \xrightarrow{f} B$) 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 $F_R = \{(A, B) : A, B \subseteq U, A \xrightarrow{f} B\}$. F_R is called the full family of functional dependencies of R . Where we write (A, B) or $A \rightarrow B$ for $A \xrightarrow{f} B$ when R, f are clear from the next context.

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

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 know [1] that if Y is an arbitrary f -family, then there is 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 relation scheme, then there is a relation R over U such that $F_R = F^+$. Such that relation in called an Armstrong relation of S .

Definition 1.5. Let be a relation, $S = \langle U, F \rangle$ be a relation scheme, YY 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 $A \xrightarrow{f} U$ ($A \rightarrow U \in F^+, (A, U) \in Y$). 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 antikeys of K , denote by K^{-1} as follows:

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

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

It is know [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 s 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 $T_R = \{A \in P(U) : \exists E_{ij} = A, \forall E_{pq} : E_{ij} \subset E_{pq}\}$. 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 $K_R = K$.

Definition 1.10. Let R be a relation, and K a Sperner system over U . We say that R present K iff $K^{-1} = T_R$, where T_R is the maximal equality system of R .

2. RESULTS

It is known that the minimal keys play important roles for the logical and structural investigation of relational datamodel. It is clear that there are relation schemes $S_1 = \langle U, F_1 \rangle, S_2 = \langle U, F_2 \rangle$ for which $K_{S_1} = K_{S_2}$ but $F_1^+ \neq F_2^+$.

Now we present the following problems to relate relations, relation schemes and sets of minimal keys.

- (1) Constructing relation for key: Given relation scheme S . Construct a relation R such that $K_R = K_S$.
- (2) Constructing relation scheme for key: Let R be a relation. Construct a relation scheme S for which $K_R = K_S$.

Let R be a relation and S a relation scheme over U , then

- (3) FD-relation key-implication problem: Decide whether $K_R \subseteq K_S$.
- (4) FD-relation key-equivalence problem: Decide whether $K_R = K_S$.

It can be seen that by Theorem 2.11, and based on (1) and (2) of Proposition 2.12 in [19] we obtain

Proposition 2.1. *The first problem has exponential complexity in the number of attributes.*

It can be seen that there are a relation R and a relation scheme $S = \langle U, F \rangle$ such that $F_R \subseteq F^+$, but $K_R \not\subseteq K_S$. Conversely, we also known that there are a relation R and a relation scheme S for which $K_R \subseteq K_S$, but $F_R \not\subseteq F^+$.

From (1), (2) of Proposition 2.12 in [19], we also obtain

Proposition 2.2. *The second problem is polynomially equivalent to finding a relation scheme S of a given Sperner system K such that $K = K_S^{-1}$.*

Clearly, by (1), (3), (9), (10) and (11) of Proposition 2.12 in [19] we can construct the exponential time algorithms solve three last problems. However, it is still unknown that the time complexities of these problems are polynomially or not. For us these problems are the interesting open problems and there are not the polynomial algorithms that solve them.

Theorem 2.3 [4]. *Let K be a Sperner system over U . We say that K is saturated if for any $A \notin K$, $\{A\} \cup K$ is not a Sperner system. If K is a saturated Sperner system then $K = K_F$ uniquely determines F , where K_F is the set of all minimal keys of an f -family F .*

By examples we see that there is K (K^{-1}) such that K (K^{-1}) is saturated, but K^{-1} (K) is not saturated.

Now we define the following notion

Definition 2.4. Let K be a Sperner system over U . We say that K is inclusive if for every $A \in K$ there is a $B \in K^{-1}$ so that $B \subset A$ we call K is embedded if for every $A \in K$ there exists a $B \in H$ such that $A \subset B$, where $H^{-1} = K$.

In [10] we proved that

Theorem 2.5. *Let K be a Sperner system over U . Denote H a Sperner system for which $H^{-1} = K$. The following facts are equivalent:*

- (1) K is saturated.
- (2) K^{-1} is embedded.
- (3) H is inclusive.

Definition 2.6. Let $S = \langle U, F \rangle$ be a relation scheme. We say that S is a k -relation scheme over U if $F = \{K_1 \rightarrow U, \dots, K_m \rightarrow U\}$, where $\{K_1, \dots, K_m\}$ is a Sperner system over U .

Clearly, if $S = \langle U, F \rangle$ is in Boyce-Codd normal form then use the algorithm for finding a minimal cover polynomial time we can a construct a k -relation scheme $S' = \langle U, F' \rangle$ such that $F^+ = F'^+$, see [11]. Conversely, we can see that an arbitrary k -relation scheme is BCNF. Now we give a class of relations and relation scheme for which above problems are solved in polynomial time.

Theorem 2.7. *Let $S = \langle U, F \rangle$ be a k -relation scheme, and R an arbitrary relation over U . Then*

- (1) *If $K = \{K_1, \dots, K_m\}$ is saturated then the first, fourth problem are solved in polynomial time in $|U|, |K|$.*
- (2) *If K_R^{-1} is saturated then the second, third, fourth problem solved on polynomial time in the size of R .*

Proof. For (1) we assume that $K = \{K_1, \dots, K_m\}$ is saturated. We set $M = \{K_i - \{a\} : a \in K_i, i = 1, \dots, m\}$ and $L = \{A \in M : A = A^+\}$, where A^+ is closure of A . Denote elements of L by A_1, \dots, A_t . We construct a relation $R \{h_0, h_1, \dots, h_t\}$, as follows: For each $a \in U$, $h_0(a) = 0$. For $i = 1, \dots, t$, $h_i(a) = 0$, if $a \in B_i$, in the converse case $h_i(a) = i$. By Theorem 2.5, Definition 2.4 and $K^{-1} = L = K_R^{-1}$. Hence, $K = K_R$ holds. It can be seen that R is constructed in polynomial time in the size of S . For problem 4 we compute K_R^{-1} of a given relation R in polynomial time. We compare K_R^{-1} with L . We decide whether $K_R = K$.

For (2): From R we compute K_R^{-1} in polynomial time in $|R|$. We set $P = \{T \cup \{a\} : a \in U, T \in K_R^{-1}\}$ and $Q = \{E \in P : (E, U) \in F_R, \forall A = E - a : A \in E, (A, U) \notin F_R\}$. Denote elements of Q by E_1, \dots, E_l . Clearly, Q is constructed in polynomial time. We set $S = \langle U, F \rangle$ with $F = \{E_1 \rightarrow U, \dots, E_l \rightarrow U\}$. By Theorem 2.5 and Definition 2.4 we obtain $Q = K_R$. Hence, $K_S = K_R$ holds.

It be seen that for problems 3 and 4 we compare Q with the family $K = \{K_1, \dots, K_m\}$. We decide whether $K_R \subseteq K_S$, and $K_R = K_S$. The theorem is proved.

Let R be a relation, and S a relation scheme over U , Denote by L_R (L_S) the family of all keys of R (S).

It is easy to see that there are relation R and relation scheme $S = \langle U, F \rangle$ such that $L_R \subseteq L_S$ ($L_S \subseteq L_R$), but $F_R \not\subseteq F^+$ ($F^+ \not\subseteq F_R$). In [20] we showed the following algorithm

Algorithm 2.8.

Input: Let R be relation, $S = \langle U, F \rangle$ a relation scheme over U .

Output: Decide whether $L_S \subseteq L_R$.

Step 1: From R we compute the equality set E_R .

Step 2: Construct the maximal equality system $T_R = \{T_1, \dots, T_m\}$ from E_R .

Step 3: By algorithm finding a closure for every T_i ($i = 1, \dots, t$) we compute T_i^+ , if $\forall i = 1, \dots, t : T_i^+ \neq U$ then $L_S \subseteq L_R$ holds. In the converse case $L_S \not\subseteq L_R$ holds.

It is known [13] that the subset delimiter complementarity (SDC) problem is co-NP-complete.

Lemma 2.9 [13]. *The following problem is co-NP-complete: Given a finite set T , a family P_1, \dots, P_n of subsets of T , and a family Q_1, \dots, Q_m of subsets of T . Decide whether for al $A \subseteq T$ there exists $P_i : P_i \subseteq A$ or there is $Q_j : A \subseteq Q_j$, where $q \leq i \leq j \leq m$.*

In [13] it is proved that if Q_1, \dots, Q_m is Sperner system over T , then SDC problem is co-NP-complete.

In [20] we showed that deciding whether $L_R \subseteq L_S$ for given a relation R and a relation scheme S is co-NP-complete, but we don't prove this results. In this Section we prove concretely it.

We show that the SDC problem is polynomially reducible to our problem.

Theorem 2.10. *The following problem is co-NP-complete:*

Give a relation R , and a relation scheme $S = \langle U, F \rangle$, decide whether $L_R \subseteq L_S$.

Proof. It can be seen that for $A \subseteq U$ we can test that A is or is not a key of S .

Consequently, we nondeterministically choose a subset A of U such that A is a key of R , but is not a key of S . It is obvious that this algorithm is nondeterministically polynomial. Thus, our problem lies in co-NP.

Now we see the SDC problem containing a set T and two families P_1, \dots, P_n and Q_1, \dots, Q_m , where Q_1, \dots, Q_m is a Sperner system over T .

It can be seen that denote $P = \{P_1, \dots, P_n\}$, $Q = \{Q_1, \dots, Q_n\}$ and

$$P' = \{P_i \in P : \exists P_j \subset P_i, i = 1, \dots, n, j = 1, \dots, n\}.$$

Clearly, P' is a set of minimal elements of P and P' is Sperner system over T . From P we compute P' in polynomial time in $|P|$ and $|T|$. It is easy to see that $\{T, P', Q\}$ is an equivalent instance of $\{T, P, Q\}$. From this we can assume that P is a Sperner system over T . We shall prove that the SDC problem is polynomial reducible to our problem.

We set $U = T$, $S = \langle U, F \rangle$, where $F = \{P_1 \rightarrow U, \dots, P_n \rightarrow U\}$.

Set $M = \{Q_i - a : i = 1, \dots, m \text{ and } a \in U\} = \{M_1, \dots, M_t\}$. We construct a relation $R = \{h_0, h_1, \dots, h_t\}$, as follows:

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

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

Clearly, R, S are constructed in polynomial time in $|T|, |P|, |Q|$. It can be seen that F_R and $S = \langle U, F \rangle$ are in Boyce-Codd normal form.

Because S is in BCNF, for all $A \subseteq U$ we have $A^+ = A$ or $A^+ = U$. By the definition of key, for each key A of S there exists a P_i such that $P_i \subseteq A$, where $1 \leq i \leq n$.

It can be seen that Q is a set of antikeys of R . Because F_R is in BCNF, we seen that for each $A \subseteq U$, $H_{FR}(A) = U$ or $H_{FR}(A) = A$, where $H_{FR}(A) = \{A \in U : (A, \{a\}) \in F_R\}$.

By the definition of antikey of R , we known that A is a key of R if and only if for all $i = 1, \dots, m$ $A \not\subseteq Q_i$.

Consequently, $L_R \subseteq L_S$ if and only if for every $A \subseteq T : \forall i = 1, \dots, m$ $A \not\subseteq Q_i$ then there is P_j so that $P_j \subseteq A$. From this we see that the SDC problem is polynomially reducible our problem. The theorem is proved.

It is easy to see that in proof of Theorem 2.10 we construct relation R and relation scheme S which are in BCNF. Clearly, in the class of BCNF relations R and BCNF relation schemes $S = \langle U, F \rangle$, then $F^+ = F_R$ iff $L_S = L_R$. We can prove that in this class the FD-relation implication problem lies in co-NP-complete.

Thus, from the proof of the Theorem 2.10 we obtain a result which stronger the result is proved [13].

Corollary 2.11. *In BCNF class of relations and relation schemes the FD-relation implication problem is co-NP-complete.*

REFERENCES

- [1] Armstrong W. W, Dependency Structures of Database Relationship, *Information Processing 74*, Holland Publ., Co., 1974, pp. 580-583.
- [2] Beeri C., Dowd M., Fagin R., Staman R., On the structure of Armstrong relations for functional dependencies, *J. ACM* **31** (1) (1984) 30-46.
- [3] Beeri C., Bernstein P. A. Computainonal problems related to the design of normal form relational schema, *ACM Trans. on Database Syst.* **4** (1) (1979) 30-59.
- [4] Burosch G., Demetrovics J., Katona G.O.H., The poset of closures as a model of changing databases, *Oder* **4** (1987) 127-142.
- [5] Czedli G., On the dependencies in the relational model of data, *EIK* **17** (2-3) (1981) 103-112.
- [6] Demetrovics J., Relacios adatmodell logikai es strukturalis vizsgalata MTA, *SZTALI Tanulmányok*, Budapest, **114** (1980) 1-97.
- [7] Demetrovics J., Gyepesi G., On the functional dependency and generalizations of it, *Acta Cybernetica Hungary* **V** (3) (1983) 295-305.
- [8] Demetrovics J., Thi. V.D., Some results about functional dependencies, *Acta Cybernetica Hungary* **VIII** (3) (1988) 273-278.
- [9] Demetrovics J., Thi. V.D., Relations and minimal keys, *Acta Cybernetica Hungary* **VIII** (3) (1988) 279-285.
- [10] Thi. V.D., Minimal keys and antikeys, *Acta Cybernetica Hungary*, Tom. 7, Fasc. 4 (1986) 361-371.
- [11] Thi. V.D., Strong dependencies and s-semilattices, *Acta Cybernetica Hungary* **VIII** (2) (1987) 195-202.
- [12] Thi V.D., Logical dependencies and irredundant relations, *Computers and Artificial Intelligence* **7** (2) (1988) 165-184.
- [13] Gottlob G., Libkin L., Investigation on Armstrong relations, dependency inference, and excluded functional dependencies, *Acta Cybernetica Hungary*, Tom. 9, Fasc. 4 (1990) 385-402.
- [14] Mannila H., Raiha K. J., Algorithms for inferring functional dependencies, *Proc. of the Thirteenth International Conference on Very Large Data Bases*, Brighton, September, 1987.

- [15] Lucchesi C.L., Osborn S.L., Candidate keys for relations, *J. Comput. Syst. Scien.* **17** (2) (1978) 270-279.
- [16] Demetrovics J., Thi V.D., On the algorithms for generating Armstrong relations and for inferring functional dependencies in the relational datamodel, *Computers Math. Applic. Great Britain* **26** (4) (1993) 43-55.
- [17] Demetrovics J., Thi V.D., Some result about normal forms for functional dependency in the relation datamodel, *Discrete Applied Mathematics* **69** (1996) 61-74.
- [18] Thi V.D., Some results about hypergraph, *J. Informatics and Control* **13** (2) (1997) 8-15.
- [19] Thi V.D., On the computational algorithms related to antikeys, *J. Informatics and Control* **14** (2) (1998).
- [20] Demetrovics J., Thi. V.D., Some problems concerning keys for relation scheme and relation in the relational datamodel, *IPL, North Holland* **46** (4) (1993) 179-183.

Tóm tắt. Quan hệ, sơ đồ quan hệ, khóa và phản khóa là những khái niệm nền tảng trong mô hình dữ liệu quan hệ. Trong bài báo này, chúng tôi trình bày một số mối liên hệ tính toán giữa quan hệ, sơ đồ quan hệ và tập các khóa tối thiểu. Một số kết quả tổ hợp khác liên quan tới chúng cũng được trình bày.

Received: October 1, 1997

*Institute of Information Technology
National Centre for Natural Sciences and Technology of Vietnam.*