Annales Univ. Sci. Budapest., Sect. Comp. **8** (1987) 35-52

KEYS, ANTIKEYS AND PRIME ATTRIBUTES*

J. DEMETROVICS and V. D. THI

Dedicated to Professor Imre Kátai on his 50^{th} birthday

Abstract. In this paper some necessary and sufficient conditions are set up for a relation to represent an arbitrarily given family of functional dependencies or a Sperner-system. Our approach is based on the study of antikeys, i.e. the maximal non-keys of a relation scheme (or a relation). An expression on the connection between minimal keys and antikeys is given. As an application, we show that the prime attribute problem for a given relation can be solved in deterministic polynomial time. The corresponding problem for relational schemes is NP-complete.

We give an improved algorithm to find a relation representing a given Sperner-system and discuss some related complexity problems.

1. Introduction

The relational data model was defined by Codd [7]. In this model a relation is a table (matrix) in which each column corresponds to an attribute and each row to an entity (record). Relations are used to describe connections among data items. One

---

* Research supported by Hungarian Foundation for Scientific Research Grant 1066

of the main concepts in the relational data model is functional dependency. Many papers dealing with the computational complexity of functional dependencies have appeared. A relational scheme is a pair

$$S = \langle \Omega,\ F = X \to Y : X, Y \subseteq \Omega \rangle,$$

where Ω is a finite set of attributes and F is a set of functional dependencies. Denote by F^+ the set of all functional dependencies that can be derived from F by using Armstrong's axioms [2].

The membership problem for a relational scheme (i.e. given a relational scheme $S = \langle \Omega, F \rangle$ and a functional dependency $X \to Y$, decide whether $X \to Y$ belongs to F^+) can be solved by an efficient polynomial time algorithm [3]. Based on this method, one can find an irredundant cover [3] in polynomial time. Maier [17] constructed a polynomial time algorithm for determining a minimum cover for a given relational scheme. In [3] it is proved that the problem of determining whether a relational scheme is in Boyce-Codd normal form is *NP*-complete.

The structure of sets of minimal keys has been thoroughly investigated, too. It was shown [20] that the number of minimal keys for a relational scheme $S = \langle \Omega, F \rangle$ can be exponential in $| \Omega |$. In [5] a relational scheme $S = \langle \Omega, F \rangle$ is constructed in which $| \Omega | = k(k-1)$, $| F | = k$ and S has $k!$ minimal keys. The equivalence of sets of minimal keys with Sperner-systems is well-known [9]. The representation of minimal keys by a given set of functional dependencies is given in [5]. An algorithm that determines the minimal keys can be constructed from it. In [11] the authors successfully investigated and constructed the minimal relations representing a given Sperner-system in some special cases.

Antikeys (i.e. maximal non-keys) play an essential role in extremal problems of the relational data model as well as in the construction of relations representing a Sperner-system [18] or in finding minimal keys [1]. Also the set of antikeys is a Sperner-system. The connections between minimal keys and antikeys are shown in [11,19].

Algorithms determining the set of all minimal keys for a relational scheme were constructed in [6,8,12,13,15,16]. It is interesting to note that the methods of C. L. Lucchesi, S. L. Osborn and M. C. Fernandez have polynomial complexity in many special cases. Thus, very frequently (especially, when there is a relatively small number of minimal keys only), these algorithms are much better than those in [6,8,13].

The prime attributes and minimal keys play important roles in the normalization process of relations. Lucchesi and Osborn [16] proved that the next two essential problems are *NP*-complete:

(1) The prime attribute problem: Given a relational scheme and an attribute A, decide whether A belongs to any minimal key.
(2) The key of cardinality problem: Given a relational scheme and an integer $m > 1$, decide whether there exists a key of cardinality less than m.

Using the *NP*-completeness of (2) it is shown in [17] that the optimal cover problem is *NP*-complete.

In Section 2 the necessary definitions are presented. In Section 3 we give some results about relations representing an arbitrarily given family of functional dependencies and/or a Sperner-system.

Beeri et al. [4] proved that problem (2) remains *NP*-complete even if the input is a fixed relation (matrix) instead of a relational scheme. In Section 3, however, we establish an expression on the connection between minimal keys and antikeys, and prove that if relations (i.e. matrices) are considered instead of relational schemes then the prime attribute problem can be solved in polynomial time. In [16] Lucchesi and Osborn proved that the key cardinality problem is polynomially transformable to the prime attribute problem. Thus, if $NP \neq P$, then this transformation is not possible for relations.

In Section 4 first we construct a combinatorial algorithm for finding a relation which represents a given Sperner-system $\mathbf{K}$. The worst-case time of this algorithm is exponential in the number of attributes and in $|\mathbf{K}|$. We show, however, that in many cases (especially, if the number of elements of $\mathbf{K}$ is small) this algorithm is polynomial in the input size. We prove that the time complexity of finding a relation which represents a given Sperner-system $\mathbf{K}$ is exponential in $|\mathbf{K}|$. Conversely, it is shown that the time complexity of finding the set of all minimal keys of a given relation R is not polynomial unless $P = NP$.

2. Definitions

In this section we give some definitions.

Let $\Omega = \{a_1, \ldots, a_n\}$ be a finite nonempty set of attributes. For each attribute a_i there is a nonempty set $D(a_i)$ of possible values of that attribute. An arbitrary finite subset of the Cartesian product $D(a_1) \times \ldots \times D(a_n)$ is called a relation over Ω. Clearly, a relation over Ω is a set of mappings

$$h : \Omega \to \bigcup_{a \in \Omega} D(a),$$

where $h(a) \in D(a)$ for all $a \in \Omega$.

DEFINITION 2.1. Let $R = \{h_1, \ldots, h_m\}$ be a relation over the finite set Ω of attributes and $A, B \subseteq \Omega$. We say that B functionally depends on A in R (denoted as $A \to B$) iff

$$(\forall h_i, h_j \in R) \ \big((\forall a \in A) \ (h_i(a) = h_j(a)) \to$$

$$\to (\forall b \in B) \ (h_i(b) = h_j(b))\big),$$

where $1 \le i, j \le m$.

Let $F_R = \{(A, B) : A \to B \text{ holds for } R\}$. F_R is called the full family of functional dependencies in R.

DEFINITION 2.2. Let Ω be a finite set, and denote $P(\Omega)$ its power set. Let $F \subseteq P(\Omega) \times P(\Omega)$. We say that F is an f-family over Ω iff for all $A, B, C, D \subseteq \Omega$:

(F1) $(A, A) \in F$;

(F2) $(A, B) \in F,\ (B, C) \in F \rightarrow (A, C) \in F$;

(F3) $(A, B) \in F,\ A \subseteq C,\ D \subseteq B \rightarrow (C, D) \in F$;

(F4) $(A, B) \in F,\ (C, D) \in F \rightarrow (A \cup C, B \cup D) \in F$.

Clearly, F_R is an f-family over Ω.

It is known [2] that if F is an arbitrary f-family over Ω, then there is a relation R such that $F_R = F$.

DEFINITION 2.3. The mapping $\mathcal{L} : P(\Omega) \rightarrow P(\Omega)$ is called a closure operation over Ω iff for every $A, B \subseteq \Omega$:

$$\begin{aligned}&(1)\ A \subseteq \mathcal{L}(A),\\ &(2)\ A \subseteq B \rightarrow \mathcal{L}(A) \subseteq \mathcal{L}(B),\\ &(3)\ \mathcal{L}(\mathcal{L}(A)) = \mathcal{L}(A).\end{aligned}$$

REMARK 2.4. Clearly, if F is an f-family, and we define $\mathcal{L}_F(A)$ as

$$\mathcal{L}_F(A) = \{b \in \Omega :\ (A, \{b\}) \in F\}$$

then $\mathcal{L}_F$ is a closure operation over Ω. Conversely, it is known [2,9] that if $\mathcal{L}$ is a closure operation, then there is exactly one f-family F over Ω so that $\mathcal{L} = \mathcal{L}_F$, where

$$F = \{(A, B) :\ A, B \subseteq \Omega,\ B \subseteq \mathcal{L}(A)\}.$$

Thus, there is a one-to-one correspondence between closure operations and f-families over Ω.

DEFINITION 2.5. Let R be a relation, $\mathcal{L}$ a closure operation over Ω. K is called a key of R (of $\mathcal{L}$) if $K \rightarrow \Omega$ ($\mathcal{L}(K) = \Omega$). K is a minimal key of R (of $\mathcal{L}$) if K is a key of R (of $\mathcal{L}$), but $B \not\to \Omega$ ($\mathcal{L}(B) \neq \Omega$) for any proper subset B of K. Denote $\mathcal{K}_R$ ($\mathcal{K}_L$) the set of all minimal keys of R (of $\mathcal{L}$). Clearly, $K_1, K_2 \in \mathcal{K}_R$ implies $K_1 \not\subset K_2$. The systems of subsets of Ω satisfying this condition are called Sperner-systems. Consequently, $\mathcal{K}_R$, $\mathcal{K}_L$ are Sperner-systems [18].

DEFINITION 2.6. Let $\mathcal{K}$ be a Sperner-system over Ω. We define the set of antikeys of $\mathcal{K}$, denoted by $\mathcal{K}^{-1}$, as follows: a set B is an antikey ($B \in \mathcal{K}^{-1}$) iff

(i) no subset of B is a key ($K \not\subseteq B$ if $K \in \mathcal{K}$) but

(ii) B is maximal with respect to this property in the sense that all proper supersets C of B (i.e. $B \subseteq C$, $B \neq C$) contain at least one key.

It is easy to see that the elements of $\mathcal{K}^{-1}$ are maximal non-keys and $\mathcal{K}^{-1}$ is a Sperner-system over Ω.

Theorem 2.7 [9,11]. *If $\mathcal{K}$ is an arbitrary Sperner-system then there is a closure operation $\mathcal{L}$ for which $\mathcal{K}_L = \mathcal{K}$.*

In this paper, we always assume that if a Sperner-system plays the role of the set of minimal keys (antikeys) then it is not empty (it is not Ω).

DEFINITION 2.8 [2]. Let F be an f-family over Ω and $(A, B) \in F$. We say that (A, B) is a maximal right side dependency of F iff

$$\forall B' \; (B \subseteq B') : \; (A, B') \in F \to B = B'.$$

Denote by $M(F)$ the set of all maximal right side dependencies of F. We say that B is a maximal side of F iff there is an A such that $(A, B) \in M(F)$. Denote by $I(F)$ the set of all maximal sides of F.

DEFINITION 2.9. Let R be a relation, F an f-family, $\mathcal{K}$ a Sperner-system over Ω. We say that R represents $F(\mathcal{K})$ iff $F_R = F$ ($\mathcal{K}_R = \mathcal{K}$).

DEFINITION 2.10. Let $R = \{h_1, \ldots, h_m\}$ be a relation over Ω. For $1 \leq i < j \leq m$ denote by E_{ij} the set $\{a \in \Omega : h_i(a) = h_j(a)\}$. We put $E_R = \{E_{ij} : 1 \leq i < j \leq m\}$. E_R is called the equality set of R.

3. Antikeys and prime attributes

First we give a necessary and sufficient condition for a relation to represent an arbitrarily given f-family.

Theorem 3.1. *Let $R = \{h_1, \ldots, h_m\}$ be a relation, F an f-family over Ω. Then R represents F iff for every $A \subseteq \Omega$*

$$\mathcal{L}_F(A) = \begin{cases} \bigcap_{A \subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij} \in E_R : \ A \subseteq E_{ij}, \\ \Omega & \text{otherwise.} \end{cases}$$

(Recall that $\mathcal{L}_F(A) = \{a \in \Omega : \ (A, \{a\}) \in F\}$ and E_R is the equality set of R.)

Proof. First we prove that in an arbitrary relation R for all $A \subseteq \Omega$

$$\mathcal{L}_{F_R}(A) = \begin{cases} \bigcap_{A \subseteq E_{ij}} E_{ij} & \text{if } \exists E_{ij} \in E_R : \ A \subseteq E_{ij}, \\ \Omega & \text{otherwise.} \end{cases}$$

Let us first assume that A is a set such that there is no $E_{ij} \in E_R$ with $A \subseteq E_{ij}$. Then for all $h_i, \ h_j \in R \ \ a \in A : h_i(a) = h_j(a)$. According to the definition of functional dependency, this implies $A \to \Omega$, and by the definition of $\mathcal{L}_{F_R}$ we obtain $\mathcal{L}_{F_R}(A) = \Omega$. Clearly,

$$\mathcal{L}_{F_R}(0) = \bigcap_{E \in E_R} E_{ij}$$

holds. If $A \neq 0$ and there is an $E_{ij} \in E$ such that $A \subseteq E_{ij}$ then we set

$$V = \{E_{ij} : A \subseteq E_{ij}, \ E_{ij} \in E_R\}$$

and

$$E = \bigcap_{E_{ij} \in V} E_{ij}.$$

It is easy to see that $A \subseteq E$ holds. If $V = E_R$ then it is obvious that $(A, E) \in F_R$.

If $V \neq E_R$ then it can be seen that for all $E_{ij} \in V$ we have

$$(\forall a \in A)\ \ (h_i(a) = h_j(a)) \to (\forall b \in B)\ \ (h_i(b) = h_j(b))$$

and for all $E_{ij} \notin V$ there is an $a \in A$ such that $h_i(a) \neq h_j(a)$. Thus, $(A, E) \in F_R$ holds.

By the definition of $\mathcal{L}_{F_R}$, we have $E \subseteq \mathcal{L}_{F_R}(A)$. Since R is a relation over Ω , we have $E \subset \Omega$. Using $A \subseteq E \subseteq \mathcal{L}_{F_R}(A)$, we obtain $(E, \mathcal{L}_{F_R}(A)) \in F_R$.

Now we assume that c is an attribute such that $c \notin E$. Then, there is an $E_{ij} \in V$ so that $c \notin E_{ij}$. This implies the existence of a pair $h_i, h_j \in R$ such that $\forall b \in E : h_i(b) = h_j(b)$ holds, but $h_i(c) \neq h_j(c)$. By the definition of functional dependency $(E \cup \{c\})$ does not depend on E. Thus, for all attributes $c \notin E$ we have $(E, E \cup \{c\}) \notin F_R$. By the definition of $\mathcal{L}_{F_R}$ we obtain

$$\mathcal{L}_{F_R}(A) = \bigcap_{E_{ij} \in V} E_{ij}.$$

By Remark 2.4 it is easy to see that $F_R = F$ holds iff $\mathcal{L}_{F_R} = \mathcal{L}_F$ does. The theorem is proved. □

DEFINITION 3.2. Let $\mathcal{L}$ be a closure operation over Ω, $N \subseteq P(\Omega)$. Let N^+ denote the set $\{\cap N' : N' \subseteq N\}$. Using the convention $\cap 0 = \Omega$ we see that N^+ always contains Ω.

Let $Z(\mathcal{L}) = \{A \subseteq \Omega : \mathcal{L}(A) = A\}$. The elements of $Z(\mathcal{L})$ are called closed sets.

Clearly, $Z(\mathcal{L})$ is closed under intersection.

It can be seen that for all E_{ij} $(E_{ij} \in E_R)$ we have $E_{ij} \in Z(\mathcal{L}_{F_R})$, i.e. $E_R^+ \subseteq Z(\mathcal{L}_{F_R})$.

By Theorem 3.1 $Z(\mathcal{L}_{F_R}) \subseteq E_R^+$ holds. We have proved

COROLLARY 3.3. Let R be a relation, F an f-family over Ω. Then R represents F iff $Z(\mathcal{L}_{F_R}) = E_R^+$ holds. □

It is known [2,9] that for an arbitrary non-empty Sperner-system $\mathcal{K}$ there is a relation R such that $\mathcal{K} = \mathcal{K}_R$.

Now we give a necessary and sufficient condition for a relation to represent a given Sperner-system. First we introduce the notion of maximal equality system.

DEFINITION 3.4. Let $R = \{h_1, \ldots, h_m\}$ be a relation over Ω, E_R is the equality set of R. Let

$$M_R = \{E_{ij} \in E_R : \ \forall E_{st} \in E_R : E_{ij} \subseteq E_{st}, \ E_{ij} \neq E_{st},$$

$$1 \leq i < j \leq m, \ 1 \leq s < t \leq m\}.$$

M_R is called the maximal equality system of R.

Theorem 3.5. *Let $\mathcal{K}$ be a non-empty Sperner-system, R a relation over Ω. Then R represents $\mathcal{K}$ iff $\mathcal{K}^{-1} = M_R$.*

Proof. It can be seen that if $\mathcal{K}$ is a non-empty Sperner-system then $\mathcal{K}^{-1}$ exists. On the other hand, $\mathcal{K}$ and $\mathcal{K}^{-1}$ are uniquely determined by each other therefore $\mathcal{K}_R = \mathcal{K}$ holds iff $\mathcal{K}_R^{-1} = \mathcal{K}^{-1}$ does. Now it suffices to show that $\mathcal{K}_R^{-1} = M_R$ holds. Clearly, F_R is an f-family over Ω. First we suppose that A is an antikey of $\mathcal{K}_R$. Obviously $A \neq \Omega$. If there is a B such that $A \subset B$ and $A \to B$, then by the definition of antikeys we have $B \to \Omega$ and $A \to \Omega$. This is a contradiction. So $A \in I(F_R)$ holds. If there is a B' such that $B' \neq \Omega$, $B' \in I(F_R)$ and $A \subset B'$, then B' is a key of R. This contradicts $B' \neq \Omega$. Hence $A \in I(F_R) - \Omega$ and $\overline{\exists} B' \ (B' \in I(F_R) - \Omega) : \ A \subset B'$.

On the other hand, according to the definition of a relation, $\Omega \notin M_R$. Clearly, $E_{ij} \in I(F_R)$. Thus, $M_R \subseteq I(F_R)$ holds. If D is a set such that $\forall C \in M_R : \ D \not\subseteq C$, then D is a key of R. Consequently, M_R is the set of maximal distinct elements of $I(F_R)$. So we obtain $A \in M_R$.

Conversely, if $A \in M_R$, then according to the definition of a relation and M_R we obtain $A \not\to \Omega$, i.e. $\forall K \in \mathcal{K}_R : K \not\subseteq A$. On the other hand, because A is a maximal equality set, for all D $(A \subset D)$ we have $D \to \Omega$, therefore by the definition of antikeys $A \in \mathcal{K}_R^{-1}$ holds. The theorem is proved. □

We mention here a result relating the sizes of $\mathcal{K}^{-1}$ and $\mathcal{K}$.

Theorem 3.6 [11]. *Let* K *be a Sperner-system over* Ω. *Let* $s(\mathcal{K}) = \min\{m : \mathcal{K} = \mathcal{K}_R,\ |R| = m,\ R$ *is a relation over* $\Omega\}$. *Then*

$$\sqrt{2\,|\mathcal{K}^{-1}|} \leq s(\mathcal{K}) \leq |\mathcal{K}^{-1}| + 1. \tag{1}$$

DEFINITION 3.7 [7]. Let R be a relation over Ω and $\mathcal{K}_R$ the set of all minimal keys of R. We say that a is a prime attribute of R iff there is a $K \in \mathcal{K}_R$ such that $a \in K$.

The following result expresses the set of prime attributes in terms of antikeys.

Theorem 3.8. *Let* $\mathcal{K}$ *be a Sperner-system over* Ω. *Then*

$$\cup\mathcal{K} = \Omega - \cap\mathcal{K}^{-1}.$$

Proof. If $c \in \cup\mathcal{K}$, then there exists a $K \in \mathcal{K}$ such that $c \in \mathcal{K}$. Let $H = K - \{c\}$. As H contains no keys, there exists an antikey $B \in \mathcal{K}^{-1}$ such that B contains H. It is clear that $c \notin B$, for otherwise we have $K \subseteq B$ which is impossible. Now we see that

$$c \in \Omega - B \subseteq \Omega - \cap\mathcal{K}^{-1}.$$

Now suppose that $c \notin \cup\mathcal{K}$ and let $B \in \mathcal{K}^{-1}$. It suffices to show that $c \in B$. Indirectly assume $c \notin B$. Then $\{c\} \cup B$ contains a key $K \in \mathcal{K}$. As $K \subseteq B$, we have $c \in K$, a contradiction. □

Based on Theorems 3.5 and 3.8 we show that when the input is a relation, then the prime attribute problem can be solved in deterministic polynomial time.

First we construct an algorithm for determining the set of prime attributes of a relation.

ALGORITHM 3.9. *Input*: $R = \{h_1, \ldots, h_m\}$, a relation over Ω.

Output: V, the set of all prime attributes of R.

Step 1: From R we construct the set $E_R = \{E_{ij} : 1 \leq i < j \leq m\}$, where $E_{ij} = \{a \in \Omega : h_i(a) = h_j(a)\}$.

Step 2: From E_R we construct the set

$$M = \{B \in P(\Omega) : \exists E_{ij} \in E_R : E_{ij} = B\}.$$

Step 3: The set $M_R = \{B \in M : \forall B' \in M : B \not\subset B'\}$ is computed.

Step 4: Construct the set $V = \Omega - \cap M_R$.

Clearly, $| M_R | \leq | M | \leq | E_R | = \binom{m}{2}$ holds, so the worst-case time of Algorithm 3.9 is polynomial in the number of rows and columns of R.

From Theorems 3.5 and 3.8 and Algorithm 3.9 the next corollary is clear.

COROLLARY 3.10. There is an algorithm that given a relation R, decides whether an attribute is prime or not, with worst-case time polynomial in the number of rows and columns of R. □

4. Finding relations and minimal keys

First we construct an algorithm for finding the set of antikeys of a given Sperner-system, as follows.

Let an arbitrary Sperner-system $\mathcal{K} = \{K_1, \ldots, K_m\}$ over Ω be given. We set $\mathcal{K}_1 = \{\Omega - \{a\} : a \in K_1\}$. It is obvious that $\mathcal{K}_1 = \{K_1\}^{-1}$.

Let us suppose that we have constructed $\mathcal{K}_q = \{K_1, \ldots, K_q\}^{-1}$ for $q < m$. We assume that $X_1, \ldots, X_t$ are the elements of $\mathcal{K}_q$ containing K_{q+1}. So $\mathcal{K}_q = F_q \cup \{X_1, \ldots, X_t\}$, where $F_q = \{B \in \mathcal{K}_q : K_{q+1} \not\subseteq B\}$.

For all i $(i = 1,\dots,t_q)$ we construct the antikeys of $\{K_{q+1}\}$ on X_i in the analogous way as $\mathcal{K}_1$, which are the maximal subsets of X not containing K_{q+1}. We denote them by $B^i_1,\dots,B^i_{r_i}$ $(i = 1,\dots,t)$. Let $\mathcal{K}_{q+1} = F_q \cup \{B^i_p : B \in F_q \to B^i_p \not\subset B,\ 1 \le i \le t,\ 1 \le p \le r_i\}$.

Theorem 4.1. *For every* q $(1 \le q \le m)$

$$\mathcal{K}_q = \{K_1,\dots,K_q\}^{-1}$$

i.e.

$$\mathcal{K}_m = \mathcal{K}^{-1}.$$

Proof. We have to prove that $\mathcal{K}_{q+1} = \{K_1,\dots,K_{q+1}\}^{-1}$. For this using the inductive hypothesis $\mathcal{K}_q = \{K_1,\dots,K_q\}^{-1}$, we show that

(a) If $B \in \mathcal{K}_{q+1}$ then B is the subset of Ω not containing K_t $(t = 1,\dots,q+1)$ and being maximal for this property, i.e. $B \in \{K_1,\dots,K_{q+1}\}^{-1}$.

(b) Every $B \subseteq \Omega$ not containing the elements K_t $(t = 1,\dots,q+1)$ and being maximal for this property is an element of $\mathcal{K}_{q+1}$.

Part (a): Let $B \in \mathcal{K}_{q+1}$. If $B \in F_q$ then B does not contain the elements K_t $(t = 1,\dots,q)$ and B is maximal for this property and at the same time $K_{q+1} \not\subseteq B$. Consequently, B is a maximal subset of Ω not containing K_t $(t = 1,\dots,q+1)$.

Let $B \in \mathcal{K}_{q+1} - F_q$. Clearly, there is a B^i_t $(1 \le i \le t_q,\ 1 \le t \le r_i)$ such that $B^i_t = B$. Our construction shows that $K_l \not\subseteq B^i_t$ $(l = 1,\dots,q+1)$. Since B^i_t is an antikey of K_{q+1} for X_i we obtain $B^i_t = X_i - \{b\}$ for some $b \in K_{q+1}$. Obviously $K_{q+1} \subseteq B^i_t \cup \{b\}$. If $a \in \Omega - X$, then by the inductive hypothesis, for $B^i_t \cup \{a,b\} = X_i \cup \{a\}$ there is a K_s $(1 \le s \le q)$ so that $K_s \subseteq B^i_t \cup \{a,b\}$. X_i does not contain $K_1,\dots,K_q$ by $X_i \in \mathcal{K}_q$. Hence, $a \in K_s$. If $K_s - \{a\} \subseteq B^i_t$, then $K_s \subseteq B^i_t \cup \{a\}$. For every K_s $(1 \le s \le q)$ with $K_s \subseteq X_i \cup \{a\}$ and $K_s \not\subseteq B^i_t$ we have $b \in K_s$. Hence, $K_s - \{a,b\} \subseteq B^i_t$. Consequently, there is a $B_1 \in F_q$ so that $B_t^i \subset B_1$. This contradicts $B \in \mathcal{K}_{q+1} - F_q$. So there is a K_s $(1 \le s \le q)$ such that $K_s \subseteq B_t^i \cup \{a\}$.

Part (b): Suppose that B is a maximal subset of Ω not containing K_t $(t = 1, \ldots, q+1)$. By the inductive hypothesis, there is a $Y \in \mathcal{K}_q$ such that $B \subseteq Y$.

The first case: If $K_{q+1} \not\subseteq Y$ then Y does not contain $K_1, \ldots$, K_{q+1}. B is a maximal subset of Ω not containing K_t $(t = 1, \ldots, q+1)$, so we have $B = Y$. $K_{q+1} \not\subseteq Y$ implies $B \in F_q$. Thus, $B \in \mathcal{K}_{q+1}$.

The second case: If $K_{q+1} \subseteq Y$ then $Y = X_i$ for some i $(1 \le i \le t)$ and $B \subseteq B_t^i$ holds for some t $(1 \le t \le r_i)$. If there is a $B_1 \in F_q$ such that $B_t^i \subset B_1$, then we also have $B \subset B_i$. By the definition of F_q it is clear that B_1 does not contain $K_1, \ldots, K_{q+1}$. This contradicts the definition of B. Hence, $B_t^i \in \mathcal{K}_{q+1}$ holds. It is easy to see that $B_t^i = B$ holds. Thus, $\mathcal{K}_{q+1} = \{K_1, \ldots, K_{q+1}\}^{-1}$. The theorem is proved. □

Let $\mathcal{K}_0 = \Omega$. We have $\mathcal{K}_q = F_q \cup \{X_1, \ldots, X_t\}$, where $1 \le q \le m-1$. Denote by l_q the number of elements of $\mathcal{K}_q$. When constructing $\mathcal{K}_{q+1}$, the worst-case time is $O(n^2(l_q - t_q)t_q)$ if $t_q < l_q$ and $O(n^2 t_q)$ if $l_q = t_q$. For the total time we derive

$$O(n^2 \sum_{q=0}^{m-1} t_q u_q),$$

where

$$u_q = \begin{cases} l_q - t_q & \text{if } l_q > t_q, \\ 1 & \text{otherwise.} \end{cases}$$

In [19] it is proved that the worst-case time of our algorithm is exponential not only in $|\Omega|$ but also in the number of elements of $\mathcal{K}$.

It can be seen that if there are only a few minimal keys (i.e. $\mathcal{K}$ is a small set) this algorithm is very effective, it runs in time polynomial in $|\Omega|$. When $l_q \le l_m$ holds $(q = 1, \ldots, m-1)$ then the number of elementary steps at most $O(n^2 \cdot |\mathcal{K}| \cdot |\mathcal{K}^{-1}|^2)$.

Thus, in these cases our algorithm finds $\mathcal{K}^{-1}$ in polynomial time in $|\Omega|$, $|\mathcal{K}|$ and $|\mathcal{K}^{-1}|$.

Now we give an algorithm that determines a relation representing a given Sperner-system.

ALGORITHM 4.2. *Input*: $\mathcal{K} = \{K_1, \ldots, K_m\}$, a Sperner-system over Ω.

Output: A relation R representing $\mathcal{K}$.

Step 1: Based on the above algorithm, we construct $\mathcal{K}^{-1}$.

Step 2: Let $\mathcal{K}^{-1} = \{B_1, \ldots, B_t\}$ be the set of antikeys of $\mathcal{K}$. Let $R = \{h_0, h_1, \ldots, h_t\}$ be a relation over Ω given as follows: for all $a \in \Omega$, $h_0(a) = 0$,

for i $(1 \le i \le t)$, $h_i(a) = \begin{cases} 0 & \text{if } a \in B_i, \\ 1 & \text{otherwise.} \end{cases}$

By Theorem 3.5 it is clear that R represents $\mathcal{K}$.

It is easy to see that the time complexity of this algorithm is the same as that of the algorithm which finds the set of antikeys.

Theorem 4.3. *The time complexity of finding a relation representing a given Sperner-system $\mathcal{K}$ is exactly exponential in the number of elements of $\mathcal{K}$.*

Proof.(1) We have Algorithm 4.2 that determines a relation representing a given Sperner-system $\mathcal{K}$, and whose worst-case time is exponential, not only in $|\Omega|$ but also in $|\mathcal{K}|$.

Now we have to prove:

(2) There is no algorithm that finds a relation representing K in time subexponential in the number of elements of $\mathcal{K}$.

To this end we give a construction. For the sake of simplicity, we assume that $n = |\Omega|$ is divisible by 3. Let $\Omega = X_1 \cup \ldots \cup X_m$ be a partition of Ω, where $m = n/3$, and $|X_i| = 3$ $(1 \le i \le m)$.

Let $\mathcal{K} = \{K : |K| = 2, K \subseteq X_i \text{ for some } i\}$. It is easy to see that $\mathcal{K}^{-1} = \{B : |B \cap X_i| = 1, \forall i\}$.

Clearly, $3^m = |\mathcal{K}^{-1}|$ holds. It is clear that $n = |\mathcal{K}|$. From

Theorem 3.6 we obtain

$$\sqrt{2 \times 3^m} \leq s(\mathcal{K}),$$

i.e. the number of rows of a minimal relation representing K 's at least $\sqrt{2 \times 3^m}$. Thus, we can always construct an example, in which the cardinality of K is not greater than $|\Omega|$, but the number of rows of any relation representing $\mathcal{K}$ is exponential in $|\mathcal{K}|$. From (1) and (2) the proof is complete. □

REMARK 4.4. In [4] it is proved that the complexity of finding an Armstrong relation for a given relational scheme is exponential in the number of attributes for all inputs.

On the contrary, Algorithm 4.2 runs in polynomial time if $|\mathcal{K}|$ is small.

As a closing remark, we mention that the complexity of finding the set of all minimal keys of a given relation R is exponential in the size of R. On the one hand, one can test all the subsets of columns in exponential time. On the other hand, we can construct a relation R such that the output (the set of minimal keys) has exponential size as follows.

R has $n+1$ rows and $2n$ columns. For $1 \leq i \leq n$ the i-th row has i in columns $2i-1$ and $2i$. All other entries are 0. It is easy to see that $x \subseteq \{1, \ldots, 2n\}$ is a minimal key iff it contains *exactly one* of the numbers $2i-1$, $2i$ for $1 \leq i \leq n$. We conclude that the number of minimal keys is 2^n.

References

[1] AHO, A. V., HOPCROFT, J. E. and ULLMAN J. D., *The Design and Analysis of Computer Algorithms*. Addison-Wesley Publ. Co., Reading MA, 1974.

[2] ARMSTRONG, W. W., Dependency structures of data base relationships. In: *Inf. Proces. 74*. North-Holland Publ. Co., Amsterdam-London, 1974, 580-583.

[3] BEERI, C. and BERNSTEIN, P. A., Computational problems related to the design of normal form relational schemes. *ACM Trans. Database Syst.* **4** (1) (1979) 30-59.

[4] BEERI, C., DOWN, M., FAGIN, R. and STATMAN, R., On the structure of Armstrong relations for functional dependencies. *J. ACM* **31** (1) (1984) 30-46.

[5] BÉKÉSSY, A. and DEMETROVICS, J., Contribution to the theory of data base relations. *Discrete Mathematics* **27** (1979) 1-10.

[6] BERNSTEIN, P. A., Normalization and functional dependencies in the relational data base model. Ph.D. Dissertation, Univ. of Toronto, 1975.

[7] CODD, E. F., A relational model for large shared data banks. *Comm. ACM* **13** (1970) 377-387.

[8] DELOBEL, C. and CASEY, R. G., Decomposition of a data base and the theory of Boolean switching functions. *IBM J. Res. Develop.* **17** (5) (1973) 374-386.

[9] DEMETROVICS, J., On the equivalence of candidate keys with Sperner-systems. *Acta Cybernetica* **4** (3) (1979) 247-252.

[10] DEMETROVICS, J. and GYEPESI, GY., On the functional dependency and some generalizations of it. *Acta Cybernetica* **5** (3) (1981) 295-305.

[11] DEMETROVICS, J., FÜREDI, Z. and KATONA, G. O. H., Minimal matrix representation of closure operations. *Discrete Applied. Math.* **11** (1985) 115-128.

[12] DEMETROVICS, J. and THI, V. D., Relations and minimal keys. *Acta Cybernetica* **8** (3) (1988) 279-285.

[13] FADOUS, R. Y., Mathematical foundations for relational data bases. Ph.D. Dissertation, Computer Science Dept. of Michigan Univ., 1975.

[14] FAGIN, R., Armstrong databases. In: *Proc. 7th IBM Symp. on Mathematical Foundations of Computer Science* (Kanagawa, Japan, 1982).

[15] FERNANDEZ, M. C., Determining the normalization level of a relation on the basis of Armstrong's axioms. *Computer and Artif. Intell.* (Czechoslovakia) **3** (1984) 495-504.

[16] LUCCHESI, C. L. and OSBORN, S. L., Candidate keys for relations. *J. Comput. Syst. Sciences* **17** (2) (1978) 270-279.

[17] MAIER, D., Minimum covers in the relational database model *J. ACM* **27** (1980) 664-674.

[18] SPERNER, E., Ein Satz über Untermengen einer endlichen Menge. *Math. Zeit.* **27** (1928) 544-548.

[19] THI, V. D., Minimal keys and antikeys. *Acta Cybernetica* **7** (4) (1986) 361-371.

[20] YU, C. T. and JOHNSON, D. T., On the complexity of finding the set of candidate keys for a given set of functional dependencies. *Inf. Proces. Lett.* **5** (4) (1976) 100-101.

(Received December 3, 1987)

JÁNOS DEMETROVICS V. D. THI

Computer and Automation Institute,
Hungarian Academy of Sciences
H-1132 Budapest, Victor Hugo u. 18-22.

HUNGARY