Acta Cybernetica, Vol. 11, No. 3, Szeged, 1994

Normal Forms and Minimal Keys in the Relational Datamodel*

J. Demetrovics †   Vu Duc Thi‡

Abstract

The normalization of relations was introduced by E. F. Codd. The main purpose of normalization is to delete undesired redundancy and anormalies. The most desirable normal forms are second normal form (2NF), third normal form (3NF) and Boyce-Codd normal form (BCNF) that have been investigated in a lot of papers. The concepts of minimal key and prime attribute (recall that an attribute is prime if it belongs to a minimal key, and nonprime otherwise) directly concern 2NF, 3NF and BCNF. This paper investigates connections between these normal forms and sets of minimal keys. Lucchesi and Osborn showed [11] that the problem to decide if an arbitrary attribute is prime is NP-complete for relation scheme. We proved [9] that a set of all nonprime attributes is the intersection of all antikeys (maximal nonkeys) and this prime attribute problem can be solved by polynomial time algorithm for relation. From these results some problems are NP-complete for relation scheme, but for relation these problems are solved by polynomial time algorithms. It is known [5] that a set of all minimal keys of a relation scheme (and a relation) is a Sperner system (sometimes it is called an antichain) and for an arbitrary Sperner system there exists a relation scheme the set of all minimal keys of which is exactly this Sperner system. In this paper the following concepts are introduced.

A Sperner system K is in 2NF (3NF, BCNF, respectively) if for each relation scheme s such that $K_s = K$ then s is in 2NF (3NF, BCNF, respectively), where K_s is a set of all minimal keys of s. This paper gives necessary and sufficient conditions for an arbitrary Sperner system is in 2NF or 3NF or BCNF. We prove that problems of deciding whether K_s is in 2NF (3NF, respectively) are NP-complete. However, we show that if a relation scheme is changed to a relation then these problems are solved by polynomial time algorithms. We give a new characterization of relations and relation schemes that are uniquely determined by their minimal keys. From this characterization we give a polynomial time algorithm deciding whether an arbitrary

*Research supported by Hungarian Foundation for Scientific Research Grant 2575.

†Computer and Automation Institute Hungarian Academy of Sciences P.O.Box 63, Budapest, Hungary, H-1502

‡Centre for Systems and Management Research National Centre for Scientific Research of Vietnam P.O.B.626 Boho, Hanoi 10000 Vietnam

relation is uniquely determined by its set of all minimal keys. Osborn [14] gives a polynomial time algorithm testing BCNF property of a given relation scheme. This paper gives a polynomial time algorithm recognizing BCNF and finding a set of all minimal keys and a minimum cover if a given relation scheme is in BCNF.

Key Words and Phrases: database, relation, relational datamodel, functional dependency, relation scheme, second normal form, third normal form, Boyce-Codd normal form, closure, closed set, minimal generator, key, minimal key, antikey.

1 Introduction

Let us give some necessary definitions and results that are used in next section.

Definition 1.1 *Let $R = \{a_1, \ldots, a_n\}$ be a nonempty finite set of attributes, $r = \{h_1, \ldots, h_m\}$ be a relation over R, and $A, B \subseteq R$. Then we say that B functionally depends on A in r (denoted $A \xrightarrow[r]{f} B$) iff*

$$(\forall h_i, h_j \in r)((\forall a \in A)(h_i(a) = h_j(a)) \Longrightarrow (\forall b \in B)(h_i(b) = h_j(b))).$$

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

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

Clearly, F_r is a set of all FDs that hold in r.

Definition 1.3 *Let R be a nonempty finite set, and denote $P(R)$ its power set. Let $y \subseteq P(R) \times P(R)$. We say that y is an f-family over R iff for all $A, B, C, D \subseteq R$*

1. $(A, A) \in y$,
2. $(A, B) \in y$, $(B, C) \in y \Longrightarrow (A, C) \in y$,
3. $(A, B) \in y$, $A \subseteq C$, $D \subseteq B \Longrightarrow (C, D) \in y$,
4. $(A, B) \in y$, $(C, D) \in y \Longrightarrow (A \cup C, B \cup D) \in y$. *Clearly, F_r is an f-family over R.*

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

Definition 1.4 *A relation scheme s is a pair $< R, F >$, where R is a set of attributes, and F is a set of FDs over R. Let F^+ be a set of all FDs that can be derived from F by the rules in Definition 1.3. 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 $B \subseteq A^+$.*

It is known [3] that there is a polynomial time algorithm which finds A^+ from A.

Clearly, if $s =< R, F >$ 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. It is obvious that all FDs of s hold in r.

Definition 1.5 *Let r be a relation, $s =< R, F >$ be a relation scheme, y be an f-family over R and $A \subseteq R$. Then A is a key of r (a key of s, a key of y) if $A \underset{r}{\xrightarrow{f}} R$ $(A \to R \in F^+, (A, R) \in y)$. 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 R.*

Definition 1.6 *Let K be a Sperner system over R. We define the set of antikeys of K, denoted by K^{-1}, as follows:*

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

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 requires at most $|R|$ elementary steps.

Definition 1.7 *Let $I \subseteq P(R)$, $R \in I$, and $A, B \in I \Longrightarrow A \cap B \in I$. Let $M \subseteq P(R)$. Denote $M^+ = \{\cap M' : M' \subseteq M\}$. We say that M is a generator of I iff $M^+ = I$. Note that $R \in M^+$ 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 R, and E_r the equality set of r, i.e. $E_r = \{E_{ij} : 1 \leq i < j \leq |r|\}$, where $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$. Let $T_r = \{A \in P(R) : \exists E_{ij} = A, \nexists E_{pq} : A \subset E_{pq}\}$. 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 R. We say that r represents K iff $K_r = K$. The following theorem is known ([8]).*

Definition 1.10 *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.*

Definition 1.11 *Let $s =< R, F >$ be a relation scheme over R. We say that an attribute a is prime if it belongs to a minimal key of s, and nonprime otherwise. $s =< R, F >$ is in*

1. *2NF if $A \to \{a\} \notin F^+$ for each $K \in K_s$, $A \subset K$, $a \notin A$, and a is nonprime.*

2. 3NF if $A \to \{a\} \notin F^+$ *for* $A^+ \neq R$, $a \notin A$, *a is nonprime.*

3. BCNF if $A \to \{a\} \notin F^+$ *for* $A^+ \neq R$, $a \notin A$.

Clearly, if s is in BCNF (3NF, respectively) then s is in 3NF (2NF, respectively).
If a relation scheme is changed to a relation we have the definition of 2NF, 3NF and BCNF for relation.

Definition 1.12 [4] *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* $H_F(A) = \{a \in R: (A, \{a\}) \in F\}$.

Theorem 1.13 [7] *Let K be a Sperner system over R. Let*

$$L(A) = \begin{cases} \bigcap_{A \subseteq B} & B \\ if \exists B \in K^{-1} : A \subseteq B & R \text{ otherwise} \end{cases}$$

and $F = \{(C, D) : D \subseteq L(C)\}$.
Then F is an f-family over $R, H_F = L$, *and* $K_F = K$. *If F' is an arbitrary f-family over R such that* $K_{F'} = K$ *then* $F \leq F'$ *holds.*

2 Results

In this section we give some results related to 2NF, 3NF, BCNF and sets of minimal keys.

Definition 2.1 *Let K be a Sperner system over R. We say that K is in 2NF (3NF, BCNF, respectively) if for every relation scheme* $s = \langle R, F \rangle$ *such that* $K_s = K$ *then s is in 2NF (3NF, BCNF, respectively).*

Now we give a necessary and sufficient condition for an arbitrary Sperner system is in 2NF.
Let K be a Sperner system over R. Denote $K_p = \{a \in R: \exists A \in K: a \in A\}$, and $K_n = R - K_p$. $K_p\,(K_n)$ is called the set of prime (nonprime) attributes of K.
Given a relation scheme $s = \langle R, F \rangle$, we say that a functional dependency $A \to B \in F$ is redundant if either $A = B$ or there is $C \to D \in F$ such that $C \subsetneq A$.

Theorem 2.2 *Let K be a Sperner system over R. Then K is in 2NF if and only if* $K_n = \emptyset$.

Proof. According to definitions of 2NF relation, 2NF Sperner system and K_n we can see that if $K_n = \emptyset$ then K is in 2NF.
Now, assume that K is in 2NF. Denote K^{-1} the set of all antikeys of K. From K, K^{-1} we construct the following relation scheme.
For each $A \subset R$ there is $B \in K^{-1}$ such that $A \subseteq B$. Denote $C = \cap\{B \in K^{-1}: A \subseteq B\}$. We set $A \to C$. Denote T the set of all such functional dependencies. Set $F = \{E \to R: E \in K\} \cup (T - Q)$, where $Q = \{X \to Y \in T: X \to Y$ is a redundant functional dependency $\}$. From Theorem 1.13 and definition of Sperner system we obtain $K_s = K$. Clearly, for each arbitrary relation scheme $s' = \langle R, F' \rangle$ such that $K_{s'} = K$ and $A \subseteq R$ we have $A^+_{s'} \subseteq A^+_s$, where $A^+_{s'} = \{a: A \to \{a\} \in F'^+\}$. We showed [9] that K_n is the intersection of all antikeys of K. Based on the

construction of $s =< R, F >$ and according to definition of 2NF Sperner system we obtain $K_n = \emptyset$. Our proof is complete.

It is easy to see that a 3NF relation scheme is in 2NF and if a set of all non-prime of arbitrary relation scheme is empty then this relation scheme is in 3NF. Consequently, Theorem 2.2 immediately implies the following corollary.

Corollary 2.3 *Let K be a Sperner system over R. Then K is in 3NF if and only if $K_n = \emptyset$.*

Definition 2.4 *Let K be a Sperner system over R. We say that K is unique if K uniquely determines the relation scheme $s =< R, F >$, i. e. for every relation scheme $s' =< R, F' >$ such that $K_{s'} = K$ we have $F^+ = F'^+$.*

From definition of BCNF Sperner system and Definition 2.4 we obtain

Proposition 2.5 *K is in BCNF iff K is unique.*

Now we introduce the following problem.

Theorem 2.6 *The following problem is NP-complete:*

Given a relation scheme s, decide whether K_s is in 2NF.

Proof. For each $a \in R$ we nondeterministically choose a subset B of R such that $a \in B$. By an algorithm finding the closure of B over s (see [3]) and based on definition of minimal key we decide whether B is a minimal key of s. From this we can decide whether a is prime of s. According to Theorem 2.2 if for every $a \in R$ a is prime then K_s is in 2NF, in the converse case K isn't in 2NF. It is obvious that this algorithm is nondeterministic polynomial. Thus, our problem lies in NP.

Now we shall show that our problem is NP-hard. It is known [11] that the prime attribute problem for relation scheme is NP-complete. Now we prove that this problem is polynomially reducible to our problem.

Let $s' =< P, F' >$ be a relation scheme over P, and $a \in P$. Without loss of generality we assume that P is not a minimal key of s', i.e. if $A \in K_{s'}$ then $A \subset P$. By a polynomial time algorithm finding a minimal key of relation scheme (see [11]) we can find a minimal key Q of s' from P and F'. Denote $T = \{l: l \in P - Q, \{l\} \rightarrow P \notin F'^+\}$. Assume that $T = \{a_1, \ldots, a_t\}$. Now we construct the relation scheme $s =< R, F >$ as follows:

$R = P \cup \{b, c, d, e_1, \ldots, e_{t-1}\}$, where $b, c, d, e_1, \ldots, e_{t-1} \notin P$ and F contains F' and the following functional dependencies:

- $\{b\} \rightarrow \{a\}$,
- $\{c, d\} \rightarrow \{b\}$,
- $Q \cup \{c\} \rightarrow R$,
- $Q \cup \{b\} \rightarrow Q \cup \{c\}$,
- $\{\{a_i, a_{i+1}, e_i\} \rightarrow R: 1 \leq i \leq t-1\}$.

It can be seen that s is constructed in polynomial time in the sizes of P and F'. According to the construction of $s =< R, F >$ and definition of minimal key and by $Q \cup \{c\} \rightarrow R$, for all $A \in K_{s'}$ we have $A \cup \{c\} \in K_s(1)$. Based on $Q \cup \{b\} \rightarrow Q \cup \{c\}$ and $\{b\} \rightarrow \{a\}$ if $A \in K_{s'}$ we obtain $(A - a) \cup \{b\} \in K_s(2)$. By $\{\{a_i, a_{i+1}, e_i\} \rightarrow R: 1 \leq i \leq t-1\}$ we have $\{a_i, a_{i+1}, e_i\}(1 \leq i \leq t-1) \in K_s$. From this and (1) $\forall a' \in P, b, c, e_1, \ldots, e_{t-1}$ are prime attributes of s. According to the construction of s and definition of 3NF relation scheme we can see that s is in 3NF. Now we prove that K_s is in 2NF iff a is prime attribute of s'.

Assume that K_s is in 2NF. According to Theorem 2.2 we can see that d is prime attribute of s. Consequently, there is a minimal key B of s such that $d \in B$.

It can be seen that $a, b, e_1, \ldots, e_{t-1} \notin B$. Since there is only functional dependency $\{c, d\} \to \{b\}$ the left side of which contains d we obtain $c \in B$. According to $\{b\} \to \{a\}$, $\{c, d\} \to \{b\}$ and (2) it is easy to see that $(B \cup a) - \{c, d\} \in K_{s'}$. Thus, a is prime attribute of s'.

Now we assume that K_s is not in 2NF. By Theorem 2.2 d is nonprime attribute of s. If $a \in A : A \in K_{s'}$ then by $\{c, d\} \to \{a\} \in F^+$ and from (2) $\{c, d\} \cup (A - a) \in K_s$ holds. This conflicts with the fact that d is nonprime attribute of s. Consequently, a is nonprime attribute of s'. The theorem is proved.

Theorem 2.6 immediately implies the following corollary

Corollary 2.7 *The problem of deciding whether K_s is in 3NF is NP-complete for given a relation scheme s.*

It is known [8] that there is a polynomial time algorithm which from a given relation r finds the maximal equality system T_r. Based on Theorem 1.10 and because the set of all nonprime attributes is the intersection of all antikeys we have the following proposition.

Proposition 2.8 *There is an algorithm that for a given relation r decides if K_r is in 2NF or 3NF. The time complexity of this algorithm is polynomial in the sizes of R and r.*

From Theorem 2.2 we immediately obtain the following corollary.

Corollary 2.9 *There is a polynomial time algorithm that decides whether a given Sperner system is in 2NF or 3NF. Let $s =< R, F >$ be a relation scheme over R, K_s is a set of all minimal keys of s. Denote K_s^{-1} the set of all antikeys of s. From Theorem 1.10 we obtain the following corollary.*

Corollary 2.10 *Let $s =< R, F >$ 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 [7] we proved the following theorem.*

Theorem 2.11 *Let $r = \{h_1, \ldots, h_m\}$ be a relation , and F an f-family over R. Then $F_r = F$ iff for every $A \in P(R)$*

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

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

Let $s =< R, F >$ be a relation scheme over R. From s we construct $Z(s) = \{X^+ : X \subseteq R\}$, 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 2.10 and Theorem 2.11 the following proposition is clear

Proposition 2.12 *Let $s =< R, F >$ be a relation scheme over R. Then*

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

It is known [5] that for given a Sperner system K there exists a relation scheme s (a relation r, respectively) such that $K_s = K$ ($K_r = K$, respectively). We say that s (r, respectively) is unique if K_s (K_r, respectively) uniquely determines s (r, respectively), i.e. K_s (K_r, respectively) is unique.

Now we give a necessary and sufficient condition for given a relation scheme is unique.

Theorem 2.13 *Let $s =< R, F >$ be a relation scheme over R. Then s is unique iff for all $a \in A$, $A \in K_s^{-1}$: $A - a = \cap\{B \in K_s^{-1}: (A - a) \subset B\}$ holds.*

Proof. It is known [4] that a Sperner system K is unique iff for all $B \subseteq A$, $A \in K^{-1}$, B is an intersection of antikeys. Denote $P_s = \{A - a: A \in K_s^{-1}, a \in A\}$.

It can be seen that if $s =< R, F >$ is unique then $B \in P_s$ implies B is an intersection of antikeys, i.e. $B = \cap\{A \in K_s^{-1}: B \subseteq A\}$.

Conversely, assume that for every $B \in P_s$ we have $B = \cap\{A \in K_s^{-1} : B \subseteq A\}(*)$. Now we shall prove the following result : $s =< R, F >$ is in BCNF iff for all $B \in P_s$, $B^+ = B(1)$ holds.

It is easy to see that if s is in BCNF then we obtain (1). Now, we assume that for each $B \in P_s$, $B^+ = B$. Suppose $C \to \{d\} \in F^+$ and $d \notin C(2)$. If $C^+ \neq R$ then by definition of antikey and Proposition 2.12 there exists an $A \in K_s^{-1}$ such that $C^+ \subseteq A$ and by (2) $d \in A$ holds. Clearly, $C \subseteq A - d$ holds. It is easy to see that $(A - d)^+ \to \{d\}$ holds. By $A - d \in P_s$ we have $(A - d)^+ \neq A - d$. This conflicts with the fact that $(A - d)^+ = A - d$. Hence, $C^+ = R$ holds, i.e. s is in BCNF.

From this result and according to Proposition 2.12 we have $N_s \subseteq (P_s \cup K_s^{-1})$. It can be seen that s is in BCNF. Based on definition of N_s and Proposition 2.12 $K_s^{-1} \subseteq N_s$ holds. According to (*) we obtain $K_s^{-1} = N_s$. Because s is in BCNF we can see that for all $B \subseteq A$, $A \in K_s^{-1}$: $B^+ = B$ holds. Thus, B is an intersection of antikeys of s. The proof is complete.

According to definition of BCNF Sperner system and based on Theorem 2.13 and Proposition 2.5 we give a necessary and sufficient condition for an arbitrary Sperner system is in BCNF.

Theorem 2.14 *Let K be a Sperner system over R. Then K is in BCNF iff for all $a \in A$, $A \in K^{-1}$: $A - a = \cap\{B \in K^{-1}: (A - a) \subset B\}$ holds.*

By a polynomial time algorithm finding a set of all antikeys of a given relation and according to Theorem 2.13 we obtain the following proposition.

Proposition 2.15 *There exists an algorithm deciding whether a given relation r is unique. The time complexity of this algorithm is polynomial in the sizes of R and r.*

Theorem 2.14 and Proposition 2.15 immediately imply the following

Proposition 2.16 *There exists a polynomial time algorithm deciding whether a set of all minimal keys of a given relation is in BCNF.*

Theorem 2.13 immediately implies the next corolllary.

Corollary 2.17 *Let K be a Sperner system over R. Then there exists a polynomial time algorithm deciding whether a Sperner system H is unique, where $H^{-1} = K$.*

Now we introduce the following problems : Given a relation scheme s (a Sperner system K, respectively), decide whether s (K, respectively) is unique.

It is obvious that these problems are equivalent to the next problems: Given a relation scheme s (a Sperner system K, respectively), decide whether K_s (K, respectively) is in BCNF.

It is unknown that these problems have polynomial time complexity. We consider these problem as interesting open problems.

Osborn [14] gives a polynomial time algorithm deciding whether a relation scheme is in BCNF. It is known [10, 12] that a relation scheme $s =< R, F >$ is in BCNF iff its minimum cover contains functional dependencies $\{K_1 \to R, \ldots, K_t \to$

$R\}$, where $K_i(1 \leq i \leq t)$ are minimal keys of s. From this the BCNF property of relation scheme also is recognized in polynomial time.

Let $s =< R, F >$ be a relation scheme over R. From rules (3) and (4) of Definition 1.3 we can see that the functional dependency $A \rightarrow \{a_1, \ldots, a_t\}$ is equivalent to the set of functional dependencies $\{A \rightarrow \{a_1\}, \ldots, A \rightarrow \{a_t\}\}$. Thus, we can assume that F only contains the functional dependencies form $A \rightarrow \{a\}$.

Definition 2.18 *Let $s =< R, F >$ be a relation scheme. We say that s is an a-relation scheme over R if $F = \{A \rightarrow \{b\}: A \neq b, \not\exists B : (B \rightarrow \{b\})(B \subset A)\}$, where $b \in R$.*

Definition 2.19 *Let $s =< R, F >$ be a relation scheme, $b \in R$. Denote $K_b = \{A \subseteq R: A \rightarrow \{b\}, \not\exists B: (B \rightarrow \{b\})(B \subset A)\}$. K_b is called the family of minimal sets of the attribute b. Clearly, $R \notin K_b$, $\{b\} \in K_b$ and K_b is a Sperner system over R.*

Algorithm 2.20 *(Finding a minimal set of the attribute b)*

Input: Let $s =< R, F >$ be a relation scheme, $A = \{a_1, \ldots, a_t\} \rightarrow \{b\}$.

Output: $A' \in K_b$

Step 0: We set $L(0) = A$

Step i+1: Set

$$L(i+1) = \begin{cases} L(i) - a_{i+1} & \text{if } L(i) - a_{i+1} \\ \rightarrow \{b\} & L(i) \text{ otherwise.} \end{cases}$$

Then we set $A' = L(t)$.

Lemma 2.21 $L(t) \in K_b$

Proof. By the induction it can be seen that $L(t) \rightarrow \{b\}$, and $L(t) \subseteq \ldots \subseteq L(0)$ (1). If $L(t) = b$, then by the definition of the minimal set of attribute b we obtain $L(t) \in K_b$. Now we suppose that there is a B such that $B \subset L(t)$ and $B \neq \emptyset$. Thus, there exists a_j such that $a_j \notin B$, $a_j \in L(t)$. According to the construction of algorithm we have $L(j-1) - a_j \not\rightarrow \{b\}$. It is obvious that by (1) we obtain $L(t) - a_j \subseteq L(j-1) - a_j(2)$. It is clear that $B \subseteq L(t) - a_j$. From (1),(2) we have $B \not\rightarrow \{b\}$. The lemma is proved.

Clearly, by the linear-time membership algorithm in [3] the time complexity of Algorithm 2.20 is $O(|R|^2|F|)$.

Algorithm 2.22 *(Finding an a-relation scheme)*

Input: Let $s =< R, F >$ be a relation scheme.

Output: an a-relation scheme $s' =< R, F' >$ such that $F'^+ = F^+$.

Step 1: By rules (3) and (4) of Definition 1.3 from s we construct $s" =< R, F" = \{A \rightarrow \{b\} : b \in R\} >$ such that $F"^+ = F^+$.

Step 2: For each $A \rightarrow \{b\} \in F"$ we use algorithm 2.20 to find a minimal set A' of attribute b over $s"$. Set $F^* = \{A' \rightarrow b : \forall b \in R\}$.

Step 3: Set $s' =< R, F' = F^* - Q) >$, where $Q = \{X \rightarrow Y \in F^*: X \rightarrow Y$ is a redundant functional dependency $\}$.

According to definition of a-relation scheme, based on Definition 2.19 and Lemma 2.21 we can see that s' is an a-relation scheme and $F'^+ = F^+$.

It can be seen that the time complexity of Algorithm 2.22 is polynomial in the sizez of R and F.

Theorem 2.23 *Let $s =< R, F >$ be a relation scheme. Then s is in BCNF if and only if there exists an a-relation scheme $s' =< R, F' >$ such that $F'^+ = F^+$ and for every $A \to \{b\} \in F'$ $A \in K_{s'}$ holds.*

Proof. Assume that s is in BCNF. By Algorithm 2.22 we can construct an a-relation scheme $s' =< R, F' >$ such that $F'^+ = F^+$. By Step 3 of this algorithm for each $A \to \{b\} \in F'$ $b \notin A$ holds. Since s' is in BCNF we have $A^+ = R$. Clearly, if there is a $C \subset A$ such that $C^+ = R$ then $C \to \{b\}$ holds. This is a contradiction. Thus, $A \in K_{s'}$ holds.

Conversely, we assume that there is an a-relation scheme $s' =< R, F' >$ such that $F^+ = F'^+$ and for every $A \to \{b\} \in F'$ $A \in K_{s'}$ holds. By Lemma 3 in [14] s' is in BCNF. Thus, s is in BCNF. Our theorem is proved.

In Theorem 2.23 we set $K = \{A: A \to \{b\} \in F'\}$. We have the following.

Proposition 2.24 $K = K_s$.

Proof. By definition of BCNF relation scheme $K_{s'} = K_s$ holds. From Theorem 2.23 $K \subseteq K_{s'}$ holds. Suppose $B \in K_{s'}, B \subset R$ and $B \notin K$. Because $K_{s'}$ is a Sperner system over R we can see that $K \cup B$ also is a Sperner system over R. It can be seen that according to definition of a-relation scheme $B^+ = B$ over s'. This conflicts with the fact that B is a minimal key of s'. The proof is complete.

Theorem 2.23 immediately implies the following.

Proposition 2.25 *Let $s =< R, F >$ be a relation scheme. Then s is in BCNF if and only if there exists an a-relation scheme $s' =< R, F' >$ such that $F'^+ = F^+$ and for every $A \to \{b\} \in F'$ A is a key of s'.*

It can be seen that based on definition of a-relation scheme, in Proposition 2.25 if $A \to \{b\} \in F'$ then A is a minimal key of s'.

Clearly, the time complexity of Algorithm 2.22 (finding an a-relation scheme) is polynomial and deciding whether a set of attributes is a key also takes polynomial time, It is known [10, 12] that a relation scheme $s =< R, F >$ is in BCNF iff its minimum cover contains functional dependencies $\{K_1 \to R, \ldots, K_t \to R\}$, where $K_i (1 \leq i \leq t)$ are minimal keys of s. We can give a polynomial time algorithm recognizing the BCNF property of arbitrary relation scheme s, and if relation scheme s is in BCNF then this algorithm finds a minimum cover and a set of all minimal keys of s.

Algorithm 2.26 *Input: Let $s =< R, F >$ be a relation scheme.*

Output: Deciding whether s is in BCNF, if s is in BCNF then finding K_s, and an a-relation scheme $s' =< R, F' >$ such that s' is a minimum cover of s.

Step 1: Use Algorithm 2.22 we construct an a-relation scheme $s" =< R, F" = \{A \to \{b\} : b \in R\} >$ such that $F"^+ = F^+$.

Step 2: If there is an $A \to \{b\} \in F"$ such that A is not a key of $s"$ then s isn't in BCNF and stop. In the converse case go to the following step.

Step 3: Set $K_s = \{A: A \to \{b\} \in F"\}$.

Step 4: Denote elements of K_s by $A_1, \ldots, A_t$. Set $F' = \{A_i \to R : 1 \leq i \leq t\}$. It can be seen that $s' =< R, F' >$ is a minimum cover of s.

3 Conclusion

Our further research will be devoted to the following problems:
Given a relation scheme s.

1. What is the time complexity of deciding whether s is in unique?

Given a Sperner system K over R.

1. What is the time complexity of deciding whether K is unique?

Acknowledgments

The authors are grateful to Dr. Uhrin Bela for useful comments to the first version of the manuscript.

References

[1] Armstrong W.W. Dependency Structures of Database Relationships. *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 (1984) pp. 30-46.

[3] Beeri C., Bernstein P.A. Computational problems related to the design of normal form relational schemas. *ACM Trans. on Database Syst.* 4 (1979) pp. 30-59.

[4] Burosch G., Demetrovics J., Katona G.O.H. The poset of closures as a model of changing databases, *Order* 4 (1987) pp. 127-142.

[5] Demetrovics J. On the equivalence of candidate keys with Sperner systems. *Acta Cybernetica* 4 (1979) pp. 247-252.

[6] Demetrovics J. Logical and structural Investigation of Relational Datamodel (in Hungarian). *MTA-SZTAKI Tanulmányok,* Budapest, 114 (1980) pp. 1-97.

[7] Demetrovics J., Thi V.D. Some results about functional dependencies. *Acta Cybernetica* 8 (1988) pp. 273-278.

[8] Demetrovics J., Thi V.D. Relations and minimal keys. *Acta Cybernetica* 8 (1988) pp. 279-285.

[9] Demetrovics J., Thi V.D. On Keys in the Relational Datamodel. *EIK* 24 (1988) 10, pp. 515-519.

[10] Gottlob G. , Libkin L. Investigations on Armstrong relations, dependency inference, and excluded functional dependencies. *Acta Cybernetica* 9 (1990) pp. 385-402.

[11] Lucchesi C.L., Osborn S.L. Candidate keys for relations. *J.Comput.Syst. Scien.* 17 (1978) pp. 270-279.

[12] Maier D. Minimum cover in the relational database model. *J.ACM* 27 (1980) pp. 664-674.

[13] Mannila H., Raiha K.J. Design by Example: An Application of Armstrong relations. *J. Comput. Syst. Scien.* 33 (1986) pp. 126-141.

[14] Osborn S.L. Testing for existence of a covering Boyce-Codd normal form. *Inform. Proc. Lett.* 8 (1979) pp. 11-14.

[15] Thi V.D. Investigation on Combinatorial Characterzations Related to Functional Dependency in the Relational Datamodel (in Hungarian). *MTA-SZTAKI Tanulmányok,* Budapest, 191 (1986) pp. 1-157. Ph.D. Dissertation.

[16] Thi V.D. Minimal keys and Antikeys. *Acta Cybernetica* 7 (1986) pp. 361-371.

[17] Thi V.D. On Antikeys in the Relational Datamodel (in Hungarian). *Alkalmazott Matematikai Lapok* 12 (1986) pp. 111-124.

Received March 8, 1993