Acta Cybernetica, Vol. 11, No. 1-2, Szeged, 1993

Some problems concerning Armstrong relations of dual schemes and relation schemes in the relational datamodel*

J. Demetrovics[†] V. D. Thi[†]

Abstract

Several papers [3,5,6,7,8,9,11,12] have appeared for investigating dual dependency. The practical meaning of dual dependency was shown in [5,6]. In this paper we give some new results concerning dual dependency. The concept of dual scheme is introduced. Some characterizations of dual scheme, such as closure, generator, generating Armstrong relation, inferring dual dependencies, irredundant cover, normal cover are studied from different aspects. We give a characterization of Armstrong relations for a given dual scheme. We prove that the membership problem for dual dependencies is solved by a polynomial time algorithm. We show that the time complexity of finding an Armstrong relation of a given dual scheme is exponential in the number of attributes. Conversely, we give an algorithm to construct a dual scheme from a given relation R such that R is Armstrong relation of it. This paper gives some polynomial time algorithms which find closure, irredundant cover, normal cover from a given dual scheme.

In the second part of this paper we present some results related to Armstrong relations for functional dependency (FD for short) in Boyce-Codd normal form. The concepts of unique relation and unique relation scheme are introduced. We prove that deciding whether a given relation R over a set of attributes U is unique is solved by a polynomial time algorithm. We show some cases in which FD-relation equivalence problem is solved in polynomial time.

Key Words and Phrases: relation, relational datamodel, dual dependency, dual scheme, generating Armstrong relation, inferring dual dependencies, membership problem, closure, closed set, irredundant cover, normal cover, minimal generator, Boyce-Codd normal form.

1 Introduction

Now we give some necessary definitions that are used in next sections. The next sections present our new results.

*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

Definition 1.1 *Let $R = \{h_1, \ldots, h_m\}$ be a relation over U, and $A, B \subseteq U$. Then we say that B dually depends on A in R denoted $A \xrightarrow[R]{d} B$) iff*

$$(\forall h_i, h_j \in R)(\exists a \in A)(h_i(a) = h_j(a)) \Longrightarrow (\exists b \in B)(h_i(b) = h_j(b)))$$

Let $D_R = \{(A, B) : A, B \subseteq U,\ A \xrightarrow[R]{d} B\}$. D_R is called the full family of dual dependencies of R. Where we write (A, B) or $A \to B$ for $A \xrightarrow[R]{d} B$ when R, d are clear from the context.

Definition 1.2 *A dual dependency (DD) over U is a statement of the form $A \to B$,where $A, B \subseteq U$. The DD $A \to B$ holds in a relation R if $A \xrightarrow[R]{d} B$ We also say that R satisfies the DD $A \to B$.*

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 a d-family over U iff for all $A, B, C, D \subseteq U$*

(1) $(A, A) \in Y$,

(2) $(A, B) \in Y$, $(B, C) \in Y \Longrightarrow (A, C) \in Y$,

(3) $(A, B) \in Y$, $C \subseteq A$, $B \subseteq D \Longrightarrow (C, D) \in Y$,

(4) $(A, B) \in Y$, $(C, D) \in Y \Longrightarrow (AUC, BUD) \in Y$.

(5) $(A, \emptyset) \in Y \Longrightarrow A = \emptyset$.

Clearly, D_R is a d-family over U.

It is known [6,7] that if Y is an arbitrary d-family, then there is a relation R over U such that $D_R = Y$.

Definition 1.4 *A dual scheme P is a pair $< U, D >$, where U is a set of attributes, and D is a set of DDs over U. Let D^+ be a set of all DDs that can be derived from D by the rules in Definition 1.3. It is easy to see that D^+ is a d-family over U.*

Clearly, if $P =< U, D >$ is a dual scheme, then there is a relation R over U such that $D_R = D^+$ (. see, [6,7]). Such a relation is called an Armstrong relation of P.

In this paper we consider the comparision of two attributes as an elementary step of algorithms. Thus, if we assume that subsets of U are represented as sorted lists of attributes, then a Boolean operation on two subsets requires at most $|U|$ elementary steps.

Definition 1.5 *Let $I \subseteq P(U)$, $U \in I$, and $A, B \in I \Longrightarrow A \cap B \in I$. 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 necessarily in M, since it is the intersection of the empty collection of sets.*

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

It is proved [7] that N is the unique minimal generator of I. Thus, for any generator N' of I we obtain $N \subseteq N'$.

Definition 1.6 *Let D be a d-family over U, and $(A,B) \in D$. (A,B) is called a maximal left-side dependency of D if $\forall A' : A \subseteq A', (A',B) \in D \Longrightarrow A' = A$. Denote by $M(D)$ the set of all maximal left-side dependencies of D. Then A is called a maximal left-side of D if there existst a B such that $(A,B) \in M(D)$. Denote by $G(D)$ the set of all maximal left-sides of D.*

Definition 1.7 *Let $G \subseteq P(U)$. We say that G is a d-semilattice over U if $\emptyset, U \in G$, $A,B \in G \Longrightarrow A \cap B \in G$.*

Theorem 1.8 [6] *Let D be a d-family over U. Then $G(D)$ is a d-semilattice over U. Conversely, if G is a d-semilattice over U, then there exists exactly one d-family D such that $G(D) = G$, where $D = \{(A,B) : \forall C \in G : A \not\subseteq C \Longrightarrow B \not\subseteq C\}$.*

Theorem 1.9 *Let K be a Sperner system over U. We define the set of antikeys of K, denoted by K^{-1}, as follows:*

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

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

2 Dual schemes

Definition 2.1 *Let R be a relation over U. Set $N_{ij} = \{a \in U : h_i(a) \neq h_j(a)\}$, and $N_R = \{N_{ij} : 1 \leq i < j \leq |R|\}$. Then N_R is called the non-equality system of R.*

According to definition of relation $\emptyset \notin N_R$.

Let $P =< U, D >$ a dual scheme over U. Then D^+ is a d-family over U, $G(D^+)$ is the set of all maximal left-sides of D^+. Clearly, $G(D^+)$ is a d-semilattice over U. Denote by $N(D^+)$ the minimal generator of $G(D^+)$.

Now we present a characterization of Armstrong relations for a given dual scheme.

Theorem 2.2 *Let $P =< U, D >$ be a dual scheme, R be a relation over U. Then R is an Armstrong relation of P if and only if $N(D^+) \subseteq N_R \cup \{\emptyset\} \subseteq G(D^+)$.*

Proof: ($\Longrightarrow$): We assume that R is an Armstrong relation of P, i.e. $D_R = D^+$. According to Theorem 1.8 we obtain $G(D_R) = G(D^+)$. Now we prove that for an arbitrary relation R $G(D_R) = (N_R - U)^+ \cup \{\emptyset\}$ holds. Because $G(D_R)$ is a d-family over U, we have $\emptyset, U \in G(D_R)$. Clearly, $U \in (N_R - U)^+$. It is obvious that $\forall N_{ij} \neq \emptyset$. We suppose that $N_{ij} \neq U$.Because for any $a \in U - N_{ij}$ we obtain $h_i(a) = h_j(a)$, but $\forall b \in N_{ij}$:$h_i(b) \neq h_j(b)$,i.e. $\{a\} \cup N_{ij} \xrightarrow[R]{d} N_{ij}$. Hence, $N_{ij} \in G(D_R)$, holds. Consequently, $N_R \subseteq G(D_R)$. Thus, we obtain $(N_R - U)^+ \cup \{\emptyset\} \subseteq G(D_R)$. Conversely, if $A \in G(D_R) - \{\emptyset, U\}$, then if we suppose that for all $h_i, h_j \in R$ then there is $a \in A$ such that $h_i(a) = h_j(a)$. So $U \xrightarrow[R]{d} A$ which contradicts the definition of A. Consequently, there is an index pair (i,j) such that $A \subseteq N_{ij}$. We set $T = \{N_{ij} : A \subseteq N_{ij}\}$. If there exists an $N_{ij} : A = N_{ij}$ then $A \in N_R$. In the converse case we set $B = \bigcap_{N_{ij} \in T} N_{ij}$. If $A \subset B$ then for all $N_{ij} \in T$ we have $A \subset N_{ij}$.So

$B \xrightarrow[R]{d} A$ which contradicts $A \in G(D_R) - \{\emptyset, U\}$.Consequently, we obtain $A = B$. Hence, $A \in (N_R - U)^+ \cup \{\emptyset\}$ holds. Thus, $G(D_R) = (N_R - U)^+ \cup \{\emptyset\}$ holds. Consequently, we have $G(D^+) = (N_R - U)^+ \cup \{\emptyset\}$. According to definition of minimal generator we obtain $N(D^+) \subseteq N_R \cup \{\emptyset\} \subseteq G(D^+)$.

($\Longleftarrow$):From $N(D^+) \subseteq N_R \cup \{\emptyset\} \subseteq G(D^+)$ we have $G(D^+) = (N_R - U)^+ \cup \{\emptyset\}$. According to above part of proof we obtain $G(D_R) = G(D^+)$. By Theorem 1.8 R is an Armstrong relation of P. The theorem is proved.

Let $P =< U, D >$ be a dual scheme. We set $H_P(A) = \{a \in U : \{a\} \to A \in D^+\}$. Let $Z(P) = \{A \in P(U) : H_P(A) = A\}$. It is easy to see that $Z(P) = G(D^+)$. Clearly, for all $A \in P(U) : A \subseteq H_P(A) = H_P(H_P(A))$ and $A \subseteq B \Longrightarrow H_P(A) \subseteq H_P(B)$.

Algorithm 2.3 *(Compute $H_P(A)$)*

Input: $P =< U, D = \{A_i \to B_i : i = 1, \dots, m\} >$ a dual scheme over U, $A \in P(U)$.

Output: $H_P(A)$

Step 1: We set $A(0) = A$.

Step $i+1$: If there is an $A_j \to B_j \in D$ such that $B_j \subseteq A(i)$ and $A_j \not\subseteq A(i)$, then we set $A(i+1) = A(i) \cup (\bigcup_{B_j \subseteq A(i)} A_j)$. In the converse case we set $H_P(A) = A(i)$.

It can be seen that there is a t such that $A = A(0) \subseteq A(1) \subseteq \ldots \subseteq A(t) = A(t+1) = \ldots$

By rules (3) and (4) in Definition 1.3 it can be seen that the DD $\{a_{i1}, \dots, a_{it}\} \to B$ is equivalent to a set of DDs $\{\{a_{i1}\} \to B, \dots, \{a_{it}\} \to B\}$. Consequently, we can assume that D only contains the DDs form $\{a\} \to B$. Clearly, if $A \neq \emptyset$ then $A \to \emptyset \notin D$.

In [2] the notion of a F-based derivation tree for functional dependency is introduced, in the analogous way we present a derivation tree for dual dependency as follows.

Definition 2.4 *Let $P =< U, D >$ be a dual scheme and D only contains the DDs form $\{a\} \to B$. The set of derivation trees (DT for short) over P is constructed as follows:*

1. *A node labeled with a is a DT,where $a \in U$.*
2. *If a is label of a leaf of DT Q and $\{a\} \to \{b_1, \dots, b_t\} \in D$. Then we replace this leaf in Q by the subtree whose root labeled with a and $b_1, \dots, b_t$ as chidren of root.An obtained tree is a DT.*
3. *Nothing else is a DT.*

Remark 2.5 *Let $P =< U, D >$ be a dual scheme and D only contains the DDs form $\{a\} \to B$. We call a sequence DDs $(d_1, \dots, d_m)$ is a derivation of a DD $E \to F$ over P if $d_m = E \to F$ and for each $i (1 \le i \le m)$ one of the following holds:*

(1) $d_i \in D$ or $d_i = A \to A$

(2) d_i is the result of applying rule (2) to two of DDs $d_1, \dots, d_{i-1}$

(3) d_i is the result of applying rule (3) to one of DDs $d_1, \dots, d_{i-1}$

(4) d_i is the result of applying rule (4) to two of DDs $d_1, \dots, d_{i-1}$.

Where rules (2),(3),(4) in Definition 1.3.

Proposition 2.6 *By Algorithm 2.3 we obtain $H_P(A) = A(t)$ and the time complexity of Algorithm 2.3 is polynomial in the size of P.*

Proof: It is easy to see that the time complexity of Algorithm 2.3 is polynomial in the size of P. Now we have to prove that $a \in A(t)$ iff $a \in H_P(A)$.

($\Longrightarrow$) : We prove by the induction. It is obvious that $a \in A(0) = A \subseteq H_P(A)$. We assume that $A(i) \subseteq H_P(A)$, and $a \in A(i+1) - A(i)$.

According to construction of Algorithm 2.3 there exists $A_j \to B_j \in D$ such that $B_j \subseteq A(i), a \in A_j - A(i)$. By (2) and (3) of Definition 1.3 we have $\{a\} \to B_j$. By $B_j \subseteq A(i)$ and (3) of Definition 1.3 $B_j \to A(i)$ holds. According to the inductive hypothesis $A(i) \to A$ holds. Consequently, by (2) of Definition 1.3 we obtain $\{a\} \to A$. Thus, $a \in H_P(A)$ holds.

($\Longleftarrow$): We can assume that D only contains the DDs form $\{a\} \to B$. By induction on the length of the derivation of $\{a\} \to F$ we can show that if $\{a\} \to F \in D^+$ then there is a DT with root labeled a and a set of leaves of this DT is a subset of F. This proof is in the analogous way as for functional dependency ,see [2], it will be omitted. From this consider and based on the notion of DT by induction on the depth of derivation trees we can show that if $a \in H_P(A)$ then $a \in A(t)$. This proof is easy, it will be omitted. Our proof is complete.

It can be seen that $A \to B \in D^+$ iff $A \subseteq H_P(B)$. From this and by Algorithm 2.3 the following proposition is clear.

Proposition 2.7 *(The membership problem)*

Let $P =< U, D >$ be a dual scheme. $X \to Y$ is a dual dependency. Then there exists a polynomial time algorithm deciding whether $X \to Y \in D^+$.

Let D be a d-family over U, $G(D)$ is the set of all maximal left-sides of D. Denote by $N(D)$ the minimal generator of $G(D)$. Denote $s(D) = min\{m: |R| = m, D_R = D\}$.

Theorem 2.8 [11] $(2|N(D)|)^{1/2} \leq s(D) \leq 2|N(D)|$.

Theorem 2.9 *(Generating Armstrong relation for a given dual scheme) The time complexity of finding Armstrong relation of a given dual scheme P is exponential in the size of P.*

Proof: Let $P =< U, D >$ be a dual scheme. We set $H_P(A) = \{a \in U : \{a\} \to A \in D^+\}$. Let $Z(P) = \{A \in P(U) : H_P(A) = A\}$. It is easy to see that $Z(P) = G(D^+)$. Thus, $N(D^+)$ is the minimal generator of $Z(P)$. First we contruct an exponential time algorithm that finds a relation R such that $D_R = D^+$. From P we compute $Z(P)$ by Algorithm 2.3. After that we construct the minimal generator of $Z(P)$. We assume that $N(D^+) = \{A_1, \ldots, A_s\}$. Construct a relation $R = \{h_1, h_2, \ldots, h_{2s-1}, h_{2s}\}$ as follows:

$\forall i = 1, \ldots, s \quad \forall a \in U: h_{2i-1}(a) = 2i - 1$

$$h_{2i}(a) = \begin{cases} 2i & \text{if } a \in A_i \\ 2i-1 & \text{otherwise.} \end{cases}$$

According to Theorem 2.2 we obtain $D_R = D^+$.

Let us take a partition $U = X_1 \cup, \ldots, \cup X_m \cup W$, where $m = [n/3]$, and $|X_i| = 3$ $(1 \leq i \leq m)$.

We set

$H = \{B: |B| = 2, B \subseteq X_i$ for some $i\}$ if $|W| = 0$,
$H = \{B: |B| = 2, B \subseteq X_i$ for some $i : 1 \leq i \leq m-1$ or $B \subseteq X_m \cup W\}$ if $|W| = 1$,
$H = \{B: |B| = 2, B \subseteq X_i$ for some $i : 1 \leq i \leq m$ or $B = W\}$ if $|W| = 2$.
It is easy to see that
$H^{-1} = \{A: |A \cap X_i| = 1, \forall i\}$ if $|W| = 0$,
$H^{-1} = \{A: |A \cap X_i| = 1, (1 \leq i \leq m-1)$ and $|A \cap (X_m \cup W)| = 1\}$ if $|W| = 1$,
$H^{-1} = \{A: |A \cap X_i| = 1, (1 \leq i \leq m)$ and $|A \cap W| = 1\}$ if $|W| = 2$.

It is clear that $n - 1 \leq |H| \leq n + 2, 3^{[n/4]} < |H^{-1}|$. We construct a dual scheme $P =< U, D = \{U \to B: B \in H\} >$. Based on Definition 1.9 and by Algorithm 2.3 we obtain $H^{-1} \subseteq N(D^+)$. By Theorem 2.8 we have $(2|N(D^+)|)^{1/2} \leq s(D^+)$. Consequently, we obtain $3^{[n/8]} < s(D^+)$. Based on the definition of $s(D^+)$ it can be seen that we always can construct a dual scheme P such that the number of rows of any Armstrong relation of P is exponential in the size of P. Our proof is complete.

Algorithm 2.10 *(Inferring dual dependencies)*

Input: a relation $R = \{h_1, \ldots, h_m\}$ over U.
Output: a dual scheme $P =< U, D >$ such that $D_R = D^+$.
Step 1: Find the non-equality system $NE_R = \{N_{ij} : 1 \leq i < j \leq m\}$, where $N_{ij} = \{a \in U : h_i(a) \neq h_j(a)\}$,
Step 2: Find the minimal generator N, where $N = \{A \in NE_R : A \neq \cap\{B \in NE_R : A \subset B\}\}$.
Denote elements of N by $A_1, \ldots, A_s$.
Step 3: For every $B \subseteq U$ if there is A_i such that $B \subseteq A_i$, we compute $C = \bigcap_{B \subseteq A_i} A_i$ and set $C \to B$. In the converse case we set $U \to B$.
Denote T the set of all such dual dependencies
Step 4: Set $D = T - Q$, where $Q = \{X \to Y \in T : X = Y$ or there is $X \to Y' \in T: Y' \subseteq Y\}$.

Clearly, according to Theorem 2.2, Algorithm 2.10 finds a relation scheme P such that a given relation R is an Armstrong relation of P.

Definition 2.11 *Let $P =< U, D >, P' =< U, D' >$ be two dual schemes. We say that P' is a cover of P if $D'^+ = D^+$. It is obvious that P also is a cover of P'.*

It can be seen that if P, P' are dual schemes over U then based on Proposition 2.7 and Algorithm 2.3 there is a polynomial time algorithm deciding whether $D^+ = D'^+$.

Definition 2.12 *Let $P =< U, D >$, $D = \{A_i \to B_i : i = 1, \ldots, m\}$ be a dual scheme. We say that P is an irredundant cover if for all $T \subset D : D^+ \neq T^+$.*

Now we give an algorithm to find an irredundant cover of a given dual scheme.

Algorithm 2.13 *(Finding an irredundant cover)*

Input : Let $P =< U, D = \{A_i \to B_i : i = 1, \ldots, m\} >$ be a dual scheme.
Output : $P' =< U, D' >$ is an irredundant cover of P.
Step 1: Set $L(1) = D$

Step (i+1) : Set $Q = L(i) - \{A_i \to B_i\}$, and

$$L(i+1) = \begin{cases} Q & \text{if } A_i \to B_i \in Q^+ \\ L(i) & \text{otherwise.} \end{cases}$$

Then we set $D' = L(m+1)$.

Proposition 2.14 $< U, L(m+1) >$ *is an irredundant cover of* P.

Proof: First we show that $< U, L(i+1) >$ is a cover of $< U, L(i) >$. If $L(i+1) = Q$ then by $A_i \to B_i \in Q^+$ we have $L(i)^+ = L(i+1)^+$. If $L(i+1) = L(i)$ it is obvious that $L(i+1)^+ = L(i)^+$. So we have $D^+ = L(1)^+ = \ldots = L(m+1)^+ = D'^+$. Now we show that $< U, D' >$ is irredundant. Suppose that there is an irredundant cover $< U, L >$ of P such that $L \subset L(m+1)$. Thus, there is a DD $A_j \to B_j \in L(m+1)$ but $A_j \to B_j \notin L$, where $1 \leq j \leq m$. From the definition of $L(j+1)$ we obtain $A_j \to B_j \notin Q^+$, where $Q = L(j) - \{A_j \to B_j\}$. Since $L(m+1) \subseteq L(j)$ it follows that $A_j \to B_j \notin Q'^+$, where $Q' = L(m+1) - \{A_j \to B_j\}$. Clearly, $Q' \subseteq Q$, $L \subseteq L(m+1) - \{A_j \to B_j\}$ hold. Consequently, $A_j \to B_j \notin L^+$. This conflicts with the fact that $L^+ = D^+$. Our proof is complete.

Let $P =< U, D >$ be a dual scheme. We can assume that the set D only contains the DDs form $\{a\} \to B$. Based on this we give the next definition

Definition 2.15 *Let* $P =< U, D >$ *be a dual scheme.* P *is called a normal dual scheme if* P *is irredundant and the following properties hold :*

(1) D *only contains the DDs form* $\{a\} \to B$, *where* $a \in U, B \in P(U)$,

(2) for all $\{a\} \to B \in D$ *and* $B' \subset B : < U, D - \{\{a\} \to B\} \cup \{\{a\} \to B'\} >$ *is not a cover of* P.

Proposition 2.16 *Let* $P =< U, D >$ *be a dual scheme. Then there is an algorithm finding a normal cover of* P. *The time complexity of it is polynomial in the size of* P.

Proof: (1) is clear. Consequently, we assume that D only contains the DDs form $\{a\} \to B$. Based on Algorithm 2.13 from P we construct an irredundant dual scheme P' which is a cover of P. Assume that $P' =< U, D' = \{\{a_i\} \to B_i : i = 1, \ldots, t\} >$, and $B_i = \{b_{i1}, \ldots, b_{ih}\}$. For each $i (1 \leq i \leq t)$ we set $E(1) = B_i$, for $j = 1, \ldots, h$

$$E(j+1) = \begin{cases} E(j) - b_{ij} & \text{if } \{a\} \to \{E(j) - b_{ij}\} \in D'^+ \\ E(j) & \text{otherwise.} \end{cases}$$

Denote $T_i = E(h+1)$. According to Algorithm 2.3 and Proposition 2.7 we compute T_i in polynomial time in the size of P'. By induction we can show that $\{a_i\} \to T_i \in D'^+$ and $\forall T \subset T_i$ we obtain $\{a_i\} \to T \notin D'^+$. This is clear and so its proof will be omitted. Now we set $P'' =< U, D'' = \{\{a_i\} \to T_i : i = 1, \ldots, t\} >$. It is easy to see that P'' is a normal cover of P. By Algorithm 2.13 and Algorithm 2.3 we can compute P'' in polynomial time in the size of P. Our proof is complete.

3 Relation schemes in BCNF

In this section we give some new results concerning relation schemes in BCNF. We show some cases in which FD-relation equivalence problem is solved by polynomial time algorithms. Now we give some necessary definitions.

Definition 3.1 *Let $R = \{h_1, \ldots, h_m\}$ be a relation over U, and $A, B \subseteq U$.*

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

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

It is easy to see that F_R satisfies the following properties:

$\forall B \subseteq A$: $A \to B \in F_R$ (pseudoreflexivity), if $A \to B \in F_R$ and $C \subseteq D$, then $\{A \cup D\} \to \{B \cup C\}$ (augmentation), if $A \to B \in F_R$ and $\{B \cup C\} \to D$, then $\{A \cup C\} \to D$(pseudotransitivity).

Definition 3.2 *A relation scheme S,or RS for short, is a pair $< U, F >$. 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 above rules. Denote $A^+ = \{a \colon A \to \{a\} \in F^+\}$. A^+ is called the closure of A over S.Denote $Z(F^+) = \{A \subseteq U \colon A^+ = A\}$.*

Clearly, in [1] if $S =< U, F >$ is a RS, then there is a relation R over U such that $F_R = F^+$. Such a relation is called an Armstrong relation of S.

Let R be a relation, $S =< U, F >$ be a RS, and $A \subseteq U$. Then A is a key of R (a key of S, *respectively*) if $A \xrightarrow[R]{f} U$ ($A \to U \in F^+$, *respectively*). A is a minimal key of R(S, *respectively*) if A is a key of R(S, *respectively*), and any proper subset of A is not a key of R(S, *respectively*). Denote K_R(K_S, *respectively*) the set of all minimal keys of R(S, *respectively*).

Clearly, K_R, K_S are Sperner systems over U.

Let R be a relation,$S =< U, F >$ be a RS.R, S are in Boyce-Codd normal form (BCNF) if for each $A \to \{a\} \in F^+$($\in F_R$, *respectively*) and $a \notin A$ then $A \to U \in F^+$($\in F_R$, *respectively*).

Definition 3.3 *Let $S =< U, F >$ be a RS. We say that S is a k-RS over U if $F = \{K_1 \to U, \ldots, K_m \to U\}$, where $\{K_1, \ldots, K_m\}$ is a Sperner system over U. It is easy to see that $K_S = \{K_1, \ldots, K_m\}$.*

It can be seen that a relation scheme $S =< U, F >$ is in BCNF iff $\forall A \subseteq U$ either $A^+ = A$ or $A^+ = U$. Clearly, if $S =< U, F >$ is in BCNF then using the algorithm for finding a minimal cover we can construct in polynomial time a k-RS $S' =< U, F' >$ such that $F^+ = F'^+$, see [10]. Conversely, it can be seen that an arbitrary k-RS is in BCNF. Consequently, we can consider a RS in BCNF as a k-RS.

Theorem 3.4 [4] *Let $S_1 = < U, F_1 >, S_2 = < U, F_2 >$ be two RS over U. Then $F_1{}^+ = F_2{}^+$ iff $Z(F_1{}^+) = Z(F_2{}^+)$, and $F_1{}^+ \subseteq F_2{}^+$ iff $Z(F_2{}^+) \subseteq Z(F_1{}^+)$.*

Theorem 3.5 [4] *Let K be a Sperner system and $S = < U, F >$ be a RS over U. Then $K_S = K$ iff*

$$\{U\} \cup K^{-1} \subseteq Z(F^+) \subseteq \{U\} \cup G(K^{-1}),$$

where $G(K^{-1}) = \{A \subseteq U : \exists B \in K^{-1} : A \subseteq B\}$.

Based on Theorem 3.5 we have

Theorem 3.6 *Let $K = \{K_1, \ldots, K_t\}$ be a Sperner system over U. Consider the relation scheme $S = (U, F)$ with $F = \{K_1 \to U, \ldots, K_t \to U\}$.*

Then $K_S = K$, and $Z(F^+) = G(K_S^{-1}) \cup \{U\}$.

Let R be a relation over U. Denote $A_R^+ = \{a \in U : A \to \{a\} \in F_R\}$, and $Z(F_R) = \{A \subseteq U : A_R^+ = A\}$.

According to Theorem 3.5 we can give examples for which there are two RSs $S_1 = < U, F_1 >, S_2 = < U, F_2 >$ such that $K_{S_1} = K_{S_2}$, but $F_1{}^+ \neq F_2{}^+$. Clearly, for relations this consider is the same.

We give the following notion.

Definition 3.7 *Let $S = < U, F >$ be a RS, R be a relation over U. We call S (R,respectively) is an unique RS (relation,respectively) if for all RS $S' = < U, F' >$ (relation R',respectively) : $K_S = K_{S'}$ ($K_R = K_{R'}$,respectively) then $F^+ = F'^+$ ($F_R = F_{R'}$,respectively).*

Proposition 3.8 *The time complexity of deciding whether a given relation R over U is unique is polynomial in the sizes of R and U.*

Proof: Let R a relation over U. By [13] from R we can compute $K_R{}^{-1}$ in polynomial time in the sizes of R and U, where K_R is a set of all minimal keys of R.

Denote elements of $K_R{}^{-1}$ by $A_1, \ldots, A_t$. Set $M_R = \{A_i - a : a \in U, i = 1, \ldots, t\}$.

Denote elements of M_R by $B_1, \ldots, B_s$. We construct a relation $R' = \{h_0, h_1, \ldots, h_s\}$ as follows:

For all $a \in U$, $h_0(a) = 0$, for each $i = 1, \ldots, s$ $h_i(a) = 0$ if $a \in B_i$, in the converse case we set $h_i(a) = i$.

By [10] R' is in BCNF and $K_R = K_{R'}$.

We construct a relation $R'' = \{l_0, l_1, \ldots, l_t\}$ as follows:

$l_0(a) = 0$ for all $a \in U$. For all $j = 1, \ldots, t$ then $l_j(a) = j$ if $a \notin A_j$,

in the converse case set $l_j(a) = 0$.

It can be seen that $K_R = K_{R''}$ and $Z(F_{R''}) = (K_R^{-1})^+$. (see Definition 1.5).

It is easy to see that M_R, R'' and R' are constructed in polynomial time in the sizes of U and R.

Based on Theorem 3.5 we see that R is unique iff $F_{R'} = F_{R''}$. Clearly,$F_{R'} = F_{R''}$ can be tested in polynomial time in the sizes of R' and R''. The proposition is proved.

Definition 3.9 [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.*

Theorem 3.10 [4] *Let $S =< U, F >$ be a RS. If K_S is a saturated Sperner system, then S is an unique RS.*

Examples show that there is a Sperner system K (K^{-1},respectively) such that K (K^{-1}, respectively) is saturated, but K^{-1} (K,respectively) is not saturated.
Now we define the next notion.

Definition 3.11 *Let K be a Sperner system over U. We say that K is inclusive, if for every $A \in K$ there is a $B \in K^{-1}$ such that $B \subset A$. We call K is embedded if for each $A \in K$ there exists a $B \in H: A \subset B$, where $H^{-1} = K$.*

Theorem 3.12 [13] *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.

Let $S =< U, F >$ be a RS in BCNF,R be a relation in BCNF. Then we say that S is an inclusive RS if K_S is inclusive and R an embedded relation if K_R^{-1} is embedded.
It can be seen that the BCNF property of S is polynomially recognizable. By [13] we can compute K_R^{-1} in polynomial time in the size of R, and based on polynomial time algorithm finding minimal cover we also construct K_S from a given BCNF relation scheme. On the other hand, by definitions of embedded,inclusive Sperner systems we obtain the following proposition.

Proposition 3.13 *Let $S =< U, F >$ be a RS ,R be a relation over U. Then*

1. *Deciding whether S is an inclusive RS is solved in polynomial time in the size of S.*

2. *There exists an algorithm deciding whether R is an embedded relation and the time complexity of it is polynomial in the sizes of U and R.*

It is easy to see that if $S =< U, F >, S' =< U, F' >$ are two RSs then deciding whether $F^+ = F'^+$ can be tested in polynomial time in the sizes of S and S'.

Now we introduct the next problem.
Let $S =< U, F >, S' =< U, F' >$ be two RSs. Decide whether $K_S = K_{S'}$.

The following proposition is clear.

Proposition 3.14 *Let S, S' be two RSs.If S is unique then deciding whether $K_S = K_{S'}$ is polynomially recognizable.*

In [10] the FD-relation equivalence problem is introduced as follows:
Let $S =< U, F >$ be a RS, R be a relation over U. Decide whether $F^+ = F_R$, i.e. R is an Armstrong relation of S.

Definition 3.15 *Let K_1, K_2 be two Sperner system over U. We set $K = K_1 \cup K_2$ and $T_K = \{A \in K: \nexists B \in K: A \subset B\}$. We say that the union $K = K_1 \cup K_2$ is equality if $\forall A_1, A_2 \in T_K: |A_1| = |A_2|$.*

Based on Definition 3.15 we give the next theorem related to the FD-relation equivalence problem .

Proposition 3.16 *Let $S =< U, F >$ be a relation scheme in BCNF and R a relation over U in BCNF. $K_S = \{A_1, \ldots, A_p\}$ ($K_R^{-1} = \{B_1, \ldots, B_q\}$) is the set of minimal keys of S (the set of antikeys of R). Then if $K_S \cup K_R^{-1}$ is equality then the FD-relation equivalence problem is solved in polynomial time in the sizes of S and R.*

Proof: Clearly, by [13] from R we compute K_R^{-1} in polynomial time in the size of R, and from S we find a k-relation scheme that is a minimum cover of S. The minimum cover is constructed in polynomial time in the size of S. We set $K = K_S \cup K_R^{-1}$. Because K is equality,we assume that $|A| = m$, and $|U| = n$. We compute the number C_n^m. Clearly, K and K^{-1} are uniquely determined by each other. By definitions of K_S and K_R^{-1} we can see that if $|T_K| \neq C_n^m$ then $K_S \neq K_R$. Thus, in BCNF class we obtain $F^+ \neq F_R$.

Now we assume that $|T_K| = C_n^m$. If there is $A_i (1 \leq i \leq p)$ such that $A_i \subseteq B_j (1 \leq j \leq q)$ then $K_S \neq K_R$. Consequently,we can assume that $A_i \not\subseteq B_j$ for all i, j. For each $j = 1, \ldots, q$ we compute B_j^+. It can be seen that for all $D \subseteq U$ D^+ is computed in polynomial time in the size of S. We set $M = \{B_j \cup \{a\}: a \in U - B_j\} = \{M_1, \ldots, M_t\}$. It is obvious that M is computed in polynomial time. If $B_j^+ \neq U$ and for all $l = 1, \ldots, t$ $M_l^+ = U$ hold then $B_j \in K_S^{-1}$ holds, otherwise we obtain $B_j \notin K_S^{-1}$. If there is a B_j: $B_j \notin K_S^{-1}$ then by the definition of antikeys $K_R \neq K_S$.We assume that for all j=1,...,q $B_j \in K_S^{-1}$. For each $i = 1, \ldots, p$ we set $N = \{A_i - \{a\}: a \in A_i\} = \{N_1, \ldots, N_s\}$. It can be seen that N is computed in polynomial time. If there is a $N_n (1 \leq n \leq s)$ such that $N_n \not\subseteq B_j$ for all $j = 1, \ldots, q$ then $A_i \notin K_R$ holds. In the converse case we obtain $A_i \in K_R$. Clearly, if there is an $A_i \notin K_R$ then $K_S \neq K_R$. We assume that for each $i = 1, \ldots, p$ we have $A_i \in K_R$. We set

$$Z = \{A_i - \{a\}: a \in A_i, i = 1, \ldots, p\},$$

$$W = \{A \in Q: A = A^+, (A \cup \{a\})^+ = U, \forall a \in U - A\},$$

$$J = \{B_j \cup \{a\}: a \in U - B_j, j = 1, \ldots, q\},$$

$$I = \{B \in J: B_R^+ = U, \{B - a\}_R^+ \neq U \forall a \in B\}.$$

Based on definition of K_S and definition of K_R^{-1} we can see that if either there is an $A \in W$ such that $A \notin K_R^{-1}$ or there exists a $B \in I$ but $B \notin K_S$ then $K_S \neq K_R$. It can be seen that W, I are constructed in polynomial time in the sizes of S, R, K_S, K_R^{-1}. Finally, we see that if for all $i = 1, \ldots, p, j = 1, \ldots, q$ $A_i \in K_R, B_j \in K_S^{-1}, W \subseteq K_R^{-1}, I \subseteq K_S$ hold then by $|T_K| = C_n^m$ and according to definition of set of minimal keys and definition of set of antikeys we obtain $K_R = K_S$. Since S, R are in BCNF we have $F_R = F^+$. The proof is complete.

Let K be a Sperner system over U. We say that K is pseudo-monotonous if for each Sperner system $K' : K \cap K' = \emptyset$ and $K \cup K'$ is a Sperner system over U then $K^{-1} \subseteq \{K \cup K'\}^{-1}$.

We say that K is a changed Sperner system if for each $H' : H' \subset H$ then there are $A \in K, B \in H'^{-1}$ such that $A \subset B$, where $H^{-1} = K$.

Proposition 3.17 *Let S be a RS in BCNF, R be a relation in BCNF. Then if either K_S is pseudo-monotonous or K_R^{-1} is changed, then FD-relation equivalence problem is solved in polynomial time in the sizez of S and R.*

Proof: First we assume that K_R^{-1} is a changed Sperner system. Based on a polynomial time algorithm finding a minimal cover,we construct a set of all minimal keys K_S. It is known [13] that from R we compute K_R^{-1} in polynomial time in the size of R.

If there are $A \in K_S$ and $B \in K_R^{-1}$ such that $A \subseteq B$, then $K_S \neq K_R$. Thus,for all $A \in K_S, B \in K_R^{-1}$ we can assume that $A \not\subseteq B$. We set $X = \{A - \{a\}: A \in K_S, a \in A\}$. If for all $C \in X, B \in K_R^{-1}$ we obtain $C \subseteq B$ then $K_S \subseteq K_R$. In the converse case we have $K_S \neq K_R$. It is easy to see that X is computed in polynomial time. We assume that $K_S \subseteq K_R$.

For each $B \in K_R^{-1}$ we compute B^+. If there is a B such that $B^+ = U$ then $K_S \neq K_R$. We assume that $B^+ \neq U$ for all $B \in K_R^{-1}$. We set $Y = \{B \cup \{a\}: B \in K_R^{-1}, a \in U - B\}$. It is obvious that Y is computed in polynomial time. If for all $D \in Y$ we have $D^+ = U$ then $K_R^{-1} \subseteq K_S^{-1}$. In the converse case we obtain $K_R^{-1} \neq K_S^{-1}$. Because K and K^{-1} are uniquely determined by each other, we have $K_R \neq K_S$. Now assume that $K_R^{-1} \subseteq K_S^{-1}$ and $K_S \subseteq K_R$. By hypothesis K_R^{-1} is a changed Sperner system. Consequently,if $K_S \subset K_R$ then there are $B \in K_R^{-1}$ and $E \in K_S^{-1}$ such that $B \subset E$. Hence, $K_R^{-1} \not\subseteq K_S^{-1}$ holds.Thus, $K_S = K_R$. Because S, R are in BCNF, we obtain $F_R = F^+$.

If S is pseudo-monotonous then the proof is the same. The proof is complete.

4 Conclusion

Our further research will be devoted to the following problems:

1. What is the time complexity of finding a dual scheme P from a given relation R such that $D^+ = D_R$
2. Given a relation scheme S and a relation R. What is the time complexity of deciding whether $K_S = K_R$.
3. Let S_1, S_2 be two relation schemes over U. What is the time complexity of deciding whether $K_{S_1} = K_{S_2}$.
4. Let S be a RS. What is the time complexity of deciding whether S is an unique RS.

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., Bernstein P.A. Computational problems related to the design of normal form relational schemas.ACM Trans. on Database syst.4,1,(1979)pp.30-59.

[3] Berman J., Blok W. Positive Boolean Dependencies. Infor.Proces. Letters 27 (1988) pp. 147-150.

[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] Czédli G. d-dependency structures in the relational model of data. Acta Cybernetica Hungary V/I (1980) pp.49-57.

[6] Czédli G. On Dependencies in the relational model of data.EIK 17 (1981) 2/3 pp.103-112.

[7] Demetrovics J. Relacios adatmodell logikai es strukturalis vizsgalata MTA-SZTAKI Tanulmanyok,Budapest, 114 (1980) pp.1-97.

[8] Demetrovics J.,Gyepesi G. On the functional dependency and some generalizations of it. Acta Cybernetica Hungary, Tom. 5, Fasc. 3 (1981), pp. 295-305.

[9] Demetrovics J.,Gyepesi G. Some generalized type functional dependencies formalized as equality set in matrices.Discrete Appl.Math. 6 (1983) pp.35-47.

[10] Gottlob G. , Libkin L. Investigations on Armstrong relations, dependency inference, and excluded functional dependencies. Acta Cybernetica Hungary Tom.9 Fasc.4 (1990) pp. 385-402.

[11] V.D.Thi. Logical dependencies and irredundant relations. Computers and Artificial Intelligence 7 (1988), pp.165-184.

[12] Thi V.D. Funkcionalis fuggoseggel kapcsolatos nehany kombinatorikai jellegu vizsgalat a relacios adatmodellben. MTA-SZTAKI Tanulmanyok, Budapest, 191 (1986) pp. 1-157. Ph.D. Dissertation.

[13] Thi V.D. Minimal keys and Antikeys.Acta Cybernetica Hungary Tom.7 Fasc.4 (1986) pp.361-371.

Received October 21, 1992