CLOSED SETS AND TRANSLATIONS OF RELATION SCHEMES

JÁNOS DEMETROVICS
Computer and Automation Institute, Hungarian Academy of Sciences

NGUYEN XUAN HUY
Institute of Computer Science and Cybernetics,
National Centre for Scientific Research of Vietnam; Hanoi, Vietnam

(Received May, 1990)

Abstract. The concept of translation of relation schemes is introduced. Some characterizations of various closed sets in relation schemes, such as generators, coatoms or antikeys are investigated from different aspects. The connections between these objects in a given relation scheme alone and in the prime and translated relation schemes are presented. It is shown that translating a relation scheme can be done in \(O(|F| \cdot |R|)\) time and testing whether a subset of attributes is an antikey can be done in \(O(|F| \cdot |R|^2)\) time, where \(|F|\) is the number of functional dependencies and \(|R|\) is the number of attributes. It is shown that computing the intersection of all antikeys is NP-complete, but computing their union can be done in polynomial time. A connection between three well-known results of Beeri et al., Demetrovics and Thi, and Ginsburg and Hull about relations representing a given relation scheme or Sperner system is presented.

Key Words and Phrases: relation scheme, functional dependency, closure, closed set, generator, coatom, translation of relation schemes, key, antikey.

1. INTRODUCTION

In the relational model of databases [7] there are many problems that require efficient algorithms for computing such objects as closures, covers, keys, antikeys, closed sets, generators etc. Several algorithms are known. For example, algorithms for computing closures in a relation scheme (RS) with functional dependencies (FD) [5] and with various types of dependencies [11,12,15,16,20], algorithms for finding keys [4,8,21,22], antikeys [14], prime attributes [14,24], and relations representing a set of FDs [6,18,19] and a closure function [9,10,13].

It is a natural observation that if we can reduce the size of the given RS, then the existing algorithms may run faster. Several aspects of the "reduction" technique are presented in the literature [2,3,5,22,23,24,26]. The kernel of these aspects is to transform a given RS to an equivalent one in some sense.

In this paper we give another approach to the problem. Our concept is called translation of relation schemes [4,11,15,16,21]. The main purpose of this concept is to transform a given RS by removing some attributes that seem to be not important for computing objects. Although, in general, the translation is not equivalent, the prime objects can be obtained from reduced objects with the help of simple operations. Let us consider an example.

Example 1.1

Let \(a = (R, F)\) be an RS, where \(R\) is the set of attributes, \(R = ABCDEFG\), and \(F\) is the set of FDs, \(F = \{AE \rightarrow D, BC \rightarrow E, E \rightarrow BC\}\). To find the set of keys \(K_a\) of \(a\) let us construct a new

---

Research supported by Hungarian Foundation for Scientific Research Grant 1066.
RS, \( e = (P, H) \) from \( a \) by removing attributes \( A, D, F \) and \( G \). Thus \( P = R - ADFG = BCE \), and \( H = \{ E \rightarrow \emptyset, BC \rightarrow E, E \rightarrow BC \} \). Now, we delete FD \( E \rightarrow \emptyset \) in \( H \). We have \( H = \{ BC \rightarrow E, E \rightarrow BC \} \). It is easy to see that RS \( e \) has two keys \( E \) and \( BC \). To obtain \( K_a \), we need only add attributes \( A, F \), and \( G \) (but not \( D \)) to every key of \( e \). Thus, \( K_a = \{ AEFG, ABCFG \} \).

Although \( F \) is an optimal set of FDs \([23]\), \( H \) uses fewer attribute symbols.

The paper is structured as follows. In Section 2 we give some results about two classes of closed sets in relation schemes: generators and coatoms. It is shown that coatoms are the maximal members in the set of generators. A graph for representing closed sets, generators and coatoms is given. Section 3 discusses some problems of keys and antikeys. It is shown that antikeys are coatoms. The connection between generators, coatoms, superkeys of cardinality 1 and antikeys is expressed by an evident formula. It is shown that the union of antikeys can be found in polynomial time. Two necessary and sufficient conditions for a set of keys to be consisted of all attributes and to be consisted of a unique empty set are given. Section 4 shows a connection between three well-known results of Beeri et al. \([6]\), Demetrovics and Thi \([14]\), and Ginsburg and Hull \([19]\) about relations representing a given RS or Sperner System. Section 5 presents some new results of the translation of RSs. The first form of representation of antikeys and the connection between closed sets, generators and coatoms of the original and translated RSs are given. An efficient algorithm for testing whether a subset of attributes is an antikey is presented. Finally, in Section 6 we list some problems of our further research.

2. RELATION SCHEMES AND CLOSED SETS

Definition 2.1

Let \( R = \{ A_1, \ldots, A_n \} \) be a nonempty finite set of symbols called \textit{attributes}. Corresponding to each attribute \( A_i \) is a nonempty set \( \text{dom}(A_i) \), \( 1 \leq i \leq n \), called the \textit{domain} of \( A_i \). Let \( D \) be the union of all attribute domains. A \textit{relation} \( r \) with the set of attributes \( R \), is a finite set of mappings \( \{ t_1, \ldots, t_p \} \) from \( R \) to \( D \) with the restriction that for each mapping \( t \in r \), \( t(A_i) \) must be in \( \text{dom}(A_i) \), \( 1 \leq i \leq n \). The mappings are called \textit{tuples}.

By the tradition appeared in the literature of Relational Database Theory, we use the following notation \([23,26]\). The attributes are denoted by the letters \( A, B, C, \ldots \), and the sets of attributes are denoted by the letters \( \ldots, X, Y, Z \). A set of attributes is written as a string of attributes, and the union of sets of attributes \( X \) and \( Y \) is written as \( XY \). The notation \( X \subseteq Y \) \((X \supseteq Y)\) means that \( X \) is a proper subset (superset) of \( Y \).

Definition 2.2

If \( r \) is a relation with the set of attributes \( R \), \( t \) is a tuple of \( r \), and \( X \) is a subset of \( R \), then \( X\text{-value of } t \), written as \( t(X) \), is the restriction of mapping \( t \) on \( X \).

Definition 2.3

A \textit{functional dependency} (FD) is a statement of the form \( X \rightarrow Y \), where both \( X \) and \( Y \) are sets of attributes. The FD \( X \rightarrow Y \) \textit{holds} in a relation \( r \) if for all tuples \( u \) and \( v \) of \( r \), \( u(X) = v(X) \) implies \( u(Y) = v(Y) \). We also say that relation \( r \) \textit{satisfies} the FD \( X \rightarrow Y \).

Definition 2.4

A \textit{relation scheme} (RS) \( a \) is a pair \((R, F)\), where \( R \) is a set of attributes and \( F \) is a set of FDs on \( R \).

Definition 2.5

Let \( a = (R, F) \) be an RS and let \( f \) be a single FD on \( R \). \( F \) \textit{implies} \( f \), written \( F \models f \), if every relation on \( R \) that satisfies all dependencies in \( F \) also satisfies \( f \).
Definition 2.6

Let \( a = (R, F) \) be an RS and let \( X \subseteq R \). The \textit{closure} of \( X \) under \( a \), written \( X^+ \), is the largest set of attributes \( Y \) such that \( F \models X \rightarrow Y \). We write \( X^+ \) for \( X^+_a \) when \( a \) is clear from the context.

The closure function has the following properties [2]:

1. \( X \subseteq X^+ \).
2. If \( X \subseteq Y \), then \( X^+ \subseteq Y^+ \).
3. \( (X^+)^+ = X^+ \).

Let \( M \) be a family of subsets of a finite set. Denote by \( \bigcup M \) (\( \bigcap M \)) the union (intersection) of all members in \( M \).

Definition 2.7

Let \( a = (R, F) \) be an RS. A set \( X \subseteq R \) is \textit{closed} under \( a \) if \( X^+ = X \).

For a given RS \( a = (R, F) \), denote by \( C_a \) the family of all closed sets under \( a \). It is easy to see that the intersection of closed sets is closed [2], and \( C_a = \{X^+ | X \subseteq R \} \).

Let \( M \) be a family of subsets of a finite set, closed under intersection. Then \( M \) contains a unique minimal subfamily \( G \) such that the members of \( G \) generate \( M \) by intersection [3,13]. Thus \( G \) is the smallest set such that \( M = \{S_1 \cap \ldots \cap S_k | k > 0 \text{ and } S_1, \ldots, S_k \in G \} \).

Definition 2.9

The members of \( G \) above are called the (intersection) \textit{generators} of \( M \).

For a given RS \( a \), denote by \( G_a \) the set of generators of \( C_a \). Note that \( R \) is in \( C_a \), but not in \( G_a \), since it is the intersection of the empty collection of sets. It follows that \( G_a \subseteq C_a - \{R\} \).

The next result will be useful later on.

Theorem 2.1

[3,13] Let \( a = (R, F) \) be an RS. The following sets are equal:

1. \( G_a \),
2. \( \{V \in C_a | V \subseteq R \text{ and if } X, Y \in C_a, X \neq V, Y \neq V, \text{ then } X \cap Y \neq V \} \),
3. \( \{V \in C_a | V \subseteq R \text{ and if } V = \cap\{X_i | X_i \in C_a, 1 \leq i \leq k \}, \text{ then } V = X_i \text{ for some } i, 1 \leq i \leq k \} \),
4. \( \{V \in C_a | V \subseteq \cap\{X | X \in C_a, V \subseteq X \} \} \).

Definition 2.10

Let \((M, \leq)\) be a finite partially ordered set (poset). An element \( m \) in \( M \) is maximal, if \( m \leq x \) and \( x \in M \) imply \( m = x \). The set of all maximal elements of \( M \) is denoted by \( \text{MAX}(M) \). It is not hard to see that for every element \( x \) in \( M \), there is an element \( m \) in \( \text{MAX}(M) \), such that \( x \leq m \).

Definition 2.11

Let \( C_a \) be the family of closed sets of a given RS \( a = (R, F) \). The members of the set \( T_a = \text{MAX}(C_a - \{R\}) \) are called coatoms of \( a \), where partial order is the set inclusion \( \subseteq \).

The next theorem gives a characterization of coatoms.

Theorem 2.2

Let \( a = (R, F) \) be an RS. Then \( T_a = \text{MAX}(G_a) \).

\textbf{Proof.} First, we show that \( T_a \subseteq G_a \). Let \( V \) be a member of \( T_a \). Then, by the definition of \( T_a \), \( V \in C_a \) and \( V \neq R \). Assume that \( V = X \cap Y \), where \( X, Y \in C_a \), \( X \neq V \) and \( Y \neq V \). Then \( V \subseteq X \) and \( V \subseteq Y \). Since \( V \) is maximal in \( (C_a - \{R\}) \), and since \( X, Y \in C_a \), it follows that \( X = Y = R \), so \( V = R \); a contradiction. Therefore, \( V \neq X \cap Y \). Hence \( V \in G_a \). Now, combining the facts that \( T_a \subseteq G_a \subseteq C_a - \{R\} \) and that \( T_a = \text{MAX}(C_a - \{R\}) \), we get \( T_a = \text{MAX}(G_a) \). \( \square \)

From the definition of closed sets, we know that \( C_a \) can be found by \( \{X^+ | X \subseteq R \} \). The next theorem gives an approach to computing generators and coatoms. Let \( a = (R, F) \) be an RS. We construct a directed graph (digraph) \( H \) whose nodes correspond to the members of \( C_a \). \( (X, Y) \) is an arc of \( H \) if \( X \supset Y \) and there is no \( Z \) in \( C_a \) such that \( X \supset Z \supset Y \). Denote by \( d(X) \) the number of arcs incident to \( X \) (in-degree).
Theorem 2.3

Let $H$ be the digraph constructed above for RS $a = (R, F)$. Then,
1. $T_a = \{X \in C_a | (R, X)$ is an arc of $H\}$.
2. $C_a = \{X \in C_a | d(X) = 1\}$.

Proof. 1. Immediate from the definition of $T_a$.
2. Since $R \in C_a$, for every $X \in C_a$ with $X \neq R$, $d(X) \geq 1$. Let $X \in G_a$ and $d(X) = p$. We must show that $p = 1$. Since $d(X) = p$, there are $p$ arcs incident to $X$, say $(Y_1, X), \ldots, (Y_p, X)$. Let $Y$ be the intersection of $Y_i, 1 \leq i \leq p$. By Theorem 2.1, $X \subseteq Y$. Suppose that $Y \subseteq Y_i$, for some $i, 1 \leq i \leq p$. Then $X \subseteq Y \subseteq Y_i$, and hence $(Y_i, X)$ is not the arc of $H$; a contradiction. Therefore, $Y = Y_i, 1 \leq i \leq p$, and so $p = 1$.

Conversely, assume $X \subseteq C_a$ and $d(X) = 1$; we shall show that $X \subseteq G_a$. Let $Y$ and $Z$ be in $C_a$ and let $X \neq Y$ and $X \neq Z$. We must show that $X \neq Y \cap Z$. For, assume not. Then $X \subseteq Y$ and $X \subseteq Z$. Since $d(X) = 1$, there is a unique incident arc to $X$, say $(W, X)$, so $X \subseteq W$, and hence $W \subseteq Y$ and $W \subseteq Z$. Thus $X \subseteq W \subseteq Y \cap Z$. This contradicts $X \neq Y \cap Z$. Therefore, the assumption that $X = Y \cap Z$ must be incorrect, and in fact, $X \neq Y \cap Z$. Thus, by Theorem 2.1, $X \in G_a$, which was to be shown.

3. KEYS AND ANTIKEYS

Definition 3.1

Let $a = (R, F)$ be an RS. Let $X \subseteq R$. $X$ is a superkey of $a$ if $X^+ = R$. $X$ is a key of $a$ if it is a superkey and $Y^+ \neq R$ for every proper subset $Y$ of $X$. The set of all keys of an RS $a$ is denoted by $K_a$. Consider the following sets of attributes. $I_a = \cap K_a$ (intersection of keys), $P_a = \cup K_a$ (prime attributes), and $N_a = R - P_a$ (non-prime attributes).

Lucchesi and Osborn [22] proved that the following problem is NP-complete (for definition of NP-completeness see [1]).

The prime attribute problem: Given an attribute $A$, decide whether it belongs to any key.

By this result, it follows that computing set $P_a$ and computing set $N_a$ are NP-complete.

Definition 3.2

Let $f$ be an FD on $R$. Denote by LEFT($f$) and by RIGHT($f$), the left and the right sides of $f$, respectively. The following result gives a formula for computing $I_a$ [4,15,16,21].

Theorem 3.1

$I_a = R - \cup \{\text{RIGHT}(f) - \text{LEFT}(f) | f \in F\}$.

It follows easily from Theorem 3.1 that $I_a$ can be found in $O(|F| \cdot |R|)$ time, where $|F|$ is the number of FDs in $F$ and $|R|$ is the number of attributes in $R$.

Definition 3.3

Let $a = (R, F)$ be an RS. Let $X \subseteq R$. $X$ is an antikey of $a$ if $X^+ \neq R$ and $Y^+ = R$ for any proper superset $Y$ of $X$. The set of all antikeys of RS $a$ is denoted by $K_a^-$. We also consider the following sets. $I_a^- = \cap K_a^-$, $P_a^- = \cup K_a^-$ and $N_a^- = R - P_a^-$. The next theorem gives another equivalent definition of antikeys.

Theorem 3.2

Given an RS $a = (R, F)$. Then $K_a^- = T_a$. 

PROOF. Consider the sets $C' = C_0 - \{R\}$ and $E = \{X \subseteq R | X^+ \neq R\}$. It is easy to see that $E$ is the family of all non-superkeys, so $R \notin E$. Moreover, $C' \subseteq E$. We shall show that for each member $X$ of $E$, there is a member $Y$ of $C'$ such that $X \subseteq Y$. Indeed, let $X$ be the given set. Consider the set $Y = X^+$. We have, $Y \neq R$, since $X \in E$. By the properties of closure function, it follows that $X \subseteq X^+ = Y$ and $Y^+ = X^{++} = X^+ = Y \neq R$, so $Y \in C'$ and $X \subseteq Y$. Now, by fact above and by the definition of antikeys, it follows that $K_a^- = \operatorname{MAX}(E) = \operatorname{MAX}(C')$. By the definition of coatoms and by Theorem 2.2, we know that $T_a = \operatorname{MAX}(C') = \operatorname{MAX}(G_a)$, and hence $K_a^- = T_a$. □

Let $S_{1a}$ denote the set of all superkeys of cardinality 1 of a given RS $a = (R, F)$, i.e. $S_{1a} = \{\{A\} | A \in R, \ A^+ = R\}$. The next theorem is extremely useful as a characterization of set $P_a^-$. 

**Theorem 3.3**

Let $a = (R, F)$ be an RS. Then,

$$P_a^- = \cup (C_a - \{R\}) = \cup G_a = \cup T_a = R - US_{1a}.$$ 

**PROOF.** $[\cup (C_a - \{R\}) \subseteq \cup G_a]$ This is because each member in $C_a - \{R\}$ is the intersection of several members of $G_a$.

$[\cup G_a \subseteq \cup T_a]$ By Theorem 2.2 we know that $T_a = \operatorname{MAX}(G_a)$, so for each member $X$ in $G_a$, there is a member $Y$ in $T_a$ such that $X \subseteq Y$. Hence $\cup G_a \subseteq \cup T_a$.

$[\cup T_a \subseteq R - US_{1a}]$ Let $A \in X$ and $X \in T_a$. By the definition of coatoms, $X \neq R$. Assume that $A$ is a superkey. Since $A$ is in $X$ and $X$ is closed, $R = A^+ \subseteq X^+ = X \subseteq R$, so $X = R$; a contradiction. Therefore, $A$ is not a superkey. Thus $A \in R - US_{1a}$.

$[R - US_{1a} \subseteq \cup (C_a - \{R\})]$ Let $A \in R - US_{1a}$. Then $A^+ \neq R$. Consider the set $X = A^+$. We have $A \in X$, $X \in C_a$ and $X \neq R$, so $X \in C_a - \{R\}$, and hence $A \in \cup (C_a - \{R\})$.

Combining the facts proved above, we have $\cup (C_a - \{R\}) = \cup G_a = \cup T_a = R - US_{1a}$. Now we show that $P_a^- = \cup T_a$. Indeed, by Theorem 3.2 we know that $K_a^- = T_a$, so $P_a^- = \cup K_a^- = \cup T_a$. □

Demetrovics and Thi [14] proved the following result.

**Theorem 3.4**

$I_a^- = N_a$.

We know that the problem of computing $N_a$ is NP-complete, it follows that the problem of computing $I_a^-$ is NP-complete. But we have the following result.

**Theorem 3.5**

Let $a = (R, F)$ be an RS. Then $P_a^-$ and $N_a^-$ can be found in $O(|F| \cdot |R|^2)$ time.

**PROOF.** From Theorem 3.3 we get $N_a^- = US_{1a}$. By a result of Beeri and Bernstein [5], we know that $X^+$ can be found in $O(|F| \cdot |R|)$ time. By the definition of $S_{1a}$, it follows that $US_{1a}$ can be found in $O(|F| \cdot |R|^2)$ time. Hence $N_a^-$ and $P_a^-$ can be found in $O(|F| \cdot |R|^2)$ time. □

We know that for any RS $a = (R, F)$, $R^+ = R$. Hence $R$ is always a superkey. Therefore, every RS has at least one key. In this paper, we assume that $F$ may have the FDs with empty left sides, i.e. FDs of the form $\emptyset \rightarrow Y$. The reason of this assumption will be discussed in Section 5. By this assumption, an RS may have the empty key. As example, if $R = AB$ and $F = \{\emptyset \rightarrow AB\}$, then $\emptyset^+ = AB = R$, and hence $\emptyset$ is a unique key in RS $a = (R, F)$. However, there are RSs, which have no antikeys. Indeed, in the example above, $K_a^- = \emptyset$. The next theorem gives an explanation for this.

**Theorem 3.6**

Let $a = (R, F)$ be an RS. Then,

1. $K_a^- = \{\emptyset\}$ if and only if $K_a = R$.
2. $K_a^- = \emptyset$ if and only if $K_a = \{\emptyset\}$. 

Note: The notation $K_a = R$ means that, if $R = \{A_1, \ldots, A_n\}$, then $K_a = \{\{A_1\}, \ldots, \{A_n\}\}$. 

Closed Sets and Translations of Relation Schemes 17
PROOF. 1.[$\Rightarrow$] If $\emptyset$ is an antikey, then $\emptyset^+ = \emptyset$, since $K_{\emptyset}^- = T_a \subseteq C_a$. Moreover, by the definition of antikeys, it follows that $\emptyset$ is a unique antikey in $a$. Thus, each attribute in $R$ is a superkey of $a$. Since $\emptyset^+ = \emptyset$, and by our assumption for the definition of RS, we know that $R \neq \emptyset$, it follows that $K_a = R$.

[\Leftarrow] If every attribute of $R$ is a key, then $\emptyset^+ \neq R$, and by the definition of antikeys, it follows that $\emptyset$ is a unique antikey in $a$.

2. [$\Rightarrow$] if $K_{\emptyset}^-$ is empty, then by the definition of antikeys, $X^+ = R$ for every subset $X$ of $R$. In particular, $\emptyset^+ = R$, and hence $K_a = \{\emptyset\}$.

[\Leftarrow] If $\emptyset$ is a key in $a$, then by the definition of keys, every subset $X$ of $R$ is a superkey, since $\emptyset \subseteq X$. It follows that $X$ is not in $K_{\emptyset}^-$. Therefore, $K_a^- = \emptyset$. $\Box$

4. RELATIONS REPRESENTING KEYS AND FUNCTIONAL DEPENDENCIES

Definition 4.1

Let $F$ and $G$ be the sets of FDs. Denote by $F^+$ the set $\{f \mid F \models f\}$. $F$ and $G$ are equivalent if $F^+ = G^+$.

Definition 4.2

Let $r$ be a relation with the set of attributes $R$. Denote by FD($r$) the set of all FDs on $R$ that hold in $r$. Clearly, $(FD(r))^+ = FD(r)$.

Definition 4.3

Let $M$ be a family of subsets of a given set. $M$ is a Sperner system [25] if $X, Y \in M$ and $X \subseteq Y$ imply $X \neq Y$.

It is easy to see that the families of keys and antikeys of a given RS are Sperner systems.

Definition 4.4

Let $r$ be a relation on $R$, and let $K$ be a Sperner system on $R$. $r$ represents $K$, if $K = K_a$, where $a$ is the RS $(R, FD(r))$.

Definition 4.5

Let $r$ be a relation on $R$, $X \subseteq R$, and let $u$ and $v$ be two tuples in $r$. Denote by $E(u, v)$ the set $\{A \in R \mid u(A) = v(A)\}$. We say that $u$ and $v$ agree exactly on $X$ if $E(u, v) = X$. Define the following sets. $agr(r) = \{E(u, v) \mid u, v \in r$ and $u \neq v\}$, and $magr(r) = \text{MAX}(agr(r))$.

Definition 4.6

Let $a = (R, F)$ be an RS and let $r$ be a relation on $R$. $r$ represents RS $a$, if $FD(r) \supseteq F^+$. $r$ exactly represents $a$, if $FD(r) = F^+$. If $r$ exactly represents $a$ then we also say that $r$ is an Armstrong relation for $a$ [6,18,19].

Our aim is to show a connection between the following well-known results.

Theorem 4.1

[9,14] Let $r$ be a relation on $R$, and $K$ be a Sperner system on $R$. Then $r$ represents $K$ if and only if,

$$magr(r) = K_a^-,$$ where $a = (R, FD(r)).$ \hspace{1cm} (4.1)

Theorem 4.2

[19] Let $r$ be a relation on $R$. Then $r$ represents RS $a = (R, F)$ if and only if,

$$agr(r) \subseteq C_a.$$ \hspace{1cm} (4.2)
Theorem 4.3

Let \( r \) be a relation on \( R \). Then \( r \) exactly represents \( RS a = (R, F) \) if and only if,

\[
G_a \subseteq \text{agr}(r) \subseteq C_a. \tag{4.3}
\]

From Theorem 3.2 and the definitions of closed sets, generators and coatoms, we know that \( K_a^- = T_a \subseteq G_a \subseteq C_a \).

Theorem 4.4

\[
(4.3) \iff (4.1) \iff (4.2)
\]

Proof. \((4.3) \Rightarrow (4.2)\) is obvious.

\((4.3) \Rightarrow (4.1)\) Let \( a = (R, F) \) and \( r \) be given, and let \( G_a \subseteq \text{agr}(r) \subseteq C_a \). We must show that \( \text{magr}(r) = K_a^- \). By Theorem 4.3, we know that \( r \) exactly represents \( a \), so \( \text{FD}(r) = F^+ \), and hence \( \text{FD}(r) \) and \( F \) are equivalent. Therefore, the RSs \( a = (R, F) \) and \( c = (R, \text{FD}(r)) \) have the same sets of keys, i.e., \( K_a = K_c \). By Definition 4.4, it follows that \( r \) represents Sperner system \( K_a \). Hence, by Theorem 4.1, \( \text{magr}(r) = K_a^- \).

\((4.2) \not\Rightarrow (4.1)\) Consider the RS \( a = (R, F) \), where \( R = ABC \) and \( F = \{A \rightarrow BC\} \). We have \( K_a = \{A\} \), \( C_a = \{ABC, BC, B, C, \emptyset\} \), and hence \( G_a = \{BC, B, C\} \) and \( K_a^- = T_a = \{BC\} \). Let \( r \) be the following relation,

\[
\begin{array}{ccc}
A & B & C \\
0 & 0 & 0 \\
1 & 1 & 0 \\
2 & 2 & 2 \\
\end{array}
\]

Then, \( \text{agr}(r) = \{C, \emptyset\} \), so \( \text{magr}(r) = \{C\} \neq K_a^- \). It follows that \( r \) does not represent \( K_a \). But \( r \) represents \( a \), since \( \text{agr}(r) \subseteq C_a \).

\((4.1) \not\Rightarrow (4.2) \text{ and } (4.1) \not\Rightarrow (4.3)\) Let \( a = (R, F) \), where \( R = ABC \) and \( F = \{A \rightarrow C, B \rightarrow C\} \). We have \( K_a = \{AB\} \), \( C_a = \{ABC, AC, BC, C, \emptyset\} \), \( G_a = \{AC, BC, \emptyset\} \) and \( T_a = K_a^- = \{AC, BC\} \). Consider the following relation \( r \),

\[
\begin{array}{ccc}
A & B & C \\
0 & 0 & 0 \\
0 & 1 & 0 \\
2 & 0 & 0 \\
2 & 3 & 3 \\
\end{array}
\]

We have \( \text{agr}(r) = \{AC, BC, A, C, \emptyset\} \), so \( \text{magr}(r) = \{AC, BC\} = K_a^- \). Thus, \( r \) represents \( K_a \). On the other hand, since \( \text{agr}(r) \not\subseteq C_a \), by Theorem 4.2, \( r \) does not represent \( a \), and, of course, it does not represent exactly \( a \).

\((4.2) \not\Rightarrow (4.3)\) Immediate from the facts that \((4.3) \Rightarrow (4.1)\) and \((4.2) \not\Rightarrow (4.1)\).

The proof is complete. \(\square\)
5. TRANSLATION OF RELATION SCHEMES

First, we present some necessary definitions and previous results.

**Definition 5.1**

Let \( a = (R, F) \) be an RS and let \( X \) be a subset of \( R \). For each object \( E \) (\( E \) may be a set of attributes, set of FDs or relation scheme itself), we construct a new object, denoted by \( E \setminus X \), by removing from \( E \) all the occurrences of symbols corresponding to attributes in \( X \) as follows.

- For each subset \( Y \) of \( R \), we define \( Y \setminus X = Y - X \), where \( Y - X \) is the set difference between \( Y \) and \( X \).
- For each FD \( f = Y \to Z \), we define \( f \setminus X = Y \setminus X \to Z \setminus X \).
- \( F \setminus X = \{ f \setminus X | f \in F \} \).
- \( a \setminus X = (R \setminus X, F \setminus X) \).

**Definition 5.2**

An RS \( e = (P, G) \) is obtained by translating RS \( a = (R, F) \) on \( X \), if \( X \subseteq R \) and \( e = a \setminus X \). That is, \( P = R - X \) and \( G = F \setminus X \).

**Theorem 5.1**

\([11,12,15,16]\) Let \( a = (R, F) \) be an RS and let \( X \) and \( Y \) be two disjoint subsets of \( R \). Then \( (XY)^+ = X(Y)^+ \setminus X \).

Note that \( X \) and \( Y \) are disjoint, since closure of \( Y \) is computed in \( a \setminus X \), which does not contain \( X \).

**Corollary 5.1**

\([16]\) Let \( a = (R, F) \) be an RS and let \( X \) and \( Y \) be two subsets of \( R \) such that \( X \subseteq Y \subseteq X^+ \).

Then \( X^+_Y = Y(X)^+ \setminus a \).

From Theorem 5.1 we obtain \( X^+_a = X(Y)^+ \setminus a \), and so \( X^+_a - X = (\emptyset)^+ \setminus a \). The problem of finding the closure of \( X \) on RS \( a \) thus is reduced to the much simpler problem of finding the closure of the empty set on RS \( a \setminus X \). It is clear that RS \( a \setminus X \) is simpler than the original RS \( a \) by the following.
- \( a \setminus X \) may have fewer attributes and shorter dependencies than \( a \).
- \( a \setminus X \) may have fewer dependencies than \( a \). Indeed, if in \( a \setminus X \) there are some trivial FDs, i.e. those of the form \( Y \to Z, Y \supseteq Z \), then we can remove them from \( a \).
- If in \( a \setminus X \) there are some FDs of the form \( \emptyset \to Y, Y \neq \emptyset \), then we can add \( Y \) to the result for \( X^+_a \), and continue to find the closure of the empty set in the new RS \( (a \setminus X) \setminus Y \). After the second translation, the FD \( \emptyset \to Y \) will be of the form \( \emptyset \to \emptyset \), so it must be removed from \( a \).
- If in \( a \setminus X \) there are some duplicate dependencies, then we can, of course, eliminate them.
- When some attributes of the closure of the empty set of \( a \setminus X \) are found, say \( Y \), then, by Corollary 5.1, we can add \( Y \) to the result for \( X^+_a \), and continue to find the closure of the empty set in the new RS \( (a \setminus X) \setminus Y = a \).

It is not hard to see from Definitions 5.1 and 5.2 that translating an RS \( a = (R, F) \) can be done in \( O(|F| \cdot |R|) \) time. Note that \( |F| \cdot |R| \) is the length of the representation of RS \( a \). Thus translating an RS can be done in linear time on the length of its representation. Note also that Theorem 5.1 was proved for a more general case, where \( F \) may contain multivalued and join dependencies besides FDs \([16]\) (for the definitions of multivalued and join dependencies see \([23,26]\).)

Let \( M \) and \( N \) be the families of subsets of a given set \( R \) and let \( Z \) be a subset of \( R \). We define \( M \oplus N \) to be \( \{ XY | X \in M, Y \in N \} \) and \( Z \oplus N \) to be \( \{ ZY | Y \in N \} \). The next theorem is known as the first form of key representation.
Theorem 5.2

\[15,16\] Let \( a = (R, F) \) be an RS and let \( X \) be a subset of \( R \). Let \( e = a \setminus X \). Then,
1. \( K_a = K_e \) if and only if \( X \subseteq N_a \).
2. \( K_a = X \oplus K_e \) if and only if \( X \subseteq I_a \).

By the results of the previous sections, we know that \( I_a \) can be found in \( O(|F| \cdot |R|) \) time by the formula \( I_a = R - \bigcup \{ \text{RIGHT}(f) - \text{LEFT}(f) | f \in F \} \), but computing \( N_a \) is NP-complete. The next theorem shows how a part of \( N_a \) can be computed in polynomial time.

Lemma 5.1

\[4,15,16,21\] Let \( a = (R, F) \) be an RS and let \( R' \) be the set \( \bigcup \{ \text{RIGHT}(f) | f \in F \} - \bigcup \{ \text{LEFT}(f) | f \in F \} \). Then,
1. \( R' \subseteq N_a \).
2. If \( X \subseteq I_a \) and \( Y \subseteq N_a \), then \((XY)^+ - X \subseteq N_a \).

Example 5.1

(Cont.) By Example 1.1 we have, \( I_a = ABCDEFG - BCDE = AFG \), \( R' = BCDE - ABCE = D \), so \((R'I_a)^+ = (ADFG)^+ = ADFG\). After translating RS \( a \) on ADFG we get RS \( e = (P, H) \), where \( P = ABCDEFG - ADFG = BCE \) and \( H = \{ BC \to E, E \to BC \} \). Since \( K_e = \{ E, BC \} \), by Theorem 5.2, it follows that \( K_a = I_a \oplus K_e = AFG \oplus \{ E, BC \} = \{ AEFG, ABCFG \} \).

Using Theorems 5.1 and 5.2 we can get the following result which is called the second form of key representation.

Theorem 5.3

\[15,16\] Let \( a = (R, F) \) be an RS. Then every key \( X \) of \( a \) can be represented in the form \( X = LY \), where \( L \) is the left side of some FD in \( F \), and \( Y \) is a key of RS \( a \setminus L^+ \).

Now, we present some new results about the change of closed sets, generators and coatoms (antikeys) in the translation.

Theorem 5.4

Let \( a = (R, F) \) be an RS and let \( X \) and \( Y \) be two disjoint subsets of \( R \). Let \( e = a \setminus X \). Then,
1. \( XY \in C_a \) if and only if \( Y \in C_e \).
2. \( XY \in G_a \) if and only if \( Y \in G_e \).
3. \( XY \in T_a \) if and only if \( Y \in T_e \).

Proof. 1.\( \Rightarrow \) If \( XY \) is closed in \( a \), then by Theorem 5.1 we have \( XY = (XY)^+ = X(Y)^+ \). Hence \( Y = Y^+ \), and so \( Y \) is closed in \( e \).

2.\( \Rightarrow \) Let \( XY \) be in \( G_a \). Then by Fact 1 above, \( Y \in C_e \), since \( G_a \subseteq C_a \). Let \( Y = Z \cup W \) for some \( Z \) and \( W \) in \( C_e \). Since \( Z \) and \( W \) are two subsets of attributes in RS \( e \), \( X \cap Z = X \cap W = \emptyset \). By Fact 1 above, \( Z \) and \( W \) are closed in \( e \). We have, \( XY = XZ \cap XW \). By Theorem 2.1, it follows that \( XY \) must be equal to \( Z \) or/and \( W \), and so \( Y \) must be equal to \( Z \) or/and \( W \). Hence \( Y \in G_e \).

3.\( \Rightarrow \) Let \( Y \in G_e \). By Fact 1, \( XY \in C_a \), since \( G_a \subseteq C_a \). Assume \( XY = Z \cup W \) for some \( Z \) and \( W \) in \( C_a \). Then \( XY \subseteq Z \) and \( XY \subseteq W \). Consider the sets \( Z' = Z \cup XY \) and \( W' = W \cup XY \). We have two partitions, \( Z = XYZ' \) and \( W = XYW' \). By Fact 1, \( Z' \) and \( W' \) are in \( C_e \). But \( Y = YZ' \cap YW' \), and since \( Y \subseteq G_e \), by Theorem 2.1, it follows that \( Y \) is equal to \( YZ' \) or/and \( YW' \). Hence \( XY \) is equal to \( Z \) or/and \( W \). Thus \( XY \in G_a \). Fact 2 is proved.

3.\( \Rightarrow \) Suppose that \( XY \) is in \( T_a \). Then, by Fact 2, \( Y \in G_e \), since \( T_a \subseteq G_a \). Suppose that \( Y \subseteq W \) for some \( W \) in \( G_e \). It follows that \( XY \subseteq XW \), and by Fact 2, \( XW \subseteq G_a \). By Theorem 2.2 we know that \( T_a = \text{MAX}(G_a) \), so \( XY = XW \). But \( X \cap Y = X \cap W = \emptyset \), hence \( Y = W \). Thus \( Y \in T_e \).
Let $Y \in T_e$. Then by Fact 2, $XY \in G_e$, since $T_e \subseteq G_e$. Suppose that $XY \subseteq Z$ for some $Z$ in $G_e$. Put $Z' = Z - XY$. We have $Z = XYZ' \in G_e$, so, by Fact 2, $YZ' \in G_e$. Hence, by Theorem 2.2, $Y = YZ'$. But $Y \cap Z' = \emptyset$, so $Z' = \emptyset$. Therefore, $Z = XY$, which shows that $XY \in T_a$. Fact 3 is proved, and this completes the proof of Theorem 5.4.

Corollary 5.2

Let $a = (R, F)$ be an RS and $X \subseteq R$. Let $e = a\backslash X$. Then $X$ is a coatom of $a$ if and only if $C_e$ contains exactly two members $\emptyset$ and $R - X$.

Proof. [\(\Rightarrow\)] Let $X$ be a coatom of $a$. Then $X \neq R$. Since in any RS, the set of all attributes is closed, it follows that $R - X \in C_e$. By an application of Fact 3 of Theorem 5.4 with $Y = \emptyset$, we have $\emptyset \in T_e$, and so $\emptyset \in C_e$, since $T_e \subseteq C_e$. If $R - X = \emptyset$, then $R = X$; a contradiction. Therefore, $R - X \neq \emptyset$. Thus $\emptyset$ and $R - X$ are two different members in $C_e$. By the definition of coatoms, and by the fact that $\emptyset \in T_e$, we have $T_e = \text{MAX}(C_e - \{R - X\}) = \{\emptyset\}$. Hence $C_e = \{\emptyset, R - X\}$.

[\(\Leftarrow\)] Assume that $C_e = \{\emptyset, R - X\}$. Then by the definition of coatoms, $T_e = \text{MAX}(C_e - \{R - X\}) = \text{MAX}(\{\emptyset\}) = \{\emptyset\}$. Now, by Fact 3 of Theorem 5.4 we have, $X\emptyset = X \in T_a$. □

Combining Theorems 3.2 and 3.6, and Corollary 5.2 we get the following.

Corollary 5.3

Let $a = (R, F)$ be an RS, $X \subseteq R$, and $e = a\backslash X$. The following are equivalent.

1. $X \in K_e$.
2. $C_e = \{\emptyset, R - X\}$.
3. $K_e = \{\emptyset\}$.
4. $K_a = R - X$.

Conditions 1 and 4 of Corollary 5.3 above give a basis for testing whether a subset of attributes is an antikey.

Algorithm 5.1

ANTIKEY.

Input: An RS $a = (R, F)$ and a subset $X$ of $R$.

Output: TRUE if $X$ is an antikey of $a$; FALSE, otherwise.

ANTIKEY($a, X$)

1. Translation: $e := (R - X, F\backslash X)$;
2. If $\emptyset^* = R - X$ or if there is an attribute $A$ in $R - X$ such that $A^* \neq R - X$ then RETURN(FALSE) else RETURN(TRUE).

Step 1 requires $O(|F| \cdot |R|)$ time. Step 2 requires $O(|F| \cdot |R|^2)$ time, since for each attribute $A$ in $R - X$, $A^*$ can be found in $O(|F| \cdot |R|)$ time. Hence algorithm ANTIKEY requires $O(|F| \cdot |R|^2)$ time.

Now, we give the first form of representation of antikeys.

Theorem 5.5

Let $a = (R, F)$ be an RS such that $K_a \neq \emptyset$. Let $X$ be a subset of $R$ and let $e = a\backslash X$. Then,

1. $K_e = K_a$ if and only if $X = \emptyset$.
2. $K_a = X \oplus K_e$ if and only if $X \subseteq I_e$.

Proof. 1. [\(\Rightarrow\)] is obvious.

[\(\Leftarrow\)] Let $K_e = K_a$. Then $K_e \neq \emptyset$, since $K_a \neq \emptyset$. Let $Y$ be in $K_a$. By Fact 3 of Theorem 5.4, $XY$ is in $K_a$, so it is in $K_a$, since $K_e = K_a$. But, by the definition of translation, $X$ is not in RS $e$, and hence $X = \emptyset$.

2. [\(\Rightarrow\)] If $K_e = X \oplus K_a$, then $X \subseteq \cap K_a = I_a$.

[\(\Leftarrow\)] Let $X \subseteq I_a$. Then each member in $K_e$ contains $X$. Since $K_e = T_e$ and $K_a = T_a$, by Fact 3 of Theorem 5.4, we have $X \oplus K_e = K_a$. □
6. CONCLUSION

Our further research will be dedicated to the following problems.

1. Demetrovics and Thi [14] constructed an algorithm for computing antikeys from a given set of keys. It is natural to form the following problem. Find algorithms for computing antikeys from a given relation scheme.

2. In this paper we have got only the first form of representation of antikeys. What is about their second form.

3. What is the connection between Armstrong relations of the original relation schemes and those of translated ones.

Acknowledgments

We would like to thank Rónyai Lajos for carefully reading earlier versions of the paper.

**This work has been written while the second author has been a visiting researcher at the Computer and Automation Institute of the Hungarian Academy of Sciences.

REFERENCES