Tạp chí Tin học và Điều khiển học, T. 12, S. 4 (1996) (101–113)

SOME MORE PROPERTIES AND REMARKS ABOUT KEYS FOR RELATION SCHEME

HO THUAN, SOUAFI SOUAD, MOHAMED BENKADA DJAMILA

Abstract. In this paper we prove some additional properties of keys and superkeys for relation schemes. Basing on these properties, some algorithms finding keys for relation schemes are improved and their complexities are estimated.

Finally, some remarks on the translations of relation schemes are also given.

1. INTRODUCTION

The relation model was first introduced by E. F. Codd in June 1970, in his famous paper *A relational model of data for large shared data banks* [1]. Its objective was to permit a formal description of the different problems encountered.

We here recall some important notions and results about the relational model.

The notation $R(A_1, A_2, ..., A_n)$, where $\Omega = \{A_1, A_2, ..., A_n\}$ will stand for the relation scheme R defined on the attributes Ω. R expresses a connection between the attributes of Ω.

An extension of a relation scheme R defined on the set of attributes Ω, is a subset of the Cartesian product of D_1, $D_2, ..., D_n$ where D_i is the domain of the possible values for the attribute A_1. The extension r of a relation scheme R is a possible realization. We also call it occurrence.

An extension r of a relation scheme can be represented as a table where each column corresponds to an attribute and each line to a tuple.

A constraint is a condition defined either on a relation (intra-relation), or between relations (inter-relation). The test of validity can be done algorithmically on the extension of a relation. Let R be a relation scheme on the set of attributes satisfied by R. Then R is also used to denote a relation scheme.

The functional dependencies (FD) are a particular class of constraints

We say that Y is functionally dependent on X, with $X, Y \subseteq \Omega$ if and only if:

$$\forall r \in R,\ \forall t,\ s \in r : t[X] = s[X] \Rightarrow t[Y] = s[Y]$$

We note $X \rightarrow Y$. We also say that X determines Y.

In the following, we will consider as constraints only the functional dependencies. Let F be a set of functional dependencies where each FD of F holds in R, $R\langle\Omega, F\rangle$ F is the relation scheme.

From a set of functional dependencies, others, can be obtained by Armstrong's axioms, and other inference rules can be derived from these axioms.

Armstrong's axioms are:

Let X, Y, $Z \subseteq \Omega$

Reflexivity : If $Y \subseteq X$, then $X \to Y$

Augmentation: If $X \to Y$ then $XZ \to YZ$

Transitivity: if $X \to Y$ and $Y \to Z$ then $X \to Z$

The following rules are easily obtained from Armstrong's axioms:

Union: If $X \to Y$ and $X \to Z$ then $X \to YZ$

Pseudo-transitivity: If $X \to Y$ and $YW \to Z$ then $XW \to Z$

Decomposition: If $X \to Y$ and $Z \subseteq Y$ then $X \to Z$

Let F be a set of functional dependencies. The closure F^+ of F is the set containing F and all the functional dependencies that can be derived from Armstrong's axioms.

Let $X \subseteq \Omega$, the closure of X with respect to a set of functional dependencies F is the set X_F^+, where:

$$X_F^+ = \{A | A \in \Omega,\ (X \to A) \in F^+\}$$

We know that $X \to Y$ is obtained by Armstrong's axioms if and only if $Y \subseteq X_F^+$, i.e.

$$X \to Y \in F^+ \Leftrightarrow Y \subseteq X_F^+$$

See [2] for a proof.

Let Ω be a set of attributes and F a set of functional dependencies over Ω

$$F = \{L_k \to R_k | L_k \cap R_k = \emptyset,\ L_k,\ R_k \subseteq \Omega\}$$
$$L = \cup L_k \text{ and } R = \cup R_k$$

Let $X \subseteq \Omega$, Beeri and Bernstein [3] proposed a linear time algorithm to compute X_F^+.

Algorithm:

1) Establish the sequence $X^{(0)}$, $X^{(1)}$, ..., as follows:

$$X^{(0)} = X$$
$$\ldots$$
$$X^{(i+1)} = X^{(i)} \cup (\cup R_k)$$
$$(L_k \to R_k) \in F$$
$$L_k \subseteq C^{(i)}$$

2) It is obvious that:

$$X^{0)} \subseteq X^{(1)} \subseteq \cdots \subseteq X^{(i)} \subseteq \cdots \subseteq \Omega$$

Since Ω is a finite set, there exists a smallest non negative integer such that

$$X^{(t)} = X^{(t+1)}$$

3) We have $X^+ = X^{(t)}$.

Let $S = \langle \Omega, F \rangle$ be a relation scheme, $\Omega = \{A_1, A_2, ..., A_n\}$. A subset X of Ω is called a key for S if X satisfies the following two conditions:

1. $(X \to A_1 A_2 ... A_n) \in F^+$ An)
2. $\forall Y,\ Y \subset X,\ (Y \to A_1 A_2 ... A_n) \notin F^+$

The subsets X of Ω satisfying the first condition are called superkeys.

Let F be a set of functional dependencies:

$$F = \{L_i \to R_i \mid i = 1, ..., k,\ L_i,\ R_i \subseteq \Omega\}$$
$$L = \cup L_i \quad i = 1, ..., k$$
$$R = \cup R_i \quad i = 1, ..., k$$

Without loss of generality, we will assume in this paper that:

$$\forall i = 1, ..., k,\ L_i \cap R_i = \emptyset$$

2. PROPERTIES

The following properties and lemmas are obvious:

Property 1. $F_1^+ \cup F_2^+ \subseteq (F_1 \cup F_2)^+$.

Property 2. $((X_{F_1}^+)_{F_2}^+ \subseteq X_{F_1 \cup F_2}^+$.

Remark: $X_{F_1 \cup F_2}^+ \subseteq ((X_{F_1}^+)_{F_2}^+$ is not always true. For example, let:

$$F_1 = \{A \to B,\ B \to C,\ D \to E\}$$
$$F_2 = \{A \to D\}$$

Suppose that: $X = \{A\}$

Then $X^+_{F_1} = \{A, B, C\}$

$(X^+_{F_1})^+_{F_2} = \{A, B, C, D\}$

$F^+_{F_1 \cup F_2} = \{A, B, C, D, E\}$.

So $F^+_{F_1 \cup F_2} \not\subseteq (X^+_{F_1})^+_{F_2}$.

Corollary 1. *If $F_1 \subseteq F_2$,, then $(X^+_{F_2})^+_{F_1} = (X^+_{F_1})^+_{F_2} = X^+_{F_2}$*

Lemma 1. $F_1 \subseteq F_2 \Rightarrow \forall i,\ X^{(i)}_{F_1} \subseteq X^{(i)}_{F_2}$

Corollary 2. $F_1 \subseteq F_2 \Rightarrow X^+_{F_1} \subseteq X^+_{F_2}$

Lemma 2. *Let $S = \langle \Omega, F \rangle$ be a relation scheme, and $X, Y \subseteq \Omega$, then*

$$(XY)^+_F = (X^+_F \cup Y)^+_F = (X \cup Y^+_F)^+_F$$

See the proof in [4].

Corollary 3. *Let $X \subseteq \Omega$*

$X \cup (\Omega - R)$ *is a superkey of* $S = \langle \Omega, F \rangle \Leftrightarrow F' = \emptyset$

With $F' = F - \{L_k \rightarrow R_k | (L_k \rightarrow R_k) \in F,\ L_k \subseteq (X \cup (\Omega - R))^+_H$.

Theorem 1. *Let $X, Y \subseteq \Omega$, $X \subseteq Y$ and F be a set of functional dependencies over Ω. We define F' a set of functional dependency as:*

$$F' = F - \{L_k \rightarrow R_k | (L_k \rightarrow R_k) \in F,\ L_k \subseteq X^+_F\},\ \textit{so}\ Y^+_F = (X^+_F \cup Y)^+_F$$

Proof. We first show that: $(X^+_F \cup Y)^+_F \subseteq Y^+_F$.

Since $X \subseteq Y$ then

$$Y^+_F = (XY)^+_F = (X^+_F \cup Y)^+_F$$

By lemma 2, from $F' \subseteq F$ we get

$$(X^+_F \cup Y)^+_F \subseteq (X^+_F \cup Y)^+_F$$

So

$$(X^+_F \cup Y)^+_F \subseteq Y^+_F$$

We show that $Y^+_F \subseteq (X^+_F \cup Y)^+_F$, by induction on the order of iteration in the algorithm for computing the closure of Y with respect to F.

For $i = 0$, $Y^{(0)}_F = Y \subseteq (X^+_F \cup Y)^+_F$.

Suppose the claim is true till i, i.e, $Y_F^{(i)} \subseteq (X_F^+ \cup Y)_F^+$.

Let $A \in Y_F^{(i+1)}$.

If $A \in Y_F^{(i)}$ then by the inductive hypothesis, we have $A \in (X_F^+ \cup Y)_F^+$.

If $A \notin Y_F^{(i)} \Rightarrow \exists(K_k \to R_k) \in F$ with $A \in R_k$ and $L_k \subseteq Y_F^{(i)}$, so by the inductive hypothesis, we have $L_k \subseteq (X_F^+ \cup Y)_F^+$.

If $L_k \to R_k) \in (F - F')$ then $L_k \subseteq X_F^+$, so we have $A \in R_k \subseteq X_F^+$ and thus $A \in (X_F^+ \cup Y)_F^+$.

If $(L_k \to R_k) \in F'$, since $L_k \subseteq (X_F^+ \cup Y)_F^+$ then $A \in R_k \subseteq (L_k)_F^+ \subseteq (X_F^+ \cup Y)_F^+$.

So $Y_F^{(i+1)} \subseteq (X_F^+ \cup Y)_F^+$. Therefore $Y_F^+ \subseteq (X_F^+ \cup Y)_F^+$

Finally: $Y_F^+ = (X_F^+ \cup Y)_F^+$.

Lemma 2. $X_{F_1}^+ \cup X_{F_2}^+ \subseteq X_{F_1 \cup F_2}^+$.

Proof. Since $F_1 \subseteq F_1 \cup F_2$ and $F_2 \subseteq F_1 \cup F_2$ then by Corollary 2, we have

$$X_{F_1}^+ \subseteq X_{F_1 \cup F_2}^+,\ X_{F_2}^+ \subseteq X_{F_1 \cup F_2}^+$$

So $X_{F_1}^+ \cup X_{F_2}^+ \subseteq X_{F_1 \cup F_2}^+$.

Remark: $X_{F_1 \cup F_2}^+ \subseteq X_{F_1}^+ \cup X_{F_2}^+$ is not always true. For example, let:

$F_1 = \{A \to B,\ B \to C,\ D \to E\}$, $F_2 = \{A \to D\}$. Suppose $X = \{A, B, C\}$.

$X_{F_2}^+ = \{A, D\}$

$X_{F_1}^+ \cup X_{F_2}^+ = \{A, B, C, D\}$

$X_{F_1 \cup F_2}^+ = \{A, B, C, D, E\}$

$X_{F_1 \cup F_2}^+ \not\subset X_{F_1}^+ \cup X_{F_2}^+$.

Theorem 2. *Let $S = \langle \Omega, F \rangle$ be a relation scheme and $X \subseteq \Omega$ a key of S. Then:*

$$(\Omega - R) \subseteq X \subseteq (\Omega - R) \cup ((L \cap R) - (\Omega - R)_F^+)$$

Proof. From a result of Thuan and Bao [5], we know that if X is a key then:

$$(\Omega - R) \subseteq X \subseteq (\Omega - R) \cup (L \cap R)$$

To show that: $(\Omega - R) \subseteq X \subseteq (\Omega - R) \cup ((L \cap R) - (\Omega - R)_F^+)$, it is sufficient to show that:

$$X \cup ((\Omega - R)_F^+ - (\Omega - R)) = \emptyset$$

Suppose not, that is $\exists X$, such that $X \cap ((\Omega - R)_F^+ - (\Omega - R)) \neq \emptyset$.

Let $A_1 \in (X \cap ((\Omega - R)^+_F - (\Omega - R)))$

$\Rightarrow A_1 \in X$ and $A \notin (\Omega - R)$ and $A_1 \in (\Omega - R)^+_F$

$\Rightarrow A_1 \in X$ and $A_1 \notin (\Omega - R)$ and $((\Omega - R) \to A_1) \in F^+$

$\Rightarrow (\Omega - R) \subseteq X - \{A_1\}$ and $((\Omega - R) \to A_1) \in F^+$

$\Rightarrow (X - \{A_1\} \to (\Omega - R)) \in F^+$ and $((\Omega - R) \to A_1) \in F^+$

$\Rightarrow (X - \{A_1\} \to A_1) \in F^+$.

Since $A_1 \in X$ and $(X - \{A_1\} \to A_1) \in F^+$ then by a lemma in [5], X is not a key, which is contradiction.

So $X \cap ((\Omega - R)^+_F - (\Omega - R)) = \emptyset$. And then:

$$(\Omega - R) \subseteq X \subseteq (\Omega - R) \cup ((L \cap R) - (\Omega - R)^+_F)$$

E.g.: Let's have the relation scheme $\langle\{a, b, c, g, h\}, F\rangle$, with:

$F = \{a \to b,\ b \to c,\ g \to h,\ h \to g\}$

$L = \{a, b, g, h\}$

$R = \{b, c, g, h\}$

$\Omega - R = \{a\}$

$(\Omega - R)^+_F = \{a, b, c\}$

$L \cap R = \{b, g, h\}$

$L \cap R - (\Omega - R)^+_F = \{g, h\} \subset (L \cap R)$

Property 3. *Let $S = \langle\Omega, F\rangle$ be a relation scheme.*

Let $L \cap R = \{A_{11}, A_{12}, ..., A_{1n}\} \neq \emptyset$ and $(\Omega - R)^+_F \neq \Omega$.

Then $\forall A_{1k} \in (L \cap R)$, $((\Omega - R) \cup (L \cap R)) - \{A_{1k}\}$ is a superkey of S.

Proof. To show that $\forall A_{1k} \in (L \cap R)$, $((\Omega - R) \cup (L \cap R)) - \{A_{1k}\}$ is a superkey of S, we will suppose the opposite, that is: $\exists A_{1p} \in (L \cap R)$ such that: $((\Omega - R) \cup (L \cap R)) - \{A_{1p}\}$ is not a superkey of S, then the attribute A_{1p} is essential to the superkey $(\Omega - R) \cup (L \cap R)$, and therefore, this attribute will exist in all the keys of S included in $(\Omega - R) \cup (L \cap R)$. Since $G = (\Omega - R)$ is the intersection of all the keys [5], then $A_{1p} \in (\Omega - R)$, but this contradicts the fact that $A_{1p} \in (L \cap R)$. So the supposition is false and:

$\forall A_{1k} \in (L \cap R)$, $((\Omega - R) \cup (L \cap R)) - \{A_{1k}\}$ is a superkey of S.

Property 4. *If $R' = R - L$ then $R'^+_F = R'$ and $(GR')^+_F = G^+_F \cup R'$.*

Proof. Since $R' = R - L$ then $\forall(L_1 \to R_i) \in F$, $L_1 \cap R' = \emptyset$.

It is obvious that $R'^+_F = R'$.

It is shown that $(GR')^+_F = (G^+_F \cup R')^+_F$, so we will show that:

$(G_F^+ \cup R')_F^+ = G_F^+ \cup R'$.

Using the algorithm for computing the closures of $X = G_F^+ \cup R'$, we have:

$X^{(0)} = G_F^+ \cup R'$

$X^{(1)} = X^{(0)} \cap (\cup R_k)$

$(L_k \to R_k) \in F$

$L_k \subseteq X^{(0)}$

Since $L_k \subseteq X^{(0)} = G_F^+ \cup R'$ and $\forall (L_i \to R_i) \in F,\ L_1 \cap R' = \emptyset$ so $L_k \subseteq G_F^+$.

Since $L_k \subseteq G_F^+$ then $R_k \subseteq G_F^+$.

Thus $\forall (L_k \to R_k) \in F$ where $L_k \subseteq G_F^+$, then $R_k \subseteq G_F^+$.

Therefore $X^{(1)} = X^{(0)}$ and $X_F^+ = G_F^+ \cup R'$

Finally $(GR')_F^+ = (G_F^+ \cup R')_F^+ = G_F^+ \cup R'$

3. REMARKS ON THE ALGORITHMS FOR FINDING KEYS

The improvements we are going to do are based on Theorems 1 and 2.

Let $S = \langle \Omega, F \rangle$ be a relation scheme.

If $X \subseteq \Omega$ a key of S then by Theorem 2:

$$(\Omega - R) \subseteq X \subseteq (\Omega - R) \cup ((L \cap R) - (\Omega - R)_F^+)$$

Since $(\Omega - R)_F^+$ is computed, then we can use it to eliminate attributes from $(L \cap R)$ without introducing a new computing.

By Theorem 1, we have:

$$Y_F^+ = (X_F^+ \cup Y)_F^+$$

with $X, Y \subseteq \Omega,\ X \subseteq Y$.

$$F' = Fv - \{L_k \to R_k | (L_k \to R_k) \in F,\ L_k \subseteq X_F^+\}$$

Let $X = (\Omega - R)$

Then $F' = F - \{L_k \to R_k | (L_k \to R_k) \in F$ and $L_k \subseteq (\Omega - R)_F^+\}$

$Y_F^+ = ((\Omega - R)_F^+ \cup Y)_F^+$

Here also, since $(\Omega - R)_F^+$ is computed, its use can only improve the algorithms.

Algorithms of Thuan and Bao for finding a key

Algorithm 1:

$X := (\Omega - R)^+_F$

if $Z = \Omega$ then $\Omega - R$ is the unique key of the scheme

else $X := (\Omega - R) \cup ((L \cap R) - Z)$

for $i := 1$ to $|(L \cup R) - Z|$ do

if $(Z \cup (X - \{A_{ti}\}))^+_F = \Omega$ then $X := X - \{A_{ti}\}$

$K := X$: $\{K$ is a key of the scheme $S\}$

Complexity: The computing of Z requires $|\Omega|.|F|$ operations.

The instruction $(Z \cup (X - \{A_{ti}\}))^+_F = \Omega$, requiring $|\Omega|.|F'|$ operations, runs $|(L \cap R) - Z|$ times.

So the complexity of the algorithm is:

$$O(|\Omega|.|F| + |(L \cap R) - Z|.|\Omega|.|F'|)$$

Algorithm of Lucchesi and Osborn for finding all the keys

An improvement of the algorithm of Lucchesi and Osborn was proposed by Thuan [6, 7]

Let $F'' = F - \{L_1 \to R_1 | (L_1 \to R_1) \in F$ and $R_1 \subseteq (R - L)\}$.

Algorithm 2: $\mathcal{K} := \{K\}$

where k is a key included in the super key $(\Omega - R) \cup (L \cap R)$ and found by the algorithm of Thuan and Bao for finding a key (see [5]).

for each key FD $(L_j \to R_j) \in F''$ where $K_i - R_j \neq K_i$ do

$T := L_j \cup (K_i - R_j)$

test := true

for each key C of $\mathcal{K}$ do

if $C \subseteq T$ then test := false

if test then $\mathcal{K} := \mathcal{K} \cup \{T'\}$

where T' is the key included in the super key T and found by the second algorithm of Thuan and Bao (see [5]).

Complexity: The computing of K in the instruction $\mathcal{K} := \{K\}$ requires:

$(|\Omega|.|F| + |L \cap R|.|\Omega|.|F|)$ operations.

The instruction $C \subseteq T$ requires $|\Omega|$ operations and runs $|\mathcal{K}|^2.|F''|$ times.

The instruction $\mathcal{K} := \mathcal{K} \cup \{T'\}$ requires $|L \cap R|.|\Omega|.|F|$ operations and runs $(|\mathcal{K}| - 1)$ times.

$$|\Omega|.|F| + |L \cap R|.|\Omega|.|F| + |\mathcal{K}|^2.|F''|.|\Omega| + (|\mathcal{K}| - 1).|L \cap R|.|\Omega|.|F| =$$
$$|\Omega|.|F| + |\mathcal{K}|.|\Omega|.(|\mathcal{K}|.(|\mathcal{K}|.|F''| + |L \cup R|.|F|).$$

So the complexity of this algorithm is:

$$O(|\Omega|.|F| + |\mathcal{K}|.|\Omega|.(|\mathcal{K}|.|F'' + |L \cap R|.|F|)$$

We propose now an improvement of this algorithm.

Algorithm 3: $\mathcal{K} := \{K\}$
where K is a key included in the super key Ω and found by the algorithm of Thuan and Bao, after been improved as presented in the algorithm 1. So, $Z = (\Omega - R)^+_F$ is known
for each key K_i of $\mathcal{K}$ do
for each FD $(L_j \to R_j) \in F''$ where $K_i - R_j \neq K_i$ do
$T := L_j \cup (K_i - R_j)$
test := true
for each key C of $\mathcal{K}$ do
if $C \subseteq$T then test = false
if test then $\mathcal{K} := \mathcal{K} \cup \{T'\}$
where T' is the key included in the super key T and found by the algorithm of Thuan and Bao in [5] improved as follows:

Improvements to the algorithm of Thuan and Bao for finding a key included in a super key:
(note that $Z = (Q - R)^+_F$ is known)
if $Tm = Q - R$ then the super key T is also a key, $T' := T$
else $T := (\Omega - R) \cup ((TcapL \cap R) - Z)$
for $i := 1$ to $|(T \cap L \cap R) - Z|$ do
if $(Z \cup (T - \{A_{ti}\}))^+_F = \Omega$ then $T := T - \{A_{t1}\}$
$T' := T$
$\{T'$ is a key of S included in the super key $T\}$

The complexity of this algorithm will then be $O(|(L \cup R) - Z|.|\Omega|.|F'|)$ instead of $O(|L \cup R|.|\Omega|.|F|)$.

Complexity: We are going to compute the complexity of the algorithm of Lucchesi and Osborn, taking into account the improvements.

The computing of K, using the improved algorithm of Thuan and Bao, in instruction $\mathcal{K} := \{K\}$ requires $(|\Omega|.|F| + |(L \cap R) - Z|.|\Omega|.|F'|)$ operations.

The instruction $C \subseteq T$ requires $|\Omega|$ operations, and runs $|\mathcal{K}|^2.|F''|$ times.

The instruction $\mathcal{K} := \mathcal{K} \cup \{T'\}$ requires $(|(L \cap R) - Z|.|\Omega|.|F'|)$ operations and runs $(|\mathcal{K}| - 1)$ times. (T' found will the improved algorithm of Thuan and Bao for finding a key included in a super key).

$|\Omega|.|F| + |(L \cap R) - Z|.|\Omega|.|F'| + |\mathcal{K}|^2.|F''|.|\Omega| + (|\mathcal{K}| - 1).|(L \cap R) - Z|.|\Omega|.|F'| =$
$|\Omega|.|F| + |\mathcal{K}|.|\Omega|.(|\mathcal{K}|.|F''| + |(L \cap R) - Z|.|F'|)$.

Thus the complexity of the algorithm is:

$O(|\Omega|.|F| + |\mathcal{K}|.|\Omega|.(|\mathcal{K}|.|F''| + |(L \cap R) - Z|.|F'|))$ instead of:
$O(|\Omega|.|F| + |\mathcal{K}|.|\Omega|.(|\mathcal{K}|.|F''| + |L \cap R|.|F|)$.

4. SOME REMARKS ON THE TRANSLATIONS OF RELATION SCHEMES

Definition and properties of the translations

Let a have relation scheme $S = \langle \Omega, F \rangle$ where:

$F = \{L_i \to R_i | L_i, R_i \subseteq \Omega, \ i = 1, ..., k\}$.

Let Z be a arbitrary subset of Ω.

We define the relation scheme $\widetilde{S} = \langle \widetilde{\Omega}, \widetilde{F} \rangle$ as follows:

$\widetilde{\Omega} = Omega - Z$ and $\widetilde{F} = \{L_i - Z \to R_i - Z) | (L_i \to R_i) \in F\}$.

The scheme $\widetilde{S}$ is called a Z-translation of the scheme S, and is noted $\widetilde{S} = S - Z$.

The FDs of the form $\emptyset \to \emptyset$ and $X \to \emptyset$ with $X \neq \emptyset$ and $X \subseteq \Omega$, resulting of the translation will be deleted from $\widetilde{F}$.

The deletion of the FDs of the form $\emptyset \to X$, with $X \neq \emptyset$ and $X \subseteq \Omega$,is impossible because semantically this FD indicates that X is always determined. We will come back with more details on this point.

Theorem 3. *Let us have the relation scheme $S = \langle \Omega, F \rangle$ and let $Z \subseteq \Omega$.*
If $\widetilde{S} = S - = Z = \langle \Omega - Z, \widetilde{F} \rangle$ then for all $X \subseteq (\Omega - Z)$, we have:

$$Z(X)^+_{\widetilde{F}} = (ZX)^+_F$$

The reader can find a proof of this theorem in [4].

Lemma 3. *If X is a super key of the scheme S then $X - Z$ is a super key of the scheme $\widetilde{S}$. And inversely Y is a super key of the scheme $\widetilde{S}$ then YZ is a super of scheme S.*

See the proof in [4].

Corollary 5. *If $\widetilde{S} = S - Z$ then*

1. $\mathcal{K}_S = \mathcal{K}_{\widetilde{S}} \Leftrightarrow Z \subseteq (\Omega - H)$ *with* $H = \cup K_i 1$ *where* $k_i \in \mathcal{K}_S$
2. $\mathcal{K}_S = Z \oplus \mathcal{K}_{\widetilde{S}} \Leftrightarrow Z \subseteq G$ *with* $G = \cap K_i$, *where* $K_i \in \mathcal{K}_S$ *and* $Z \oplus \mathcal{K}_{\widetilde{S}} = \{ZK_i | K_i \in \mathcal{K}_{\widetilde{S}}\}$.

See a proof in [4].

Corollary 6. *If* $Z \subseteq (\Omega - H)$ *and* $\widetilde{S} = (S - G) - Z$ *then* $\mathcal{K}_S = G \oplus \mathcal{K}_{\widetilde{S}}$.

See a proof in [4].

Lemma 4. *Let* $S = \langle \Omega, F \rangle$ *be a relation scheme and let* K *be a key of* S *then:*

$$\forall Z \subseteq K \quad Z_F^+ \cap (K - Z) = \emptyset$$

Proof. Suppose that $\exists Z \subseteq K$ where $Z_F^+ \cap (K - Z) \neq \emptyset$, then $\exists A \in Z_F^+ \cap (K - Z)$ so $A \in Z_F^+$ and $A \in K$ and $A \notin Z$.

Since $A \notin Z$ then $Z \subseteq (K - \{A\})$ and since $A \in Z_F^+ \subseteq (K - \{A\})_F^+$, so $((K - \{A\}) \to \{A\}) \in F^+$ therefore K is not a key, which is a contradiction.

Thus we have showed that $\forall Z \subseteq K$, $Z_F^+ \cap (K - Z) = \emptyset$.

Property 5. *Let* $S = \langle \Omega, F \rangle$ *be a relation scheme and let* $Y \subseteq \Omega$.

Let ZY_F^+, *We define* $\widetilde{S}$, $\widetilde{S} = S - Z = \langle \widetilde{\Omega}, \widetilde{F} \rangle$.

Then $\widetilde{F}$ *can't contain and FD of the form* $\emptyset \to X$ *with* $X \neq \emptyset$.

Proof. If $F = \{L_i \to R_i | R_i \subseteq \Omega,\ i = 1, ..., k\}$ then:

$\widetilde{F} = \{L_i - Z \to R_i - Z | (L_i \to R_i) \in F\}$.

Let $(L_i \to R_i) \in F$ where $L_i - Z = \emptyset$, then $L_i \subseteq Z = Y_F^+$, so $(Y \to L_i) \in F^+$.

Thus, we obtain by transitivity: $(Y \to R_i) \in F^+$ therefore $R_i \subseteq Z = Y_F^+$ and so $R_i - Z = \emptyset$.

Remark: Let $S = \langle \Omega, F \rangle$ be a relation scheme.

$F = \{L_i \to R_i | R_i \subseteq \Omega,\ i = 1, ..., k\}$

Let $\widetilde{S} = S - G - Z$, with $Z \subseteq \Omega - H$, $G = \Omega - R$.

When searching keys the FDs of the form $\emptyset \to X$ must be not be deleted from $\widetilde{F}$.

Proof. Let $L_i \to R_i$, with $L_i \subseteq G$ and $R_i \not\subset G$, the FD that will be the form:

$\emptyset \to R_i - G - Z$ in $\widetilde{F}$, with $R_i - G - Z \neq \emptyset$.

We are going to show that deleting this FD from $\widetilde{F}$ can involve in finding super keys instead of keys.

Let $\widetilde{K}$ be a key of the scheme $\widetilde{S} = \langle \widetilde{\Omega}, \widetilde{F} \rangle$.

Let us take an attribute A_j, $A_j \in (R_i - G - Z)$ where $\forall l = 1, ..., n$, $l \neq i$, $A_j \notin R_l$, that A_j is not contained in any other right side but the right side of $L_i \rightarrow R_i$.

Suppose that the FD $\emptyset \rightarrow R_i - G - Z$ is deleted from $\widetilde{F}$.

Since $A_j \in \widetilde{\Omega}$, then $A_j \in \widetilde{\Omega} - \widetilde{R}$, where $\widetilde{R} = \cup R_p$

$$(L_p \rightarrow R_p) \in \widetilde{F}$$

So A_j belongs to each keys of $\widetilde{S} = \langle \widetilde{\Omega}, \widetilde{F} \rangle$ and in particular to $\widetilde{K}$.

By Corollary 6, since $Z \subseteq \Omega - H$ then $\mathcal{K}_S = G \oplus \mathcal{K}_{\widetilde{S}}$, A_j will belong to each key of $S = \langle \Omega, F \rangle$.

Let $K = G\widetilde{K}$.

We have $G \rightarrow L_i$, $L_i \rightarrow R_i$, $R_i \rightarrow A_j$, so $G \rightarrow A_j$,

moreover $A_j \notin G$ because $A_j \in R_i$ and $G \rightarrow GA_j$.

By reflexivity $\widetilde{K} - \{A_j\} \rightarrow \widetilde{K} - \{A_j\}$.

By applying the union rule, we obtain:

$$G(\widetilde{K} - \{A_j\}) \rightarrow GA_j(\widetilde{K} - \{A_j\})$$

that is $K - \{A_j\} \rightarrow K$ showing that KK is a super key and not a key.

So the FDs of the form $\emptyset \rightarrow X$, $X \neq \emptyset$, must not be deleted when searching keys.

E.g.: Let us have the relation scheme $S = \langle \Omega, F \rangle$ with:

$\Omega = \{a, b, c, d, e, f, g, h\}$

$F = \{ac \rightarrow bg,\ b \rightarrow acd,\ h \rightarrow dfg, adeh \rightarrow bcf,\ abc \rightarrow d,\ cf \rightarrow aeg\}$

$L = \{a, b, c, d, e, f, h\}$

$R = \{a, b, c, d, e, f, g\}$

$G = \Omega - R = \{h\}$

Let $Z = \{g\} \subseteq \Omega - H$

Let $\widetilde{S} = S - G - Z$

$\widetilde{\Omega} = \{a, b, c, d, e, f\}$

$\widetilde{F} = \{ac \rightarrow b,\ b \rightarrow acd,\ \emptyset \rightarrow fd,\ ade \rightarrow bcf,\ abc \rightarrow d,\ cf \rightarrow ae\}$

If we delete the FD $\emptyset \rightarrow fd$, then when searching the keys of $\widetilde{S}$ we will find the following set $\{cf, ade, ace, be, bf\}$, and thus the keys of S would be $\{cfh, adeh, aceh, beh, bfh\}$ while it is obvious that cfh is a super key of S, as well as $adeh$, $aceh$, beh and bfh.

However, if the FD $\emptyset \rightarrow fd$ is not deleted while searching the keys of $\widetilde{S}$, we will find the set $\{b, c, ae\}$ and thus $\{bh, ch, aeh\}$ are the keys of S.

Two remarks about the summary of [4]

1. In step 2 (page 95) the phrase:

"Eliminate from F' functional dependencies of the form:

$\emptyset \to \emptyset,\ \emptyset \to X,\ X \to \emptyset\ (X \neq \emptyset)$"

must be replaced by this one:

"Eliminate from F' functional dependencies of the form:

$\emptyset \to \emptyset,\ X \to \emptyset\ (X \neq \emptyset)$"

because by the use of $Z = (GR')^+_S$ to define

$S' = \langle \Omega', F' \rangle = S - Z$

so in S', the functional dependency $\emptyset \to X$ dose not exist (See Property 5).

2. In example 3.2 (page 96), there is errors in calculus.

We have $Z = (78)^+ = 7846$, instead of $Z = (78)^+ = 78$.

So the final result is the following: $\mathcal{K}_{\widetilde{S}} = \{2, 3, 15\}$; $\mathcal{K}_S = \{28, 38, 158\}$.

Acknowledgement

We are grateful to the referees for their comments, which have helped to improve the presentation of this paper.

REFERENCES

1. E. F. Codd, *A relational model of data for large shared data banks*, Comm., ACM, Vol. 13, No. 5, June 1970.
2. J. D. Ullman, *Principles of Database Systems*, Pitman, London, 1981.
3. C. Beeri and P. A. Bernstein, *Computational problem related to the design of normal form relation schemes*, Trans. on DBSys. 4:1, March 1979.
4. J. Demetrovics, Ho Thuan, Nguyen Xuan Huy, and Le Van Bao, *Translation of relation schemes, balanced relation schemes and the problem of key representation*, J. Inf. Process. Cybern. EIK, **23** (2–3) (1987), 81 - 97.
5. Ho Thuan and Le Van Bao, *Some results about keys of relation schemes*, Acta Cybernetica, Hungary **7** (1) (1985), 99 - 113.
6. J. Demetrovics and Ho Thuan, *Keys and superkeys for relation schemes*, Computer and Artificial intelligence, **5** (6) (1986), 511 - 519.
7. Ho Thuan, *Some remarks on the algorithm of Lucchesi and Osborn*, Acta Cybernetica, Hungary **8** (2) (1987).

Institute of informatic
University of Sidi Bel Abbes, Algeria

Received February 12, 1996