# tanulmányok

184/1986

MTA Számitástechnikai és Automatizálási Kutató Intézet Budapest

*Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete*
*Computer and Automation Institute, Hungarian Academy of Sciences*

# CONTRIBUTION TO THE THEORY OF RELATIONAL DATABASES

*HO THUAN*

Tanulmányok 184/1986
Studies 184/1986

A kiadásért felelős:

*KEVICZKY LÁSZLÓ*

Főosztályvezető:

*DEMETROVICS JÁNOS*

*(HO THUAN researcher worker, on leave from Institute of Cybernetics and Computer Sciences Hanoi, Vietnam)*

ISBN 963 311 213 3

ISSN 0237-0131

## Abstract

The relational data model was introduced by Codd [1]. Since his fundamental paper was published, the theory of relational databases has been the subject of an intensive research during the past decade.

In this work some new results about keys and superkeys for relation schemes, about the theory of translations for relation schemes and about the structure of minimum covers are presented.

## Table of Contents

## 0. Introduction

The relational model of data was introduced by Codd [1]. Since his fundamental paper was published, the theory of relations for data bases has been the subject of an intensive research during the past decade.

The paper of Delobel and Casey [2] can be considered as the first major study on the functional dependencies.

Significant advances in the theory were made by Armstrong [15] and shortly thereafter, nearly simultaneously, by Fagin, [3], Beeri, Fagin and Howard, [4], Rissanen, [5], and Aho, Beeri, and Ullman [6].

Nowadays the field is under an intense process of development.

In Hungary, J. Demetrovics and his colleagues also have important contributions to the theory of relations for databases, specially to combinatorial aspects of the theory. [7,8,9,17,18].

In this work we present in a systematic way some selected new results concerning the theory of relational data bases. These results either have been published or will appear in [26-38].

This work consists of three chapters. In Chapter 1 we present some results concerning keys and superkeys for the relation scheme $S=<\Omega,F>$. Namely, a necessary condition under which a subset X of $\Omega$ is a key, a simple explicit formula for computing the intersection of all keys for S, sufficient conditions under which a relation scheme has exactly one key, sufficient conditions for a superkey in a special family to be a key, three algorithms for the key finding and key recognition problems and so on ...

Chapter 2 is devoted to the so-called theory of translations of relation schemes. The concept of a translation of relation scheme seems to be useful in the sense that it can reduce a relation scheme to a simpler one, i.e., a relation scheme with a smaller number of attributes and with shorter functional dependencies so that the key-finding problem becomes less cumbersome.

On the other hand, from the set of keys of the new relation scheme obtained by this transformation, the corresponding keys of the original relation scheme can be found by a single "translations".

In this chapter we present the main results about the translation of relation schemes, give a classification of relation schemes, investigate the so-called balanced relation scheme and nontranslatable relation scheme and prove a theorem for key representation. In connection with these results, general scheme for the transformation of an arbitrary relation scheme into a balanced relation scheme and for the finding of all its keys are proposed.

In Chapter 3 results about the structure of minimum covers will be presented.

The nonredundant and minimum covers have been investigated in depth by Bernstein [21], Maier [22], Ausiello et al. [23], and several useful properties of them have been proved and used in various problems in the logical design of data bases.

But few attention has been paid to the study of invariants concerning the attribute sets of the left and right sides of these covers. Moreover, the structure of right sides of FDs in minimum covers has not been investigated.

In this chapter we establish the relationship between the notion of direct determination and FD-graph, prove some well known and new results concerning direct determination, prove some additional invariants for covers and nonredundant covers, study the structure for right sides of FDs in minimum covers. Basing upon these results an algorithm for finding the "quasi optimal" cover (in the sense of effective and economical memory management) is proposed.

This work has been written while the author has been a visiting researcher at the Computer and Automation Institute of the Hungarian Academy of Sciences during the years 1985-1986. The author has the chance to work in the research group on the theory of relational data bases under the direction of Pr. Dr. János Demetrovics.

I am indebted to him for several useful discussions and for his excellent advice and support.

I would like to express my sincere thanks to my vietnamese colleagues Le Van Bao, Nguyen Xuan Huy, Tran Thai Son, Dinh thi Ngoc Thanh of the Institute of Informatics and Cybernetics, Hanoi Vietnam and of course to Prof. J. Demetrovics, for that,with great pleasure, they allow me to use some our common results in this work.

Finally, special thanks are due to Drs. A. Békéssy and B. Uhrin and all members of the Computer Sciences division of the Computer and Automation Institute for their help and encouragement.

# 1. KEYS AND SUPERKEYS FOR RELATION SCHEMES

## § 1.1. Introduction

In relational data base design, functional dependencies, in general, and keys for relation scheme in particular play on important role.

Basing upon these notions, the normalization theory has been the subject of an intensive research during the past decade.

In this chapter, we present some results concerning keys and superkeys for the relation scheme $S=\langle\Omega,F\rangle$ : a necessary condition under which a subset $X$ of $\Omega$ is a key, some sufficient conditions under which a superkey in a special family is a key, a simple explicit formula for computing the intersection of all keys for S, sufficient condition under which a relation scheme has exactly one key, a criterion for which an attribute is a non prime one and some other results.

Basing on these results, some effective algorithms are proposed for the finding of keys and for the key recognition problems.

Finally some remarks improving the performance of the algorithm of Lucchesi and Osborn [11] are also given.

Some of above results are published in [26-31].

## § 1.2. Basic definitions

In this section we give some basic definitions and notation concerning the relational data model ([12]; see also [13]).

Throughout this work, when we speak about a set of tuples the word relation is used, while speaking about structural description of sets of tuples we use the word relation scheme [14]. With this approach, a relation is an instance of a relation scheme.

A <u>relation</u> involving the set of attributes $\Omega=\{A_1,A_2,\ldots,A_n\}$ is a subset of the cartesian product Dom $(A_1)$ x Dom $(A_2)$ x...x Dom $(A_n)$ where Dom $(A_i)$ - the domain of $A_i$ - is the set of possible values for that attribute. The elements of the relation are called <u>tuples</u> and will be denoted by <t>.

A <u>constraint</u> involving the set of attributes $\{A_1,A_2,\ldots,A_n\}$ is a predicate on the collection of all relations on this set. A relation $R(A_1,A_2,\ldots,A_n)$ fulfils the constraint if the value of the predicate for R is "true".

We shall restrict ourselves to the case of functional dependencies.

A <u>functional dependency</u> (abbr. FD) is a sentence denoted by $f{:}X \to Y$, where f is the name of the FD and X and Y are sets of attributes. A functional dependency $f{:}\ X \to Y$ holds in R ($\Omega$) where X and Y are subsets of $\Omega$, if for every tuples u and $v \in R$, $u[x] = v[x]$ implies $u[Y] = v[Y]$ ($u[x]$ denotes the projection of the tuple u on X).

Let F be a set of functional dependencies. A relation R defined over the attributes $\Omega = \{A_1, A_2, \ldots, A_n\}$ is said to be an <u>instance</u> of the <u>relation scheme</u> $S = \langle \Omega, F \rangle$ iff each FD $f \in F$ holds in R.

The following Armstrong's inference rules are sound and complete for FDs[1] [15].

For every $X, Y, Z \subseteq \Omega$,

A1. (Reflexivity): if $Y \subseteq X$ then $X \to Y$.

A2. (Augmentation): if $X \to Y$ then $X \cup Z \to Y \cup Z$.

A3. (Transitivity): if $X \to Y$ and $Y \to Z$ then $X \to Z$

From the Armstrong's axioms the following two rules are easily derived:

Union rule: if $X \to Y$ and $X \to Z$ then $X \to Y \cup Z$

Decomposition rule: if $X \to Y$ and $Z \subseteq Y$ then $X \to Z$.

---

[1] In fact we use here a system of axioms which is equivalent to that of Armstrong.

Let F be a given set of FDs. The closure $F^+$ of F is the set of all FDs that can be derived from the FDs in F by repeated applications of Armstrong's axioms.

It is shown in [13] that

$$(X \to Y) \in F^+ \quad \text{iff} \quad Y \subseteq X^+,$$

where

$$X^+ = \{A_i \mid (X \to A_i) \in F^+\}$$

is by definition the closure of X w.r.t. F.

In the following, instead of $(X \to Y) \in F^+$ and $X \cup Y$, we shall write $X \overset{*}{\to} Y$ and XY respectively.

There is a linear-time algorithm in the length of the description of the FDs, proposed by Beeri and Bernstein [10] for computing the closure $X^+$ of a given set X (w.r.t. F):

1) Establish the sequence $X^{(o)}, X^{(1)}, \ldots,$ as follows:

$$X^{(o)} \equiv X.$$

Suppose $X^{(i)}$ is computed, then

$$X^{(i+1)} = X^{(i)} \cup Z^{(i)}$$

where

$$Z^{(i)} = \bigcup_{\substack{X_j \subseteq X^{(i)},\, Y_j \not\subseteq X^{(i)} \\ (X_j \to Y_j) \in F}} Y_j$$

2) In view of the construction, it is obvious that

$$X^{(o)} \subseteq X^{(1)} \subseteq X^{(2)} \subseteq \cdots$$

Since $\Omega$ is a finite set, there exists a smallest non negative integer t such that

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

3) We then have

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

Two subsets X and Y of $\Omega$ are said to be equivalent under a set of FDs F, written $X \leftrightarrow Y$, if

$$X \xrightarrow{*} Y \quad \text{and} \quad Y \xrightarrow{*} X.$$

It is easy to show that

$$X \leftrightarrow Y \quad \text{iff} \quad X^{+} = Y^{+}.$$

<u>Keys for a relation scheme</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme and let X be a subset of $\Omega$.

X is a key for S if it satisfies the following two conditions:

(i) $\quad X \xrightarrow{*} \Omega$ ,

(ii) $\quad \nexists X' \subset X: X' \xrightarrow{*} \Omega$ .

The subset X which satisfies only (i) is called a <u>superkey</u> for S.

It is clear that

X is a superkey for S iff $X^{+}=\Omega$.

## § 1.3. Preliminary results

We are now in a position to prove some lemmas which will be needed in the sequel.

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

$$\Omega=\{A_1,A_2,\ldots,A_n\},$$

$$F=\{L_i \to R_i \mid L_i,R_i \subseteq \Omega,\ i=1,2,\ldots,m\}.$$

Without loss of generality, throughout this work we use only sets of FDs in the natural reduced form, i.e. those which satisfy the following conditions:

(i) $L_i \cap R_i=\emptyset \qquad \forall\, i \neq j$

(ii) $L_i \neq L_j \qquad$ if $\qquad i \neq j$.

Let us denote

$$L = \bigcup_{i=1}^{m} L_i , \qquad R = \bigcup_{i=1}^{m} R_i$$

$K_S =\{K \mid K \text{ is a key for } S\}$

$C_i =\Omega \setminus L_i^+ , \qquad i=1,2,\ldots,m$ ;

$I =\{i \mid \text{there is no } j \text{ such that } L_i \supset L_j\}$

$\{1,2,\ldots,m\}$ .

It is obvious that for every $j \in \{1,2,\ldots,m\}$, $L_j\ C_j$ is a superkey for S.

We have the following lemmas.

Lemma 1.3.1.

Let $S=<\Omega,F>$ be a relation scheme, $X,Y\subseteq\Omega$. Then

$$(XY)^+ = (X^+Y)^+ = (XY^+)^+ \qquad (1.1)$$

Proof

It is sufficient to prove that

$$(X^+Y)^+ = (XY)^+ .$$

By the definition of the closure $X^+$ of X, it is obvious that

$$X^+ \supseteq X .$$

Hence

$$X^+Y \supseteq XY.$$

By the algorithm for the finding of the closure, we have

$$(X^+Y)^+ \supseteq (XY)^+ . \qquad (1.2)$$

On the other hand, from

$$X \xrightarrow{*} X^+ ,$$

we have

$$XY \xrightarrow{*} X^+Y ,$$

or equivalently,

$$X^+Y \subseteq (XY)^+ .$$

It follows that:

$$(X^+Y)^+ \subseteq ((XY)^+)^+ = (XY)^+ . \qquad (1.3)$$

Combining (1.2) and (1.3), we obtain (1.1). The proof is complete.

Lemma 1.3.2.

For any $i \in I$,

$L_i$ is a key for S if and only if $C_i = \emptyset$.

Proof.

If part: If $C_i = \emptyset$, i.e. $L_i^+ = \Omega$, then $L_i$ is a superkey for S. Since $i \in I$, it follows that for all $X \subset L_i$, we have

$$X^+ = X \subset L_i,$$

showing that $L_i$ is a key for S. The only if part is straight-forward.

Lemma 1.3.3.

Let K be any key for $S = \langle \Omega, F \rangle$.

Then $Z^+ \cap (K \setminus Z) = \emptyset$ for all $Z \subsetneq K$.

Proof.

Denote $Y = Z^+ \cap (K \setminus Z)$ .

It is clear that $Y \subseteq Z^+$, $Y \subseteq K$ and

$$Y \cap Z = \emptyset.$$

Therefore we can write

$$K = Z \,|\, Y \,|\, X \qquad \text{(a partition of K)}$$

and, by Lemma 1.3.1, we have

$$(ZX)^+ = (Z^+X)^+ = (Z^+YX)^+ \quad (ZYX)^+ = \Omega.$$

Since K is a key, so ZX=K, showing that

$Y = Z^+ \cap (K \setminus Z) = \emptyset$.

Lemma 1.3.4.

Let $S=<\Omega,F>$ be a relation scheme.

If $A \overline{\in} L$ and $X \overset{*}{\rightarrow} Y$ then

$$X \setminus \{A\} \overset{*}{\rightarrow} Y \setminus \{A\}.$$

Proof

From $X \overset{*}{\rightarrow} Y$ it follows that there exists a derivation sequence

$$\{L_{i_1} \rightarrow R_{i_1}, L_{i_2} \rightarrow R_{i_2}, \ldots, L_{i_p} \rightarrow R_{i_p}\}$$

such that

$$\begin{aligned} & X \supseteq L_{i_1} \\ & XR_{i_1} \supseteq L_{i_2} \\ & \ldots \qquad\qquad\qquad (1.4) \\ & XR_{i_1} R_{i_2} \ldots R_{i_{p-1}} \supseteq L_{i_p} \\ & XR_{i_1} R_{i_2} \ldots R_{i_p} \supseteq Y \end{aligned}$$

.

Since $A \notin L = \bigcup_{j=1}^{m} L_j$, from (1.4) we have

$$X\setminus\{A\} \supseteq L_{i_1}$$

$$(X\setminus\{A\})\, R_{i_1} \supseteq L_{i_2}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$(X\setminus\{A\})\, R_{i_1} R_{i_2} \ldots R_{i_p} \supseteq Y\setminus\{A\} \ ,$$

showing that

$$X\setminus\{A\} \overset{*}{\rightarrow} Y\setminus\{A\}.$$

<u>Lemma 1.3.5.</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme, $X\subseteq\Omega$. If $A\in X$ and $X\setminus A \overset{*}{\rightarrow} A$[+)] then X is not a key for S.

<u>Proof</u>.

By the hypothesis of the lemma

$$X\setminus A \overset{*}{\rightarrow} A.$$

On the otherhand, it is obvious that:

$$X\setminus A \overset{*}{\rightarrow} X\setminus A.$$

Applying the union rule, we obtain

$$X\setminus A \overset{*}{\rightarrow} X.$$

Since $A\in X$, it is obvious that $X\setminus A \subset X$, showing that X is not a key. The proof is complete.

---

[+)] Here and in the following $X\setminus A$ stands for $X\setminus\{A\}$.

Lemma 1.3.6.

Let $S=\langle\Omega,F\rangle$ be a relation scheme. Then any key $K$ for $S$ has the following form

$$K = L_iX_i$$

where $X_i \subseteq C_i$ , $i \in I$.

Proof

Let $K_S$ be the set of all keys for $S$ and $K \in K_S$.
If $K=\Omega$, then obviously

$$K=L_iX_i \qquad \forall\, i \in I.$$

If $K \subset \Omega$, then by the algorithm for the finding of the closure $K^+$ of $K$ w.r.t. $F$, there exists $L_j$ such that $L_j \subseteq K$.

Consequently, there is $i \in I$ such that $L_i \subseteq K$.
Thus $K = L_iX_i$ , $i \in I$.
Now we have to prove that $X_i \subseteq C_i$. By Lemma 1.3.1. we have

$$L_i^+ X_i \subseteq (L_i^+X_i)^+ = (L_iX_i)^+ = K^+ = \Omega = L_i^+C_i. \qquad (1.5)$$

By lemma 1.3.3:

$$L_i^+ \cap (K \setminus L_i) = L_i^+ \cap X_i = \emptyset .$$

On the other hand, it is clear the

$$L_i^+ \cap C_i = \emptyset.$$

Hence, from (1.5) we have:

$$X_i \subseteq C_i.$$

The proof is complete.

Remark 1.3.1

Lemma 1.3.6 still holds if the set $I$ is replaced by the set $\{1,2,\ldots,m\}$.

## § 1.4. Necessary condition under which a subset X of $\Omega$ is a key.

In this section we investigate the necessary condition under which a subset X of $\Omega$ is a key and prove a theorem which will be used as a basis for the design of algorithms to find keys for a relation scheme.

Theorem 1.4.1.

Let S $=\langle\Omega,F\rangle$ be a relation scheme and X be a key of S.

Then

$$\Omega\setminus R \subseteq X \subseteq (\Omega\setminus R) \cup (L\cap R).$$

Proof

We shall begin by showing that

$$\Omega\setminus R \subseteq X.$$

First we observe that $X^{+}\subseteq XR$. Since X is a key, obviously $X^{+}=\Omega$. Hence $XR=\Omega$. This implies that

$$\Omega\setminus R \subseteq X.$$

To complete the proof it remains to show that:

$$X \subseteq (\Omega\setminus R) \cup (L\cap R). \qquad (1.6)$$

It is clear that

$$X \subseteq \Omega = (\Omega \setminus R) \cup (L \cap R) \cup (R \setminus L) . \qquad (1.7)$$

To obtain (1.6), we have only to prove that

$$X \cap (R \setminus L) = \emptyset .$$

Assume the contrary, that there exists an attribute $A \in X$, $A \in R$ and $A \bar{\in} L$. Since X is a key, we have $X \overset{*}{\to} \Omega$. Since $A \bar{\in} L$, we refer to Lemma 1.3.4 to deduce

$$X \setminus \{A\} \overset{*}{\to} \Omega \setminus \{A\}$$

On the other hand, from $A \bar{\in} L$, and $L \subseteq \Omega$, we have $L \subseteq \Omega \setminus A$. Hence $\Omega \setminus A \overset{*}{\to} L$.

Applying the transitivity rule for the sequence $X \setminus A \overset{*}{\to} \Omega \setminus A \overset{*}{\to} L \overset{*}{\to} R \overset{*}{\to} A$ (since $A \in R$), we obtain

$$X \setminus A \overset{*}{\to} A \text{ with } A \in X.$$

By virtue of Lemma 1.3.5, this contradicts the hypothesis that X is a key. Thus we have proved that if X is a key, then $X \cap (R \setminus L) = \emptyset$.

From (1.7) we deduce that

$$X \subseteq (\Omega \setminus R) \cup (L \cap R) .$$

The proof is complete.

Theorem 1.4.1 is illustrated by Fig.1.1 where X is an arbitrary key for the relation scheme $S = \langle \Omega , F \rangle$.

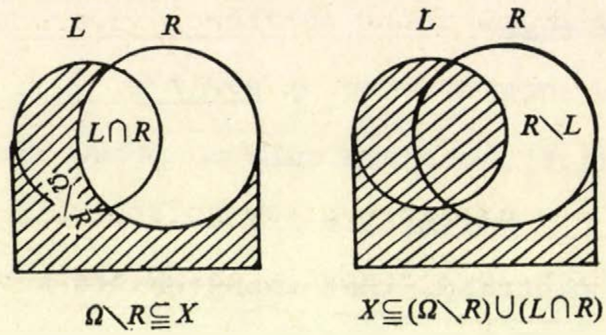



Fig.1.1

In view of Theorem 1.4.1., it is easily seen that the keys for $S=<\Omega,F>$ are different only on the attributes of $L\cap R$. In other words, if $X_1$ and $X_2$ are two different keys for S, then

$$X_1 \setminus X_2 \subset L\cap R \qquad \text{and} \qquad X_2 \setminus X_1 \subset L\cap R.$$

Let $K_S$ denote the set of all keys for S, and $\mathcal{S}(Z)$ the maximal cardinality Sperner system on a set Z [16].

As immediate consequences of Theorem 1.4.1. and results in [17], [18], we have the following corrolaries.

Corollary 1.4.1

Let $S=<\Omega,F>$ be a relation scheme. Then

$$\# K_S \le \# \mathcal{S}(L\cap R) = C_h^{[h/2]}$$

where $h=\#(L\cap R)$ is the cardinality of $L\cap R$.

<u>Corollary 1.4.2</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme and X be a key of S.

Then

$$\#(\Omega\setminus R)\leq\# X\leq\#(\Omega\setminus R)+\#(L\cap R).$$

<u>Corollary 1.4.3</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme. If $R\setminus L\neq\emptyset$ then there exists a key X for S suth that $X\neq\Omega$ (non trivial key). Moreover $R\setminus L\subseteq\Omega\setminus H$, where $H=\bigcup_{K\in K_S} K$ is the union of all keys for S.

<u>Corollary 1.4.4</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme. If $L\cap R=\emptyset$ then $\# K_S=1$ and $\Omega\setminus R$ is the unique key for S.

It is natural to ask whether the results formulated in Theorem 1.4.1 can be improved. The answer is affirmative as it is showed by the following lemma and Theorem.

<u>Lemma 1.4.1</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme and X be a key for S. Then

$$X \cap R \cap (L \setminus R)^{+} = \emptyset .$$

<u>Proof</u>

Suppose the statement is not true. Then there exists an attribute A such that

$$A \in X \cap R \cap (L \setminus R)^{+} .$$

Thus $A \in X$, $A \in R$, $L \setminus R \overset{*}{\to} A$. Since $A \in R$, it follows that $A \bar{\in} (L \setminus R)$. On the other hand, it is clear that

$$L \setminus R \subseteq \Omega \setminus R .$$

Taking into account Theorem 1.4.1, we get

$$L \setminus R \subseteq \Omega \setminus R \subseteq X .$$

Thus

$$L \setminus R \subseteq X \setminus A \quad (\text{since } A \bar{\in} L \setminus R).$$

It follows that

$$X \setminus A \overset{*}{\to} L \setminus R \overset{*}{\to} A$$

where $A \in X$.

By Lemma 1.3.5, this contradicts the hypothesis that X is a key for S. The proof is complete.

We define

$$\mathfrak{a}(L,R) = (L \setminus R)^{+} \cap (L \cap R) .$$

It is clear that

$$\mathfrak{a}(L,R) \subseteq (L \setminus R)^{+} \cap R$$

From this

$X \cap \mathfrak{a}(L,R)=\emptyset$ for every $X \in K_S$. Combining with Theorem 1.4.1, the following theorem is immediate:

<u>Theorem 1.4.2</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme, and X be any key for S. Then

$$(\Omega \setminus R) \subseteq X \subseteq (\Omega \setminus R) \cup ((L \cap R) \setminus \mathfrak{a}(L,R)).$$

The following example where $\mathfrak{a}(L,R) \neq \emptyset$, shows that Theorem 1.4.2 is nontrivial.

<u>Example 1.4.1</u>

$\Omega=\{A,B,H,G,Q,M,N,V,W\}$

$F = \{A \to B,\ B \to H,\ G \to Q,\ V \to W,\ W \to V\}$

From this we have

$L = ABGVW;\quad R = BHQVW;\quad L \cap R = BVW;$

$L \setminus R = A\ G;\quad (L \setminus R)^+ = AGBHQ;$

$\mathfrak{a}(L,R) = (L \setminus R)^+ \cap (L \cap R) = \{B\} \neq \emptyset.$

<u>Remark 1.4.1</u>

It is worth noticing that

$$(\Omega \setminus R)^+ = (\Omega \setminus (L \cup R)) \cup (L \setminus R)^+.$$

Therefore, if X is a key for S then obviously:

$$X \cap R \cap (\Omega \setminus R)^{+} = X \cap R \cap (L \setminus R)^{+} = \emptyset,$$

and

$$(\Omega \setminus R) \cup \{(L \cap R) \setminus (\Omega \setminus R)^{+}\} = (\Omega \setminus R) \cup \{(L \cap R) \setminus a(L,R)\}$$

<u>Remark 1.4.2</u>

Using Theorem 1.4.2, the Corollaries 1.4.1, 1.4.2 and 1.4.3, deduced from Theorem 1.4.1 above, can be improved, as well.

<u>Theorem 1.4.3</u>

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

$$L \cap R = \{A_{t_1}, A_{t_2}, \ldots, A_{t_h}\} \subseteq \{A_1, \ldots, A_n\} = \Omega .$$

Let us define

$$K(1) = (\Omega \setminus R) \cup (L \cap R),$$

$$K(i+1) = \begin{cases} K(i) \setminus A_{t_i} & \text{if} \quad K(i) \setminus A_{t_i} \overset{*}{\to} A_{t_i} \\ K(i) & \text{if} \quad K(i) \setminus A_{t_i} \overset{*}{\not\to} A_{t_i} \end{cases} ,$$

with $i=1,2,\ldots,h$.

Then $K(h+1)$ is a key for $S=\langle\Omega,F\rangle$.

<u>Proof</u>

We shall begin with showing that

$$K(i+1) \overset{*}{\to} K(i) .$$

Two cases can occur:

a) If $K(i) \setminus A_{t_i} \not\xrightarrow{*} A_{t_i}$ then from the definition of $K(i+1)$, we have

$$K(i+1) = K(i)$$

and it is obvious that

$$K(i+1) \xrightarrow{*} K(i).$$

b) If $K(i) \setminus A_{t_i} \xrightarrow{*} A_{t_i}$, we have

$$K(i+1) = K(i) \setminus A_{t_i}.$$

On the other hand, it is obvious that

$$K(i) \setminus A_{t_i} \xrightarrow{*} K(i) \setminus A_{t_i}.$$

Applying the union rule, we get:

$$K(i) \setminus A_{t_i} \xrightarrow{*} K(i).$$

Therefore

$$K(i+1) \xrightarrow{*} K(i).$$

So we have

$$K(h+1) \xrightarrow{*} K(h) \xrightarrow{*} \ldots \xrightarrow{*} K(1).$$

From the above definition of $K(i+1)$, it is clear that

$$K(h+1) \subseteq K(h) \subseteq \ldots \subseteq K(1).$$

We are now in a position to prove the theorem.

As an immediate consequence of Theorem 1.4.1, $K(1) = (\Omega \setminus R) \cup (L \cap R)$ is a superkey for S. On the other hand

$$K(h+1) \xrightarrow{*} K(1)$$

showing that K(h+1) is a superkey for S too. To complete the proof, it remains to show that K(h+1) is a key.

Assume it is not. Then there would exist a key $\bar{X}$ for S such that $\bar{X} \subset K(h+1)$, and using the result of Theorem 1.4.1, we have

$$\Omega \setminus R \subseteq \bar{X} \subset K(h+1) \subseteq (\Omega \setminus R) \cup (L \cap R).$$

Clearly, there exists

$$A_{t_j} \in K(h+1) \cap (L \cap R) \setminus \bar{X}$$

with $1 \leq j \leq h$.

From the definition of K(j+1), we find

$$K(j) \setminus A_{t_j} \overset{*}{\not\rightarrow} A_{t_j}.$$

Since $K(h+1) \subset K(j)$, it follows that

$$K(h+1) \setminus A_{t_j} \overset{*}{\not\rightarrow} A_{t_j}.$$

On the other hand $\bar{X} \subseteq K(h+1) \setminus A_{t_j}$.

Therefore

$$\bar{X} \overset{*}{\not\rightarrow} A_{t_j}$$

which conflicts with the fact that $\bar{X}$ is a key for $S=\langle\Omega,F\rangle$.

The proof is complete.

## § 1.5. The intersection of all keys for a relation scheme

In this section we establish a simple explicit formula for computing the intersection of all keys for a relation scheme $S=\langle\Omega,F\rangle$, and a criterion under which an attribute $A_i \in \Omega$ is a non-prime one. Finally, another characterization for the intersection of all keys for a relation scheme is also given.

Let us denote by

$$G = \bigcap_{K \in K_S} K,$$

the intersection of all keys for a relation scheme $S=\langle\Omega,F\rangle$.

First, we prove the

Lemma 1.5.1.

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

Then

$$G \cap R = \emptyset.$$

Proof

It is sufficient to prove that for each $A \in R$ there exists a key $K$ for $S$ such that $A \notin K$.

In fact, from $A \in R$ we deduce that A belongs to some $R_i$. Consider the functional dependency

$$L_i \to R_i, \qquad (L_i \cap R_i = \emptyset).$$

Hence $A \bar{\in} L_i$.

It is easily seen that

$$L_i \cup \{\Omega \setminus (L_i \cup R_i)\} \overset{*}{\to} \Omega,$$

and

$$A \bar{\in} L_i \cup \{\Omega \setminus (L_i \cup R_i)\},$$

showing that $L_i \cup \{\Omega \setminus (L_i \cup R_i)\}$ is a superkey for S. This superkey includes a key K such that $A \bar{\in} K$.

Hence $G \cap R = \emptyset$.

<u>Theorem 1.5.1</u>

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

Then

$$G=\Omega \setminus R.$$

<u>Proof</u>

As an immediate consequence of Lemma 1.5.1 we have

$$G \subseteq \Omega \setminus R.$$

On the other hand, by Theorem 1.4.1, it is easily seen that

$$\Omega \setminus R \subseteq G.$$

Hence

$$G=\Omega\setminus R.$$

The proof is complete.

Theorem 1.5.2

Let $S=\langle\Omega,F\rangle$ be a relation scheme and let $A\in\Omega$.

Suppose that the following conditions hold for all $L_i$, $i=1,2,\ldots,m$

(i) $A\in L_i \Longrightarrow L_i\setminus A \overset{*}{\rightarrow} A$,

(ii) $A\notin L_i \Longrightarrow A\in L_i^+$ .

Then A is a non-prime attribute, that is $A\notin H$ where $H=\bigcup_{K\in K_S} K$ is the union of all keys for S.

Proof

The proof is by contradiction. Assume the contrary that $A\in H$. Then there would exist a key K for S such that $A\in K$, and an $L_j$ such that $L_j\subseteq K$.

(1) If $A\in L_j$, then by the hypothesis of the theorem (condition (i)), we have

$$L_j\setminus A \overset{*}{\rightarrow} A$$

Consequently

$$K\setminus A \overset{*}{\rightarrow} L_j\setminus A \overset{*}{\rightarrow} A,$$

which, by Lemma 1.3.5, contradicts the fact that K is a key.

(2) If $A \notin L_j$, then by condition (ii) of the theorem, we have $A \in L_j^+$.

Since $A \notin L_j$,

$$L_j \subseteq K \setminus A$$

Hence

$$K \setminus A \overset{*}{\rightarrow} L_j \overset{*}{\rightarrow} A,$$

which contradicts the fact that K is a key. Thus $A \notin H$. The proof is complete.

Example 1.5.1

$$\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$$

$$F = \{A_1 \rightarrow A_3A_5,\ A_3A_4 \rightarrow A_1A_6,\ A_1A_5A_6 \rightarrow A_3A_4\}$$

It is easy to verify that $A_5$ satisfies all conditions of Theorem 1.5.2.

Therefore $A_5 \notin H$.

Theorem 1.5.3

Let $S = \langle \Omega, F \rangle$ be a relation scheme, and G be the intersection of all keys for S. Then

$$G^+ \setminus G \subseteq \Omega \setminus H.$$

In other words $G^+ \setminus G$ consists of only non-prime attributes.

Proof

First, we prove that

$$(G^+\setminus G)\cap K=\emptyset \text{ for every } K\in K_S.$$

If it were not true, there would exist a key $K_i$ and an attribute $A_j$ such that

$$A_j\in G^+,\ A_j\notin G,\ A_j\in K_i \text{ where } K_i\in K_S.$$

It follows that:

$$A_j\in G^+\cap(K_i\setminus G),\ G\subsetneq K_i.$$

This means

$$G^+\cap(K_i\setminus G)\neq\emptyset,$$

a contradiction, by virtue of Lemma 1.3.3.

Hence

$$(G^+\setminus G)\cap(\bigcup_{K\in K_S} K)=\emptyset,$$

Or equivalently

$$G^+\setminus G\subseteq\Omega\setminus H.$$

<u>Definition 1.5.1</u>

An attribute $A_j\in\Omega$ is said to be a deterministic one w.r.t $S=\langle\Omega,F\rangle$, if for every $(L_i\to R_i)\in F$, $A_j\in R_i$ implies $A_j\in L_i$. In other words, $A_j$ is a deterministic attribute iff whenever it belongs to the right hand side of some FD, it must also belong to the left hand side of this FD.

Let us denote by D the set of all deterministic attributes w.r.t. $S=\langle\Omega,F\rangle$.

The following theorem establishes the relation between the set of deterministic attributes D and G - the intersection of all keys for S.

Theorem 1.5.4

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

Then

$$D=G.$$

Proof

First we prove that $D\subseteq G$. Suppose that $A\in D$ and there exists a key $K\in K_S$ such that $A\notin K$.

Since $K^+=\Omega$, so $A\in K^+$. By the algorithm for finding the closure of a set of attributes w.r.t. F, there exists an index t and some FD $(L_i \rightarrow R_i)$ in F such that

$$L_i\subseteq K^{(t)},\ A\notin L_i,\ A\in R_i.$$

This contradicts the fact that A is a deterministic attribute.

Hence, $A\in D$ implies $A\in K$, $\forall K\in K_S$. In other words, $A\in G$. Consequently $D\subseteq G$.

To complete the proof, it remains to show that $G\subseteq D$. Were this false, there would exist an attribute $A\in G$ and $A\notin D$. This means $A\in R_i$ for some i. (From $L_i\cap R_i=\emptyset$, it follows that $A\notin L_i$).

We arrive to a contradiction, since $A\in G=\Omega\setminus R$ implies that $A\notin R_i$ for every $i=1,2,\dots,m$. The proof is complete.

## § 1.6. Relation schemes that have exactly one key

Theorem 1.6.1

Let $S=\langle\Omega,F\rangle$ be a relation scheme. Suppose that the following condition holds

$$\forall i\ (R_i\cap L \neq \emptyset \Longrightarrow L_i\cap R = \emptyset).$$

Then S has exactly one key and $\Omega\setminus R$ is this unique key.

Proof.

Let $C=\Omega\setminus(L\cup R)$.

Since $L \overset{*}{\rightarrow} R$, we have

$$L\cup C \overset{*}{\rightarrow} L\cup R\cup C=\Omega.$$

Let $I=\{i \mid R_i\cap L \neq \emptyset\}$

Evidently

$$\bigcup_{i\in I} L_i\cap R = \emptyset \tag{1.8}$$

and

$$L\cap R \subseteq \bigcup_{i\in I} R_i \tag{1.9}$$

It is obvious that

$$\bigcup_{i\in I} R_i \overset{*}{\rightarrow} L\cap R.$$

On the other hand we have

$$\bigcup_{i\in I} L_i \overset{*}{\rightarrow} \bigcup_{i\in I} R_i$$

Clearly we have together with (1.9)

$$\bigcup_{i\in I} L_i \overset{*}{\rightarrow} L\cap R.$$

From (1.8), we have

$$\bigcup_{i \in I} L_i \subseteq L \setminus R.$$

Hence

$$L \setminus R \overset{*}{\rightarrow} \bigcup_{i \in I} L_i \overset{*}{\rightarrow} L \cap R.$$

It follows that

$$L \setminus R \overset{*}{\rightarrow} (L \setminus R) \cup (L \cap R) = L.$$

Using $L \cup C \overset{*}{\rightarrow} \Omega$, we have

$$(L \setminus R) \cup C \overset{*}{\rightarrow} \Omega,$$

showing that $(L \setminus R) \cup C = \Omega \setminus R$ is a superkey for S.

By Theorem 1.4.1, $S=\langle\Omega,F\rangle$ has $(\Omega \setminus R)$ as the unique key.

<u>Theorem 1.6.2</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme, and X be a superkey for S.

If $X \cap R = \emptyset$ then X is the unique key for S.

<u>Proof</u>

From $X \cap R = \emptyset$, it is obvious that

$$X \subseteq \Omega \setminus R.$$

Since X is a superkey for S, there exists a key $\bar{X} \subseteq X$.

Using Theorem 1.4.1, clearly

$$\Omega \setminus R \subseteq \bar{X} \subseteq X \subseteq \Omega \setminus R$$

showing that $\Omega \setminus R$ is the unique key for S.

Theorem 1.6.3

Let $S=\langle\Omega,F\rangle$ be a relation scheme, and X be a superkey for S.

Then X is a unique key for S iff $X\cap R = \emptyset$.

Proof

The sufficiency of this theorem is essentially Theorem 1.6.2. We have only to prove the necessity. Let X be the unique key for S. Then, by Theorem 1.5.1,

$$X = G = \Omega\setminus R,$$

showing that $X\cap R = \emptyset$.

Theorem 1.6.4

Let $S=\langle\Omega,F\rangle$ be a relation scheme with $L\cap R \neq \emptyset$. Then $(\Omega\setminus R)\cup(L\cap R)$ is not a key for S.

Proof

Assume the contrary that $(\Omega\setminus R)\cup(L\cap R)$ is a key for S.

By Theorem 1.4.1, it is obvious that $K=(\Omega\setminus R)\cup(L\cap R)$ is the unique key for S and K must be equal to G. On the other hand

$$K = (\Omega\setminus R)\cup(L\cap R) \neq (\Omega\setminus R) = G,$$

a contradiction. The proof is complete.

## § 1.7. A special family of superkeys

In this section we prove some additional properties of keys and superkeys for relation schemes which can be used for the design of algorithms for the finding of keys for relation scheme. We mainly deal with the special family of superkeys for S, namely the family

$$\{L_iC_i \mid i=1,2,\ldots,m\}.$$

Recall that

$$C_i=\Omega\setminus L_i^+,\ i=1,2,\ldots,m\ .$$

We begin with the following lemma.

Lemma 1.4.1

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

Then $\forall\ i\neq j,\ i,j\in\{1,2,\ldots,m\}$, $L_i(C_i\cap L_jC_j)$ is a superkey for S.

Proof

In the case $C_i=\emptyset$, we have

$$L_i(C_i\cap L_jC_j)=L_i\ .$$

But in that case, it is obvious that $L_i$ is a superkey. We now consider the case $C_i\neq\emptyset$. First, we will prove that if $C_i\neq\emptyset$ then

$$C_i\cap L_jC_j\neq\emptyset,\ \ \forall j\neq i.$$

In fact, assume the contrary that

$$C_i \cap L_jC_j = \emptyset \quad \text{with some } i \neq j.$$

It follows that:

$$(C_i \cap L_j) \cup (C_i \cap C_j) = \emptyset.$$

On the other hand

$$C_i = (C_i \cap L_j) \cup (C_i \cap C_j) \cup (C_i \cap (L_j^+ \setminus L_j)) = C_i \cap (L_j^+ \setminus L_j),$$

showing that

$$C_i \subseteq L_j^+ \setminus L_j.$$

Thus

$$\Omega \setminus C_i \supseteq \Omega \setminus (L_j^+ \setminus L_j)$$

or

$$L_i^+ \supseteq L_jC_j$$

The last set inclusion shows that $L_i$ is a super-key, a contradiction. Therefore, if $C_i \neq \emptyset$ then

$$C_i \cap L_jC_j \neq \emptyset.$$

Now, it is clear that

$$L_i \overset{*}{\to} L_i^+$$

$$C_i \cap L_jC_j \overset{*}{\to} C_i \cap L_jC_j.$$

Consequently,

$$L_i(C_i \cap L_jC_j) \overset{*}{\to} L_i^+(C_i \cap L_j) \quad (C_i \cap C_j).$$

On the other hand, we have:

$$L_j = (L_j \setminus C_i) \quad (L_j \cap C_i) \subseteq L_i^+(C_i \cap L_j)$$

$$C_j = (C_j \setminus C_i) \quad (C_j \cap C_i) \subseteq L_i^+(C_i \cap C_j)$$

Hence

$$L_i^+(C_i \cap L_j)\ (C_i \cap C_j) \supseteq L_j C_j .$$

Finally, we have

$$L_i(C_i \cap L_j C_j) \overset{*}{\rightarrow} L_j C_j$$

Showing that $L_i(C_i \cap L_j C_j)$ is a superkey for S.

<u>Lemma 1.7.2</u>

Let K be any key for $S=\langle\Omega,F\rangle$ having the form

$$K=L_i X, \quad X \subseteq C_i$$

Then there exists $j_o \neq i$ such that

$$K \subseteq L_i(C_i \cap L_{j_o} C_{j_o}).$$

<u>Proof</u>

Assume the contrary that

$$L_i X \nsubseteq L_i(C_i \cap L_j C_j),\ \forall j \neq i,$$

or, equivalently

$$X \nsubseteq C_i \cap L_j C_j, \quad \forall j \neq i.$$

Then, for all $j \neq i$ there exists an attribute

$$A_{i_j} \in (L_j^+ \setminus L_j) \cap X .$$

Obviously we have:

$$L_i X \overset{*}{\rightarrow} L_i R_i X.$$

Then there must exist p such that

$$L_p \subseteq L_i R_i X$$

(Otherwise $L_iX \overset{*}{\not\to} \Omega$, a contradiction)

Let $A_{i_p} \in (L_p^+ \setminus L_p) \cap X$ and let

$$X' = X \setminus \{A_{i_p}\}.$$

Since $A_{i_p} \notin L_p$, so $L_p \subseteq L_iR_iX'$. Therefore, it is easy to see that

$$L_iX' \overset{*}{\to} L_iR_iX' \overset{*}{\to} L_iR_iL_pR_pX' \overset{*}{\to} L_iR_iL_p^+X' .$$

Moreover $A_{i_p} \in L_p^+$ .

Consequently,

$$L_iX' \overset{*}{\to} L_iX \overset{*}{\to} \Omega,$$

showing that $L_iX$ is not a key, a contradiction.

The proof is complete.

<u>Corollary 1.7.1.</u>

The family

$$\{L_i(C_i \cap L_jC_j) \mid j \neq i, \quad 1 \leq i,\ j \leq m\}$$

can be used to find all keys for the relation scheme $S = \langle\Omega, F\rangle$.

<u>Remark 1.7.1</u>

Lemmas 1.7.1 and 1,7,2 have been proved (perhaps by different methods) and used to design an interesting algorithm to find all keys for any relation scheme [19].

Theorem 1.7.1

Let $S=<\Omega,F>$ be a relation scheme. Suppose that the following conditions hold:

(i) $L_i(C_i \cap L_jC_j) = L_iC_i, \ \forall i=1,2,\ldots,m,$

(ii) $L_i \cap R_j = \emptyset \ \ \forall j \neq i.$

Then $L_iC_i$ is a key for S.

Proof

First, from condition (i) we can prove that for every $\bar{X} \subseteq C_i$, $L_i\bar{X}$ is not a superkey for S.
In fact, since $C_i \cap L_jC_j = C_i$, $\forall j \neq i$, it follows that

$$C_i \cap (L_j^+ \setminus L_j) = \emptyset, \ \forall j.$$

Therefore, if $A \in C_i$ then

$$\{A\} \cap (L_j^+ \setminus L_j) = \emptyset, \ \forall j.$$

Let A be any element of $C_i$ and $X=C_i \setminus \{A\}$. It is easy to see that

$$L_iX \overset{*}{\rightarrow} L_iR_iX \ .$$

Since $L_iR_i \cap C_i = \emptyset$ (because $L_iR_i \subseteq L_i^+$), $A \in C_i$, $A \notin X$, it follows that

$$A \notin L_iR_iX.$$

Now, suppose that there exists

$$L_h \subseteq L_iR_iX, \ h \neq i.$$

Obviously $A \notin L_h$ and

$$L_iX \overset{*}{\rightarrow} L_iR_iX \overset{*}{\rightarrow} L_iR_iL_hR_hX.$$

It is clear that $A \notin R_h$, otherwise $A \in (L_h^+ \setminus L_h)$, a contradiction. By repeating the same reasoning, we can prove that

$$L_i X \overset{*}{\not\rightarrow} \Omega,$$

showing that for every $\bar{X} \subset C_i$, $L_i \bar{X}$ is not a superkey for S.

In other words, $L_i C_i$ contains only a key (or keys) of the form $L_i' C_i$ with $L_i' \subseteq L_i$.

By condition (ii), we have

$$L_i \setminus R = L_i \subseteq L \setminus R.$$

On the other hand, from Theorem 1.4.1,

$$L \setminus R \subseteq \Omega \setminus R \subseteq K, \quad \forall K \in K_S.$$

This shows that $L_i C_i$ is a key for S. Q.E.D.

<u>Corollary 1.7.2</u>

If $S = <\Omega, F>$ has a key $K = L_i X$ with $X \subset C_i$, then there exists $j_o \neq i$ such that

$$L_i (C_i \cap L_{j_o} C_{j_o}) \subset L_i C_i$$

<u>Corollary 1.7.3</u>

If $L_i (C_i \cap L_j C_j) = L_i C_i$, $\forall j \neq i$,

then

$$C_i \subseteq H = \bigcup_{K \in K_S} K.$$

In other words, $C_i$ consits of only prime attributes.

Corollary 1.7.4

If $|C_i|=1\ \forall i=1,2,\dots,m$ then $L_jC_j$ is a key for S iff there is no q, $q\neq j$, such that $L_jC_j \supset L_qC_q$.

Theorem 1.7.2

Let $S=\langle\Omega,F\rangle$ be a relation scheme, $L_iZ$ be a key for S,

$$L_i \longleftrightarrow L_j, \quad L_i \cap Z = L_j \cap Z = \emptyset,$$
$$L_j \cap R_h = \emptyset \ \forall h \neq j.$$

Then $L_jZ$ is a key for S.

Proof

It is easy to see that if $L_iZ$ is a key for S and $L_i \longleftrightarrow L_j$ then $L_jZ$ is a superkey for S.
In fact, we have

$$L_jZ \overset{*}{\rightarrow} L_iZ \overset{*}{\rightarrow} \Omega.$$

Moreover, we can prove that for every $Z' \subset Z$, $L_jZ'$ is not a superkey for S. Assume the contrary that $L_jZ'$ is a superkey for S with $Z' \subset Z$.
It is clear that

$$\Omega = (L_jZ')^+ = (L_j^+Z')^+ = (L_i^+Z')^+ = (L_iZ')^+$$

showing that $L_iZ$ is not a key, a contradiction.

The condition $L_j \cap R_h = \emptyset, \forall h$ implies that $L_j \cap R = \emptyset$.
Hence $L_j \subseteq L \setminus R$.
Moreover, again by Theorem 1.4.1

$$L \setminus R \subseteq \Omega \setminus R \subseteq K \quad \forall K \in K_S$$

showing that $L_j Z$ is a key for S.

Theorem 1.7.3

Let $S=<\Omega,F>$ be a relation scheme; $X,Y,Z \subseteq \Omega$, $X \cap Z = Y \cap Z = \emptyset$. Suppose that the following condition hold:

(i) $X \longleftrightarrow Y$

(ii) for every $X' \subset X$ with $|X'| = |X| - 1$
there exists $Y' \subset Y$ such that $Y' \longleftrightarrow X'$ ,

(iii) for every $Y' \subset Y$ with $|Y'| = |Y| - 1$
there exists $X' \subset X$ such that $X' \longleftrightarrow Y'$ .

Then ZX is a key iff ZY is a key.

Proof

We begin to prove the "only if" part.
Suppose that ZX is a key.
Since $X \longleftrightarrow Y$, following the proof of theorem 1.7.2, YZ is a superkey for S while $YZ'$ is not for every $Z' \subset Z$.
In other words, YZ contains only a key (or keys) of the form $Y'Z$ with $Y' \subseteq Y$.

Now, we shall prove that for every $\bar{Y} \subset Y$, $\bar{Y}\,Z$ is not a superkey for S.

The proof is by contradiction.

Let $Y'Z$ is a superkey for S with $\bar{Y} \subseteq Y' \subset Y$ where $|Y'| = |Y| - 1$. Taking the condition (iii) into account we get

$$\Omega = (Y'Z)^{+} = ((Y')^{+}Z)^{+} = ((X')^{+}Z)^{+} = (X'Z)^{+}$$

where $X' \subset X$, $X' \longleftrightarrow Y'$,

showing that XZ is not a key, a contradiction.

Similarly, we can prove the "if part". The proof is complete.

<u>Corollary 1.7.5</u>

Let $S = \langle \Omega, F \rangle$ be a relation scheme, $L_i \longleftrightarrow L_j$, $|L_i| = |L_j| = 1$, $L_i \cap Z = L_j \cap Z = \emptyset$. Then $L_j Z$ is a key for S iff $L_i Z$ is a key.

<u>Proof</u>

It is easy to verify that all conditions of theorem 1.7.3 are satisfied.

<u>Example 1.7.1</u>

We take up again the example in [11]. According to our notation, we have

$\Omega=\{C,I,N,P,T\}$ *)

$F=\{N \to I,\ I \to N,\ NC \to PT,\ PT \to C\}$

It is easy to see that $N \longleftrightarrow I$. So, using the algorithm of Lucchesi and Osborn, after the keys IPT and IC have been found, we can add immediately to the set $K_S$ of found keys two new keys NPT and NC.

Theorem 1.7.4

Let $S=\langle\Omega,F\rangle$ be a relation scheme, and $L_iZ$ is a key for S with $Z \cap L_i = \emptyset$.

If $\quad Z \subset C_j,\quad L_j \longleftrightarrow L_i,$

and

$$L_j(C_j \cap L_h C_h) = L_j C_j,\quad \forall h \neq j$$

then S has no key including $L_j$.

Proof

The condition $L_j \longleftrightarrow L_i$ implies that $L_jZ$ is a superkey for S.

From $Z \subset C_j$, it follows that $L_jC_j$ is not a key.

From $L_j(C_j \cap L_h\ C_h) = L_jC_j$ and $L_jC_j$ is not a key, by corollary 1.7.2, we conclude that S has no key including $L_j$. Q.E.D.

---

*) C,I,N,P,T stand for Course, ID-number, Name, Professor, and Time respectively.

## § 1.8. Three algorithms

Basing upon Theorems 1.4.1. and 1.4.3, we now propose some algorithms for the key searching and key recognition problems. It is worth recalling that:

(i) X is superkey for $S=<\Omega,F>$ iff $X^+=\Omega$;

(ii) $X \overset{*}{\rightarrow} Y$ iff $Y \subseteq X^+$.

### Algorithm 1.

Algorithm for finding one key for the relation scheme $S=<\Omega,F>$ , where

$$\Omega = \{A_1, A_2, \ldots, A_n\},$$

$$F = \{L_i \rightarrow R_i \mid L_i, R_i \subseteq \Omega,\ i=1,2,\ldots,m\},$$

$$L = \bigcup_{i=1}^{m} L_i, \qquad R = \bigcup_{i=1}^{m} R_i,$$

$$L \cap R = \{A_{t_1}, A_{t_2}, \ldots, A_{t_h}\}.$$

The block schema of the Algorithm 1 is presented in Fig. 1.2

### Example 1.8.1

The following example illustrates the performance of Algorithm 1.

Let $S=<\Omega,F>$ be a relation scheme, where

$$\Omega = \{A,B,C,D,E,G\}$$

$$F = \{B \rightarrow C,\ C \rightarrow B,\ A \rightarrow GD\}$$

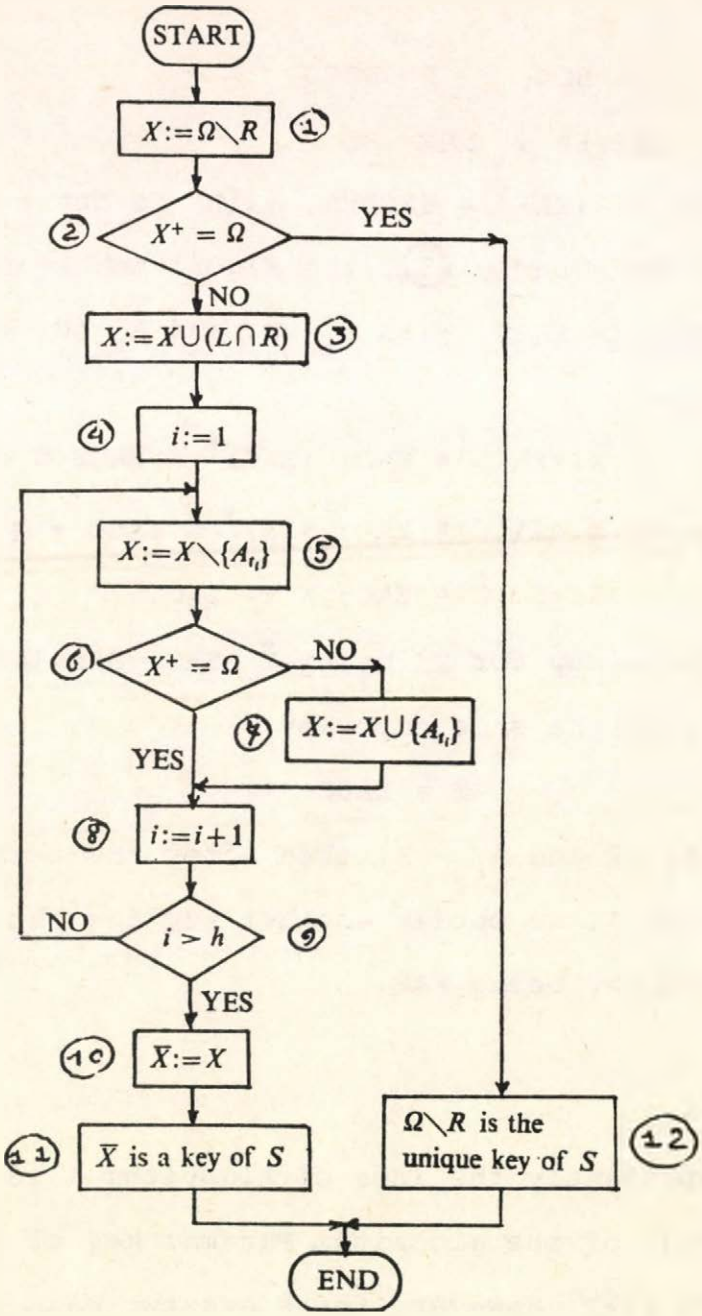

START
X:=Ω\R
1
2
X⁺ = Ω
YES
NO
X:=X∪(L∩R)
3
4
i:=1
X:=X\{A_{t_i}}
5
6
X⁺ = Ω
NO
7
X:=X∪{A_{t_i}}
YES
8
i:=i+1
NO
i > h
9
YES
10
X̄:=X
11
X̄ is a key of S
Ω\R is the unique key of S
12
END

Fig.1.2

We have

$$L = BCA, \quad R = BCGD$$

$$\Omega \setminus R = EA, \quad L \cap R = BC.$$

Since $(\Omega\setminus R)^{+} = (EA)^{+} = EAGD \neq \Omega$, $(\Omega\setminus R)$ is not a key of $S=\langle\Omega,F\rangle$. From the bloc 3, the algorithm begins with the superkey $X = EABC$. With $A_{t_1}=B$, and $A_{t_2}=C$, we have the sequence

$$X:=X\setminus\{B\} = EAC;\ (EAC)^{+} = EACBGD = \Omega;$$

$$X:=X\setminus\{C\} = EA\ ;\ (EA)^{+} = EAGD \neq \Omega,$$

$$X:=X\cup\{C\} = EAC;\ \bar{X} := EAC.$$

We obtained a key for S, being $\bar{X} = EAC$. Similarly, if we start with the same superkey

$$X = EABC$$

but with $A_{t_1}=C$ and $A_{t_2}=B$, then after the termination of Algorithm 1, we obtain another key for the relation scheme $S=\langle\Omega,F\rangle$, being EAB.

<u>Remark 1.8.1.</u>

Independently the idea of Algorithm 1 is quite near to that of the algorithm Minimal key of Lucchesi and Osborn [11]. However, there are two main differences:

1) Algorithm 1 is much more detailed and more easy for implementation.
2) Algorithm 1 takes Theorem 1.4.1 into account and

there-fore only require $O(|F||\Omega||L\cap R|)$ elementary operations (comparison of two attribute names) while algorithm Minimal key require $O(|F||\Omega|^2)$ elementary operations.

(Here $|F|$ denote the cardinality of the set F).

Therefore, as will be shown in the next section, Algorithm 1 can be used together with Algorithm 2 to improve the performance of the second algorithm of Lucchesi and Osborn to find all keys for a relation scheme.

Algorithm 2.

This is an algorithm for finding one key for the relation scheme $S=\langle\Omega,F\rangle$ that is included in a given superkey X.

Suppose that $\bar{X}$ is a key included in X. Then $\bar{X}\subseteq X$.

On the other hand, from Theorem 1.4.1.:

$$\Omega\setminus R\subseteq\bar{X}\subseteq(\Omega\setminus R)\cup(L\cap R).$$

Therefore

$$\bar{X}\subseteq(\Omega\setminus R)\cup(X\cap(L\cap R)).$$

Thus we can start with the superkey

$$(\Omega\setminus R)\cup(X\cap(L\cap R))$$

for finding a key included in a given superkey X. It is easily seen that Algorihtm 2 (see Fig.1.3 ) is similar to Algorithm 1 but block 3 is replaced by the assignment

$$X := (\Omega \setminus R) \cup (X \cap (L \cap R))$$

with $X \cap (L \cap R) = \{A_{\ell_1}, \ldots, A_{\ell_s}\}$ and there are, in addition, some non significant modifications.

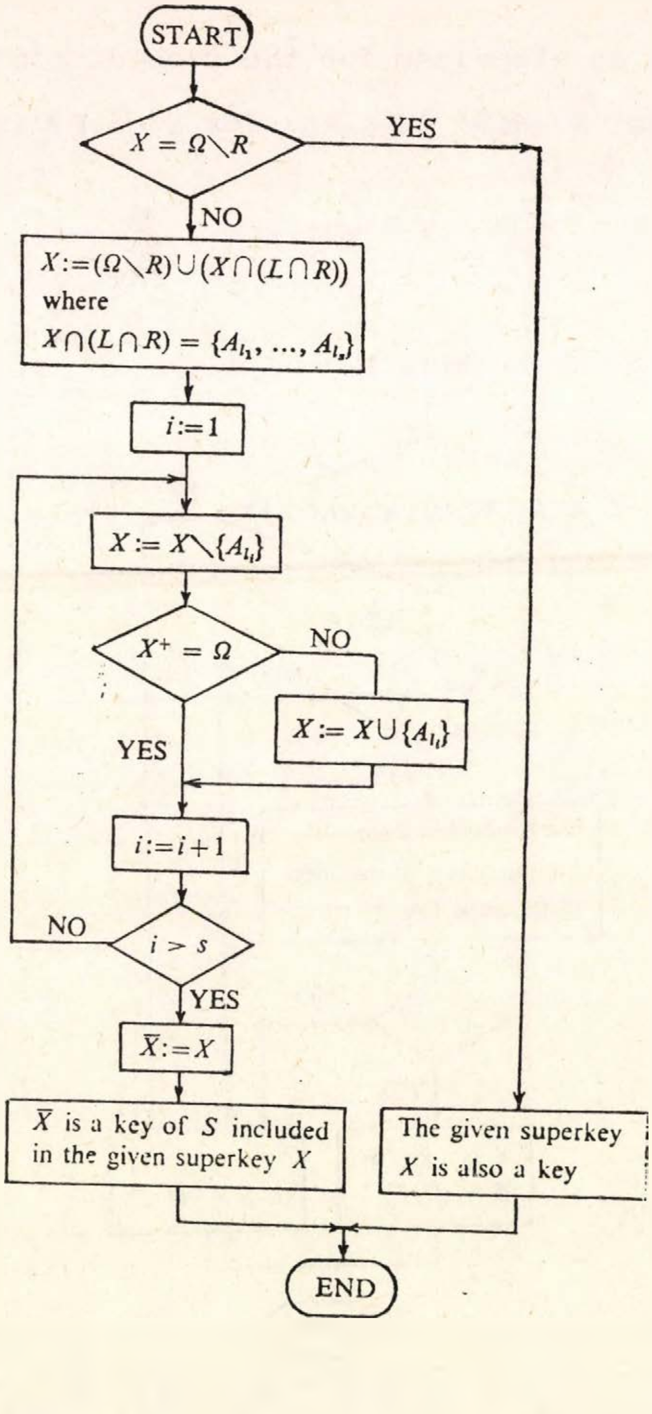

START
$X = \Omega \setminus R$
YES
NO
$X := (\Omega \setminus R) \cup (X \cap (L \cap R))$
where
$X \cap (L \cap R) = \{A_{l_1}, \ldots, A_{l_s}\}$
$i := 1$
$X := X \setminus \{A_{l_i}\}$
$X^+ = \Omega$
NO
$X := X \cup \{A_{l_i}\}$
YES
$i := i+1$
NO
$i > s$
YES
$\bar{X} := X$
$\bar{X}$ is a key of $S$ included in the given superkey $X$
The given superkey $X$ is also a key
END

Fig.1.3

Algorithm 3.

This is an algorithm for the recognition whether a given subset X ($X \subseteq \Omega$) is a key for $S = \langle \Omega, F \rangle$ (see Fig.1.4).

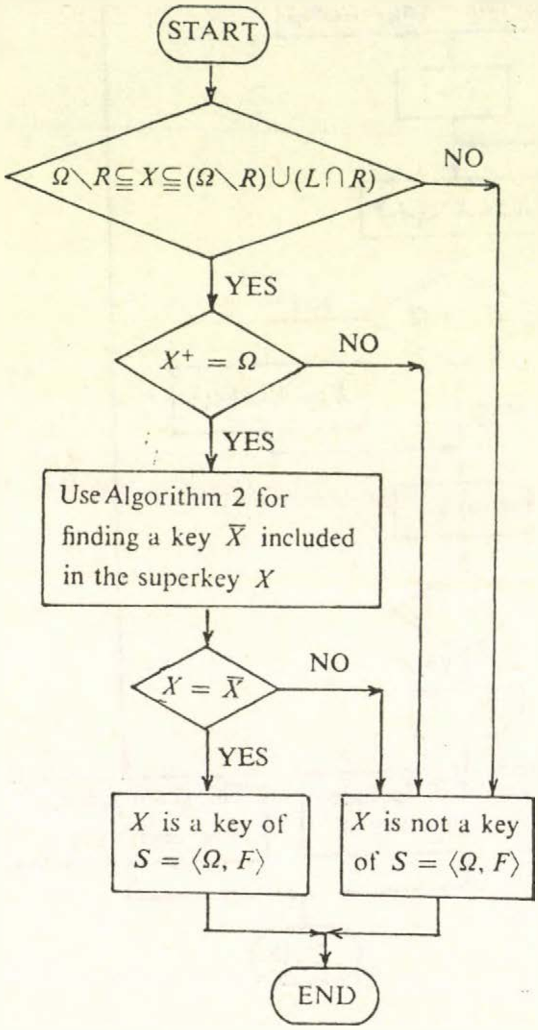



Fig.1.4

## § 1.9. Some remarks on the algorithm of Lucchessi and Osborn

In [11] C.L. Lucchesi and S.L. Osborn gave a very interesting algorithm to determine the set of all keys for any relation scheme $S=\langle\Omega,F\rangle$. The algorithm has time complexity

$$O\ (|F|\,|K_S|\,|\Omega|\,(|K_S|+|\Omega|)),$$

(in our notation), i.e. it is polynomial in $|\Omega|$, $|F|$ and $|K_S|$.

We reproduce here this algorithm with some modifications in accordance with our notation.

Algorithm OL1

Set of all keys for $S=\langle\Omega,F\rangle$;

Comment $K_S$ is the set of keys being accumulated in a sequence which can be scanned in the order in which the keys are entered;

```
K_S ← {Key*) (Ω,F,Ω)};
for each K in K_S do
    for each FD (L_i → R_i) in F do
        T ← L_i ∪ (K\R_i);
        test ← true;
        for each J in K_S do
            if T includes J then test ← false;
            if test then K_S ← K_S ∪ {Key(Ω,F,T)}
    end
end;
return K_S.
```

The following simple remarks can be used to improve in some cases the performance of the algorithm of Lucchesi and Osborn.

Remark 1.9.1

To find the first key for $S=\langle\Omega,F\rangle$, instead of $\Omega$, it is better to use the superkey $(\Omega\setminus R)\cup(L\cap R)$ and algorithm 1 in § 1.8 and instead of the algorithm key $(\Omega,F,T)$, it is better to use algorithm 2 (§ 1.8) for finding one key for S included in a given superkey T.

---

*) Key $(\Omega,F,X)$ is the algorithm which determines a key for $S=\langle\Omega,F\rangle$ that is a subset of a specified superkey X.

Remark 1.9.2

In § 1.4. we have proved that

$$R\setminus L \subseteq \Omega\setminus H,$$

i.e. $R\setminus L$ consists only of non-prime attributes.

Therefore if $R_i \subseteq R\setminus L$ then

$$R_i \cap K = \emptyset, \forall K \in K_S.$$

and $$L_i \cup (K\setminus R_i) \supseteq K.$$

That means, when computing $T=L_i \cup (K\setminus R_i)$, We can neglect all FDs $L_i \to R_i$ with $R_i \subseteq R\setminus L$, for every $K \in K_S$.

Let us denote

$$\overline{\overline{F}} = F\setminus\{L_j \to R_j \mid L_j \to R_j \in F \quad \text{and } R_j \subseteq R\setminus L\}$$

Remark 1.9.3

With a fixed K in $K_S$, it is clear that if

$$K \cap R_i = \emptyset \quad \text{then} \quad L_i \cup (K\setminus R_i) \supseteq K.$$

In that case it is not necessary to check whether T includes J for each J in $K_S$.

So, it is better to compute T by the following way:

$$T = (K\setminus R_i) \cup L_i.$$

Remark 1.9.4

The algorithm of Lucchesi and Osborn is particularly effective when the number of keys for $S=\langle\Omega,F\rangle$ is small.

But, what information we need to conclude that the number of keys for $S=\langle\Omega,F\rangle$ is small? There is no general answer for all the cases and it is shown in [20] that the number of keys for a relation scheme $S=\langle\Omega,F\rangle$ can be factorial in $|F|$ or exponential in $|\Omega|$, and that both of these upper bounds are attainable. However, it is shown (in § 1.4, Corollary 1.4.1) that

$$|K_S| \le C_h^{[h/2]} ,$$

where h is the cardinality of $L\cap R$. Thus if $L\cap R$ has only a few elements then it is a good criterion for saying that S has a small number of keys.
In the case $L\cap R = \emptyset$, $\Omega\setminus R$ is the unique key for $S=\langle\Omega,F\rangle$ as pointed out in § 1.4, Corollary 1.4.4.

Example 2.

Let us return once more to the example in [11, Appendix I].

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

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

It is clear that for this relation scheme

$$L\cap R = \emptyset,$$

and it has exactly one key, namely aceg.

Taking into account the Remarks 1.9.1, 1.9.2, 1.9.3 the above algorithm can be modified as follows:

Algorithm OL2.

```
Set of all keys for S=<Ω,F>;
        K_S ← {Algo 1*) (Ω,F,(Ω\R)∪(L∩R))}
        for each K in K_S do
           for each FD (L_i → R_i) in F̄ such that
                               K\R_i ≠ K do
               T ← (K\R_i)∪L_i;
               test ← true;
               for each J in K_S do
                  if T includes J then test ← false;
               if test then K_S ← K_S∪{Algo2*) (Ω,F,T)}
           end
        end;
        return K_S.
```

---

*) Algo 1 and Algo 2 refer to Algorithm 1 and Algorithm 2 in § 1.8 respectively.

# 2. TRANSLATIONS OF RELATION SCHEMES

## § 2.1. Introduction

In this chapter we shall be concerned with the theory of so-called translations of relation schemes. Starting from a given relation scheme, translations make possible to obtain simpler relation schemes, i.e. those with a less number of attributes and with shorter functional dependencies so that the key finding problem becomes less cumbersome, etc...

On the other hand, from the set of keys of the relation scheme obtained in this way, the corresponding keys of the original scheme can be found by a single "translation".

In § 2.2 we introduce the notion of Z-translation of relation scheme, give a classification of the relation schemes and investigate the characteristic properties of some special classes of Z-translations.

In § 2.3 some subsets of $\Omega^{(o)}$ -the set of all non prime attributes for a relation scheme $S=\langle\Omega,F\rangle$ are described. They will be used in the reduction process for relation schemes.

In § 2.4, the properties of relation schemes belonging to the class $\mathcal{L}_4$ called balanced relation schemes, are investigated.

In § 2.5 the problem of key representation will be formulated and solved. A general scheme to transform an arbitrary relation scheme into a balanced relation scheme and to find all its keys will be presented too.

Finally in § 2.6 we study some properties of the so-called nontranslatable relation scheme.

Most of the results presented in this chapter are published in [7],[8], [38]

## § 2.2. Translation of relation schemes

Definition 2.2.1

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

$$\Omega=\{A_1,A_2,\dots,A_n\}$$

is the set of attributes,

$$F=\{L_i \to R_i \mid L_i,R_i\subseteq\Omega;\ i=1,2,\dots,m\}$$

is the set of functional dependencies (FD) and $Z\subseteq\Omega$ be an arbitrary subset of $\Omega$.

We define a new relation scheme

$$\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle \text{ as follows:}$$

$$\tilde{\Omega}=\Omega\setminus Z\ (=\bar{Z}),$$

$$\tilde{F}=\{L_i\setminus Z \to R_i\setminus Z \mid (L_i \to R_i)\in F,\ i=1,2,\dots,m\}$$

Then $\tilde{S}$ is said to be obtained from S by a Z-translation, and the notation

$$\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle= S-Z =\langle\Omega,F\rangle-Z$$

is used.

Remark 2.2.1

1) Depending on the characteristic properties of the class Z chosen, the corresponding class of translations has its own characteristic features.

2) From the above definition, it is clear that, after the transformation, $\tilde{F}$ can contain the FDs of

the following form:

(i) $\emptyset \rightarrow \emptyset$;

(ii) $X \rightarrow \emptyset$ where $X \subseteq \tilde{\Omega}$, $X \neq \emptyset$;

(iii) $\emptyset \rightarrow X$ where $X \subseteq \tilde{\Omega}$, $X \neq \emptyset$.

However, by the algorithm for the finding the closure $X^+$ of the subset $X \subseteq \Omega$ w.r.t.F (see § 1.2), we observe that the omission of FDs of the form (i) and (ii) in $\tilde{F}$ does not change $K_{\tilde{S}}$, the set of all keys for $\tilde{S}$. Later, we will show that all FDs of the form (iii) can be omitted too.

Definition 2.2.2

Let $S=\langle\Omega,F\rangle$ be a relation scheme, and $K_S$ be the set of all keys for S. We define a partition of $\Omega$ as follows:

$\Omega=\Omega^{(o)} \cup \Omega^{(1)} \cup \Omega^{(2)}$, such that

$\Omega^{(i)} \cap \Omega^{(j)} = \emptyset$; $i \neq j$; $i,j \in \{0,1,2\}$

where

$$\Omega^{(2)} = G = \bigcap_{K \in K_S} K;$$

$$\Omega^{(1)} = (\bigcup_{K \in K_S} K) \setminus G = H \setminus G;$$

$$\Omega^{(o)} = \Omega \setminus H.$$

Sometimes, for the sake of simplicity, the notation

$$\Omega = \Omega^{(0)} | \ \Omega^{(1)} | \ \Omega^{(2)} = \Omega^{(0)} |_H$$

is also used.

<u>Definition 2.2.3.</u>

Let $\Omega$ be the universe of attributes,

$$X \subseteq \Omega, \ \mathfrak{M} \subseteq 2^{\Omega}, \ \mathfrak{N} \subseteq 2^{\Omega}.$$

we define

$$X \circledast \mathfrak{M} = \{XY \mid Y \in \mathfrak{M}\}$$

$$\mathfrak{M} \circledast \mathfrak{N} = \{YZ \mid Y \in \mathfrak{M}, \ Z \in \mathfrak{N}\}$$

Here XY stands for $X \cup Y$.

Now, we give a classification of relation schemes as follows:

$\mathcal{L}_0 = \{<\Omega,F> \mid <\Omega,F> \text{ is a relation scheme}\};$

$\mathcal{L}_1 = \{<\Omega,F> \mid <\Omega,F> \in \mathcal{L}_0 \text{ and } \Omega = L \cup R\};$

$\mathcal{L}_2 = \{<\Omega,F> \mid <\Omega,F> \in \mathcal{L}_0 \text{ and } L \subseteq R = \Omega\};$

$\mathcal{L}_3 = \{<\Omega,F> \mid <\Omega,F> \in \mathcal{L}_0 \text{ and } R \subseteq L = \Omega\};$

$\mathcal{L}_4 = \{<\Omega,F> \mid <\Omega,F> \in \mathcal{L}_0 \text{ and } L = R = \Omega\}.$

Following this classification, it is easiliy seen that:

$\alpha)\ \mathcal{L}_4 \subseteq \mathcal{L}_3 \subseteq \mathcal{L}_1 \subseteq \mathcal{L}_0$ ;

$\beta)\ \mathcal{L}_4 \subseteq \mathcal{L}_2 \subseteq \mathcal{L}_1 \subseteq \mathcal{L}_0$ ;

$\gamma)\ \mathcal{L}_4 = \mathcal{L}_2 \cap \mathcal{L}_3$ .

Figure 2.1 shows the hierarchy of classes $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4$.

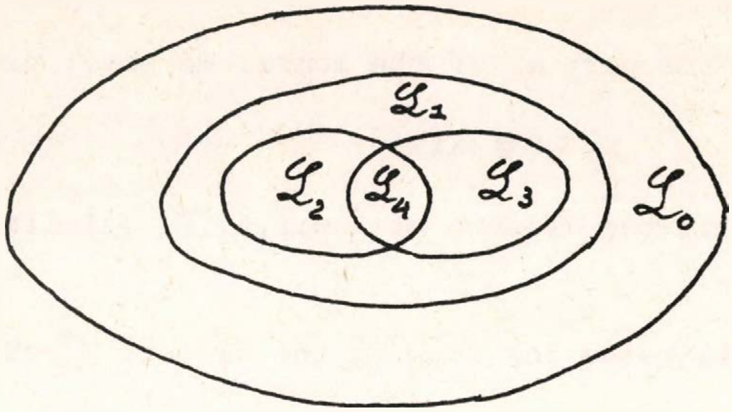



Fig. 2.1

The next lemma is fundamental for this chapter.

Lemma 2.2.1

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

and $\tilde{S}=S-Z$, $Z\subseteq\Omega$.

Then

a) $X \overset{*}{\underset{F}{\rightarrow}} Y$ implies $X\setminus Z \overset{*}{\underset{\tilde{F}}{\rightarrow}} Y\setminus Z$,

b) $X \overset{*}{\underset{\tilde{F}}{\rightarrow}} Y$ implies $X\cup Z \overset{*}{\underset{F}{\rightarrow}} Y\cup Z$,

where $X \overset{*}{\underset{F}{\rightarrow}} Y$ means $(X \rightarrow Y)\in F^+$ and similarly, $X \overset{*}{\underset{\tilde{F}}{\rightarrow}} Y$ means $(X \rightarrow Y)\in \tilde{F}^+$.

<u>Proof</u>

For the part a) of the lemma, we shall prove that

$$X_F^+ \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^+ \qquad (2.2.1)$$

where $X_F^+$ is the closure of $X$ w.r.t. $F$. (similarly for $(X\setminus Z)_{\tilde{F}}^+$).

By the algorithm for finding the closure $X^+$ of $X$ [13; see also § 1.2], with $X_F^{(o)}=X$, $(X\setminus Z)_{\tilde{F}}^{(o)} = X\setminus Z$, we have

$$X_F^{(o)} \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^{(o)} \; .$$

Suppose that

$$X_F^{(i)} \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^{(i)}, \qquad (2.2.2)$$

we shall prove that (2.2.2) holds for (i+1) as well.

Indeed we have

$$X_F^{(i+1)} \setminus Z = (X_F^{(i)} \cup (\bigcup_{\substack{L_j \subseteq X_F^{(i)} \\ (L_j \to R_j) \in F}} R_j ))\setminus Z =$$

$$= (X_F^{(i)} \setminus Z) \cup (\bigcup_{\substack{L_j \subseteq X_F^{(i)} \\ (L_j \to R_j) \in F}} R_j \setminus Z) \subseteq$$

$$\subseteq (X\setminus Z)_{\tilde{F}}^{(i)} \cup (\bigcup_{\substack{L_j \subseteq X_F^{(i)} \\ (L_j \to R_j) \in F}} (R_j \setminus Z)),$$

(by virtue of the inductive assumption (2.2.2)).

On the other hand, from $L_j \subseteq X_F^{(i)}$ and the inductive assumption (2.2.2), we have:

$$L_j \setminus Z \subseteq X_F^{(i)} \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^{(i)} \quad .$$

Consequently,

$$X_F^{(i+1)} \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^{(i)} \cup \Big( \bigcup_{L_j \subseteq X_F^{(i)}} (R_j \setminus Z)\Big) \subseteq$$

$$\subseteq (X\setminus Z)_{\tilde{F}}^{(i+1)}$$

Thus (2.2.1) has been proved.

Now, it is well known that

$$X \overset{*}{\underset{F}{\rightarrow}} Y \iff Y \subseteq X_F^+$$

Hence, from $X \overset{*}{\underset{F}{\rightarrow}} Y$, we have

$$Y\setminus Z \subseteq X_F^+ \setminus Z \subseteq (X\setminus Z)_{\tilde{F}}^+ \ ,$$

showing that

$$X\setminus Z \quad \overset{*}{\underset{\tilde{F}}{\rightarrow}} \quad Y\setminus Z .$$

Similarly, for the part b) of the lemma, we shall prove by induction that

$$X_{\tilde{F}}^+ \cup Z \subseteq (X\cup Z)_F^+ \qquad (2.2.3)$$

By the algorithm for finding the closure $X^+$ of X we have

$$X_{\tilde{F}}^{(o)} \cup Z \subseteq (X\cup Z)_F^{(o)}$$

Suppose that

$$X_{\tilde{F}}^{(i)} \cup Z \subseteq (X\cup Z)_F^{(i)} \quad , \qquad (2.2.4)$$

we shall prove that (2.2.4) also holds for (i+1). Indeed, we have:

$$X_{\tilde{F}}^{(i+1)} \cup Z = X_{\tilde{F}}^{(i)} \cup \Big( \bigcup_{\substack{L_j \setminus Z \subseteq X_{\tilde{F}}^{(i)} \\ (L_j \setminus Z \to R_j \setminus Z) \in \tilde{F}}} (R_j \setminus Z) \Big) \cup Z =$$

$$= (X_{\tilde{F}}^{(i)} \cup Z) \cup \Big( \bigcup_{L_j \setminus Z \subseteq X_{\tilde{F}}^{(i)}} (R_j \setminus Z) \Big) \subseteq$$

$$\subseteq (X \cup Z)_F^{(i)} \cup \Big( \bigcup_{L_j \setminus Z \subseteq X_{\tilde{F}}^{(i)}} R_j \Big)$$

(by the virtue of the inductive assumption (2.2.4)).

On the other hand, from $L_j \setminus Z \subseteq X_{\tilde{F}}^{(i)}$ and from (2.2.4) we have

$$L_j \subseteq X_{\tilde{F}}^{(i)} \cup Z \subseteq (X \cup Z)_F^{(i)} .$$

Consequently,

$$X_{\tilde{F}}^{(i+1)} \cup Z \subseteq (X \cup Z)_F^{(i)} \cup \Big( \bigcup_{L_j \setminus Z \subseteq X_{\tilde{F}}^{(i)}} R_j \Big) \subseteq$$

$$\subseteq (X \cup Z)_F^{(i+1)}$$

Thus (2.3.3) has been proved.

From $X \overset{*}{\underset{\tilde{F}}{\to}} Y$ we have $Y \subseteq X_{\tilde{F}}^{+}$.

Hence

$$Y \cup Z \subseteq X_{\tilde{F}}^{+} \cup Z \subseteq (X \cup Z)_F^{+} ,$$

showing that

$$X \cup Z \overset{*}{\underset{F}{\to}} Y \cup Z .$$

The proof is complete.

We are now in a position to prove the following theorems.

<u>Theorem 2.2.1</u>

Let $S=<\Omega,F>$ be a relation scheme,

$$Z\subsetneq G, \quad \tilde{S}=<\tilde{\Omega},\tilde{F}>= S-Z.$$

Then $X$ is a key for $\tilde{S}$ if and only if $X\cap Z=\emptyset$ and $XZ$ is a key for $S$.

<u>Proof</u>

We first prove the necessity.

Suppose that $X$ is a key for $\tilde{S}$. Obviously, $X\subsetneq\tilde{\Omega}$. Therefore $X\cap Z=\emptyset$.

Since $X$ is a key for $\tilde{S}$, we have

$$X \overset{*}{\underset{F}{\to}} \tilde{\Omega}.$$

Taking Lemma 2.2.1 into account, we get

$$XZ \overset{*}{\underset{F}{\to}} \tilde{\Omega}Z =\Omega,$$

showing that $XZ$ is a superkey for $S$. Assume that $XZ$ is not a key for $S$, then there would exist a key $\bar{X}$ for $S$ such that

$$Z\subsetneq\bar{X}\subsetneq XZ$$

(The validity of the first inclusion is due to the fact $Z \subseteq G$ - the intersection of all keys for S).

Consequently, there would exist $X_1 \subset X$ such that

$$\bar{X} = X_1 Z, \; X_1 \cap Z = \emptyset.$$

Since $\bar{X}$ is supposed to be a key for S,

$$X_1 Z \overset{*}{\underset{F}{\to}} \Omega.$$

Using lemma 2.2.1, clearly

$$X_1 Z \setminus Z \overset{*}{\underset{\tilde{F}}{\to}} \Omega \setminus Z,$$

that is

$$X_1 \overset{*}{\underset{\tilde{F}}{\to}} \tilde{\Omega}.$$

This contradicts the hypothesis that X is a key for $\tilde{S}$. Thus XZ is a key for S.

We now turn to the proof of sufficiency. Suppose that $X \cap Z = \emptyset$ and XZ is a key for S. We have to show that X is a key for $\tilde{S}$. Since XZ is a key for S, we have

$$XZ \overset{*}{\underset{F}{\to}} \Omega$$

By virtue of lemma 2.2.1, we get

$$XZ \setminus Z \overset{*}{\underset{F}{\to}} \Omega \setminus Z$$

Consequently (from $X \cap Z = \emptyset$):

$$X \underset{\tilde{F}}{\to} \tilde{\Omega},$$

showing that X is a superkey for $\tilde{S}$. Assume that X is not a key for $\tilde{S}$. Then, there would exist a key $\bar{X}$ for $\tilde{S}$ such that

$$\bar{X} \subset X \quad \text{and} \quad \bar{X} \overset{*}{\underset{\tilde{F}}{\to}} \tilde{\Omega}$$

Applying Lemma 2.2.1, it follows:

$$\bar{X}Z \overset{*}{\underset{F}{\to}} \tilde{\Omega}\ Z = \Omega,$$

where

$$\bar{X}Z \subset XZ.$$

This contradicts the fact that XZ is a key for S.

Hence X is a key for $\tilde{S}$.

The proof is complete

Theorem 2.2.2

Let $S=<\Omega,F>$ be a relation scheme, $Z\subseteq\Omega$, $Z\cap H=\emptyset$ and

$$\tilde{S} = <\tilde{\Omega},\tilde{F}> = S-Z.$$

Then X is a key for $\tilde{S}$ if and only if X is a key for S.

Proof

First, observe that if X is a superkey for S then after removing from X some non prime attributes, the remaining part of X is also a superkey for S. In other words, if X is a superkey for S, then with all $Z\subseteq\Omega^{(o)}$ (equivalently $Z\cap H=\emptyset$), $X'=X\setminus Z$ is also a superkey for S.

Now we begin to prove the only if part of the theorem.

Suppose that X is a key for $\tilde{S}$.

Obviously

$$X \overset{*}{\underset{F}{\to}} \tilde{\Omega}.$$

By virtue of lemma 2.2.1, we have

$$XZ \overset{*}{\underset{F}{\to}} \tilde{\Omega}Z = \Omega,$$

showing that XZ is a superkey for S. In view of the above observation, we find that X is also a superkey for S.

Assume that there exists a key $\bar{X}$ for S such that $\bar{X} \subset X$.

Applying Lemma 2.2.1, we have

$$\bar{X} \setminus Z \overset{*}{\underset{F}{\to}} \Omega \setminus Z$$

or

$$\bar{X} \overset{*}{\underset{\tilde{F}}{\to}} \tilde{\Omega}.$$

This contradicts the fact that X is a key for $\tilde{S}$.

Hence X is a key for S too.

<u>The if part</u>.

Suppose that X is a key for S. We have to prove that X is also a key for $\tilde{S}$. We have, by the definition of a key

$$X \overset{*}{\underset{F}{\to}} \Omega.$$

Applying lemma 2.2.1

$$X \setminus Z \overset{*}{\underset{\tilde{F}}{\to}} \Omega \setminus Z = \tilde{\Omega}.$$

Since $Z \cap H = \emptyset$, it follows $Z \cap X = \emptyset$. Consequently

$$X \overset{*}{\underset{\tilde{F}}{\to}} \tilde{\Omega},$$

showing that X is a superkey for $\tilde{S}$.

Now, assume the contrary that X is not a key for $\tilde{S}$.
Then there would exist a key $\bar{X}$ for $\tilde{S}$ such that $\bar{X} \subset X$.
Obviously

$$\bar{X} \overset{*}{\underset{\tilde{F}}{\rightarrow}} \tilde{\Omega}.$$

We invoke Lemma 2.2.1 to deduce

$$\bar{X}Z \overset{*}{\underset{F}{\rightarrow}} \tilde{\Omega}Z = \Omega,$$

showing that $\bar{X}Z$ is a superkey for S.
Since $Z \cap H = \emptyset$, using again the observation at the beginning of this proof, we find that $\bar{X}$ is a superkey for S, a contradiction.
Hence X is a key for $\tilde{S}$.
The proof is complete.

According to our notation it is easily seen that both Theorems 2.2.1 and 2.2.2 can be formulated in the form of a single theorem as follows:

<u>Theorem 2.2.3</u> [33]

Let $S=\langle\Omega,F\rangle$ be a relation scheme, $Z \subseteq \Omega$, and $\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle=S-Z$.
Then:
(i) $K_S=K_{\tilde{S}}$ iff $Z \subseteq \Omega^{(o)}$
(ii $K_S=Z \oplus K_{\tilde{S}}$ iff $Z \subseteq G$.

Basing upon Theorem 2.2.3, in the following we investigate only the class of Z-translations with

$Z\neq\emptyset$, $Z=Z_1\cup Z_2$, $Z_1\cap Z_2=\emptyset$, $Z_1\subseteq G$, $Z_2\cap H=\emptyset$.

Bearing this in mind, if

$$\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle = S-Z,\ S=\langle\Omega,F\rangle,$$

then applying Theorems 2.2.2 and 2.2.1 consecutively one after another to the $Z_2$-translation and the $Z_1$-translation, we have: X is a key for $\tilde{S}$ if and only if $X\cap Z = \emptyset$ and $XZ_1$ is a key for S.

For the sake of convenience we use in the sequel the notation

$$\langle\Omega,F\rangle \xrightarrow[\rho=(Z,Z_1)]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where the meaning of $\rho$ is obvious. To continue, let us recall some results in § 1.4.

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

$$\Omega=\{A_1,A_2,\dots,A_n\},$$

$$F=\{L_i \to R_i \mid L_i,R_i\subseteq\Omega,\ i=1,2,\dots,m\}$$

As usual, let us denote by

$$L=\bigcup_{i=1}^{m} L_i, \qquad R=\bigcup_{i=1}^{m} R_i.$$

Then, the necessary condition under which X, a subset of $\Omega$, is a key for S is that

$$\Omega\setminus R\subseteq X\subseteq(\Omega\setminus R)\cup(L\cap R).$$

For $V\subseteq\Omega$, we denote by $\bar{V}=\Omega\setminus V$. It is easily seen that:

$$\overline{L\cup R}\subseteq\Omega\setminus R\subseteq G;$$

$$L\setminus R\subseteq\Omega\setminus R\subseteq G;$$

$$R\setminus L\subseteq\bar{H}$$

Consequently $(R\setminus L)\cap H=\emptyset$.

Moreover, we have the following lemma:

<u>Lemma 2.2.2</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme, $Z\subseteq G$ where $G$ is the intersection of all keys for $S$.

Then $$(Z^{+}\setminus Z)\cap H=\emptyset.$$

<u>Proof</u>

Assume the contrary that

$$(Z^{+}\setminus Z)\cap H\neq\emptyset.$$

Then, there would exist an attribute $A\in Z^{+}$, $A\notin Z$ and $A\in H$. Consequently, there exists a key $X$ for $S=\langle\Omega,F\rangle$ such that $A\in X$. Since $Z\subseteq X$, $A\in Z^{+}$ and $A\notin Z$, we infer that

$$Z\subseteq X\setminus A.$$

Hence $X\setminus A \overset{*}{\rightarrow} Z \overset{*}{\rightarrow} Z^{+} \overset{*}{\rightarrow} A$, with $A\in X$.

This contradicts the fact that $X$ is a key for $S$ /see Lemma 1.3.5 in § 1.3/. The proof is complete. It is worth noticing that Theorem 1.5.3 is only a special case of Lemma 2.2.2. From the results just mentioned above, the following theorems are obvious.

<u>Theorem 2.2.4</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_0$,

$$\langle\tilde{\Omega},\tilde{F}\rangle = \langle\Omega,F\rangle - \overline{L\cup R}$$

Then

$$\langle\Omega,F\rangle \xRightarrow[\rho=(\overline{L\cup R},\ \overline{L\cup R})]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

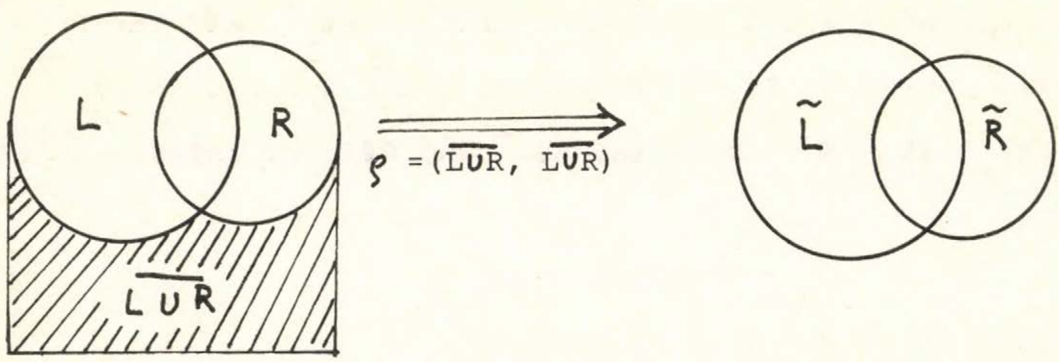
where $\langle\tilde{\Omega},\tilde{F}\rangle\in\mathcal{L}_1$.

<u>Proof</u>

As pointed out above, $\overline{L\cup R}\in G$. Applying Theorem 2.2.1 to the Z-translation $\tilde{S}=S-Z$ with $Z=\overline{L\cup R}$, we have

$$\langle\Omega,F\rangle \xRightarrow[\rho=(\overline{L\cup R},\ \overline{L\cup R})]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

Theorem 2.2.4 is illustrated by Fig. 2.2.




$\langle\Omega,F\rangle \in \mathcal{L}_0$ $\qquad\qquad$ $\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-\overline{L\cup R}\in\mathcal{L}_1$

Fig. 2.2

<u>Example 2.2.1</u>

Let be given $S=\langle\Omega,F\rangle$ with

$$\Omega=\{a,b,c,d,e\},\ F=\{c\to d,\ d\to e\}$$

we have $\overline{L\cup R} = ab$

Consider

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-ab.$$

Obviously

$$\tilde{\Omega}=\{c,d,e\},\ \tilde{F}=\{c\to d,\ d\to e\}$$

It is easily seen that c is the unique key for $\langle\tilde{\Omega},\tilde{F}\rangle$.

Hence abc is the unique key for $\langle\Omega,F\rangle$.

<u>Theorem 2.2.5</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_1$ ,

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-(\overline{L\cup R}\cup(L\setminus R)).$$

Then

$$\langle\Omega,F\rangle \xRightarrow[\rho=(\overline{L\cup R}\cup(L\setminus R),\ \overline{L\cup R}\cup(L\setminus R))]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

with

$$\langle\tilde{\Omega},\tilde{F}\rangle\in\mathcal{L}_2$$

<u>Proof</u>

It is clear that

$$Z = \overline{L\cup R}\cup(L\setminus R) =\Omega\setminus R\in G.$$

The Theorem 2.2.5 now follows from applying Theorem

2.2.1 to the Z-translation $\tilde{S}$=S-Z. Theorem 2.2.5 is illustrated by figure 2.3.

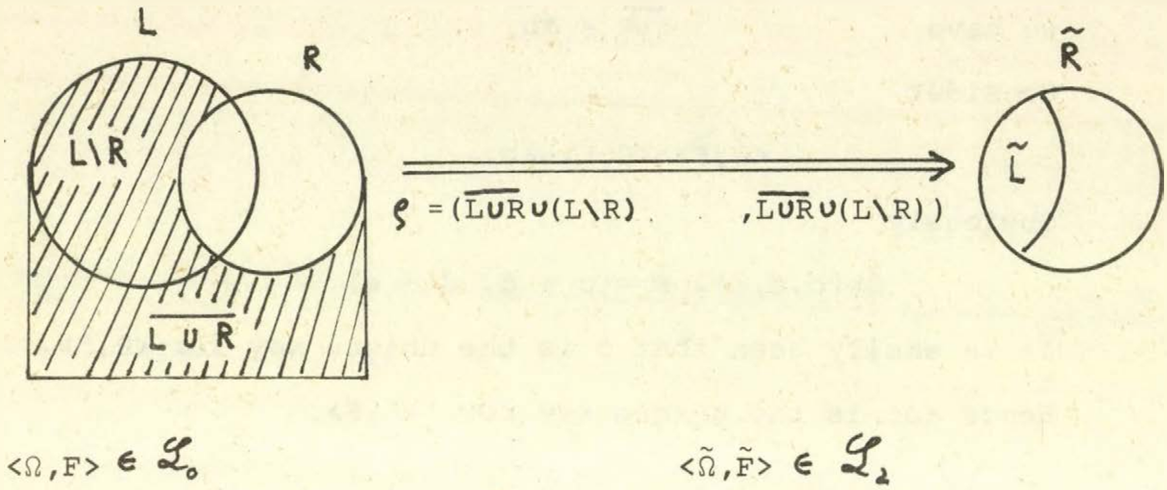



Fig. 2.3.

<u>Theorem 2.2.6</u>

Let S=$\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_0$

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-(\overline{L \cup R} \cup ( R\setminus L )).$$

Then

$$\langle\Omega,F\rangle \xrightarrow[\rho = (\overline{L \cup R} \cup (R\setminus L), \overline{L \cup R})]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

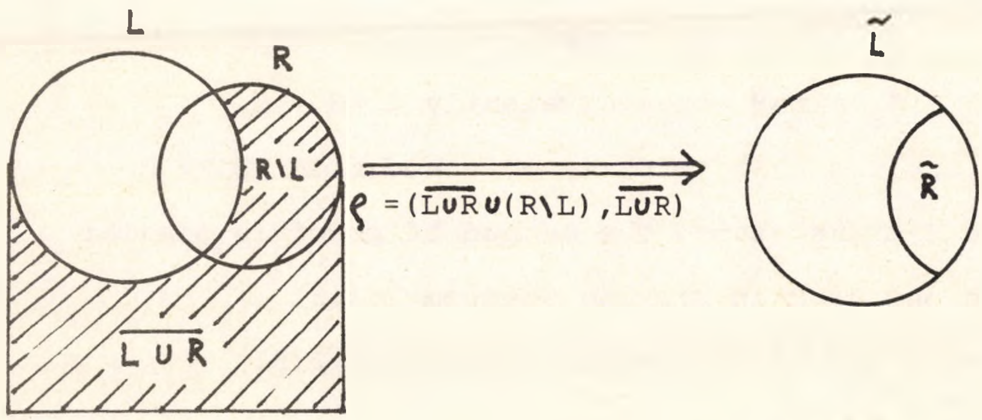
where $\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_3$

Proof

As remarked above, $R\setminus L \subseteq \bar{H}$.

Let $Z = \overline{L\cup R} \cup (R\setminus L) = Z_1 \cup Z_2$, where $Z_1 = \overline{L\cup R} \subseteq G$, $Z_2 = R\setminus L$, $Z_2 \cap H = \emptyset$. The Theorem 2.2.6 now follows from sequential applications of Theorems 2.2.2 and 2.2.1 to the $Z_2$-translation $S' = S - Z_2$ and the $Z_1$-translation $\tilde{S} = S' - Z_1$ respectively.

Theorem 2.2.6 is illustrated by Fig. 2.4.




$\langle\Omega,F\rangle \in \mathcal{L}_0$ $\qquad\qquad$ $\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_3$

Fig 2.4

Theorem 2.2.7

Let $S = \langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_0$,

$\langle\tilde{\Omega},\tilde{F}\rangle = \langle\Omega,F\rangle - (\overline{L\cup R} \cup (L\setminus R) \cup (R\setminus L))$.

Then

$$\langle\Omega,F\rangle \xRightarrow[\varrho = (\overline{L\cup R}\,\cup(L\setminus R)\,\cup(R\setminus L)\,,\ \overline{L\cup R}\,\cup(L\setminus R))]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where

$$\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_4$$

<u>Proof</u>

Let $Z = \overline{L\cup R}\cup(L\setminus R)\cup(R\setminus L) = Z_1 \cup Z_2$,

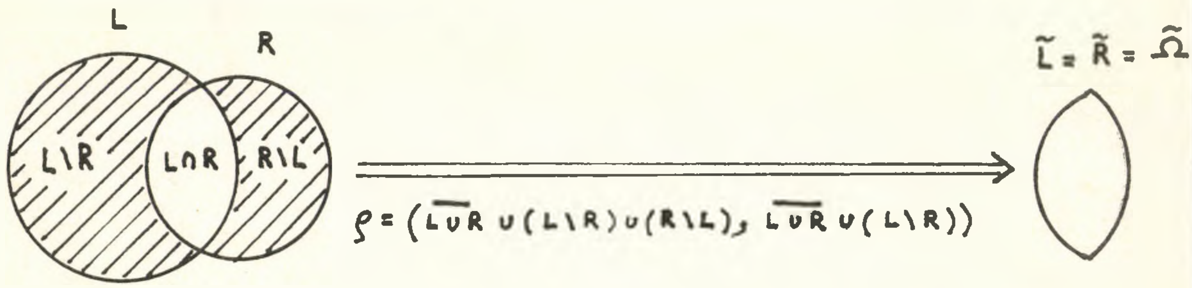
where

$Z_1 = \overline{L\cup R}\cup(L\setminus R) = \Omega\setminus R \subseteq G$,

$Z_2 = R\setminus L \subseteq \bar{H}$ or equivalently $Z_2 \cap H = \emptyset$.

It is obvious that $\langle\tilde{\Omega},\tilde{F}\rangle$ is obtained from $\langle\Omega,F\rangle$ by the Z-translation. The method of proof is similar to the one used in proving Theorem 2.2.6.

Theorem 2.2.7 is illustrated by figure 2.5.




$\langle\Omega,F\rangle \in \mathcal{L}_0$ $\qquad\qquad$ $\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_4$

Fig. 2.5

Similarly, we can prove the following theorems.

Theorem 2.2.8

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_1$,

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-(L\setminus R).$$

Then

$$\langle\Omega,F\rangle \underset{\rho=(L\setminus R,\ L\setminus R)}{\Longrightarrow} \langle\tilde{\Omega},\tilde{F}\rangle$$

where $\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_2$

Theorem 2.2.8 is illustrated by Fig. 2.6.

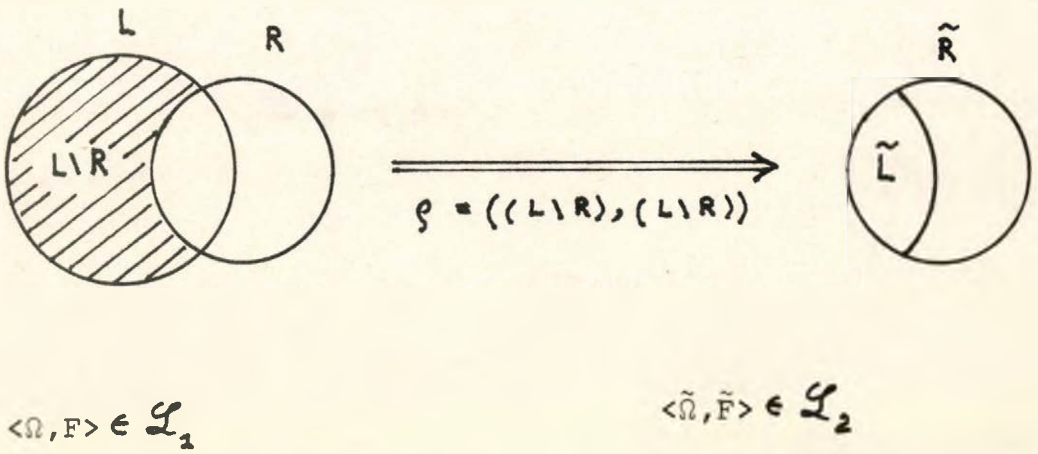



Fig. 2.6.

Theorem 2.2.9

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_1$

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-(R\setminus L).$$

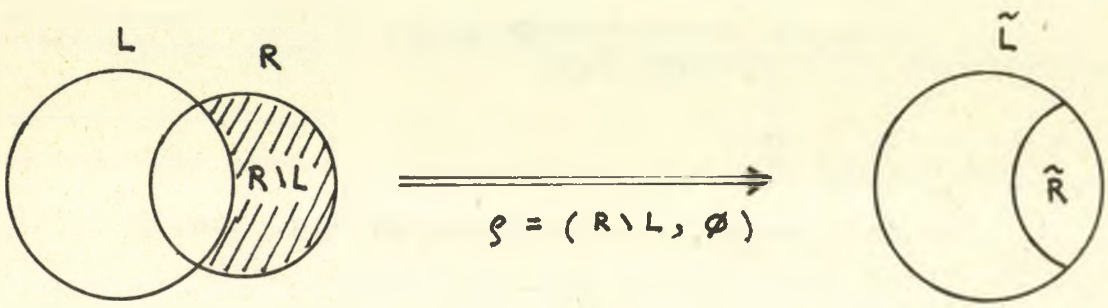
Then

$$\langle\Omega,F\rangle \xrightarrow[\varrho=(R\setminus L,\ \emptyset)]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where

$$\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_3 .$$

Theorem 2.2.9 is illustrated by Fig. 2.7.




$\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_3$

$\langle\Omega,F\rangle \in \mathcal{L}_1$

Fig. 2.7

<u>Theorem 2.2.10</u>

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_1$

$$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-((L\setminus R)\cup(R\setminus L)) .$$

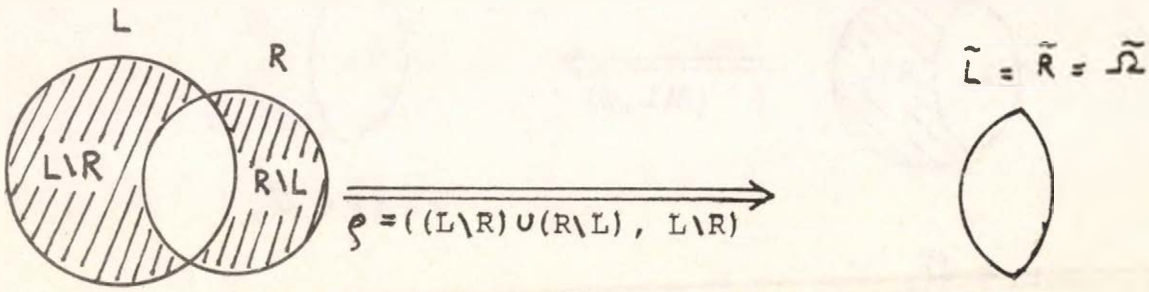
Then

$$\langle\Omega,F\rangle \xrightarrow[\varrho=((L\setminus R)\cup(R\setminus L),\ L\setminus R)]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where

$$<\tilde{\Omega},\tilde{F}> \in \mathcal{L}_4$$

Theorem 2.2.10 is illustrated by Fig. 2.8.




$<\Omega,F> \in \mathcal{L}_1$ $\qquad\qquad$ $<\tilde{\Omega},\tilde{F}> \in \mathcal{L}_4$

Fig. 2.8.

<u>Theorem 2.2.11</u>

Let $<\Omega,F>$ be a relation scheme in $\mathcal{L}_2$

$$<\tilde{\Omega},\tilde{F}> = <\Omega,F> - (R\setminus L).$$

Then

$$<\Omega,F> \underset{\rho=(R\setminus L,\emptyset)}{\Longrightarrow} <\tilde{\Omega},\tilde{F}>$$

where

$$<\tilde{\Omega},\tilde{F}> \in \mathcal{L}_4.$$

Theorem 2.2.11 is illustrated by Fig. 2.9.

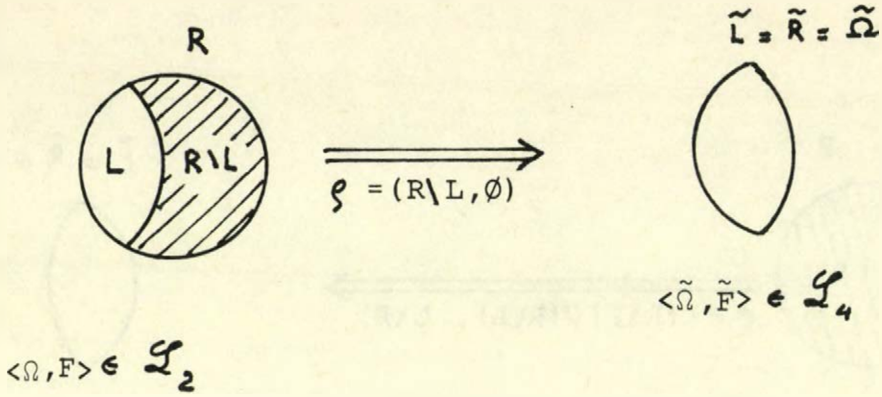



Fig. 2.9

<u>Theorem 2.2.12</u>

Let $\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_3$

$$\langle\tilde{\Omega},\tilde{F}\rangle = \langle\Omega,F\rangle - (L \setminus R)\,.$$

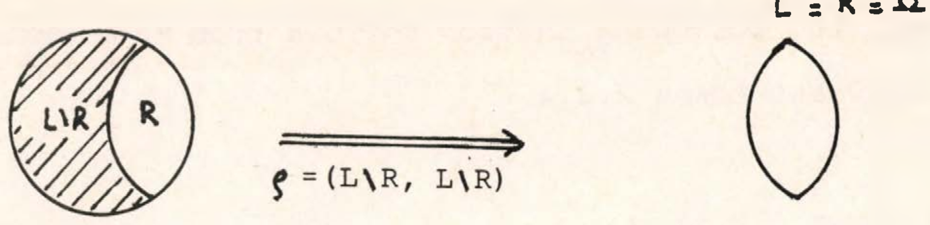
Then

$$\langle\Omega,F\rangle \xRightarrow[\rho = (L \setminus R,\; L \setminus R)]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where

$$\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_4$$

Theorem 2.2.12 is illustrated by Fig. 2.10.




$\langle \Omega, F \rangle \in \mathcal{L}_3$ $\qquad\qquad$ $\langle \tilde{\Omega}, \tilde{F} \rangle \in \mathcal{L}_4$

Fig. 2.10

Combining Theorems 2.2.4 - 2.2.12, we have the diagram of translations of relation schemes as illustrated by figure 2.11.

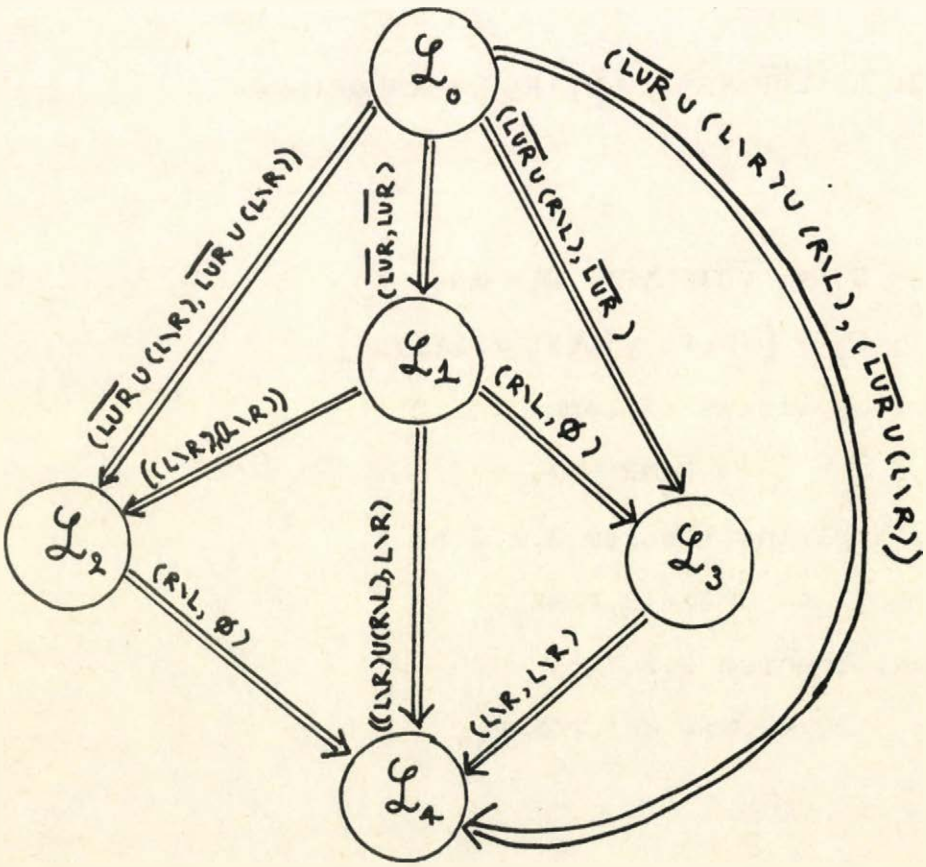



Fig. 2.11

Now, the following theorem follows from Theorems 2.2.1 2.2.2 and Lemma 2.2.2.

Theorem 2.2.13

Let $S=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_0$

$\langle\tilde{\Omega},\tilde{F}\rangle=\langle\Omega,F\rangle-\{\overline{L\cup R}\cup(L\setminus R)^{+}\cup(R\setminus L)\}$.

Then

$$\langle\Omega,F\rangle \xRightarrow[\varrho=(\overline{L\cup R}\cup(L\setminus R)^{+}\cup(R\setminus L),\ \overline{L\cup R}\cup(L\setminus R))]{} \langle\tilde{\Omega},\tilde{F}\rangle$$

where

$$\langle\tilde{\Omega},\tilde{F}\rangle \in \mathcal{L}_4$$

Proof

Put $Z=\overline{L\cup R}\cup(L\setminus R)\cup[(L\setminus R)^{+}\setminus(L\setminus R)]\cup(R\setminus L)=$

$=Z_1\cup Z_2$,

where

$$Z_1 = \overline{L\cup R}\cup(L\setminus R)=\Omega\setminus R \cup G,$$

$$Z_2 = [(L\setminus R)^{+}\setminus(L\setminus R)]\cup(R\setminus L).$$

Clearly, by virtue of Lemma 2.2.2,

$$Z_2\cap H = \emptyset.$$

Now, by applying Theorem 2.2.2 to

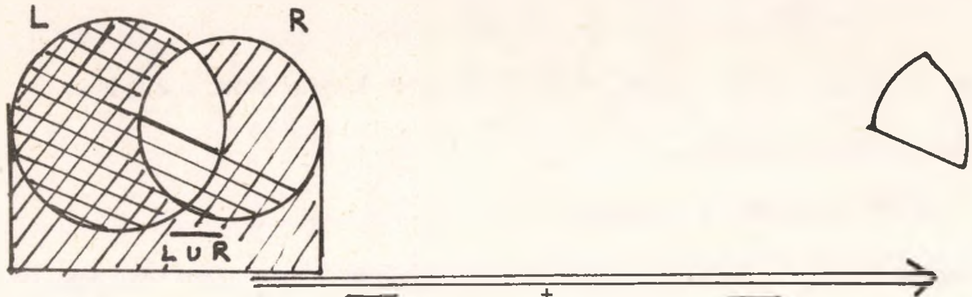
$$\langle\Omega',F'\rangle=\langle\Omega,F\rangle-Z_2,$$

and then, Theorem 2.2.1 to

$$\langle\tilde{\Omega},\tilde{F}\rangle= \langle\Omega',F'\rangle-Z_1,$$

the proof of Theorem 2.2.13 is immediate.

Theorem 2.2.13 is illustrated by Fig. 2.12




$$\rho = (\overline{L\cup R} \cup (L\setminus R)^{+} \cup (R\setminus L),\ \overline{L\cup R}\cup(L\setminus R))$$

$\langle\Omega, F\rangle \in \mathcal{L}_0$ $\qquad\qquad$ $\langle\tilde{\Omega}, \tilde{F}\rangle \in \mathcal{L}_4$

Fig. 2.12

From the just mentioned results, we have the following diagram of translations of relation schemes (Fig. 2.13)

<u>Example 2.2.2</u>

Let $\Omega$ = abhgqmnvwkl,

F={a → b, b → h, g → q, kv → w, w → vl}

we have

L=abgkvw; R=bhqwvl; R\L=hql;

L\R=kga; $(L\setminus R)^{+}$ =kgabhq; $\overline{L\cup R}$=mn;

$$(R\setminus L)\cup(L\setminus R)^{+}\cup(\overline{L\cup R}) = mnkgabhql\ .$$

$$<\tilde{\Omega},\tilde{F}> = <\Omega,F> - mnkgabhql =$$

$$= <wv, \{v \to w,\ w \to v\}> .$$

It is easily seen that v and w are keys for $<\tilde{\Omega},\tilde{F}>$.

On the other hand

$$(\overline{L\cup R})\cup(L\setminus R) = mnkga.$$

Consequently, mnkgav and mnkgaw are keys for $<\Omega,F>$.

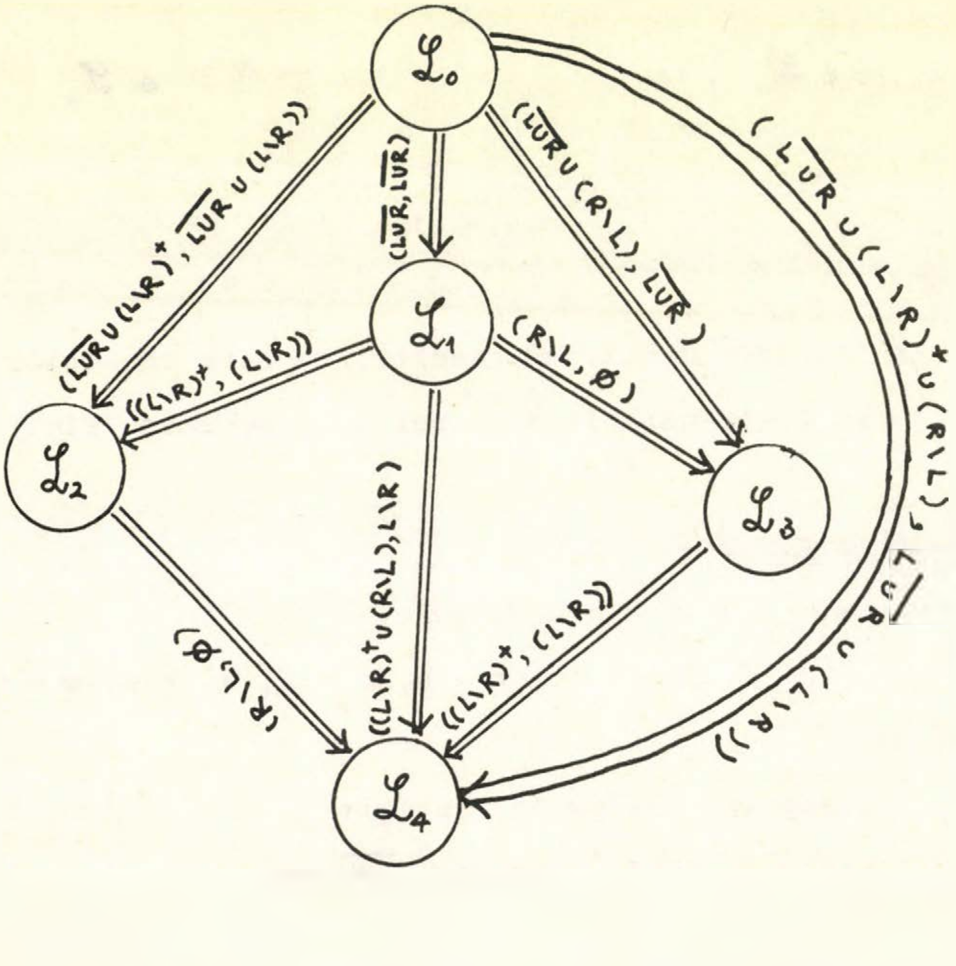



Fig. 2.13

## §. 2.3 Subsets of $\Omega^{(o)}$

By the nature NPC of the problem [11], in opposition to G, we have not got the explicit expression for the set $\Omega^{(o)}$ (equivalently, for H - the union of all keys for S). Recall that $\Omega^{(o)}=\Omega\setminus H$ is the set of all non-prime attributes for S.

However in §1.4 it is shown that

$$R'=R\setminus L\subseteq\Omega^{(o)} .$$

The aim of this section is twofold. First we shall prove that, after applying a Z-translation to a relation scheme $S=<\Omega,F>$, we can delete in the obtained relation scheme $\tilde{S}=<\tilde{\Omega},\tilde{F}>=<\Omega,F>-Z$ all FDs of the form $\emptyset\longrightarrow X$ $(X\neq\emptyset)$, while preserving $K_{\tilde{S}}$ - the set of all keys for $\tilde{S}$.

Secondly, we present a method for extending a given subset of $\Omega^{(o)}$ to a greater one. In doing so results in § 2.2 can be improved.

We begin with showing the following lemma

Lemma 2.3.1

Let $S=<\Omega,F>$ be a relation scheme.

Then

$$R'' = \bigcup_{L_i\subseteq G} R_i\subseteq\Omega^{(o)} .$$

Proof

If $A \in R''$ then there exists $(L_i \rightarrow R_i) \in F$ such that $A \in R_i$, $A \notin L_i$, $L_i \subseteq G$. (Recall that F is in natural reduced form, i.e. $L_i \cap R_i = \emptyset, \forall i=1,2,\dots,m$, and

$L_i \neq L_j$ if $i \neq j$).

Let K be an arbitrary key for S. We shall show that $A \notin K$. Assume the contrary that $A \in K$. From $A \notin L_i$ and $L_i \subseteq K$, we have

$$L_i \subseteq K \setminus \{A\} = K'.$$

Obviously:

$$L_i \rightarrow R_i \overset{*}{\rightarrow} \{A\}$$
$$K' \overset{*}{\rightarrow} L_i$$

Consequently

$$K' \overset{*}{\rightarrow} \{A\}$$

Combining with $K' \overset{*}{\rightarrow} K'$, we get

$$K' \overset{*}{\rightarrow} K' \cup \{A\} = K.$$

This contradicts the fact that K is a key.

Hence, $\forall K \in K_S : A \notin K$, i.e. $A \in \Omega^{(o)}$.

Corollary 2.3.1

$$\bigcup_{L_i=\emptyset} R_i \subseteq \bigcup_{L_i \subseteq G} R_i \subseteq \Omega^{(o)}$$

The proof is obvious.

This corollary shows that we can eliminate from a relation scheme all FDs of the form $\emptyset \to R_i$, while preserving its set of all keys.

The following lemma gives us a constructive way for extending a given subset of $\Omega^{(o)}$.

<u>Lemma 2.3.2</u>

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

For every $X \subseteq G$, $Y \subseteq \Omega^{(o)}$, we have

$$(XY)^+ \setminus X \subseteq \Omega^{(o)}.$$

<u>Proof</u>

If $A \in (XY)^+ \setminus X$ then $A \in (XY)^+$ and $A \notin X$

Suppose that $A \notin \Omega^{(o)}$.

Obviously $A \notin Y$.

Since $A \in (XY)^+$, so $XY \overset{*}{\to} A$.

From $A \notin X$, $A \notin Y$, it follows that

$$A \notin XY.$$

Since $A \notin \Omega^{(o)}$, there exists a key $K \in K_S$ such that $A \in K$.

Let $K' = K \setminus \{A\}$, $K' \subset K$.

It is clear that

$$YXK' \overset{*}{\to} K'\{A\} = K,$$

showing that $XYK'$ is a superkey for S. After removing from $XYK'$ the subset $Y \subseteq \Omega^{(o)}$, $XK'$ is still a superkey for S.

On the other hand, from $X \subseteq G \subseteq K$, $A \notin X$, we have

$$X \subseteq K \setminus \{A\} = K',$$

showing that $XK' = K'$ is a superkey for S.

This contradicts the fact that K is a key for S.

Hence we must have $A \in \Omega^{(o)}$. Since A is arbitrary, so

$$(XY)^{+} \setminus X \subseteq \Omega^{(o)}.$$

The proof is complete.

<u>Corollary 2.3.2</u>

$$(GR')^{+} \setminus G \subseteq \Omega^{(o)}$$

<u>Proof</u>

By direct use of Lemma 2.3.2 with

$$X = G, \quad Y = R' = R \setminus L \subseteq \Omega^{(o)}.$$

<u>Example 2.3.1</u>

We consider one example in which

$$R' \subset (GR')^{+} \setminus G \subseteq \Omega^{(o)}$$

and so, showing that our Lemma 2.3.2 is non trivial.

Let

$$\Omega = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9$$

$$F = \{137 \to 2,\ 27 \to 134,\ 1238 \to 49,\ 7 \to 23,$$
$$1458 \to 236,\ 368 \to 159\}$$

we have:

$$L = \bigcup_{i=1}^{6} L_i = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8$$

$$R = \bigcup_{i=1}^{6} R_i = 1\ 2\ 3\ 4\ 5\ 6\ 9; \quad R' = R \setminus L = 9$$

$$G = \Omega \setminus R = 7\ 8$$

$$(GR')^{+} = (7\ 8\ 9)^{+} = 7\ 8\ 9\ 1\ 2\ 3\ 4.$$

$$(GR')^{+} \setminus G = 12349 \supset 9.$$

Results in this section will be used in the next section to improve the results in § 2.2.

## § 2.4 The balanced relation scheme

Definition 2.4.1

The relation scheme $S=\langle\Omega,F\rangle$ is called balanced if the following conditions hold:

(i) $\bigcup_{i=1}^{m} L_i = \bigcup_{i=1}^{m} R_i = \Omega$;

(ii) $L_i \cap R_i = \emptyset$, $\forall i=1,2,\ldots,m$;

(iii) $\forall i,j=1,2,\ldots,m$, $i\neq j$ implies $L_i\neq L_j$.

where

$\Omega = \{A_1,A_2,\ldots,A_n\}$

$F = \{L_i \rightarrow R_i \mid L_i,R_i \subseteq \Omega,\ i=1,2,\ldots,m\}$.

In other words a balanced relation scheme is a relation scheme in $\mathcal{L}_4$ and in the natural reduced form.

From Definition 2.4.1, we can prove the following simple properties of a balanced relation scheme (b.r.s.).

Proposition 2.4.1

Let $S=\langle\Omega,F\rangle$ be a b.r.s.

Then:

1. $G=\emptyset$;
2. If $|\Omega|\leq 1$ then $K_S=\{\emptyset\}$;

3. $\emptyset \notin K_S$ iff $|K_S| \geq 2$ ;

4. $\forall Z \subseteq \Omega$, $S-Z$ is a b.r.s.

Proof

1. By the definition of a b.r.s., we have

$$G=\Omega\setminus R=\Omega\setminus\Omega=\emptyset.$$

2. If $\Omega=\emptyset$, it is obvious that $K_S=\{\emptyset\}$.

The case $\Omega=\{A\}$.

From (i) (def 2.4.1), we have

$$R=L=\Omega=\{A\}.$$

From (ii) (def. 2.4.1), F contains only two FD: $\{A\}\rightarrow\emptyset$ and $\emptyset\rightarrow\{A\}$, showing that $\emptyset$ is the unique key for S.

3. Suppose $|K_S| \geq 2$. Then $\emptyset \notin K_S$, since otherwise $\emptyset$ will be the unique key for S.

Conversely, suppose that $\emptyset \notin K_S$. Then $K_S$ has at least two elements, since otherwise, if $K_S=\{K\}$ then from $G=K$ and $G=\emptyset$ it follows that $K=\emptyset$, a contradiction.

4. This property is straightforward.

Theorem 2.4.1

Let $S=\langle\Omega,F\rangle$ be an arbitrary given relation scheme (not necessary be in natural reduced form), where

$$\Omega=\{A_1,A_2,\ldots,A_n\},$$

$$F=\{L_i\rightarrow R_i \mid L_i,R_i \subseteq \Omega,\ i=1,2,\ldots,m\}.$$

Then there exists a b.r.s $\tilde{S}=<\tilde{\Omega},\tilde{F}>$ such that $K_S = G \oplus K_{\tilde{S}}$, where G is the intersection of all keys for S.

Proof

Without loss of generality, we can always assume that, for the relation scheme S,

$$L_i \cap R_i = \emptyset, \quad i=1,2,...,m.$$

(Otherwise, we replace S by $S_1=<\Omega,F_1>$, where $F_1=\{L_i \rightarrow R_i \setminus L_i \mid (L_i \rightarrow R_i) \in F, \ i=1,2,...,m\}$.
It is easy to show that $F^+=F_1^+$ [13] and therefore $K_S=K_{S_1}$).

We construct the b.r.s. as follows:

1. Compute

$$L = \bigcup_{i=1}^{m} L_i ; \quad R = \bigcup_{i=1}^{m} R_i ; \quad R'=R \setminus L ;$$

$$G = \Omega \setminus R ; \quad Z = (GR')^+ .$$

(It is worth noticing that

$$Z=(GR')^+ = G \cup [(GR')^+ \setminus G]$$
$$= Z_1 \cup Z_2 ,$$

where $Z_1 = G$,

$$Z_2 = [(GR')^+ \setminus G] \subseteq \Omega^{(o)}$$

(see § 2.3)).

Now, consider the relation scheme

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

where $\Omega' = \Omega \setminus Z$ ,

$F' = \{ L'_i \longrightarrow R'_i \mid i=1,2,\ldots,m \}$

with $L'_i = L_i \setminus Z$, $R'_i = R_i \setminus Z$.

It is obvious that:

$$L'_i \cap R'_i = \emptyset \ , \ i=1,2,\ldots,m \ ;$$

$$V = \bigcup_{i=1}^{m} L'_i = L \setminus Z \quad \text{and } W = \bigcup_{i=1}^{m} R'_i = R \setminus Z.$$

2. We shall prove that: $V \subseteq \Omega' \subseteq W \subseteq V$ to deduce that $V = \Omega' = W$.

Indeed, if $A \in V$, then $A \in L$ and $A \notin Z$.

It is obvious that $A \in \Omega$.

Consequently $A \in \Omega \setminus Z = \Omega'$.

Hence $V \subseteq \Omega'$.

Now let $A \in \Omega' = \Omega \setminus Z$.

It follows that $A \notin Z$.

Since $A \notin Z = (GR')^{+} \supseteq GR'$, so

$$A \notin G \text{ and } A \notin R'.$$

By virtue of $G = \Omega \setminus R$, we find that $A \in R$.

From $A \in R$ and $A \notin Z$, we deduce $A \in R \setminus Z$. Therefore $\Omega' \subseteq W$.

Finally, if $A \in W = R \setminus Z$, then $A \in R$ and $A \notin Z$. Arguing as above, we get

$$A \notin G \text{ and } A \notin R' (= R \setminus L) \cdot$$

Since $A \in R$ and $A \notin R \setminus L$, we deduce $A \in L$.

From $A \in L$ and $A \notin Z$, we have $A \in L \setminus Z$, showing that $W \subseteq V$.

Thus we have shown: $L' = R' = \Omega'$.

3. If there are several FDs in $F'$ which the same left side, we can replace them by a FD which has the left side as the common one and its right side is the union of the right sides of the relevant FDs.

It is easy to see that the above transformation does not change the closure of $F'$, and thus, the set $K_S$, too.

Denote by $\tilde{S}$, the relation scheme obtained from $S'$ after performing the above substitutions. It is clear that $\tilde{S}$ is the desired balanced relation scheme, and by theorem 2.2.3

$$K_S = G \oplus K_{\tilde{S}}.$$

## § 2.5. The problem of key representation

First, we give another characterization of Z-translation of relation schemes, formulated in form of Theorem 1.1 in [33].
Here we provide another proof of this theorem.

Theorem 2.5.1.

Let $S=<\Omega,F>$ be a relation scheme, and $Z\subseteq\Omega$.
If $\tilde{S} = S-Z = <\tilde{\Omega},\tilde{F}>$, $(\tilde{\Omega}=\Omega\setminus Z=\bar{Z})$ then for every $X\subseteq\bar{Z}$
we have

$$Z\,(X)^{+}_{\tilde{F}} = (ZX)^{+}_{F}\ . \qquad (2.5.1)$$

Formula (2.5.1) expresses the relationship between closures in the source relation S and in the target one $\tilde{S}$.

Proof

First we prove that $(ZX)^{+}_{F} \subseteq Z(X)^{+}_{\tilde{F}}$ .
Let $A\in(ZX)^{+}_{F}$ .
If $A\in ZX$ then obviously $A\in Z(X)^{+}_{\tilde{F}}$ . We have only to consider the case $A\notin ZX$, i.e. $A\notin Z$ and $A\notin X$.
From $A\in(XZ)^{+}_{F}$ , we have

$$ZX \ \overset{*}{\underset{F}{\rightarrow}}\ A. \qquad (2.5.2)$$

By virtue of Lemma 2.2.1, (2.5.2) implies

$$ZX\setminus Z \ \underset{\tilde{F}}{\rightarrow}\ A\setminus Z,$$

or

$$X \overset{*}{\underset{\tilde{F}}{\rightarrow}} A,$$

showing that $A \in (X)^{+}_{\tilde{F}}$.

Hence $A \in Z\ (X)^{+}_{\tilde{F}}$

Thus we have proved that

$$(ZX)^{+}_{F} \subseteq Z\ (X)^{+}_{\tilde{F}}\ .$$

To complete the proof, it remains to prove that

$$Z(X)^{+}_{\tilde{F}} \subseteq (ZX)^{+}_{F}$$

Let $A \in Z\ (X)^{+}_{\tilde{F}}$ .

Just like the above reasoning, we have only to consider the case

$$A \notin ZX, \text{ i.e. } A \notin Z \text{ and } A \notin X.$$

From $A \in Z\ (X)^{+}_{\tilde{F}}$ and $A \notin Z$ we get

$$A \in (X)^{+}_{\tilde{F}}\ , \text{ i.e.}$$

$$X \overset{*}{\underset{\tilde{F}}{\rightarrow}} A. \tag{2.5.3}$$

By virtue of Lemma 2.2.1, (2.5.3) implies

$$ZX \overset{*}{\underset{F}{\rightarrow}} Z\{A\},$$

or, equivalently

$$Z\{A\} \subseteq (ZX)^{+}_{F}\ ,$$

showing that $A \in (ZX)^{+}_{F}$ .

Thus $Z\ (X)^{+}_{\tilde{F}} \subseteq (ZX)^{+}_{F}$ .

Combining these two results we get the required equality (2.5.1).

The proof is complete.

Definition 2.5.1

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

$\Omega=\{A_1,\ldots,A_n\}$

$$F=\{L_i \to R_i \mid i=1,2,\ldots,m\}.$$

Let us denote by

$$\mathcal{L}_S=\{L_i \mid i=1,2,\ldots,m\},$$

the set of all left sides of F.

Construct the directed graph $\mathcal{G}_S$ as follows:

(1) $\mathcal{L}_S$ is the set of nodes of $\mathcal{G}_S$;

(2) $(L_i,L_j)$ is an arc of $\mathcal{G}_S$ iff $L_i \supset L_j$ and there is no $L_k$ such that $L_i \supset L_k \supset L_j$.

Let $\bar{\mathcal{L}}_S$ is the set of all terminal nodes of $\mathcal{G}_S$, i.e., nodes for which the outdegree is equal to zero. The members of $\bar{\mathcal{L}}_S$ are called minimal left sides of S.

Lemma 2.5.1

Let $L_i$ be an arbitrary element of $\mathcal{L}_S$, and

$$\tilde{S}= S-(L_i)_F^+ .$$

Then the elements of $L_i \oplus K_{\tilde{S}}$ are superkeys for S.

Proof

Let $Z= (L_i)_F^+$ . Then $\forall \tilde{K} \in K_{\tilde{S}}$ ,

$$(L_i\tilde{K})^+_F = ((L_i)^+_F\ \tilde{K})^+_F = (Z\tilde{K})^+_F = Z\ (\tilde{K})^+_{\tilde{F}} =$$
$$= Z\bar{Z} = \Omega,$$

(by virtue of Lemma 1.3.1 and Theorem 2.5.1), showing that $L_i\tilde{K}$ is a superkey for S.

<u>Theorem 2.5.2.</u> (key representation).

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

Then each key for S can be represented in the form:

$$K=L_i\tilde{K},$$

where $L_i$ is a minimal left side of S, i.e. $L_i\in\bar{\mathcal{L}}_S$ and $\tilde{K}$ is a key for the relation scheme $\tilde{S} = S-(L_i)^+_F$ .

<u>Proof</u>

Let K be any key for S, i.e. $K\in K_S$.

If $K=\Omega$ then, of course, K contains all elements of $\mathcal{L}_S$.

If $K\neq\Omega$, so $K\in K^+_F=\Omega$. That means, there exists $L_j\in\mathcal{L}_S$ such that $L_j\subseteq K$ and $R_j\setminus K\neq\emptyset$. (This follows from the algorithm to find the closure of a set of attributes w.r.t. F).

Starting from the node $L_j$ of the graph $\mathcal{G}_S$, we move along the arcs until a node $L_i\in\bar{\mathcal{L}}_S$ is reached. Obviously $L_i\subseteq L_j$. Thus we have proved that:

$$\forall K\in K_S,\ \exists L_i\in\bar{\mathcal{L}}_S \text{ such that } L_i\subseteq K.$$

Let $Z=(L_i)^+_F$.

If $Z=\Omega$, so, by lemma 1.3.2, $L_i$ is a key for S and we have $K=L_i\emptyset$. But in that case $\tilde{S}=S-Z=S-\Omega=\langle\emptyset,\{\emptyset\to\emptyset\}\rangle$, and clearly $\emptyset$ is a key for $\tilde{S}$.

If $Z\subset\Omega$ we can write

$$K=L_i|\tilde{K},\text{ i.e. } K=L_i\cup\tilde{K},\ L_i\cap\tilde{K}=\emptyset.$$

We shall prove that $\tilde{K}$ is a key for $\tilde{S}$.

By Lemma 1.3.3 we have

$$(L_i)_F^+\cap(K\setminus L_i)=Z\cap\tilde{K}=\emptyset,\ (\tilde{K}=K\setminus L_i)$$

Consequently

$$\tilde{K}\subseteq\bar{Z}=\tilde{\Omega}.$$

Moreover, again from Lemma 1.3.3,

$$\tilde{K}=K\setminus L_i=K\setminus(L_i)_F^+ .$$

Therefore, by Lemma 2.2.1, $\tilde{K}$ is a superkey for $\tilde{S}$.

Now, suppose that there is $\tilde{K}'\subset\tilde{K}$ and $\tilde{K}'$ is a key for $\tilde{S}$.

Again, by Lemma 2.2.1, $Z\tilde{K}'=(L_i)_F^+\tilde{K}'$ is a superkey for S.

Thus, using Lemma 1.3.1, we get

$$\Omega=(Z\tilde{K}')_F^+=((L_i)_F^+\tilde{K}')_F^+=(L_i\tilde{K}')_F^+$$

showing $L_i\tilde{K}'$ is a superkey for S. On the otherhand it is clear that $L_i\tilde{K}'\subset K$.

This contradicts the fact that K is a key for S. Hence $\tilde{K}'=\tilde{K}$, i.e. $\tilde{K}$ is a key for $\tilde{S}$. The proof is complete.

<u>Remark 2.5.1</u>

Lemma 1.3.6 in § 1.3 can be considered as an immediate

consequence of Theorem 2.5.2.

Remark 2.5.2.

In general, the converse of Theorem 2.5.2 is not true. It is quite possible that there exists $L_i \in \bar{\mathcal{L}}_S$ that is not contained in any key for S, as shown by the following example.

Example 2.5.1.

Let be given

$$\Omega = 1\ 2\ 3\ 4\ 5$$

$$F = \{24 \to 35,\ 15 \to 4,\ 53 \to 124,\ 25 \to 134\}$$

we have

$$\mathcal{L}_S = \bar{\mathcal{L}}_S = \{24, 15, 53, 25\}.$$

The graph $\mathcal{G}_S$ consists of all disjoint nodes (Fig 2.14).

$$\begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ 24 & 15 & 53 & 25 \end{array}$$

Fig. 2.14

Direct computation shows that:

$$(24)^+ = (53)^+ = (25)^+ = \Omega.$$

Therefore 24,53 and 25 are keys for S. On the other hand:

$$(15)^+ = 154 \neq \Omega.$$

It is clear that 15 is not contained in any key for S because 152 contains the key 25 and 153 contains the key 35.

Corollary 2.5.1

Let $L_i \in \bar{\mathcal{L}}_S$, $\tilde{S}=S-(L_i)_F^+$ .

If $K \in L_i \oplus K_{\tilde{S}}$ and except $L_i$, K does not contain any other minimal left side $L_j \in \bar{\mathcal{L}}_S$ with $j \neq i$, then K is a key for S.

Proof

By virtue of Lemma 2.5.1, K is a superkey for S. Suppose that $K' \subseteq K$ and $K'$ is a key for S. We shall prove that $K'=K$. Since $L_i$ is the unique element of $\bar{\mathcal{L}}_S$, contained in K, so $K'$ contains at most only $L_i$. If $K'$ does not contain $L_i$ then

$$(K')_F^+ = K' \neq \Omega.$$

Thus $K'$ must contain $L_i$.

We have $K'= L_i \tilde{K}'$.

Since $K \in L_i \oplus K_{\tilde{S}}$, $K=L_i\tilde{K}$, $\tilde{K} \in K_{\tilde{S}}$.

From $L_i\tilde{K}'=K' \subseteq K=L_i\tilde{K}$, and $L_i \cap \tilde{K}' = L_i \cap \tilde{K} = \emptyset$, we deduce $\tilde{K}' \subseteq \tilde{K}$. Since $\tilde{K}'$ and $\tilde{K}$ are keys for $\tilde{S}$, so $\tilde{K}'=\tilde{K}$. Thus $K'=K$.

We are now ready to present a general scheme to transform an arbitrary relation scheme (in natural reduced form) into a balanced relation scheme and to find all its keys.

Let be a given relation scheme

$$S=\langle\Omega,F\rangle$$

where

$$\Omega=\{A_1,A$$

$$F=\{L_i \to R_i \mid L_i,R_i \subseteq \Omega, i=1,2,\dots,m\}$$

<u>Step 1</u>

Compute $L=\bigcup_{i=1}^{m} L_i$; $R=\bigcup_{i=1}^{m} R_i$;

$$R'=R\setminus L;\ G=\Omega\setminus R;$$

$$Z=(GR')^{+}$$

<u>Step 2</u>

Define $\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle= S-Z$

where

$$\tilde{\Omega}=\Omega\setminus Z;$$

$$\tilde{F}=\{L_i\setminus Z \to R_i\setminus Z \mid i=1,2,\dots,m\}.$$

Eliminate from $\tilde{F}$ all FDs of the form:

$$\emptyset \to \emptyset, \emptyset \to X, X \to \emptyset \quad (X\neq\emptyset).$$

Thus we obtain the b.r.s $\tilde{S}=\langle\tilde{\Omega},\tilde{F}\rangle$.

Step 3

Find all keys for $\tilde{S}$.

Construct:

$\mathcal{L}_{\tilde{S}}$ the set of all left sides of $\tilde{F}$; the graph $\mathcal{G}_{\tilde{S}}$;

$\bar{\mathcal{L}}_{\tilde{S}}$ - the set of minimal left sides of $\tilde{F}$.

Let $\bar{\mathcal{L}}_{\tilde{S}} = \{\tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_k\}$.

Compute $Z_i = (\tilde{L}_i)^+_{\tilde{F}}$ , $i=1,2,\ldots,k$ .

If $Z_i = \tilde{\Omega}$ then $\tilde{L}_i \in K_{\tilde{S}}$ .

Denote by $I=\{j \mid Z_j \neq \tilde{\Omega}\} \subseteq \{1,2,\ldots,k\}$

For $Z_j \neq \tilde{\Omega}$, consider the b.r.s.

$$\mathcal{S}_j = \tilde{S} - Z_j \quad \forall j \in I .$$

Repeat the step 3 for the relation schemes $\mathcal{S}_j$. Suppose that at some moment we found all keys for $\mathcal{S}_j$, $j \in I$.

$$K_{\mathcal{S}_j} = \{K_1^{(j)}, K_2^{(j)}, \ldots, K_{s_j}^{(j)}\}, \ \forall j \in I.$$

To complete the set $K_{\tilde{S}}$, we perform as follows:

Consider sequentially the sets $L_j K_t^{(j)}$, for each $j \in I$, $t = 1,2,\ldots,s_j$.

(i) If $L_j K_t^{(j)}$ contains a key already found of $K_{\tilde{S}}$, so we omit it;

(ii) If $L_j K_t^{(j)}$ contains no element of $\bar{\mathcal{L}}_{\tilde{S}}$ but $L_j$ then $L_j K_t^{(j)} \in K_{\tilde{S}}$ .

(iii) Otherwise, use algorithm 3 in §1.8 to check whether $L_j K_t^{(j)}$ is a key.

Step 4

Compute $K_S = G \oplus K_{\tilde{S}}$ .

Remark 2.5.3.

Alternatively, to find all keys for $\tilde{S}$ (Step 3), we can use algorithm of Lucchesi and Osborn [11] or algorithm of M.C. Fernandez [19] for instance.

Example 2.5.2

Let be given $S=<\Omega,F>$, where

$$\Omega = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8,$$

$F=\{13 \to 27,\ 2 \to 134,\ 8 \to 746,\ 1458 \to 236,\ 213 \to 4$

$36 \to 157\}$

Step 1 $L = 1234568$ ; $R = 1234567$ ; $R' = R \setminus L = 7$;

$G = \Omega \setminus R = 8$; $Z = (78)^+ = 78$.

Step 2 $\tilde{S} = <\tilde{\Omega},\tilde{F}>$, where $\tilde{\Omega} = \Omega \setminus Z = 123456$,

$\tilde{F} = \{13 \to 2,\ 2 \to 134,\ 213 \to 4,\ 145 \to 236,\ 36 \to 15\}$.

Step 3 Find all keys for $\tilde{S}$.

$\mathcal{L}_{\tilde{S}} = \{13,\ 2,\ 213,\ 145,\ 36\}$; $\bar{\mathcal{L}}_{\tilde{S}} = \{2,\ 13,\ 145,\ 36\}$

The graph $\mathcal{G}_{\tilde{S}}$ is shown in Fig. 2.15

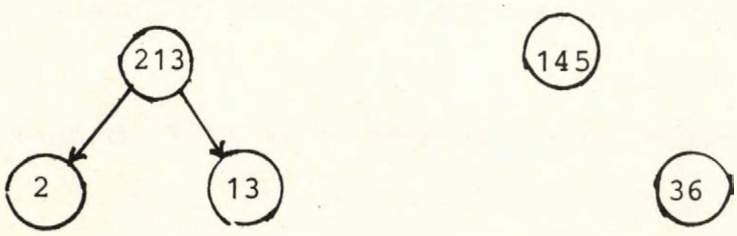



Fig. 2.15

We have:

$(36)^+_{\tilde F} = 123456 = \tilde\Omega \Rightarrow 36 \in K_{\tilde S}$ ; $(145)^+_{\tilde F} = \tilde\Omega \Rightarrow 145 \in K_{\tilde S}$

$(13)^+_{\tilde F} = 1234 \neq \tilde\Omega$; $(2)^+_{\tilde F} = 1234 \neq \tilde\Omega$.

Since $(13)^+_{\tilde F} = (2)^+_{\tilde F}$ , we have only to consider the b.r.s.

$\mathcal{S} = S - (13)^+_{\tilde F} = < 56, \{5 \to 6,\ 6 \to 5\}>$.

It is easily seen that $K_{\mathcal{S}} = \{5,6\}$

Now, consider sequentially the elements of the two following sets: $13 \oplus K_{\mathcal{S}} = \{135,\ 136\}$ and $2 \oplus K_{\mathcal{S}} = \{25, 26\}$.

Since $136 \supseteq 36$ being a key already found, so $136 \bar\in K_{\tilde S}$ .

Since each of 135, 25, 26 contains exactly one minimal left side of $\tilde S$, so they are keys for $\tilde S$.

Thus $K_{\tilde S} = \{36,\ 145,\ 135,\ 25,\ 26\}$

<u>Step 4</u> $K_S = G \oplus K_{\tilde S} = \{368,\ 1458,\ 1358,\ 258,\ 268\}$.

## § 2.6. Nontranslatable relation scheme

In this section we investigate some properties of the so-called nontranslatable relation scheme.

Definition 2.6.1

Let $S = \langle\Omega,F\rangle$ be a relation scheme. S is called translatable if there exists two subsets $Z, Z_1 \subseteq \Omega$ such that:

(i) $Z \neq \emptyset$, $Z_1 \subseteq Z$

(ii) X is a key for $\langle\tilde{\Omega},\tilde{F}\rangle$ iff $X \cap Z = \emptyset$ and $X \cup Z_1$ is a key for $\langle\Omega,F\rangle$, where $\langle\tilde{\Omega},\tilde{F}\rangle = \langle\Omega,F\rangle - Z$ .

Otherwise S is called nontranslatable.

Theorem 2.6.1

Let $S = \langle\Omega,F\rangle$ be a translatable relation scheme with Z and $Z_1$ defined as above (def. 2.6.1)

Then

$$H \setminus G = \tilde{H} \setminus \tilde{G}$$

where H and G (similarly for $\tilde{H}$ and $\tilde{G}$) are the union and intersection of all keys for S ($\tilde{S}$) respectively.

Proof

Let $\langle\tilde{\Omega},\tilde{F}\rangle = \langle\Omega,F\rangle - Z$ .

Since X is a key for $<\tilde{\Omega},\tilde{F}>$ iff $X\cap Z = \emptyset$ and $X\cup Z_1$ is a key for S, it follows that:

$$H=\tilde{H}\cup Z_1,\ Z_1\cap\tilde{\tilde{H}} = \emptyset$$

$$G=\tilde{G}\cup Z_1,\ Z_1\cap\tilde{\tilde{G}} = \emptyset.$$

Hence

$$H\setminus G = (\tilde{H}\cup Z_1)\setminus(\tilde{G}\cup Z_1) =$$

$$= ((\tilde{H}\cup Z_1)\setminus Z_1)\setminus\tilde{G} = \tilde{H}\setminus\tilde{G}.$$

(by virtue of $\tilde{H}\cap Z_1=\emptyset$).

Combining Theorems 2.1.1, 2.1.2 with Theorem 2.6.1, the following theorem is immediate.

Theorem 2.6.2.

Let $S = <\Omega,F>$ be a relation scheme.

$<\Omega,F>$ is nontranslatable iff $H=\Omega$ and $G=\emptyset$.

Theorem 2.6.3.

Let $S=<\Omega,F>$ be a relation scheme,

$\tilde{S}=<\tilde{\Omega},\tilde{F}> =<\Omega,F>-(G\cup\bar{H})$

where $\bar{H} =\Omega\setminus H$.

Then:

a) $<\Omega,F> \xRightarrow[\rho=(G\cup\bar{H},G)]{} <\tilde{\Omega},\tilde{F}>$;

b) $<\tilde{\Omega},\tilde{F}>$ is nontranslatable;

c) $<\tilde{\Omega},\tilde{F}>\in \mathcal{L}_4$ .

<u>Proof</u>

Let $Z = G \cup \bar{H} = Z_1 \cup Z_2$,

where $Z_1 = G$, $Z_2 = \bar{H}$ (clearly $Z_2 \cap H = \emptyset$).

Hence part a) of the theorem is obvious. To prove b) we have only to show that

$$\tilde{G} = \emptyset \quad \text{and} \quad \tilde{H} = \tilde{\Omega} .$$

From a) it is clear that X is a key for $\tilde{S}$ iff $X \cap G = \emptyset$ and $X \cup G$ is a key for S.

Therefore,

$$G = G \cup \tilde{G},\ G \cap \tilde{G} = \emptyset,$$

$$H = G \cup \tilde{H},\ G \cap \tilde{H} = \emptyset.$$

Hence

$$\tilde{G} = G \setminus G = \emptyset,$$

and

$$\tilde{H} = H \setminus G$$

On the otherhand, we have

$$\tilde{\Omega} = \Omega \setminus (G \cup \bar{H}) = (\Omega \setminus \bar{H}) \setminus G =$$

$$= H \setminus G = \tilde{H}.$$

To prove c) we have to show that

$$\tilde{L} = \tilde{R} = \tilde{\Omega},$$

where $\tilde{L}$ ($\tilde{R}$) is the union of all left (right) sides of all FDs in $\tilde{F}$ respectively.

It is known [see § 1.5] that

$$\tilde{\Omega} \setminus \tilde{R} = \tilde{G}.$$

Since $\tilde{G} = \emptyset$, we have $\tilde{R} = \tilde{\Omega}$. To complete the proof, it remains to show that

$$\tilde{L} = \tilde{\Omega}.$$

Were this false, there would exist an $A \in \tilde{\Omega} \setminus \tilde{L}$. Since $\tilde{R} = \tilde{\Omega}$, we have

$$A \in \tilde{R} \text{ and } A \notin \tilde{L}.$$

From $\tilde{\Omega} = \tilde{H}$, there exists a key X for $\tilde{S}$ such that $A \in X$.

Obviously $X \overset{*}{\to} \tilde{\Omega}$.

Since $A \notin \tilde{L}$, it follows from lemma 1.3.4, that

$$X \setminus A \overset{*}{\to} \tilde{\Omega} \setminus A.$$

From $A \notin \tilde{L}$, it follows that

$$\tilde{L} \subseteq \tilde{\Omega} \setminus A.$$

From this

$$X \setminus A \overset{*}{\to} \tilde{\Omega} \setminus A \overset{*}{\to} \tilde{L} \overset{*}{\to} \tilde{R} \overset{*}{\to} A.$$

This contradicts the fact that X is a key for $\tilde{S}$.

(see § 3.1, Lemma 1.3.5).

Hence $\tilde{L} = \tilde{\Omega}$.

The proof is complete.

From the proof of c) we conclude that all nontranslatable relation schemes are in $\mathcal{L}_4$.

<u>Theorem 2.6.4</u>

Let S $=\langle\Omega,F\rangle$ be a relation scheme in $\mathcal{L}_4$, satisfying the following conditions

(i) $L_i \cap R_i = \emptyset$, $i=1,2,\dots,m$;

(ii) For each $L_i$, $i=1,2,\dots,m$ there exists a key $X_i$ such that $L_i \subseteq X_i$.

Then $\langle\Omega,F\rangle$ is a nontranslatable relation scheme.

<u>Proof</u>

By virtue of Theorem 2.6.2, we have to prove that $H=\Omega$, and $G=\emptyset$. In fact, from $\langle\Omega,F\rangle \in \mathcal{L}_4$, we have $L=R=\Omega$.

From the hypothesis of the theorem, we get:

$$\Omega=L=\bigcup_{i=1}^{m} L_i \subseteq \bigcup_{i=1}^{m} X_i \subseteq H \subseteq \Omega.$$

Consequently $H=\Omega$.

On the other hand, from $G=\Omega\setminus R$, and $\Omega=R$, we have $G=\emptyset$.

The proof is complete.

# 3. STRUCTURE OF MINIMUM COVERS

## § 3.1 Introduction

In most studies concerning covers for functional dependencies (abbr. FD), we usually start from a set F of FDs over

$\Omega=\{A_1,A_2,\ldots,A_n\}$,

$F=\{L_i \rightarrow R_i \mid L_i,R_i \subseteq \Omega,\ i=1,2,\ldots,m\}$

and try to find a shorter representation for F, i.e. a new set F′ of FDs with either a fewer number of FDs or a less total size such that F and F′ imply the same set of FDs.

So doing, several algorithms concerning relational databases which start with a smaller cover will run faster.

The nonredundant and minimum covers have been investigated in depth by different approaches in [21], [22],[23], and several useful properties of them have been proved and used in various problems in the logical design of databases.

But few attention is paid to the study of invariants concerning the attribute sets of the left and right sides of these covers. Moreover, as pointed out by D. Maier [22], for minimum covers the problem is what sort of transformations can be found for right sides of FDs. This problem was not investigated.

In § 3.2 we define several kinds of minimality for covers and recall some basic results.

In § 3.3 we establish the relationship between the notions of direct determination and FD-graph. Some well known and new results as well concerning direct determination will be proved.

In § 3.4. we prove some additional invariants for covers and nonredundant covers.

Finally, in § 3.5 we study the structure for right sides of FDs in minimum covers. And basing upon these results, an algorithm for finding a "quasi optimum" cover (in the sense of economical memory requirement) is proposed.

## § 3.2 Basic definitions and results

As usual, we will only consider sets of FDs in natural reduced form (see § 1.3) and we assume that all attributes are chosen from some fixed universe $\Omega$.

### Definition 3.2.1

Two sets of FDs over $\Omega$

$$F_1 = \{L_i^{(1)} \rightarrow R_i^{(1)} \mid i=1,2,\ldots,m_1\}$$

and

$$F_2 = \{L_i^{(2)} \rightarrow R_i^{(2)} \mid i=1,2,\ldots,m_2\}$$

are said equivalent, written $F_1 \equiv F_2$, if $F_1^+ = F_2^+$ .
If $F_1 \equiv F_2$ then $F_i$ is a cover for $F_j$ with $i,j \in \{1,2\}$, $i \neq j$.

### Definition 3.2.2

A set F of FDs is nonredundant if there is no proper subset F′ of F with $F' \equiv F$.
If such F′ exists, F is redundant. $F_1$ is a nonredundant cover for $F_2$ if $F_1$ is a cover for $F_2$ and $F_1$ is nonredundant.

Let F be a set of FDs over $\Omega$ and let $X \rightarrow Y$ be a FD in F. Attribute A is said extraneous in $X \rightarrow Y$ if

$(F \setminus \{X \rightarrow Y\}) \cup \{X \setminus A \rightarrow Y \setminus A\})^{+} = F^{+}$.

Definition 3.2.3. [24]

Let F be a set of FDs over $\Omega$ and let $X \rightarrow Y$ be in F.

$X \rightarrow Y$ is left reduced if X contains no attribute A extraneous in $X \rightarrow Y$.

$X \rightarrow Y$ is right reduced if Y contains no attribute A extraneous in $X \rightarrow Y$.

$X \rightarrow Y$ is reduced if it is left-reduced and right reduced and $Y \neq \emptyset$.

A set F of FDs is left reduced (right reduced, reduced) if every FD in F is left reduced (respectively right-reduced, reduced).

Definition 3.2.4

Two sets of attributes X and Y are equivalent under a set of FDs F, written $X \leftrightarrow Y$, if $X \rightarrow Y$ and $Y \rightarrow X$ are in $F^{+}$.

Definition 3.2.5. [22]

Given a set of FDs F with $X \rightarrow Y$ in $F^{+}$.

X directly determines Y under F, written $X \overset{\rightarrow}{\rightarrow} Y$, if $X \rightarrow Y \in [F \setminus E_F(X)]^{+}$, where $E_F(X)$ is the set of all FDs in F with left sides equivalent to X.

That is, no FDs with left sides equivalent to X are used to derive $X \longrightarrow Y$.

<u>Lemma 3.2.1</u> [13]

Given sets of FDs $F_1$ and $F_2$ over $\Omega$.

$F_1 \equiv F_2$ iff $F_1 \subseteq F_2^+$ and $F_2 \subseteq F_1^+$.

Let $|T|$ denote the cardinality of a set T. Let $\bar{E}_F$ be the collection of all non empty $E_F$ (X)'s. (That is, X is equivalent to some left side of an FD in F).

<u>Lemma 3.2.2</u> [21]

If G and F are equivalent, nonredundant sets of FDs and there is an FD $X \rightarrow W$ in G, then there is an FD $Y \rightarrow Z$ in F with $X \leftrightarrow Y$ under F.

<u>Definition 3.2.6</u>

A set of FDs F is <u>minimum</u> if there is no set G with fewer FDs than F such that $G \equiv F$.

<u>Definition 3.2.7</u>

A set of FDs F is <u>optimal</u> if there is no set of FDs G with fewer attribute symbols such that $G \equiv F$. (Repeated symbols are counted as many times as they occur).

Theorem 3.2.1 [22]

Given equivalent minimum sets of FDs F and G.

$|E_F\ (X)| = |E_G(X)|$ for any X.

Thus the size of equivalence classes in $\bar{E}_F$ is the same for all minimum F with the same closure.

Remark 3.2.1 [22]

Let F and G both minimum, and look at $E_F(X)$ and $E_G(X)$

| $E_F(X)$ | $E_G(X)$ |
|---|---|
| $X_1 \rightarrow \bar{X}_1$ | $Y_1 \rightarrow \bar{Y}_1$ |
| $X_2 \rightarrow X_2$ | $Y_2 \rightarrow \bar{Y}_2$ |
| . | . |
| . | . |
| . | . |
| $X_p \rightarrow \bar{X}_p$ | $Y_p \rightarrow \bar{Y}_p$ . |

Then for every $X_i$ in $e_F(X)$ *) there is exactly one $Y_j$ in $e_G(X)$ such that $X_i \dot{\rightarrow} Y_j$ and $Y_j \dot{\rightarrow} X_i$. This relationship allows $X_i$ to be substituted for $Y_j$ without changing the closure of G and $Y_j$ for $X_i$ in F since one left side can still be derived from the other after the substitution.

Moreover we can arrange (number the Fds) such that

---

*) $e_F(x)$ is the set of left sides of $FD_s$ in $E_F(x)$.

the following relationship between $e_F(X)$ and $e_G(X)$ holds:

$$X_i \overset{\bullet}{\longleftrightarrow} Y_i \quad \forall i = 1,2,\ldots,p \ .$$

Thus, without loss of generality, in studying the structure of right sides of FDs in minimum covers, we can assume that $E_F(X)$ and $E_G(X)$ have the following form (i.e. $e_F(X) = e_G(X)$).

| $E_F(X)$ | $E_G(X)$ |
|---|---|
| $X_1 \rightarrow \bar{X}_1$ | $X_1 \rightarrow \bar{Y}_1$ |
| $X_2 \rightarrow \bar{X}_2$ | $X_2 \rightarrow \bar{Y}_2$ |
| $\vdots$ | $\vdots$ |
| $X_p \rightarrow \bar{X}_p$ | $X_p \rightarrow \bar{Y}_p$ |

where F and G are equivalent minimum covers.

<u>Theorem 3.2.2.</u> [25]

Let $F = \{X_i \rightarrow Y_i \mid i=1,2,\ldots,m\}$ be a set of FDs over $\Omega$, and $\mathfrak{F}$ be the set of all FDs $X \rightarrow Y$ such that there is a sequence of FDs in F

$$\{X_{i_j} \rightarrow Y_{i_j},\ j=1,2,\ldots,k,\ k\geq 0\}$$

with

$$X \supsetneq X_{i_1}$$

$$XY_{i_1} \supsetneq X_{i_2}$$

$$\cdots$$

$$XY_{i_1} Y_{i_2} \ldots Y_{i_k} \supsetneq Y.$$

Then $\mathcal{F}$ is the smallest full family of FDs that contains F, and each FD $X_{i_j} \to Y'_{i_j}$, $Y'_{i_j} \subsetneq Y_{i_j}$ is said to be used in the Armstrong's derivation sequence in F for $X \to Y$.

<u>Definition 3.2.8</u> [23]

Given a set of FDs F on $\Omega$, the FD-graph $G_F$=<V,E> associated with F is the graph with node labeling function $w: V \to P(\Omega)$ and arc labeling function $w': E \to \{0,1\}$ such that.

(i) for every attribute $A \in \Omega$, there is a node in V labeled A (called simple node);

(ii) for every dependency $X \to Y$ in F where $\|X\| \not> 1$, there is a node in V labeled X (called a compound node);

(iii) for every dependency $X \to Y$ in F where $Y = A_1 \ldots A_k$ there are arcs labeled o (full arcs) from the node labeled X to the nodes labeled $A_1, \ldots, A_k$;

(iv) for every compound node i in V labeled $A_1 \ldots A_k$

there are arcs labeled 1 /dotted arcs/ from the node i to all simple nodes (component nodes of i) labeled $A_1, \dots, A_k$.

The set of full arcs /dotted arcs/ is denoted $E_o$ ($E_1$).

<u>Definition 3.2.9</u> [23]

Given an FD-graph $G_F = \langle V,E \rangle$ and two nodes $i,j \in V$, a (directed) FD-path $\langle i,j \rangle$ from i to j is a <u>minimal subgraph</u> $\bar{G}_F = \langle \bar{V},\bar{E} \rangle$ of $G_F$ such that $i,j \in \bar{V}$ and either $(i,j) \in \bar{E}$ or one of the following possibilities holds:

(a) j is a simple node and there exists a node k such that $(k,j) \in \bar{E}$ and there is an FD-path $\langle i,k \rangle$ included in $\bar{G}_F$ (graph transitivity).

(b) j is a compound node with component nodes $m_1, \dots, m_r$ and there are dotted arcs $(j,m_1) \dots, (j,m_r)$ in $\bar{G}_F$ and r FD-paths $\langle i,m_1 \rangle, \dots, \langle i,m_r \rangle$ included in $\bar{G}_F$ (graph union).

<u>Definition 3.2.10.</u> [23]

The closure of an FD-graph $G_F = \langle V,E \rangle$ is the graph $G_F^+ = \langle V,E^+ \rangle$, labeled on the nodes and on the arcs, where the set V is the same as in $G_F$, while the set $E^+ = (E^+)_o \cup (E^+)_1$ is defined in the following way:

$(E^+)_1 = \{(i,j) \mid i,j \in V$ and there exists a dotted FD-path $<i,j>\}$;

$(E^+)_0 = \{(i,j) \mid i,j \in V,\ (i,j) \notin (E^+)_1$ and there exists a full FD-path $<i,j>\}$.

Definition 3.2.11 [23]

Two nodes i,j in an FD-graph are said equivalent if the arcs (i,j) and (j,i) both belong to the closure of $G_F$. Furthermore a node i of $G_F$ is said to be equivalent to a node j of $G_{\bar{F}}$ where $G_{\bar{F}}$ is a cover of $G_F$ (i.e. $F^+ = \bar{F}^+$) if i,j are equivalent in some cover of $G_F$.

Theorem 3.2.2 [23]

Let $G_F = <V,E>$ be the FD-graph associated with the set F of FDs, and let $G_F^+ = <V,E^+>$ be its closure. An arc (i,j) is in $E^+$ if and only if $w(i) \rightarrow w(j)$ is in $F^+$.

Theorem 3.2.4 [23]

A nonredundant FD-graph $G_F = <V,E>$ is minimum if and only if it has no superfluous nodes.

Recall that a node $i \in V$ is superfluous if there exists a dotted FD- path $<i,j>$ where j is a node of V equivalent to i.

## § 3.3 Direct determination and FD-graph

As shown in § 3.2, the notion of direct determination was introduced by D. Maier [22] to study the structure of minimum covers. Using direct determination he showed it is possible to find covers with the smallest number of FDs in polynomial time.

In [23], G. Ausiello et al. presented an approach which is based on the representation of the set of FDs by FD-graph (a generalization of graphs). Such a representation provides a unified frame-work for the treatment of various properties and for the manipulation of FDs. However, the notion of direct determination in its relationship with FD-graph is not explicitly presented. In this section, we establish the relation between FD-graph and direct determination, and prove some well-known and new properties concerning direct determination. First it is worth giving a few comments on the definition of an FD-graph (Definition 3.2.8).

### Remark 3.3.1

The Definition 3.2.8 is reasonable and concise in the sense that the FD-graph $G_{\bar{F}}$ includes all the "meaningful parts" of the closure of the set F of FDs.

On the otherhand, with the FD-graph, we can provide a simple and unified treatment of all properties of sets of FDs.

Following the definition of a FD-graph, it is clear that <u>every compound node has at least one outgoing full arc</u>.

However, in [23,p.755] we found the following observation:

"Finally we may observe that by definition of FD-path, a compound node without outgoing full arcs can only be either a source or a target node of FD-paths to which it belongs"!

Part (ii) of Lemma 1 [23,p.757] touches the same problem. Let us see it:

"(ii) If $G_{\bar{\Sigma}}$ be a subgraph of $G_{\Sigma}$ such that all arcs in $E-\bar{E}$ are dotted (i.e., $G_{\Sigma}$ may contain compound nodes not in $G_{\bar{\Sigma}}$ but no more full arcs) and $(i,j)$ is in $(E^{+})_{0}$ $[(E^{+})_{1}]$, then $(i,j)$ is in $(\bar{E}^{+})_{0}$ $[(\bar{E}^{+})_{1}]$". (where $i,j$ are two nodes belonging to both $V$ and $\bar{V}$).

It is obvious that, strictly following the Definition 3.2.8 there is no possibility that $G_{\Sigma}$ may contain compound nodes not in $G_{\bar{\Sigma}}$ but no more full arcs. And it is easy to show that under these conditions the subgraph $G_{\Sigma}$ coincides with $G_{\Sigma}$. In that case, part (ii) of Lemma 1 is trivial.

How to overcome these difficulties? A natural way is to think that a FD-graph $G_F$=<V,E> associated with F is defined by Definition 3.2.8 precisely to an arbitrary finite number of different compound nodes which do not correspond to the left side of any FD in F, together with the dotted arcs from each of them to its corresponding component nodes.

In our opinion, the view just presented above must be mentioned explicitly after introducing the definition of the FD-graph.

In so doing, according to the necessity, we can freely add to an FD-graph some new compound nodes without outgoing full arcs if it makes easy to prove a certain required property.

In fact, this technique was often used by the authors of [23].

By the above reasons, it would be better to remove part (ii) from Lemma 1 in [23], changing it into a remark.

Definition 3.3.1

Given an FD-graph $G_F$=<V,E> and a node $i \in V$ with at least a full outgoing arc. A strong component of $G_F$ with representative node i is a maximal set of pairwise equivalent nodes which contains i, denoted by SC (i).

Notice that every node in SC(i) has at least one full outgoing arc.

The following lemma is obvious.

<u>Lemma 3.3.1</u>

Given an FD-graph $G_F$=<V,E>, a node $i \in V$, its corresponding strong component SC(i) and two nodes j, k such that j is equivalent to i. (j not necessarily belongs to SC(i), i.e. j can be a compound node without outgoing full arc that we add it to the FD-graph. The same situation can happen with the node k too).

Then $w(j) \rightarrow w(k)$ if and only if there exists a dotted FD-path <j,k> containing no full outgoing arc from any node of SC(i). In other words, the dotted FD-path $\langle j,k \rangle$ contains no intermediate nodes that are nodes in SC(i). In that case, for sake of simplicity, we write $\langle j \xrightarrow{SC(i)} k \rangle$

<u>Example 3.3.1</u>

Given $\Omega$ = ABCDEIH

$F = \{A \to BCH,\ BC \to A,\ AD \to EI,\ EA \to ID\}$

It is easy to verify that

$$E_F\ (AD) = \{AD \to EI,\ AE \to DI\}$$

and

$$BCD \longleftrightarrow AD$$

The corresponding FD-graph $G_F$ with an added node BCD (without outgoing full arc) is shown in Fig. 3.1

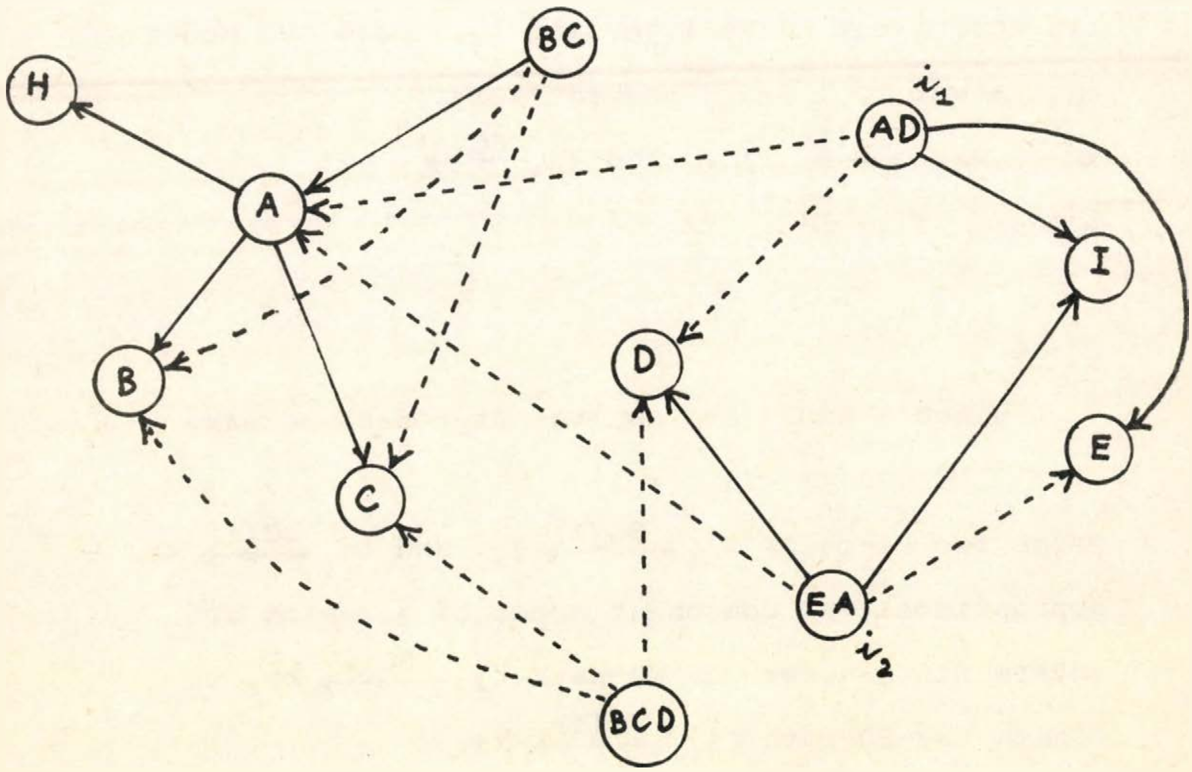



Fig 3.1

$SC(i_1) = \{i_1, i_2\}$ where $w(i_1) = AD$

$w(i_2) = EA$.

We find that

$$BCD \rightarrow H$$

$$BCD \rightarrow AD$$

<u>Lemma 3.3.2</u>

Given an FD-graph $G_F = \langle V, E\rangle$, two equivalent nodes $i, j \in V$ and $i_1, j_1$ are two nodes equivalent to $i$ and $j$ respectively.

If $\langle i_1 \xrightarrow{SC(i)} j_1\rangle$ and $\langle j_1 \xrightarrow{SC(j)} k\rangle$

then $\langle i_1 \xrightarrow{SC(i)} k\rangle$.

<u>Proof</u>

Since $i$ and $j$ are equivalent nodes, we have

$$SC(i) = SC(j) \ .$$

Merge two FD-paths $\langle i_1 \xrightarrow{SC(i)} j_1\rangle$ and $\langle j_1 \xrightarrow{SC(i)} k\rangle$ appropriately at component nodes of $j_1$ which are intermediate nodes of FD-path $\langle j_1 \xrightarrow{SC(i)} k\rangle$, we obtain the FD-path $\langle i_1 \xrightarrow{SC(i)} k\rangle$.

In other words, from

$w(i) \longleftrightarrow w(i_1) \longleftrightarrow w(j_1)$, $w(i_1) \rightarrow w(j_1)$ and $w(j_1) \rightarrow w(k)$, we have $w(i_1) \rightarrow w(k)$. Notice that the above lemma corresponds to [22, Lemma 5].

Example 3.3.2

Take up again the Example 3.3.1 /Fig 3.1/ we have $BCD \to AD$,

and $AD \to H$.

Since A is the unique component node of AD that is an intermediate node on the FD-path $\langle AD \xrightarrow{SC(i_1)} H\rangle$, we will merge two FD-paths $\langle BCD,AD\rangle$ and $\langle AD,H\rangle$ at A to obtain the FD-path $\langle BCD,H\rangle$ such that $BCD \to H$.

Lemma 3.3.3

Given an FD-graph $G_F = \langle V,E\rangle$, $i \in V$ is a node having at least one outgoing full arc* and $i_o$ is equivalent to i ($i_o$ can be an added node to the FD-graph without outgoing full arc). Then there exists $j \in SC(i)$ such that $\langle i_o \xrightarrow{SC(i)} j\rangle$.

Proof

Suppose that $i_o \notin SC(i)$. Otherwise, take $j = i_o$ and the lemma is proved. Consider the dotted FD-path $\langle i_o,i\rangle$. In the case, there is no intermediate node in $\langle i_o,i\rangle$ that is node of SC(i) then i is the node to be found. Otherwise, suppose $i_1 \in SC(i)$ is an intermediate node of $\langle i_o,i\rangle$. Now we have only to consider the FD-path

---

* i.e. corresponds to some left side of a FD in F.

$\langle i_o, i_1\rangle$. Repeat the above reasoning for $\langle i_o, i_1\rangle$. Finally, we will find the required j such that $\langle i_1 \xrightarrow{SC(i)} j\rangle$. Q.E.D.

Notice that the above lemma corresponds to [22 , Lemma 6].

Lemma 3.3.4

Let $G_F = \langle V,E\rangle$ be a minimum FD-graph (i.e. F is minimum), and $i \in V$ is a node with at least one outgoing arc. Then in SC(i) there exists no $j_1, j_2$, $j_1 \neq j_2$ such that $\langle j_1 \xrightarrow{SC(i)} j_2\rangle$.

Proof

Assume the contrary that there exists $j_1, j_2 \in SC(i)$, $j_1 \neq j_2$ such that there is a dotted FD-path from $j_1$ to $j_2$. Since $j_1$ is equivalent to $j_2$, $j_1$ is a superfluous node. We arrive to a contradiction. (See Theorem 3.2.4)

Notice that the above lemma corresponds to [22, Lemma 7].

Definition 3.3.2

An FD-graph $G_F$ is nonredundant if F is non-redundant.

Given two FD-graphs $G_{F_1}$ and $G_{F_2}$, $G_{F_2}$ is a cover of $G_{F_1}$ if $F_2$ is a cover of $F_1$.

Lemma 3.3.5

Given two nonredundant FD-graphs $G_{F_1}$ and $G_{F_2}$, where $G_{F_2}$ is a cover of $G_{F_1}$.

$G_{F_1} = \langle V_1, E_1 \rangle$, $G_{F_2} = \langle V_2, E_2 \rangle$

Let $i_1$ and $i_2$ be two equivalent nodes in $V_1$ and $V_2$ respectively with at least one outgoing full arc, $(p_2, q_2)$ be a full arc of $E_2$ with $p_2 \bar{\in} SC^{(2)}(i_2)$.*)

If $(i_1, p_2) \in E_2^+$ then $\langle p_2 \xrightarrow[\not\;]{SC^{(1)}(i_1)} q_2 \rangle$.

Proof

Since $(i_1, p_2) \in E_2^+$, by Theorem 3.2.3, there is a FD-path in $G_{F_1}$ from $i_1$ to $p_2$.

Now assume the contrary that the FD-path in $G_{F_1}$ from $p_2$ to $q_2$ has an intermediate node $j_1 \in SC^{(1)}(i_1)$.

The presence of the FD-path $\langle j_1, i_1 \rangle$ shows that $p_2$ is equivalent to $i_1$, i.e. $p_2 \in SC^{(2)}(i_2)$, a contradiction.

Q.E.D.

---

*) $SC^{(1)}$ and $SC^{(2)}$ refer to $G_{F_1}$ and $G_{F_2}$ respectively.

Theorem 3.3.1

With the same assumptions as in Lemma 3.3.5, if we replace in $G_{F_1}$ all nodes belonging to $SC^{(1)}(i_1)$ together with their corresponding outgoing arcs by all nodes in $SC^{(2)}(i_2)$ together with their corresponding outgoing arcs, then the new FD-graph is a cover of $G_{F_1}$.

Proof

We have only to prove that for every full arc $(j_1,k_1) \in E_1$ with $j_1 \in SC^{(1)}(i_1)$ there is a FD-path $\langle j_1,k_1 \rangle$ in the new FD-graph.
By the Lemma 3.3.5 we have just the required result.

Remark 3.3.2

Theorem 3.3.1 can be formulated in an another form as follows:
If $F_1,F_2$ are nonredundant and equivalent sets of FDs, then

$$F_1 \equiv \{F_1 \setminus E_{F_1}(X)\} \cup E_{F_2}(X) \equiv \{F_2 \setminus E_{F_2}(X)\} \cup E_{F_1}(X).$$

We close this section with the following useful lemma:

Lemma 3.3.6

Let $V \to W$ be an FD in $F^+$ and let $X \to Y$ be an FD in F that participates in the Armstrong's derivation sequence for $V \to W$.

Then we have

$V \to X,\ VY \to W \in (F \setminus \{X \to Y\})^+$.

<u>Proof</u>

Let $G_F = \langle V,E \rangle$ be the FD-graph associated with F. From $V \to W$ in $F^+$ it follows that there is an FD-path $\langle i,j \rangle$ from i to j, where $w(i)=V$, $w(j)=W$.

Since $X \to Y \in F$ takes part in the derivation sequence for $V \to W$, the nodes p and q with $w(p)= X$ and $w(q)=Y$ are intermediate nodes on $\langle i,j \rangle$. It is clear that the FD-paths $\langle i,p \rangle$ and $\langle q,j \rangle$ contains no outgoing full arcs from node p.

Q.E.D.

<u>Example 3.3.3</u>

Reconsider the Example 3.3.1 (Fig. 3.1) We have

$$BCD \to H \in F^+ ,$$

$(BC \to A) \in F$ participates in the derivation sequence for $BCD \to H$.

It is clear that:

$BCD \to BC \in (F \setminus \{BC \to A\})^+$ and corresponds to the FD-path $\langle BCD , BC \rangle$;

$BCDA \to H \in (F \setminus \{BC \to A\})^+$ and corresponds to the FD-path $\langle BCDA, H \rangle$.

## § 3.4. Some additional invariants of covers for functional dependencies

Let F be a set of FDs on $\Omega$.

Let us denote by

$$AF = \{L_i \longrightarrow R_i \mid (L_i \longrightarrow R_i) \in F \text{ and } |L_i| = 1\}$$

the set of all FDs in F with left side consists of only one attribute, and by

$$\mathcal{L}(AF) = \{A \in L_i \mid (L_i \to R_i) \in AF\} \subseteq \Omega$$

We have the following lemma:

Lemma 3.4.1

Let $F_1$ and $F_2$ be two equivalent sets of FDs on $\Omega$.

$$F_1 = \{L_i^{(1)} \to R_i^{(1)} \mid i = \overline{1,k_1}\},$$

$$F_2 = \{L_i^{(2)} \to R_i^{(2)} \mid i = \overline{1,k_2}\}.$$

Then

$$\mathcal{L}(AF_1) = \mathcal{L}(AF_2).$$

Proof

The proof is by contradiction.

Without loss of generality, suppose that there exists

$$L_j^{(1)} \equiv A_{i_j} \in \mathcal{L}(AF_1) \setminus \mathcal{L}(AF_2).$$

It is easy to show that

$$(L_j^{(1)} \rightarrow R_j^{(1)}) \notin F_2^+ .$$

In fact, it is obvious that

$$(L_j^{(1)})_{F_2}^+ = L_j^{(1)} .$$

On the other hand, we have

$$L_j^{(1)} \cap R_j^{(1)} = \emptyset,$$

($F_1, F_2$ are in natural reduced form)

Hence $R_j^{(1)} \not\subseteq (L_j^{(1)})_{F_2}^+$ ,

Showing that

$$L_j^{(1)} \rightarrow R_j^{(1)} \notin F_2^+ ,$$

a contradiction. The lemma is proved.

<u>Example 3.4.1</u>

Let be given

$$\Omega = ABCDE$$

$$F_1 = \{A \rightarrow BC,\ AD \rightarrow CE\},$$

$$F_2 = \{A \rightarrow B,\ B \rightarrow C,\ AD \rightarrow CE\}.$$

We have

$$\mathfrak{L}(AF_1) = \{A\},$$

$$\mathfrak{L}(AF_2) = \{A,B\} \neq \mathfrak{L}(AF_1).$$

Hence $F_1 \not\equiv F_2$.

Lemma 3.4.2

Let be given two equivalent sets of FDs on $\Omega$.

$F_1 = \{L_i^{(1)} \to R_i^{(1)} \mid i=\overline{1,k_1}\}$,

$F_2 = \{L_i^{(2)} \to R_i^{(2)} \mid i=\overline{1,k_2}\}$.

Then

$$R(F_1) = R(F_2)$$

where $$R(F_j) = \bigcup_{i=1}^{k_j} R_i^{(j)} \;, \; j=1,2 \;.$$

Proof

We first show that $R(F_1) \subseteq R(F_2)$.

Let $A \in R(F_1)$.

It follows that there exists

$$L_i^{(1)} \to R_i^{(1)} \text{ with } R_i^{(1)} = AX \;.$$

Since $F_1 \equiv F_2$, we have

$$(L_i^{(1)} \to AX) \in F_2^+$$

or, equivalently

$$AX \subseteq (L_i^{(1)})^+_{F_2} = L_i^{(1)} \cup \Big(\bigcup_{j\in\alpha} R_j^{(2)}\Big) \;.$$
$$\alpha \subseteq \{1,2,\ldots,k_2\}$$

On the other hand

$$AX \cap L_i^{(1)} = \emptyset,$$

showing that $A \in R(F_2)$.

Similarly, we can show that $R(F_2) \subseteq R(F_1)$.

Hence $R(F_1) = R(F_2)$.

Example 3.4.2

Let be given

$$\Omega = ABCDE$$

$$F_1 = \{A \to BC,\ AD \to CE\}$$

$$F_2 = \{A \to BD,\ AD \to CE\}$$

We have

$$R\ (F_1) = BCE \neq BCDE = R\ (F_2).$$

Hence $F_1 \not\equiv F_2$.

Remark 3.4.1

Lemma 3.4.1 is equivalent to the assertion that for a given FD-graph $G_F = \langle V,E \rangle$ associated with the set of FDs F, all covers of $G_F$ have the same set of simple nodes without outgoing arcs [23].

Theorem 3.4.1

Let $F_1$ and $F_2$ be two equivalent and nonredundant sets of FDs on $\Omega$,

$$F_j = \{L_i^{(j)} \to R_i^{(j)} \mid i = \overline{1,k_j}\},\ j=1,2\ .$$

Then

$$L(F_1) \setminus R(F_1) = L(F_2) \setminus R(F_2).$$

Where

$$L(F_j) = \bigcup_{i=1}^{k_j} L_i^{(j)}\ ,\ j=1,2$$

$$R(F_j) = \bigcup_{i=1}^{k_j} R_i^{(j)}\ ,\ j=1,2.$$

Proof

First we prove that

$$L(F_1)\setminus L(F_2) \subseteq R(F_1).$$

Let $A \in L(F_1) \setminus L(F_2)$, i.e.

$A \in L(F_1)$ and $A \bar{\in} L(F_2)$.

Then there exists

$$(L_i^{(1)} \rightarrow R_i^{(1)}) \in F_1$$

with $L_i^{(1)} = AX,\ X \neq \emptyset$.

(This follows from $A \bar{\in} L(F_2)$ and Lemma 3.3.1).

Since $F_1$ is nonredundant, it follows that $L_i^{(1)} \rightarrow R_i^{(1)}$ must participate in some derivation sequence for some $(L_h^{(2)} \rightarrow R_h^{(2)}) \in F_2$ . (see Theorem 3.2.2).

So we have

$$L_h^{(2)} \supseteq L_{i_1}^{(1)}$$

$$L_h^{(2)} R_{i_1}^{(1)} \supseteq L_{i_2}^{(1)}$$

$$\cdots \qquad \cdots$$

$$L_h^{(2)}\ R_{i_1}^{(1)}\ R_{i_2}^{(1)}\ \ldots\ R_{i_t}^{(1)} \supseteq L_i^{(1)} = AX,$$

$$L_h^{(2)}\ R_{i_1}^{(1)}\ R_{i_2}^{(1)}\ \ldots\ R_{i_t}^{(1)} R_i^{(1)} \supseteq L_{i_{t+2}}^{(1)}$$

$$\cdots \qquad \cdots$$

$$L_h^{(2)}\ R_{i_1}^{(1)}\ \ldots\ R_{i_t}^{(1)}\ R_i^{(1)} R_{i_{t+2}}^{(1)} \ldots R_{i_s}^{(1)} \supseteq R_h^{(2)}.$$

so

$$A \in L_h^{(2)}\ R_{i_1}^{(1)}\ \ldots\ R_{i_t}^{(1)} \quad .$$

Since $A \notin L(F_2)$, it is obvious that $A \in R(F_1)$.

Thus we have proved that

$$L(F_1) \setminus L(F_2) \subseteq R(F_1).$$

Similarly, we can prove that

$$L(F_2) \setminus L(F_1) \subseteq R(F_2).$$

On the other hand, by Lemma 3.4.2,

$$R(F_2) = R(F_1).$$

Consequently,

$$\begin{aligned} L(F_1)\setminus R(F_1) &= \{[L(F_1)\setminus L(F_2)]\setminus R(F_1)\} \\ &\quad \cup \{[L(F_1)\cap L(F_2)]\setminus R(F_1)\} = \\ &= [L(F_1)\cap L(F_2)]\setminus R(F_1) = \\ &= [L(F_1)\cap L(F_2)]\setminus R(F_2) = L(F_2)\setminus R(F_2). \end{aligned}$$

The theorem is completely proved.

Example 3.4.3

Take up again the Example 3.4.2

$$\Omega = ABCDE$$

$$F_1 = \{A \to BC,\ AD \to CE\}$$

$$F_2 = \{A \to BD,\ AD \to CE\}$$

We have

$$L(F_1)\setminus R(F_1) = AD \neq A = L(F_2)\setminus R(F_2) .$$

Hence $F_1 \not\equiv F_2$.

<u>Theorem 3.4.2</u>

Let $F_1$ and $F_2$ be two equivalent and nonredundant sets of FDs over $\Omega$,

$$F_j = \{L_i^{(j)} \rightarrow R_i^{(j)} \mid i=\overline{1,k_j}\},\ j=1,2.$$

Then

$$L(F_1) \cup R(F_1) = L(F_2) \cup R(F_2).$$

<u>Proof</u>

We first prove that

$$L(F_1) \cup R(F_1) \subseteq L(F_2) \cup R(F_2).$$

By Lemma 3.4.2, we have

$$R(F_1) = R(F_2) \subseteq L(F_2) \cup R(F_2).$$

We have to prove

$$L(F_1) \subseteq L(F_2) \cup R(F_2).$$

Following the proof of Theorem 3.4.1 we have

$$L(F_1) \setminus L(F_2) \subseteq R(F_1).$$

But $\quad R(F_1) = R(F_2).$

Therefore

$$L(F_1) \setminus L(F_2) \subseteq R(F_2)$$

Hence

$$L(F_1) \subseteq L(F_2) \cup R(F_2).$$

Thus we have proved

$$L(F_1) \cup R(F_1) \subseteq L(F_2) \cup R(F_2).$$

Similarly, we can prove that

$$L(F_2) \cup R(F_2) \subseteq L(F_1) \cup R(F_1).$$

Combining these two results, we get

$$L(F_1) \cup R(F_1) = L(F_2) \cup R(F_2).$$

Q.E.D.

Theorem 3.4.3

Let $F_1$ and $F_2$ be any two equivalent, nonredundant and left reduced sets of FDs on $\Omega$

$$F_j = \{L_i^{(j)} \to R_i^{(j)} \mid i = \overline{1,k_j}\}, \quad j = 1,2 .$$

Then $L(F_1) = L(F_2)$.

Proof

Assume the contrary that

$$L(F_1) \neq L(F_2).$$

Without loss of generality, let

$$A \in L(F_1) \text{ and } A \bar{\in} L(F_2).$$

It follows that there exists

$L_i^{(1)} \to R_i^{(1)}$ with $L_i^{(1)} = AX$, $X \neq \emptyset$.

Since $A \bar{\in} L(F_2)$, by Lemma 1.3.4 (see §1.3), and from

$$AX \to R_i^{(1)} \in F_2^+ ,$$

we have

$$X \to R_i^{(1)} \in F_1^+ \quad (\text{since } F_2^+ = F_1^+).$$

This means that in $F_1$ we can replace $AX \to R_i^{(1)}$ by $X \to R_i^{(1)}$ without altering $F_1^+$.
We arrive to a contradiction because, by the hypothesis of Theorem 3.4.3, $F_1$ is left reduced.
Thus we have $L(F_1)=L(F_2)$. Q.E.D.

Remark 3.4.2

(i) Basing on the results of this section we can conclude that, after removing all extraneous attributes, the sets $L(F)$ and $R(F)$ are the same for all equivalent sets of FDs on $\Omega$.

(ii) The invariants just have been established can be used, for instance, as a simple criterion to check whether two sets of FDs are not equivalent.

## § 3.5 Structure of minimum covers

In [22] the notion of equivalent classes of left sides $E_F(X)$ has been introduced by D. Maier, and it is shown that for any equivalent minimum sets of FDs $F_1$ and $F_2$, $|E_{F_1}(X)| = |E_{F_2}(X)|$ for any X. (see Theorem 3.2.1 and Remark 3.2.1). D. Maier also proved that for each FD $X_i \to \overline{X}_i \in E_{F_1}(X)$ there exists a unique $Y_i \to \overline{Y}_i \in E_{F_2}(X)$ such that $X_i \leftrightarrow Y_i$. Therefore $Y_i$ (resp $X_i$) can be substituted for $X_i$ (resp $Y_i$) without changing the closure of $F_1$ (resp. $F_2$), i.e.

$$[\{F_1 \setminus (X_i \longrightarrow \overline{X}_i)\} \cup (Y_i \longrightarrow \overline{X}_i)] \equiv F_1$$

and

$$[\{F_2 \setminus (Y_i \longrightarrow \overline{Y}_i)\} \cup (X_i \longrightarrow \overline{Y}_i)] \equiv F_2.$$

So, the structure of left sides of FDs in minimum covers has been described quite well. In this section we investigate the structure of right sides of FDs in equivalent minimum covers, and try to find a certain sort of transformations for right sides. In studying the structure of right sides of FDs in minimum covers, by the results of D. Maier just mentioned above, we can assume that all equivalent minimum covers have the same set of left sides.

Denoted by $LE_F(X)$ and $RE_F(X)$ the sets of attributes in left and right sides of FDs in $E_F(X)$ respectively and instead of $[F\setminus\{X \to \bar{X}\}]\cup\{X \to \bar{X}\setminus Z_o\}$, sometimes for sake of simplicity, we write $F\setminus\{X \to Z_o\}$ if $Z_o \subseteq \bar{X}$ and $X \to \bar{X} \in F$.

We begin with the following fundamental theorem.

<u>Theorem 3.5.1</u>

Let $F_1$ and $F_2$ be two equivalent minimum sets of FDs, $X_1 \to \bar{X}_1 \in E_{F_1}(X)$ ,

$$Z_o \subseteq \bar{X}_1 \text{ and } Z_o \cap RE_{F_2}(X) = \emptyset.$$

Then there exists Z such that

$$X_1Z \leftrightarrow X_1Z_o \in [F_1\setminus \{X_1 \to Z_o\}]^+ .$$

$Z_o$ and Z are said to be equivalent <u>via</u> $X_1$

<u>Proof</u>

Since $X_1 \to \bar{X}_1 \in E_{F_1}(X)$, there exists

$$X_1 \to \bar{Y}_1 \in E_{F_2}(X).$$

First of all, we show that $X_1 \to \bar{X}_1$ must participate in the Armstrong's derivation sequence for $X_1 \to \bar{Y}_1$ and vice versa. Assume the contrary that $X_1 \to \bar{X}_1$ does not participate in the derivation sequence for $X_1 \to \bar{Y}_1$. So

$$X_1 \to \bar{Y}_1 \in [F_1 \setminus X_1 \to \bar{X}_1]^+$$

Since $F_1$ and $F_2$ are nonredundant, $X_1 \to \bar{Y}_1$ cannot be derived from $F_2 \setminus E_{F_2}(X)$. On the other hand, by Theorem 3.3.1,

$$F_2 \equiv F_1 \equiv \{F_1 \setminus E_{F_1}(X)\} \cup E_{F_2}(X) \equiv \{F_2 \setminus E_{F_2}(X)\} \cup E_{F_1}(X)$$

it follows that there exists $X_k \to \bar{X}_k \in E_{F_1}(X)$ that participates in the derivation sequence for $X_1 \to \bar{Y}_1$.

By Lemma 3.3.6 we have

$$X_1 \to X_k \in [F_1 \setminus \{X_1 \longrightarrow \bar{X}_1,\ X_k \longrightarrow \bar{X}_k\}]^+.$$

But, in that case, we have

$$(F_1 \setminus \{X_1 \to \bar{X}_1,\ X_k \to \bar{X}_k\}) \cup \{X_k \to \bar{X}_1 \bar{X}_k\} \equiv F_1$$

which contradicts the fact that $F_1$ is minimum.
Thus $X_1 \to \bar{X}_1$ must participate in the Armstrong's sequence for $X_1 \to \bar{Y}_1$, and in turn, $X_1 \to \bar{Y}_1$ must participate in the Armstrong's derivation sequence for $X_1 \to \bar{X}_1$. By Lemma 3.3.6, we have:

$$X_1 \bar{X}_1 \to \bar{Y}_1 \in \{F_1 \setminus (X_1 \to \bar{X}_1)\}^+ \quad \text{and}$$

$$X_1 \bar{Y}_1 \to \bar{X}_1 \in \{F_2 \setminus (X_1 \longrightarrow \bar{Y}_1)\}^+.$$

Now we can split $X_1 \to \bar{X}_1$ into $X_1 \longrightarrow Z_o$ and $X_1 \to \bar{X}_1 \setminus Z_o$. If $X_1 \to Z_o$ does not participate in the Armstrong's derivation sequence for $X_1 \bar{Y}_1 \to \bar{X}_1$ then $\bar{Y}_1$ is the required Z. Indeed, in that case we have

$$X_1 \bar{Y}_1 \to Z_o \in [\{F_1 \setminus (X_1 \longrightarrow \bar{X}_1)\} \cup \{X_1 \to (\bar{X}_1 \setminus Z_o)\}]^+$$

Moreover,

$X_1\bar{X}_1 \to Y_1 \in \{F_1 \setminus (X_1 \to \bar{X}_1)\}^+ \subseteq [\{F_1 \setminus (X_1 \to \bar{X}_1)\} \cup \{X_1 \to (x_1 / Z_o)\}]^+$,
$X_1 \to (\bar{X}_1 \setminus Z_o) \in [\{F_1 \setminus (X_1 \to \bar{X}_1)\} \cup \{X_1 \to \bar{X}_1 \setminus Z_o\}]$ and
$X_1 Z_o \to X_1 \bar{X}_1$ can be derived from $X_1 \to \bar{X}_1 \setminus Z_o$, so
using the transitivity rule, we get:

$$X_1 Z_o \to \bar{Y}_1 \in [\{F_1 \setminus (X_1 \to \bar{X}_1)\} \cup \{X_1 \to \bar{X}_1 \setminus Z_o\}]^+$$

Now consider the case where $X_1 \to Z_o$ participates in the derivation sequence in $F_1$ for $X_1\bar{Y}_1 \to \bar{X}_1$. Since $X_1\bar{Y}_1 \to \bar{X}_1 \in F_2^+$, it can be derived from FDs in $F_2$, which in turn can be derived from FDs in $F_1$.

So there exists at least an FD $X_2 \to \bar{Y}_2$ in $F_2$ such that $X_2 \to \bar{Y}_2$ participates in the derivation sequence in $F_2 \setminus (X_1 \to \bar{Y}_1)$ for $X_1\bar{Y}_1 \to \bar{X}_1$, and $X_1 \to Z_o$ will participate in the Armstrong's derivation sequence in $F_1$ for $X_2 \to \bar{Y}_2$.

By virtue of Lemma 3.3.6, we have:

$$X_1\bar{Y}_1 \to X_2 \quad \text{and } X_2 \to X_1 \in F_1^+.$$

And from $X_1 \to \bar{Y}_1 \in F_2 \subseteq F_1^+$, we conclude that: $X_2 \longleftrightarrow X_1$.
So $(X_2 \to \bar{Y}_2) \in E_{F_2}(X)$.

By the hypothesis of the theorem,

$$Z_o \cap RE_{F_2}(X) = \emptyset.$$

It follows that

$$X_2 \to \bar{Y}_2 \neq X_1\bar{Y}_1 \to \bar{X}_1.$$

Moreover, by Lemma 3.3.6 , we have

$$X_1\bar{Y}_1\bar{Y}_2 \to Z_o \in [F_2 \setminus \{X_1 \to \bar{Y}_1, X_2 \to \bar{Y}_2\}]^+$$

Two cases can be happen:

If $X_1 \to Z_o$ does not participate in the derivation sequence for $X_1\overline{Y}_1\overline{Y}_2 \to Z_o$ in $F_1$, we prove that $\overline{Y}_1\overline{Y}_2$ is the Z to be found.

Indeed, we have $X_1\overline{Y}_1\overline{Y}_2 \to Z_o \in [F_1 \setminus (X_1 \longrightarrow Z_o)]$.

Moreover, using an argument similar to the one given at the beginning of the proof, we can prove that $X_1 \longrightarrow \overline{X}_1$ must participate in the Armstrong's sequence for $X_1 \longrightarrow \overline{Y}_j$, $j=1,2,\ldots, p$. So, from $X_1Z_o \longrightarrow \overline{Y}_i \in [F_1 \setminus (X_1 \longrightarrow Z_o)]^+$, $i=1,2$, we have: $X_1Z_o \longrightarrow \overline{Y}_1\overline{Y}_2 \in [F_1 \setminus (X_1 \longrightarrow Z_o)]^+$ .

Now consider the case where $X_1 \to Z_o$ participates in the Armstrong's derivation sequence in $F_1$ for $X_1\overline{Y}_1\overline{Y}_2 \to Z_o$.

Similarly as before, there exists

$X_3 \to \overline{Y}_3 \in [E_{F_2}(X) \setminus \{X_1 \to \overline{Y}_1,\ X_2 \to \overline{Y}_2\}]$

such that $X_3 \to \overline{Y}_3$ participates in the Armstrong's derivation sequence for $X_1\overline{Y}_1\overline{Y}_2 \to Z_o$.

The process continues. But as $|E_{F_2}(X)| < +\infty$ so it must be finished at step h, where $h \leq |E_{F_2}(X)|$.

At that moment, we have

$$X_1(\overline{Y}_1\overline{Y}_2\ldots\overline{Y}_h) \to Z_o,$$

and $X_1 Z_o \to (\overline{Y}_1\ldots\overline{Y}_h) \in [F_1 \setminus \{X_1 \to Z_o\}]^+$

The proof is complete.

The theorem shows that attributes of the right side of an FD $X_i \to \overline{X}_i$ belonging to $E_F(X)$ in a minimum cover F can be divided in two classes. The first class consists of invariant attributes, that is, they must appear in right sides of FDs in every $E_{F_j}(X)$ where $F_j$ is any minimum cover equivalent to F. The second class consists of all attributes that belong to a set $Z_o \subseteq \bar{X}_i$, where $X_i \to \bar{X}_i \in E_F(X)$ such that $\exists Z \neq Z_o$ and $X_i Z_o \leftrightarrow X_i Z \in [\{F \setminus (X_i \to Z_o\}]^+$. For attributes of the second class, we have:

<u>Corollary 3.5.1.</u>

With $Z_o$ and Z as determined in Theorem 3.5.1., we can replace $Z_o$ in $(X_1 \to \bar{X}_1) \in F$ by Z and doing so, we obtain an equivalent minimum cover.

<u>Proof</u>

The proof is straight-forward.
From $X_1 \to Z_o \in F$ and $X_1 Z_o \to Z \in [F \setminus X_1 \to Z_o]$ we obtain $X_1 \to Z$ and conversely, from $X_1 \to Z$ and $X_1 Z \to Z_o \in (F \setminus \{X_1 \to Z_o\})^+$ we obtain $X_1 \to Z_o$.

<u>Remark 3.5.1.</u>

From Theorem 3.5.1 and Corollary 3.5.1, we found the transformation rules for right sides of FDs in equivalent minimum covers as follows:
First, there are attributes of right sides that can be

replaced by equivalent sets via left sides. Such transformations can be done in a single FD too. Second, there are attributes of right sides that are invariant (i.e. always present) and only change places in right sides of the equivalence class. In that case, transformation must be done simultaneously in several FDs of the equivalence class.

Corollary 3.5.2.

Let $Z_o \subseteq \bar{X}_1$ where $X \to \bar{X}_1 \in E_{F_1}(X)$ and $Z_o \subseteq R \setminus L$. Then in any minimum cover $F_2 \equiv F_1$, we have

$$Z_o \subseteq RE_{F_2}(X).$$

Proof.

First observe that if there exists Z such that $(X_1 Z_o \to Z) \in F^+$, then there exists an FD $(Y \to W) \in F$ such that $Y \cap Z_o \neq \emptyset$.

Thus $$Z_o \not\subseteq R \setminus L$$

Therefore all attributes in $R \setminus L$ belong to the first class, i.e., invariant in equivalence classes of equivalent minimum covers.

Let $E_F(X)$ is the set of following FDs:

$$X_1 \to \tilde{X}_1\tilde{X}_1'$$
$$X_2 \to \tilde{X}_2\tilde{X}_2'$$
$$\cdots$$
$$X_k \to \tilde{X}_k\tilde{X}_k' \quad ,$$

where $\tilde{X}_i \subseteq LE_F(X)$ and $\tilde{X}_i' \cap LE_F(X) = \emptyset$.

We can split $E_F(X)$ into

$$X_1 \to \tilde{X}_1$$
$$X_2 \to \tilde{X}_2$$
$$\cdots$$
$$X_k \to \tilde{X}_k$$
$$X_1 \to \tilde{X}_1'$$
$$X_2 \to \tilde{X}_2'$$
$$\cdots$$
$$X_k \to \tilde{X}_k' \ .$$

Consider the first k FDs. They can be replaced by the following FDs while not altering the closure.

$$X_{i_1} \longrightarrow X_{i_2}$$
$$X_{i_2} \longrightarrow X_{i_3}$$
$$\cdots$$
$$X_{i_k} \longrightarrow X_{i_1}$$

where $(i_1, i_2, \ldots, i_k)$ is a permutation of $(1, 2, \ldots, k)$.

Let A be any attribute in $RE_F$ (X). Then, there exists i such that $A \in \tilde{X}_i \tilde{X}'_i$ . If $A \in \tilde{X}_i \subseteq LE_F(X)$, then there exists j such that $A \in X_j$.
With any p, $1 \leq p \leq k$, $p \neq j$, we can construct a new cover equivalent to F by replacing $E_F(X)$ by:

$$
\begin{aligned}
X_p &\longrightarrow X_j \tilde{X}'_p \\
X_j &\longrightarrow X_{i_3} \tilde{X}'_j \\
&\cdots \\
X_{i_{k-1}} &\longrightarrow X_{i_k} \tilde{X}'_{i_{k-1}} \\
X_{i_k} &\longrightarrow X_p \tilde{X}'_{i_k}
\end{aligned}
$$

where ( p, j, $i_3$, $i_4$, ..., $i_k$ ) is a permutation of ( 1, 2, ..., k )

After reducing right sides, if A is an invariant attribute, A must belong to the right side of the FD that has p as index.

If $A \in \tilde{X}'_i \not\subseteq LE_F(X)$ then with any p, $1 \leq p \leq k$, we can construct a new minimum cover equivalent to F by replacing $E_F(X)$ by:

$$
\begin{aligned}
X_1 &\rightarrow X_2 \tilde{X}'_2 \\
X_2 &\rightarrow X_3 \tilde{X}'_3 \\
&\cdots \\
X_{i-1} &\rightarrow X_i \tilde{X}'_p \\
&\cdots \\
X_p &\rightarrow X_{p+1} \tilde{X}'_i \\
&\cdots
\end{aligned}
$$

. . .

$$X_k \longrightarrow X_1 \tilde{X}'_1$$

After reducing right sides, if A is an invariant attribute, A must belong to the right side of the FD in $E_F(X)$ that has p as index.

For these reasons, we say that invariant attributes in right sides can be distributed enough freely in right sides of FDs in $E_F(X)$.

However, FDs in an equivalence class must satisfy the following property.

<u>Property 3.5.1</u>

Let $E_F(X) = \{X_j \to \bar{X}_j \mid j=\overline{1,k}\}$.

Then $\forall\, i\ \exists\, j$ such that

$$X_i \to X_j \in [\{F \setminus E_F(X)\} \cup \{X_i \to \bar{X}_i\}]^+$$

<u>Proof</u>

Let be given arbitrary i ,p. By Theorem 3.2.2. we have

$$X_i \supseteq Z_{h_1}$$

$$X_i W_{h_1} \supseteq Z_{h_2}$$

$$\cdots$$

$$X_i W_{h_1} \ldots W_{h_q} \supseteq X_p$$

where $(Z_{h_t} \longrightarrow W_{h_t}) \in F$, $t=\overline{1,q}$.

Choose $Z_{h_l} \to W_{h_l} \in E_F(X)$ with l as small as possible. With such $Z_{h_l}$, we have the required result. In the case, such $Z_{h_l}$ does not exist, $X_p$ is the required $X_j$.

Thus, in spite of the relative arbitrary distribution, each right side of an FD in the equivalence class must carry enough information such that, together with FDs in $F \setminus E_F(X)$, an FD of the form $(X_i \to X_j) \in F^+$ can be derived, ensuring the equivalence of left sides.

Finally, using the results just mentioned in this section, we can introduce the notion of "quasioptimal" cover. In [22] Maier defined the optimal cover (see Def. 3.2.7) and shown that the optimal cover problem is NP-complete.

However from the point of view of effective memory management this does not mean that there is no problem to be discussed, even in the case this optimal cover is found.

Consider k FDs of $E_F(X)$ in a minimum cover.

$E_F(X) = \{X_i \to \tilde{X}_i \tilde{X}'_i \mid i=\overline{1,k}\}$.

If we replace $E_F(X)$ by

$$\{X_1 \to X_2,\ X_2 \to X_3, \ldots, X_k \to X_1,\ X_1 \to \bigcup_{i=1}^{k} \tilde{X}'_i\}$$

then, for k first FDs we have only to take care of (to manage) their left sides, in opposite of the case of $E_F(X)$ we must manage both of left and right sides of all its FDs.

Moreover with $Z_o \subseteq \bigcup_{i=1}^{k} \tilde{X}'_i$, if there exists $Z \subseteq LE_F$ such that

$$X_i Z_o \leftrightarrow X_i Z \in [F \setminus \{X_i \to Z_o\}]^+,$$

then we can replace $X_1 \to \bigcup_{i=1}^{k} \tilde{X}'_i$ by the FD: $X_1 \to \bigcup_{i=1}^{k} \tilde{X}'_i \setminus Z_o$ and still obtain an equivalent cover.

We close this chapter with an algorithm to find the "quasi optimal" cover, in the above sense.

Input: The set G of FDs.

Output: The "quasi optimal" cover F with $F^+ = G^+$

Method

Step 1: From G, find the minimum cover $F_1$ = MINIMIZE [see 22].

We obtain the equivalence classes $E_{F_1}(X^1), \ldots, E_{F_1}(X^s)$ with the corresponding sets of FDs.

$$\{X_i^1 \to \bar{X}_i^1 \mid i=\overline{1,k_1}\}, \ldots, \{X_i^s \to \bar{X}_i^s \mid i=\overline{1,k_s}\}.$$

Step 2: For each equivalence class $E_{F_1}(X^l)$, $l=1,2,\ldots,s$

Set $M_l = RE_{F_1}(X^l) \setminus L$ *) .

For $i=\overline{1,k_l}$ , consider $X_i^l \to \bar{X}_i^l$ .

---

*) L is the union of all left sides of FDs in $F_1$

$\forall A \in \bar{X}_i^l$ ,

If $A \in LE_{F_1}(X^l)$ we omit it,

If $A \bar{\in} LE_{F_1}(X^l)$ then for each $Z \subseteq LE_{F_1}(X^l) \setminus X_i^l$ check whether

$$(X_i^l Z \to A) \in [F \setminus \{X_i^l \to A\}]^+ \quad ?$$

if <u>true</u> then we omit it

if <u>false</u> then $M_l = M_l \cup \{A\}$.

<u>Step 3:</u> At the end of Step 2 we obtain $M_l$ for each equivalence class, and F is the following cover:

$$F = \bigcup_{l=1}^{s} \{X_1^l \to X_2^l , X_2^l \to X_3^l, \ldots, X_{k_l}^l \to X_1^l, X_1^l \to M_l\}$$

REFERENCES

[1] Codd, E.F., A relational model of data for large shared data banks, Comm. ACM, Vol 13, N°6, June 1970 pp. 377-387.

[2] Delobel, C., Casey, R.G., Decompositon of a Data base and the theory of Boolean switching functions, IBM J. Res. Dec. 17, 1973, pp. 374-386.

[3] Fagin, R., Multivalued dependencies and a new normal form for relational databases, ACM Trans on Database Systems, Vol.2, No 3, Sept. 1977, pp. 262-278.

[4] Beeri, C., Fagin, R, and Howard, H.J., A complete axiomatization of functional and multivalued dependencies in data base relations, ACM SIGMOD Int. Conf. on Management of Data, Toronto, Canada, 1977, pp. 47-61.

[5] Rissanen, J., Independent component of relations, ACM Trans. on Data base Systems, Vol. 2, No 4., Dec. 1977.

[6] Aho, A.V., Beeri, C., and Ullman, J., The theory of foins in relational databases, Proc. 19th IEEE Symp. on Foundations of Comp. Sci., 1977.

[7] Békéssy, A., Demetrovics, J., et.al., On the number of maximal dependencies in a database of fixed order, Discrete Mathematics 30 (1980) pp. 83-88.

[8] Demetrovics, J. and Gyepesi, G., On the functional dependency and some generalization of it, Acta Cybernetica 5(1981), 295-305.

[9] Demetrovics, J., Füredi, Z., and Katona, G.O.H, Minimum matrix representation of closure operations, Discrete Applied mathematics 11 (1985), 115-128.

[10] Beeri, C. and Bernstein, P.A., Computational problems related to the design of normal forms for relational schemes. ACM Transactions on Database Systems, Vol. 4, No 1, March 1979.

[11] Lucchesi, C.L., Osborn, S.L.. Candidate keys for relations. J. of Computer and System Sciences, 17 1978, pp. 270-279.

[12] Date, C.J.: An introduction to Databases Systems, Addison. Wesley Publishing Company, London, 1982.

[13] Ullman, J.D.: Principles of Database Systems. 2nd ed. Computer Science Press. Potomac, Md., 1982.

[14] Delobel, C.: Theoretical aspects of modeling in relational database. Rapport de recherche No 177, July 1979, Université scientifique et medicale et Institut national polytechnique de Grenoble.

[15] Armstrong, W.W.: Dependency structures of database relationships. Information Processing 74, North Holland Publishing Company, 1974, pp. 580-583.

[16] Sperner, E.: Ein Satz über Untermengen einer endlichen Menge. Math. Z. 27(1928) 544-548.

[17] Demetrovics, J.: On the number of candidate keys, Information processing letter, vol. 7. number 6, October 1978.

[18] Békéssy, A. Demetrovics, J.: Contributions to the theory of database relations, Discrete Math., 27(1979), 1-10.

[19] Fernandez, M.C.: Determining the normalization level of a relation on the basis of Armstrong's axioms. Computers and Artificial Intelligence, 3 (1984), pp. 495-504.

[20] Osborn, S.L.: Normal forms for relational databases. Ph. D. Dissertation, University of Waterloo, 1977.

[21] Bernstein, P.A.: Synthesizing third normal form relations from functional dependencies. ACM Trans. Database Syst. 1,4(Dec. 1976), 277-298.

[22] Maier, D.: Minimum covers in the relational databse model. J. ACM 27 (oct. 1980), 664-674.

[23] Ausiello, G. et al.: Graphs algorithms for functional dependency manipulation. J. ACM 30 (Oct. 1983), 752-766.

[24] Maier, D.: The theory of relational data bases, Computer Science Press, Potomac, Md, 1983.

[25] Armstrong, W.W.: On the generation of dependency structures of relational data bases. Publ. 272 Université de Montreal 1977.

[26] Ho Thuan and Le van Bao, Some results about keys of relational schemas, Acta Cybernetica, Tom 7, Fasc. 1, Szeged 1985. pp. 99-113.

[27] Le van Bao and Ho Thuan, On some properties of keys for relation scheme, Preprint series, No 9, 1983, Hanoi, Institute of Mathematics and Institute of Computer Sciences and Cybernetics.

[28] Le van Bao and Ho Thuan, Sufficient conditions for which a relation scheme has precisely one key. Preprint Series, No 7, 1983, Hanoi.

[29] Demetrovics, J. and Ho Thuan, Some additional properties of keys for relation scheme, MTA SZTAKI, Közlemények 34/1986

[30] Demetrovics, J. and Ho Thuan, Keys and super-keys for relation scheme, Proceedings of the KNVVT Conference, May 5-9. 1986. Budapest.

[31] Ho Thuan, Some remarks on the algorithm of Lucchesi and Osborn, MTA SZTAKI, Közlemények 35/1986 (To appear).

[32] Ho Thuan and Le van Bao, Translation of relational schemas, MTA SZTAKI, Közlemények 30/1984 pp. 7-36.

[33] Demetrovics, J., Ho Thuan et al., Balanced relation scheme and the problem of key representation, MTA SZTAKI, Közlemények 32/1985 pp. 51-80.

[34] Ho Thuan, Direct determination and FD-graph, MTA SZTAKI, Közlemények 34/1986

[35] Ho Thuan, Some invariants of covers for functional dependencies, MTA SZTAKI Közlemények 34/1986.

[36] Tran Thai Son, Dinh thi ngoc Thanh and Ho Thuan, Some results about the structure of minimum covers in the relational database model, MTA SZTAKI Közlemények 35/1986 /To appear/

[37] Ho Thuan, Tran thai Son and Dinh thi ngoc Thanh, Structure of nonredundant and minimum covers, /To appear/.

[38] J. Demetrovics, Ho Thuan, Nguyen xuan Huy, Le van Bao, Translations of relation schemes, balanced relation scheme and the problem of key representation. EIK, Berlin, 1986. /To appear/

[39] M. Arató, A. Benczur, Dynamic placement of records and the classical occupancy problem, Comp. and Math. with Appls., 7/1981/, 173-185.

[40] M. Arató, A. Benczur, A general treatment of rearrangement problems in a linear storage, Performance Evaluation, 2/1982/, 108-117.

[41] A. Benczur, Problems in modelling of data base performance, Perspectives in Developing Computer Technics, Moscow, URSS, 1983, 85-98.

[42] A. Benczur, J. Stahl, On updating a large-scale data-system, Alkalmazott Matematikai Lapok, 10/1984/, 1-13. /In Hungarian/.

1985-BEN MEGJELENTEK:

166/1985 Radó Péter: Információs rendszerek számitógépes tervezése

167/1985 Studies in Applied Stochastic Programming I. Szerkesztette: Prékopa András /utánnyomás/

168/1985 Böszörményi László - Kovács László - Martos Balázs - Szabó Miklós: LILIPUTH

169/1985 Horváth Mátyás: Alkatrészgyártási folyamatok automatizált tervezése

170/1985 Márkus Gábor: Algoritmus mátrix alapu logaritmus kiszámitására kriptográfiai alkalmazásokkal

171/1985 Tamás Várady: Integration of free-form surfaces into a volumetric modeller

172/1985 Reviczky János: A számitógépes grafika terület-kitöltő algoritmusai

173/1985 Kacsukné Bruckner Livia: Mozgáspálya generálás bonyolult geometriáju felületek 2 1/2D-s NC megmunkálásához

174/1985 Bolla Marianna: Mátrixok spektrálfelbontásának és szinguláris felbontásának módszerei

175/1985 Hannák László, Radó Péter: Adatmodellek, adatbázis-filozófiák

176/1985 Számitógépes képfeldolgozási és alakfelismerési kutatók találkozója.
Szerkesztette: Csetverikov Dmitirj,
Főglein János és Solt Péter

177/1985 Gyárfás András: Problems from the world surrounding perfect graphs

178/1985 PUBLIKÁCIÓK'84
Szerkesztette: Petróczy Judit

1986-BAN EDDIG MEGJELENTEK:

| | |
|---|---|
| 179/1986 | Terlaky Tamás: Egy véges criss-cross módszer és alkalmazásai |
| 180/1986 | K.N. Čimev: Separable sets of arguments of functions |
| 181/1986 | Renner Gábor: Kör approximációja a számitógépes geometriai tervezésben |
| 182/1986 | Proceedings of the Joint Bulgarian-Hungarian Workshop on "Mathematical Cybernetics and Data Processing" Scientific Station of Sofia University, Giulecica /Bulgaria/, May 6-10, 1985 /Editors: J. Denev, B. Uhrin/ Vol I |
| 183/1986 | Proceedings of the Joint Bulgarian-Hungarian Workshop on "Mathematical Cybernetics and Data Processing" Scientific Station of Sofia University, Giulečica /Bulgaria/, May 6-10, 1985 /Editors: J. Denev, B. Uhrin/ Vol II |

12

Készült az Országos Széchényi Könyvtár
Sokszorositó üzemében, Budapest
21 A/5 ív terjedelemben, 320 példányban,
munkaszám: 86 300
Felelős vezető: Rosta Lajosné