International Journal of Computer and Information Sciences, Vol. 10, No. 4, 1981

Theta-Join and Normalization[1]

Sudhir K. Arora[2] and Surya R. Dumpala[3]

Received May 1981; revised September 1981

Normalization and dependency theory is used for the logical design of relational data bases. Historically this theory has been based on the operations projection and natural join. Some work has been reported in the literature for the operations union and splitting. This paper extends this theory to the operation theta-join.

KEY WORDS: Relational data bases; normal forms; dependencies; inference rules; relational algebra; data base design.

1. INTRODUCTION

The process of normalization has so far been confined to the operations projection and natural join. Very little work has been done to extend this process to other operations. A beginning was made in [11] and [24] where the operations union and splitting are considered. This paper extends the normalization process to theta-join. We study the operator multivalued dependency (OMVD) which is based on theta-join and propose the modified theta-join, projection normal form (MTP/NF).

In sec. 2 we acquaint the reader with some background in this area; in sec. 3 we define OMVD and show its semantic usefulness; in sec. 4 we propose MTP/NF; in sec. 5 we study inference rules for OMVD; in sec. 6 we note some limitations of this work and in sec. 7 we offer some concluding remarks.

[1] This research was supported in part by the Science and Engineering Research Board Grant Number 214–7248.

[2] Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan 48202.

[3] Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.

2. BACKGROUND

The theoretical design of a relational data base must involve a consideration of the following three concepts:

1. Dependencies;
2. Normalization;
3. Operations.

We consider each one of them separately.

Dependencies are a systematic way of dealing with semantic situations that arise in data. For many years, Functional Dependency was the only known dependency. However, the need for other dependencies was felt when it became apparent that FD is not adequate to model many semantically useful situations in data. (See for example Ref. 23). In the last three years, several new dependencies have appeared in the literature. The various dependencies in the literature are:

1. Functional Dependency (FD),[5]
2. Multivalued Dependency (MVD),[9]
3. Mutual Dependency (MUD),[18]
4. Hierarchical Dependency (HD),[8]
5. Mixed Dependency,[2]
6. CoDependency (COD),[2]
7. Template Dependency, (TD)[21]
8. Transitive Dependency, (TrD)[19]
9. General Dependency (GD),[14]
10. Subset Dependency (SD),[22]
11. Boolean Dependency (BD),[10]
12. Generalized Mutual Dependency (GMD),[16]
13. Implication Dependency (ID),[12]
14. Null Multivalued Dependency (NMVD),[15]
15. Extended Functional Dependency (EFD),[25]
16. Algebraic Dependency (AD),[26]
17. Join Dependency (JD).[20]

It should be noted that many of these dependencies have no easy semantic interpretation and may be of theoretical interest only.

Codd,[5,6] proposed normalization for a relational data base for two main reasons: first it allowed the data base to be viewed as a collection of tables and second, it permitted the definition of a small class of primitive operators that were capable of manipulating relations to obtain all necessary logical connections among attributes. Further, normalization eliminated some undesirable effects,[5,13] in the relational scheme when operations were carried out on them. These were "Insertion anomaly," "Deletion anomaly," and "Consistency on update." Another way to view normalization is as follows. We have pointed out earlier that dependencies are a systematic way to model knowledge and ensure integrity in a data base. Unless we control the pattern of dependencies in the relations of a data base semantically undesirable effects (anomalous behavior) occur in the data base. Normalization pocess can be thought of as a way of controlling this pattern of dependencies. Here we do not distinguish between a relational, network or hierarchical data base, although the normal forms presented in the literature and in this paper are all concerned with the relational model. Extending the normalization process to other models is an open problem.

Another concept that is implicit in all the normal forms is that of operations on the data vase. Historically, only projection and natural join entered the normalization process. Hence, most normal forms are based on these two operations. However, in the last two years union and splitting (defined by Fagin[11] as the opposite of union) have entered this process (PSJU/NF and (3, 3)NF (See Refs. 11, 24). These are the beginnings of a new area in normalization theory.

3. OPERATOR MULTIVALUED DEPENDENCY (OMVD)

In this section we define OMVD and show its semantic usefulness. Two definitions for OMVD are given: one based on theta-join and the other based on the pattern of tuples in a relation. It should be noted that these two methods for defining dependencies have been widely used in the literature. (For example see Refs. 9, 21, 26, 20).

Definition

In a relation $R[A, B, C, D]$ if $(a_i b_1 c_j d_1)$ $(a_k b_2 c_l d_2)$ are tuples then R must have $(a_i b_1 c_l d_2)$ and $(a_i b_2 c_j d_1)$ if $a_i = a_k$ and/or $(a_k b_2 c_j d_1)$ and $(a_i b_1 c_j d_2)$ if $c_j = c_l$ and $(a_i \theta c_j)$ must be true for all possible a_i and c_j in R. Then R satisfies the OMVD $(A \theta C) \to \to B \mid D$ where θ may be $=$, $\neq$, $>$, $<$, $\geqslant$, $\leqslant$ or any binary relation L between A and C. Here A, B, C, and D are mutually disjoint sets of attributes.

Definition

In a relation $R[A, B, C, D]$ if

$$R[A, B, C, D] = R[A, B][A \theta C] R[C, D] \text{(Theta-Join)}$$

where θ may be $>$, $<$, $=$, $\geqslant$, $\leqslant$, $\neq$ or an arbitrary binary relation L, between A and C, then the *Operator Multivalued Dependency* (OMVD) $A\ \theta\ C) \rightarrow \rightarrow B \mid D$ holds in $R[A, B, C, D]$. Here A, B, C, and D are mutually disjoint sets of attributes.

We now give examples of semantic situations which to the best of our knowledge can be modeled easily by an OMVD only.

1. Consider the relation,

$$R\ [\text{CREDIT-LIMIT, AGENT, PRODUCT-PRICE}].$$

W wish to model the following constraint on R.
An agent must represent all products with price less than his credit limit.
In R the following OMVD holds.

$$(\text{PRODUCT-PRICE} < \text{CREDIT-LIMIT}) \rightarrow \rightarrow \phi \mid \text{AGENT}.$$

2. Consider the relation,

$$R\ [\text{COMPANY, PRODUCT, AGENT, SALESMAN}].$$

We wish to model the following constraint on R.
When a company is associated with an agent, then every salesman working for this agent must represent every product of that company.
In R the following OMVD holds.

$$(\text{COMPANY}\ L\ \text{AGENT}) \rightarrow \rightarrow \text{PRODUCT} \mid \text{SALESMAN}.$$

We get one OMVD for each binary relation L.

3. Consider the relation,

$$\text{NETWORK}[\text{PATH-FLOW-RATES, PATH, ITEM-FLOW-RATES, ITEM}].$$

We wish to model the following constraint.
If a path flow rate is greater than an item flow rate then the item must be channelled through the path.
In NETWORK the following OMVD holds.

$$(\text{PATH-FLOW-RATES} > \text{ITEM-FLOW-RATES}) \rightarrow \rightarrow \text{PATH} \mid \text{ITEM}.$$

4. MODIFIED THETA-JOIN PROJECTION NORMAL FORM (MTP/NF)

We propose a normal form to control the pattern of OMVDs in a relation. But first we show by an example that OMVDs cause anomalous behavior of a relational data base. For the semantic example 2 in sec. 3 we have the following instance of R:

$$R[\text{COMPANY, PRODUCT, AGENT, SALESMAN}] = \begin{array}{cccc} c_1 & p_1 & a_1 & b_1 \\ c_1 & p_2 & a_1 & b_2 \\ c_2 & p_1 & a_2 & b_3 \\ c_2 & p_3 & a_2 & b_4 \\ c_1 & p_1 & a_1 & b_2 \\ c_1 & p_2 & a_1 & b_1 \\ c_2 & p_1 & a_2 & b_4 \\ c_2 & p_3 & a_2 & b_3 \end{array}$$

Here, if a user deletes the tuple $(c_1\, p_2\, a_1\, b_1)$, the constraint will be violated because company, c_1 makes product p_2 and is associated with agent a_1. But the salesman b_1 does not represent product p_2. Again, if a user inserts the tuple $(c_1\, p_3\, a_1\, b_2)$, the constraint is violated because the salesman b_1 does not represent product p_3. Here the operations on the data base affect more than the one tuple addressed by the user, causing anomalous behavior.

These problems would not be there if we decompose R as follows:

$$R[\text{COMPANY, PRODUCT}] = \begin{array}{cc} c_1 & p_1 \\ c_1 & p_2 \\ c_2 & p_1 \\ c_2 & p_3 \end{array}$$

$$R[\text{COMPANY, AGENT}] = \begin{array}{cc} c_1 & a_1 \\ c_2 & a_2 \end{array}$$

$$R[\text{AGENT, SALESMAN}] = \begin{array}{cc} a_1 & b_1 \\ a_1 & b_2 \\ a_2 & b_3 \\ a_2 & b_4 \end{array}$$

We have proposed the Theta-Join Projection Normal Form (TP/NF) in an earlier paper.[3]

Definition

A relation, $R[A, B, C, D]$, is in *Theta-Join, Projection Normal Form* (*TP/NF*) if whenever a nontrivial OMVD $(A\ \theta\ B) \rightarrow\ \rightarrow C \mid D$ holds in R then so does $AB \rightarrow X$ for every column name X in R. Here A, B, C, D are mutually disjoint sets of attributes.

Here a trivial OMVD is one in which θ is "=" and one of the attribute sets on the right hand side is null.

This definition of TP/NF leads to the following problem not envisaged earlier. In $R[A, B, C, D]$ let $(A\ \theta\ B) \rightarrow\ \rightarrow C \mid D$, $AB \rightarrow C$ and $AB \rightarrow D$, then R is in TP/NF. Consider an instance of R;

$$R = \begin{matrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \end{matrix}$$

Insert the tuple $a_1\, b_3\, c_2\, d_3$ where $(a_1\ \theta\ b_3)$ is valid. Now $AB \rightarrow C$ and $AB \rightarrow D$ but $(A\ \theta\ B) \not\rightarrow\ \not\rightarrow C \mid D$ because $a_1\, b_1\, c_2\, d_1$ and $a_1\, b_3\, c_1\, d_3$ are not in R. Hence even though R is in TP/NF and the insert operation preserves the key, the OMVD is violated. On the other hand, if we preserve to OMVD then $AB \not\rightarrow C$. We define Modified Theta-Join, Projection Normal Form (MTP/NF) to take care of this problem.

Definition

A relation $R[A, B, C, D]$ is in Modified Theta-Join Projection Normal Form (MTP/NF) if whenever a nontrivial OMVD $(A\ \theta\ B) \rightarrow\ \rightarrow C \mid D$ holds in R then so does $A \rightarrow X$ and $B \rightarrow X$ for every column name X in R. Here A, B, C, D are mutually disjoint sets of attributes.

In the above example the insertion of $a_1\, b_3\, c_2\, d_3$ would be disallowed because $A \rightarrow C$ is violated.

5. INFERENCE RULES FOR OMVD

Inference rules for a dependency are necessary to be able to derive all possible dependencies implied by a given set of dependencies. Considerable research effort has been devoted to discovering inference rules for the other dependencies in the literature [see for example Refs. 1, 4, 7, 8]. In this section we propose to study 15 inference rules for OMVD.

We extend the definition of OMVD to the case where the sets of attributes are not mutually disjoint.

Definition

In a relation, $R[X_1, X_2, A, B]$ where $X_1, X_2 A$, B are not necessarily disjoint, the OMVD.

$$(X_1\ \theta\ X_2) \rightarrow\ \rightarrow A \mid B$$

holds if

$$R[X_1, X_2, A, B] = R[X_1, A][X_1 \, \theta \, X_2] \, R[X_2, B]$$

where θ is any operation $>$, $<$, $\geqslant$, $\leqslant$, $\neq$, $=$ or any binary relation L between the values of the attribute sets X_1 and X_2.

We extend the concept of a loss less decomposition of a relation as follows. A relation $R[X_1, X_2, A, B]$ is decomposable in a loss-less way into its projections $R[X_1, A]$ and $R[X_2, B]$ if

$$R[X_1, X_2, A, B] = R[X_1, A][X_1 \, \theta \, X_2] \, R[X_2, B]$$

i.e. if OMVD, $(X_1 \, \theta \, X_2) \rightarrow \rightarrow A \mid B$ holds in R.

Note

1. In a relation $R[X_1, X_2, A, B]$ with nondisjoint attribute sets X_1, X_2, A, and B, we cannot conveniently specify a tuple. For example $(x_1 \, x_2 \, a \, b)$ is a "tuple" in which some attribute values are duplicated merely because x_1, x_2, a and b are not mutually disjoint. However if we remove the duplicates from $(x_1 \, x_2 \, a \, b)$ we do get a unique tuple of R. We shall refer to a tuple with duplicate values as, "tuple".
2. We can define the following special OMVD's.

$$(\phi \, \theta \, \phi) \rightarrow \rightarrow A \mid B \Leftrightarrow R[A, B] = R[A] \, XR[B].$$

$$(X \, \theta \, \phi) \rightarrow \rightarrow A \mid B \Leftrightarrow R[X, A, B] = R[X, A] \, XR[B].$$

where X denotes Cartesian Product.

3. In a relation $R[X_1, X_2, A, B]$,

$$U = X_1 \cup X_2 \cup A \cup B$$

$$\bar{A} = U - A$$

Similarly in a "tuple" $(x_1 \, x_2 \, a \, b)$ of R,

$$u = x_1 \cup x_2 \cup a \cup b$$

$$\bar{a} = u - a$$

4. In a relation $R[X_1, X_2, A, B]$, $(X_1 \, \theta \, X_2)^*$ is true if every "tuple" in $R[X_1, X_2]$ satisfies $(X_1 \, \theta \, X_2)$ and no such new "tuple" can be added to $R[X_1, X_2]$ without adding a new element to $S_R[X]$ or $S_R[X_2]$. It is trivially true that if $(X_1 \, \theta \, X_2) \rightarrow \rightarrow A \mid B$ then $(X_1 \, \theta \, X_2)^*$ is true.

5. In a relation $R[X_1, X_2, A, B]$ we specify a "tuple" as $(x_1\, x_2\, a\, b)$ where x_1, x_2, a, b are values taken by sets of attributes X_1, X_2, A and B, respectively.

OMVD1 (Reflexivity)

In a relation $R[X_1, X_2]$ if $A \subseteq X_1$ and $B \subseteq X_2$ and $(X_1\, \theta\, X_2)^*$ then $(X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid B$

Proof

In $R[X_1, X_2]$,

$$X_1 \rightarrow A$$

$$X_2 \rightarrow B$$

and $(X_1\, \theta\, X_2)^*$

$\therefore$ By FD-OMVD 1 (Proved later)

$$(X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid B$$

OMVD2 (Computation)

In a relation $R[X_1, X_2, A, B]$,

$$(X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid B \text{ if and only if}$$

$$(X_1\, \theta\, X_2) \rightarrow \rightarrow B \mid A$$

This rule is not true as is demonstrated by the following example.

$$R = \begin{matrix} x_{11} & x_{21} & a_1 & b_1 \\ x_{11} & x_{21} & a_2 & b_1 \\ x_{12} & x_{21} & a_3 & b_1 \end{matrix} \qquad L = \begin{matrix} x_{11} & x_{21} \\ x_{12} & x_{21} \end{matrix}$$

Here $(X_1\, .L.\, X_2) \rightarrow \rightarrow A \mid B$ is true
For $(X_1\, .L.\, X_2) \rightarrow \rightarrow B \mid A$ to be true,

$$R = \begin{matrix} x_{11} & x_{21} & a_1 & b_1 \\ x_{11} & x_{21} & a_2 & b_1 \\ x_{11} & x_{21} & a_1 & b_1 \\ x_{12} & x_{21} & a_1 & b_1 \\ x_{12} & x_{21} & a_2 & b_1 \\ x_{12} & x_{21} & a_3 & b_1 \end{matrix}$$

OMVD3 (Augmentation)

In a relation $R[X_1, X_2, A, B]$ if $(X_1 \theta X_2) \rightarrow \rightarrow A \mid B$, $W' \subseteq W$, $W'' \subseteq W$ then $(X_1\, W \theta X_2\, W) \rightarrow \rightarrow A W' \mid B W''$. Here $W \subseteq U$.

Proof

The operation 'θ' can be $=$, $\neq$ or L to be meaningful for this inference rule. In a tuple of R, if $(X_1 \theta X_2)$ holds where θ is one of the above three, then obviously $(X_1\, W \theta X_2\, W)$ also holds. Now,

$$R = R[X_1, A][X_1 \theta X_2] R[X_2, B] \tag{1}$$

Assume,

$$R \neq R[X_1, W, A][X_1\, W \theta X_2\, W] R[X_2, W, B] = R_1 \tag{2}$$

Let $(x_1\, a_1\, x_2\, b_1)$ be a "tuple" in R and not in R_1. It can be rewritten as $(x_1\, w\, a_1\, x_2\, w\, b_1)$ where,

$$(x_1\, w\, a_1) \in R[X_1, W, A]$$

and

$$(x_2\, w\, b_1) \in R[X_2, W, B]$$

In (2) therefore, $(x_1\, w\, a_1\, x_2\, w\, b_1) \in R_1$-contradiction.
Similarly a "tuple" in R_1 must belong to R, because if,

$$(x_1\, w\, a_1\, x_2\, w\, b_1) \in R_1$$

$$(x_1\, w\, a_1) \in R[X_1, w, A]$$

and

$$(x_2\, w\, b_1) \in R[X_2, W, B]$$

i.e.

$$(x_1\, a_1) \in R[X_1, A]$$

and

$$(x_2\, b_1) \in R[X_2, B]$$

$$\therefore \quad (x_1\, a_1\, x_2\, b_1) \in R$$

But

$$(x_1\, a_1\, x_2\, b_1) = (x_1\, w\, a_1\, x_2\, w\, b_1)$$

$\therefore$ $$(x_1\, w\, a_1\, x_2\, w\, b_1) \in R\text{—a contradiction.}$$

Hence the rule is true.

OMVD4 (Transitivity)

In a relation $R[X_1, X_2, A, C, D]$ if,

$(X_1\, \theta\, X_2) \to \to A \mid CD$, $(A'\, \theta_1\, A'') \to \to C \mid D$,

$U = X_1 \cup X_2 \cup A \cup C \cup D = A \cup C \cup D$ and $A' \cup A'' = A$ then,

$(X_1\, \theta\, X_2) \to \to AC \mid D$

Proof

Let $(X_1\, \theta X_2) \to \to A \mid CD$ and $(A'\, \theta_1\, A'') \to \to C \mid D$ but $(X_1\, \theta\, X_2) \not\to \not\to AC \mid D$.

Let

$$R = R[X_1, A][X_1\, \theta\, X_2]\, R[X_2, C, D] \tag{3}$$

$$= R[A', C][A'\, \theta_1\, A'']\, R[A'', D] \tag{4}$$

$$\neq R[X_1, A, C][X_1\, \theta\, X_2]\, R[X_2, D] = R_1 \tag{5}$$

A "tuple" in R always belongs to R_1 because R_1 is the join of projections of R. Let

$$(x_1\, a_1\, c_1\, x_2\, d_1) \in R_1$$
$$\notin R$$

But from (3), OMVD5 and (5),

$$R[X_1, X_2, A, D] = R[X_1, A][X_1\, \theta\, X_2]\, R[X_2, D] = R_1[X_1, x_2, A, D] \tag{6}$$

The equation (6) is true because $(X_1\, \theta\, X_2)^*$ is true in both R and R_1.

$\therefore$ $(a_1\, d_1) \in R[A, D]$, $(x_1\, a_1\, c_1) \in R[X_1, A, C]$ and $(a_1) \in R[A]$.

But from (4) (c_1), occurring in a "tuple" of R with $(a_1) \in R[A]$, also occurs with every (d_j) which occurs in some "tuple" with (a_1) in R.

$\therefore$ $$(a_1\, c_1\, d_1) \in R[A, C, D]$$

But $A \cup C \cup D = U$

$$\therefore \qquad (a_1\, c_1\, d_1) = (x_1\, a_1\, c_1\, x_2\, d_1) \in R\text{—a contradiction.}$$

Hence the rule is true.

OMVD5 (Embedded Operator Multivalued Dependency)

In a relation $R[X_1\ X_2, A, B]$ if $(X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid B$ then $(X_1\, \theta\, X_2) \rightarrow \rightarrow A' \mid B'$ in the appropriate projection R' of R and the latter is an Embedded OMVD where $A' \subseteq A$ and $B' \subseteq B$.

Proof

Let

$$R = R[X_1, A][X_1\, \theta\, X_2]\, R[X_2, B] \tag{7}$$

and

$$R' \neq R[X_1, A'][X_1\, \theta\, X_2]\, R[X_2, B'] = R_1 \tag{8}$$

Any "tuple" in R must have its projection in R_1 because R_1 is a join of projections of R. Let

$$(x_1, a'_1\, x_2\, b_1' \in R_1$$

$$\notin R'$$

$$\therefore \qquad (x_1\, a'_1) \in R[X_1, A']$$

$$(x_2\, b_1') \in R[x_2, B']$$

$$(\exists b_1'')(b_1'\, b_1'' \in R[B]) \wedge (x_2 b_1'\, b_1'' \in R[X_2, B])$$

and

$$(\exists a_1'')(a_1'\, a_1'' \in R[A]) \wedge (x_1 a_1'\, a_1'' \in R[X_1, A]).$$

$$\therefore \qquad (x_1\, a'_1\, a_1'' x_2\, b_1'\, b_1'') \in R$$

$$\therefore \qquad (x_1\, a_1'\, x_2\, b_1') \in R'\text{—a contradiction.}$$

Hence the rule is true.

OMVD6

In a relation $R[X_1, X_2, A, B]$ if,

$$(X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid B \text{ then } (X_1\, \theta\, X_2) \rightarrow \rightarrow A \mid \bar{A}$$

Proof

Let

$$R = R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, B] \tag{9}$$

$$\neq R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, \bar{A}] = R_1 \tag{10}$$

Every "tuple" in R is also in R_1 because R_1 is the join of projections of R.
Let

$$(x_1\,a_1\,x_2\,\bar{a}_1) \in R_1$$

$$\notin R$$

$\therefore\ (x_1\,a_1) \in R[X_1, A]$

Case 1 Assume $\bar{A} \supseteq B$.

$\therefore\ (x_2\,b) \in R[x_2, B]$ where b is a projection of $\bar{a}_1$.

$\therefore\ (x_1\,a_1\,x_2\,b) = (x_1\,a_1\,x_2\,\bar{a}_1) \in R$—a contradiction.

Case 2 Assume $\bar{A} \subset B$.

Let $B = \bar{A}B'$ where $B' \subseteq A$ and $B' \cap \bar{A} = \phi$

From (1) and OMVD5,

$R = R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, \bar{A}, B']$

$= R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, \bar{A}]$—a contradiction

Case 3 Assume $\bar{A} \cap B = \phi$

Then A B. i.e., $B = A' \subseteq A$

$\therefore$ From(9),

$R = R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, A']$

and $(x_1\,a_1\,x_2) = (x_1\,a_1\,x_2\,\bar{a}_1) \in R$ (Because $x_2 \in R[x_2]$)

Case 4

Assume $\bar{A} \cap B = P \neq \phi$

Let $B = PB'$ where $P \cap B' = \phi$ and $B' \subseteq A$

From (9) on OMVD5,

$R = R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, P, B']$

$= R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, P]$

Let $\tilde{a}_1 = (pb')$

$\therefore\ (x_2\, p) \in R[X_2, P]$

$\therefore\ (x_1\, a_1\, x_2\, p) = (x_1\, a_1\, x_2\, \tilde{a}_2) \in R$—a contradiction

Hence the rule is true.

OMVD7

In a relation $R[X_1, X_2, A, C]$ if,

$(X_1\, \theta\, X_2) \rightarrow\ \rightarrow A \mid \bar{A}$ and $(X_1\, \theta\, X_2) \rightarrow\ \rightarrow C \mid \bar{C}$ then

1. $(X_1\, \theta\, X_2) \rightarrow\ \rightarrow A \cup C \mid \overline{A \cup C}$
2. $(X_1\, \theta\, X_2) \rightarrow\ \rightarrow A \cap C \mid \overline{A \cap C}$

Proof

Let

$$R = R[X_1, A', A_c][X_1\, \theta\, X_2]\, R[X_2, \tilde{A}', \tilde{A}_c] \tag{11}$$

$$= R[X_1, A_c, \tilde{A}_c][X_1\, \theta\, X_2]\, R[X_2, A', A'] \tag{12}$$

where

$$A = A'A_c, \qquad \bar{A} = \tilde{A}\tilde{A}_c$$

$$C = A_c \tilde{A}_c$$

$$A \cap C = A_c$$

$\bar{A} \cap C = \tilde{A}_c$

Let

1. $R \neq R[X_1, A', A_c, \tilde{A}_c][X_1\, \theta\, X_2]\, R[X_2, \tilde{A}'] = R_1$
2. $R \neq R[X_1, A_c][X_1\, \theta\, X_2]\, R[X_2, A', \tilde{A}', \tilde{A}_c] = R_1$

A "tuple" in R belongs to all R_1 because R_1 are joins of projections of R. To prove that a "tuple" in any R_1 is also in R

1. Let $(x_1\, a'\, a_c\, \tilde{a}_c\, x_2\, \tilde{a}') \in R_1$

$$\therefore\ (x_1\, a'\, a_c) \in R[X_1, A', A_c]$$

To prove $(x_2\, \tilde{a}'\, \tilde{a}_c) \in R[X_2, \tilde{A}', \tilde{A}_c]$—then from (11) we have the result.

But, $(x_2\, \tilde{a}') \in R[X_2, \tilde{A}']$, $(x_1\, \tilde{a}_c) \in R[X_1, \tilde{A}_c]$
$(x_1, x_2) \in R[X_1, X_2]$

From (12) and OMVD5,
$R[X_1, \tilde{A}_c][X_1\,\theta\,X_2]\,R[X_2, \tilde{A}'] = R[X_1, \tilde{A}_c, X_2, \tilde{A}'] = A$ Projection of R

i.e. Any a' which occurs with x_2 occurs with all $\tilde{a}_c$ which occur with x_1 whenever $(x_1\,\theta\,x_2)$ is true.

$\therefore\ (x_2\,\tilde{a}'\,\tilde{a}_c) \in R[X_2, \tilde{A}', \tilde{A}_c]$

Hence $R = R_1$

2. Let $(x_1\,a'\,\tilde{a}_c\,a_c\,x_2\,\tilde{a}') \in R_1$

$\therefore\ (x_2\,\tilde{a}'\,\tilde{a}_c) \in R[X_2, \tilde{A}', \tilde{A}_c]$

To prove, $(x_1\,a'\,a_c) \in R[X_1, A', A_c]$—then from (11) we have the result.

But

$$(x_1\,a_c) \in R[X_1, A_c]$$
$$(x_1\,x_2) \in R[X_1, X_2]$$
$$(x_2\,a') \in R[X_2, A']$$

From (12) and OMVD5,

$R[X_1, A_c][X_1\,\theta\,X_2]\,R[X_2, A'] =$ Appropriate Projection of R i.e. An a_c which occurs with x_1 also occurs with all a' which occurs with x_2 in projections of R whenever $(x_1\,\theta\,x_2)$ is true.

$$\therefore\ (x_1\,a'\,a_c) \in R[X_1, A', A_c]$$
$$\therefore\ R = R_1$$

OMVD8

In a relation $R[X_1, X_2, A, C]$ if,
$(X_1\,\theta\,X_2) \rightarrow\rightarrow \bar{A} \mid A$ and $(X_1\,\theta\,X_2) \rightarrow\rightarrow C \mid \bar{C}$ then
$(X_1\,\theta\,X_2) \rightarrow\rightarrow \overline{A - C} \mid A - C$

Proof

Let

$$R = R[X_1, \tilde{A}', \tilde{A}_c][X_1\,\theta\,X_2]\,R[X_2, A', A_c] \tag{13}$$
$$= R[X_1, A_c, \tilde{A}_c][X_1\,\theta\,X_2]\,R[X_2, A', \tilde{A}'] \tag{14}$$

where

$$A = A' A_c \qquad \bar{A} = \tilde{A}' \tilde{A}_c$$
$$C = A_c \tilde{A}_c \qquad A \cap C = A_c$$
$$\bar{A} \cap C = \tilde{A}_c$$

Let,

$$R \neq R[X_1, A_c, \tilde{A}', \tilde{A}_c][X_1 \,\theta\, X_2]\, R[X_2, A'] = R_1$$

A "tuple" in R belongs to R_1 because R_1 is join of projections of R. To prove that a "tuple" in R_1 also belongs to R

Let $(x_1\, a'\, a_c\, \tilde{a}_c\, x_2\, \tilde{a}') \in R_1$

$\therefore \quad (x_1\, a_c\, \tilde{a}_c) \in R[X_1,\ A_c, \tilde{A}_c]$

To prove

$(x_2\, a'\, \tilde{a}') \in R[X_2, A', \tilde{A}']$—then from (14) we have the result.

But,

$$(x_1\, x_2) \in R[X_1, X_2]$$
$$(x_2\, a') \in R[X_2, A']$$
$$(x_1\, \tilde{a}') \in R[X_1, \tilde{A}']$$

From (13) and OMVD5,

$R[X_1, \tilde{A}'][X_1 \,\theta\, X_2]\, R[X_2, A'] =$ Appropriate Projection of R

i.e. An $\tilde{a}'$ which occurs with x_1 also occurs with all a' which occur with x_2 in projections of R whenever $(x_1 \,\theta\, x_2)$ is true.

$$\therefore \quad (x_2\, a'\, \tilde{a}') \in R[X_2, A', \tilde{A}']$$
$$\therefore \quad R = R_1$$

OMVD9

In a relation $R[X_1, X_2, A, B]$ if

$(X_1 \,\theta\, X_2) \rightarrow \rightarrow A \mid B$ then

$(X_1 \,\theta\, X_2) \rightarrow \rightarrow A\, X' \mid B\, X''$ where $X'\, X_1\, X_2$

and $X'' \subseteq X_1 \cup X_2$

Proof

Let

$$R = R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, B]$$
$$\neq R[X_1, A, X'][X_1\,\theta\,X_2]\,R[X_2, B, X''] = R_1$$

Every "tuple" in R also belongs to R_1 because R_1 is a join of projections of R.

Let

$$(x_1\, a_1\, x'\, x_2\, b_1\, x'') \in R_1$$
$$\notin R$$

But

$(x_1\, a_1) \in R[X_1, A]$

$(x_2\, b_1) \in R[X_2, B]$

$\therefore$ $(x_1\, a_1\, x_2\, b_1) \in R$

and

$$(x_1\, a_1\, x_2\, b) = (x_1\, a_1\, x'\, x_2\, b_1\, x'')\text{—a contradiction}$$

Hence the rule is true.

OMVD10

In a relation $R[X_1\; X_2, A, B]$ if

$$(X_1\,\theta\,X_2) \to\; \to A \mid B \text{ then } (X_1\,\theta\,X_2) \to\; \to AB \mid \phi$$

This rule is not true because of the following example:

$$R = x_1\, a_1\, b_1\, x_3 \qquad R[X_1, X_2] = x_1\, x_3 = L$$
$$x_1\, a_1\, b_2\, x_3 \qquad\qquad x_1\, x_2$$
$$= R[X_1, A][X_1\,\theta\,X_2]\,R[X_2, B]$$
$$\neq R[X_1, A, B][X_1\,\theta\,X_2]\,R[X_2]$$

OMVD11

In a relation $R[X_1, X_2, A, B]$ if,

$$(X_1 \,\theta\, X_2) \to\,\to A \mid B \text{ then}$$
$$(X_1 \,\theta\, X_2) \to\,\to A \cup P \mid ((B - P) = B') \text{ where}$$
$$P \subseteq A \cap B$$

Proof

Let

$$\begin{aligned} R &= R[X_1, A][X_1 \,\theta\, X_2]\, R[X_2, B] \\ &\neq R[X_1, A, P][X_1 \,\theta\, X_2]\, R[X_2, B'] = R_1 \end{aligned} \tag{15}$$

If a "tuple" belongs to R, it definitely belongs to R_1 because R_1 is the join of projections of R.

Let $(x_1\, a\, p\, x_2\, b') \in R_1$
$\notin R$
$\therefore \quad (x_1\, a) \in R[X_1, A]$
and $(x_2\, b') \in R[X_2, B']$
But from (15) and OMVD5,
$R = R[X_1, A][X_1 \,\theta\, X_2]\, R[X_2, B']$
because $U = x_1 \cup A \cup x_2 \cup B'$
$\therefore \quad (x_1\, a\, x_2\, b') = (x_1\, a\, p\, x_2\, b') \in R$—a contradiction.
Hence the rule is true.

FD-OMVD1

In a relation $R[X_1, X_2, A, B]$,
if $X_1 \to A$ and $X_2 \to B$ and $(X_1 \,\theta\, X_2)^*$ then

$$(X_1 \,\theta\, X_2) \to\,\to A \mid B$$

Proof

Let $X_1 \to A$, $X_2 \to B$, $(X_1 \,\theta\, X_2)^*$
Assume $(X_1 \,\theta\, X_2) \not\to\not\to A \mid B$.
i.e. $R \neq R[X_1, A][X_1 \,\theta\, X_2]\, R[X_2, B] = R_1$

Any "tuple" in R belongs to R_1 because R_1 is a join of projections of R

Let a "tuple",

$(x_1\, x_2\, a\, b) \in R_1$

$\notin R$

$\therefore \quad (x_1\, a) \in R[X_1, A], (X_2\, b) \in R[X_2, B]$

Also $(X_1\, \theta\, X_2)^*$ is true in R

$\therefore \quad (x_1\, x_2) \in R[X_1, X_2]$

But in R, $X_1 \to A$ and $X_2 \to B$

Hence x_1 always occurs in a "tuple" with a and x_2 always occurs in a "tuple" with b

$\therefore \quad (x_1\, x_2\, a\, b) \in R$—contradiction

Hence the rule is true.

FD-OMVD2

In a relation $R[X_1, X_2, A, B]$ if,

$(X_1\, \theta\, X_2) \to\; \to A \mid B$, $Y \to A'$, $Y \cap A = \phi$

and $A' \subseteq A$ then $X_1 X_2 \to A'$

Proof

Let

$$R = R[X_1,\ A][X_1\, \theta\, X_2]\, R[X_2, B] \tag{16}$$

and $Y \to A'$, $Y \cap A = \phi$ and $A' \subseteq A$

and let

$$X_1 X_2 \nrightarrow A' \qquad \text{Let } A = A'\, A''$$

i.e. $(x_1\, x_2\, a_1'\, a_1''\, b_1)$ and $(x_1\, x_2\, a_2'\, a_2''\, b_2)$ are both in R

i.e. $(x_1\, a_1'\, a_1'')$ and $(x_1\, a_2'\, a_2'') \in R[X_1, A]$

$(x_2\, b_1)$ and $(x_2\, b_2) \in R[X_2, B]$

Hence from (16),

$(x_1\, x_2\, a_1'\, a_1''\, b_2)$ and $(x_1\, x_2\, a_2'\, a_2''\, b_1) \in R$

But $Y \to A'$ and attributes of Y must be from $\bar{A}$, i.e. in the above tuples Y-values must be from $(x_1\, x_2\, b_i)$ where $i = 1, 2$. It is obvious that whatever may be Y from $\bar{A}$, $Y \to A'$ is violated, for example $Y = X_1\, B$.

$(x_1\, b_1\, a_1')$ and $(x_1\, b_1\, a_2')$ in the projections of R will violate $Y \to A'$.

Only if $Y \cap A \neq \phi$ can the FD $Y \rightarrow A'$ be sometimes preserved. For example $Y = A'' B$ and $A'' B \rightarrow A'$ is not violated in this case.
Hence our assumption leads to a contradiction.
Hence the rule is true.

FD-OMVD3

In a relation $R[X_1, X_2, A, B]$ if,
$(X_1 \,\theta\, X_2) \rightarrow \rightarrow A \mid B$ then
$(X_1\, X_2) \rightarrow A \cap B$.

Proof

Let

$$R = R[X_1, A][X_1 \,\theta\, X_2]\, R[X_2, B] \tag{17}$$

and

$$x_1 x_2 \nrightarrow A \cap B$$

i.e. There are two tuples in R
$(x_1\, x_2\, x_a\, x_b)$ and $(x_1\, x_2\, x'_a\, x'_b)$
where $x_a \cap x_b \neq x'_a \cap x'_b$
$\therefore$ By (17) we must have tuples

$$(x_1\, x_2\, x_a\, x'_b) \quad \text{and} \quad (x_1\, x_2\, x'_a\, x_b) \text{ in } R.$$

These tuples cannot exist because

$$(x_a \cap x_b) \neq x'_a \cap x'_b$$

So (17) cannot be true—a contradiction
$\therefore$ $(X_1 X_2) \rightarrow A \cap B$.

MVD-OMVD1

In a relation R
$X \rightarrow \rightarrow A$ if and only if

$$(X = X) \rightarrow \rightarrow A \mid \bar{A}$$

Proof If

Let $X \rightarrow \rightarrow A$
i.e. $R = R[X, A][X = X]\, R[X, (U - X - A)]$
i.e. $R = R[X, A][X = X]\, R[X, \bar{A}] (\because\ X \cup \bar{A} = X \cup (U - X - A))$
i.e. $(X = X) \rightarrow \rightarrow A \mid \bar{A}$
Only if follows similarly.

5. SOME LIMITATIONS

1. Semantics of OMVD may sometimes be impossible to preserve under update operations. We take an earlier example where in $R[A, B, C, D]$, $(A\ \theta\ B) \rightarrow \rightarrow C \mid D$, $A \rightarrow X$ and $B \rightarrow X$ for every column name X in R and A, B, C, D are mutually disjoint. Consider the instance;

$$R = \begin{matrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \end{matrix}$$

   Insert the tuple $a_3 b_3 c_3 d_3$. Now if θ is "=" or "L", MTP/NF is preserved along with the OMVD, $(A\ \theta\ B) \rightarrow \rightarrow C \mid D$. For any other θ (Say ">") the OMVD may be violated because $(A\,\theta B)^*$ is not preserved.

2. A set of complete inference rules for OMVD's may not be possible. We state this conjecture because complementation, which is a very important rule for MVD's(17) [MENDELZON 79b].

6. CONCLUDING REMARKS

In this paper the normalization of a relational data base has been extended to the operation theta-join. A new normal form MTP/NF has been proposed. Relevant examples show the semantic usefulness of MTP/NF and its underlying dependency OMVD. Fifteen inference rules involving OMVD, MVD, and FD have been studied. There appears to be no complete set of inference rules for OMVD. This conjecture is based on the fact that complementation rule is not true for OMVD, while it appears in every complete set of rules for MVD. It should be noted that other dependencies based on theta-join can now be defined and studied. We leave that as a problem for future research.

REFERENCES

1. W. W. Armstrong, "Dependency Structures of Data Base Relationships," *Proc. of IFIP '74*, 580–583 (North-Holland, New York, 1974).
2. S. K. Arora and K. C. Smith, "Well Connected Relations in Dependency Structutes of Data Bases," *Proc. of Conference 1979* CIPS/DPMA/FIQ, 95–101 (1979).
3. S. K. Arora and K. C. Smith, "A Normal Form Based on Theta-Join and Projection for Relational Data Bases," *IEEE COMPCON Fall 1980*, 295–300 (1980).
4. C. Beeri, R. Fagin, and J. H. Howard, "A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations," *ACM SIGMOD*, 47–61 (Toronto, 1977).

5. E. F. Codd, "Further Normalization of the Data Base Relational Model," *Courant Computer Science Symposium: Data Base Systems*, Ed. R. Rustin, **6**:33–64 (Prentice Hall, New York, 1971).
6. E. F. Codd, "Recent Investigations in Relational Data Base Systems," *IFIP 74* (North Holland Pub. Co., New York, 1974).
7. C. Delobel and R. G. Casey, "Decomposition of a Data Base and the Theory of Boolean Switching Functions," *IBM J. of Res. Develop.*, **17** (5):374–386 (1973).
8. C. Delobel, "Normalization and Hierarchical Dependencies in a Relational Data Model," *ACM TODS*, **3** (3):201–222 (1978).
9. R. Fagin, "Multivalued Dependencies and a new Normal Form for Relational Databases," *ACM TODS*, **2** (3):262–278 (1977).
10. R. Fagin, "Functional Dependencies in a Relational Database and Propositional Logic," *IBM J. Res. Develop.*, 534–544 (1977).
11. R. Fagin, "Normal Forms and Relational Operators," *ACM SIGMOD*, 153–160 (Boston, 1979).
12. R. Fagin, "Horn Clauses and Database Dependencies," IBM Res. Rep., Rj 2741, (1980).
13. I. J. Heath, "Unacceptable File Operations in a Relational Data Base," Proc. ACM SIGFIDET Workshop on Data Description, Access and Control, 19–33 (1971).
14. D. Janssens and J. Paredaens, "General Dependencies," (University of Antwerp) *Res. Rep.* 79–35 (1979).
15. Y. E. Lein, "Multivalued Dependencies with Null Values in Relational Databases," *Proc. 5th Int. Conf. on VLDB*, 61–66 (Brazil, 1979).
16. A. O. Mendelzon and D. Maier, "Generalized Mutual Dependencies and the Decomposition of Database Relations," *Proc. 5th Int. Conf. on VLDB*, 75–82 (Brazil, 1979).
17. A. O. Mendelzon, "On Axiomatizing Multivalued Dependencies in Relational Databases," *J. of ACM*, **26** (1):37–44 (1979).
18. J. M. Nicolas, "Mutual Dependencies and Some Results on Undecomposable Relations," *VLDB 78*, 13–15 (Berlin, 1978).
19. J. Paredaens and H. Gallaire, "Transitive Dependencies in a Database Scheme," *R.A.I.R.O. Informatique/Computer Science*, **14** (2):149–163 (1980).
20. J. Rissanen, "Theory of Relations for Dara Bases—A Tutotial Survey," *Proc. 7th Symp. on Math. Foundations of Computer Science*, Lecture Notes in Computer Science, **64**:537–551 (Springer-Verlag, 1978).
21. F. Sadri and J. D. Ullman, "A Complete Axiomatization for a Large class of Dependencies in Relational Databases," *Proc. 12th Annual ACM Symp. on Theory of Computing* (1980).
22. Y. Sagiv and Walecka, "Subset Dependencies as an Alternative to Embedded Multivalued Dependencies," Dept. of Computer Science, University of Illinois at Urbana-Champaign, (UIUCDCS-R-79-980), (UILU-ENG 79 1732), (July 1979).
23. H. A. Schmid and J. R. Swenson, "On the Semantics of the Relational Data Model," *ACM SIGMOD International Conference on Management of Data*, 211–223 (1975).
24. J. M. Smith, "A Normal Form for Abstract Syntax," *Proc. VLDB*, 156–162 (West Berlin, 1978).
25. Y. Vassiliou, "Functional Dependencies and Incomplete Information," *Proc. VLDB*, 260–269 (1980).
26. M. Yannakakis and C. M. Papadimitrou, "Algebraic Dependencies," *M.I.T. Research Report* (1980).