INFORMATION PROCESSING 74 – NORTH-HOLLAND PUBLISHING COMPANY (1974)

DEPENDENCY STRUCTURES OF DATA BASE RELATIONSHIPS*

W. W. ARMSTRONG

Département d'Informatique, Université de Montréal
Montréal, Québec, Canada

An axiomatic description of dependency structures in relationships is presented. The axioms specify those families of dependencies "A determines B" which can hold for some relationship. Families of maximal dependencies, i.e., those where A is a minimal set of attributes determining the given B and where B is the set of all attributes determined by A are also characterized axiomatically. The ∩-semilattice of the B in maximal dependencies is shown to determine the dependency structure completely. Families of subsets of attributes which can be the candidate keys of some relationship are completely characterized. Some insight into recent work on use of boolean functions in data base decomposition is provided.

1. INTRODUCTION

The concept of a relationship has been developed into a powerful means for representing formatted data at a logical level suitable for exchange between information systems and for use in applications programs [1,2,3]. The user of an information management system finds this tool very natural, for he has long been accustomed to the presentation of data in the form of tables. Any relationship, though defined abstractly in terms of sets and functions, can be concretely displayed as a table consisting of lines, each one of which represents an assignment of values to the attributes named in the column headings (e.g. 'NAME','AGE','HEIGHT', etc.). The data base administrator and the system implementor find it easy to interact with such a logically transparent information structure. It leaves them great freedom to choose economical data and storage structures, depending upon the relative frequency of access of various parts of the data and the characteristics of the machines being used.

In this paper, we examine the mathematically possible families of dependencies of the form "A determines B" which can hold in some relationship R. "A determines B" means that, given a table of R and the values assigned to the attributes of A in any line or lines, it is possible to uniquely determine the values of the attributes of B in these lines.

Our hope is that this examination will permit workers in the area of relational models of data to use precise concepts and terminology to define and manipulate dependencies involved when complex relationships are broken down into collections of simpler relationships in various normal forms. Symptoms of a deficiency in this area are not hard to find. E.F. Codd states in [3] that he finds it necessary to give numerous examples to explain and motivate his normal forms and their many subtle ramifications. The fact that it is necessary to use examples of relationships to explain the nature of a restriction on a relationship implies that we need a means of describing the restrictions themselves. This can be accomplished by conceptually separating the relationships existing at a particular moment of time from the possible classes of dependency structures which are imposed upon those relationships by the meanings of the attributes in a given data base. In general, a restriction will require that at least certain dependencies be present in a relationship at any time, although fortuitous or unimportant dependencies may exist too (as for example in a one-line table, wherein given the values of any attributes or none at all, we can determine all attribute values).

After a brief introduction to the relational model of data in section 2, and the definition of dependency structure in section 3, we introduce in section 4 a completely general axiomatic characterization of full families of dependencies, which includes all dependency structures holding for relationships R. In order to justify the axioms by constructing a relationship which actually has precisely the dependencies of any axiomatically defined full family, we must discuss "maximal" dependencies first, in section 5. It is shown in section 6 that the sets B belonging to maximal dependencies are closed under intersection. A set of generators of these B (under intersection) serves to construct the required relationship in section 7. We give a characterization in section 8 of the possible structures for candidate keys of a relationship. In section 9 we apply the results of the present paper to derive new and simpler proofs of some theorems of Casey and Delobel [5]. Section 10 contains some conclusions.

2. THE RELATIONAL MODEL OF DATA

Let α be any finite set. Its elements will be called attributes. (In some other contexts, it would be more appropriate to refer to them as variables.) With each attribute b we associate a set D_b called in data base terminology the domain of b. (This choice of terminology is unfortunate, since D_b will not be used as the domain of a function in the mathematical sense.) We are interested in evaluations of the attributes of α, which are functions

$$f : \alpha \to D ,$$

where $D = \bigcup_{b \in \alpha} D_b$, and where for all $b \in \alpha$ $f(b)$ is an element of D_b. We define a relationship over α to be any finite set of such evaluations f. It is not hard to see how these evaluations can be represented as lines in a table where each column is associated one-to-one with an element of α.

The difference between a relationship and the less general concept of a relation is simply the (implicit) choice of $\alpha = \{1,2,3,\ldots,n\}$ for the latter. That is to say, in the case of a relation, whether the attributes have names or not, there is an ordering of the attributes and their domains: $D_1, D_2, \ldots, D_n$. Then $f \in D_1 \times \ldots \times D_n$ is just a finite

*This paper was prepared with assistance from the National Research Council of Canada, operating grant A5634.

sequence of values. The arbitrary imposition of an ordering on all sets of attributes is an unnecessary requirement from the theoretical point of view. Furthermore, the use of relationships rather than relations is of great practical value in that a certain attribute can be referred to by its name (e.g. "HEIGHT") rather than by its position in an ordering of all attributes (e.g. column 3 in a table). The latter would be difficult for the user to remember if he were dealing with a data base containing several relations with many attributes, and would be subject to change whenever the data base is restructured.

We do not need to discuss partially defined functions f in this paper.

3. DEPENDENCY STRUCTURES

If α is a set of attributes, an ordered pair (A,B) with $A \subseteq \alpha$, $B \subseteq \alpha$ will be referred to as a dependency. $\mathcal{L} = \{(A,B) | A \subseteq \alpha,\ B \subseteq \alpha\}$ is the set of all dependencies over α. If R is a relationship, we say that the dependency (A,B) holds in R iff for all $f_1, f_2 \in R$ $f_1|A = f_2|A$ implies $f_1|B = f_2|B$. (The vertical bar denotes restriction of a function to a subset of its domain. "Domain" here is in the mathematical sense, i.e. a subset of α. We consider two functions to be equal if they have the same domain and have equal values at all points of the domain.) We shall symbolize (A,B) holding in R by $A \to_R B$.

The dependency structure of the relationship R is the family

$$\mathfrak{F}_R = \{(A,B) | A \subseteq \alpha\ ,\ B \subseteq \alpha\ ,\ A \to_R B\} \quad .$$

We shall now proceed to describe these dependency structures axiomatically by introducing the concept of a full family of dependencies over α. The equivalence of these two concepts will be shown in section 7.

4. FULL FAMILIES OF DEPENDENCIES

In order to define certain sets of dependencies, we introduce a partial ordering $\geqslant$ in the set $\mathcal{L}$ of all dependencies over α as follows: for $(A,B) \in \mathcal{L}$ and $(A',B') \in \mathcal{L}$ we define: $(A,B) \geqslant (A',B')$ iff $A \subseteq A'$, $B \supseteq B'$. We might read this as "the dependency (A,B) is at least as informative as the dependency (A',B')".

With this structure $\mathcal{L}$ becomes a lattice [4] wherein the l.u.b. of two elements (A,B) and (A',B') is $(A \cap A', B \cup B')$ and their g.l.b. is $(A \cup A', B \cap B')$. We note this only in passing, since we shall not make use of the lattice operations.

Definition: Let $\mathfrak{F}$ be a subset of $\mathcal{L}$, and let $\to$ be defined by: $A \to B$ iff $(A,B) \in \mathfrak{F}$. Then $\mathfrak{F}$ will be called a full family of dependencies over α if

($\mathfrak{F}$1) $A \to A$,
($\mathfrak{F}$2) if $A \to B$ and $B \to C$ then $A \to C$,
($\mathfrak{F}$3) if $A \to B$ and $(A,B) \geqslant (A',B')$ then $A' \to B'$,
($\mathfrak{F}$4) if $A \to B$ and $A' \to B'$ then $A \cup A' \to B \cup B'$,

where each condition is universally quantified over all $A,B,C,A',B' \subseteq \alpha$.

Theorem 1: The dependency structure of any relationship R is a full family of dependencies.
Proof: We must verify ($\mathfrak{F}$1) to ($\mathfrak{F}$4) when $\to$ is replaced by $\to_R$. This is very easy and is left to the reader.□

The proof of the converse of Theorem 1, which states that any full family of dependencies over α is the dependency structure of some relationship R over α with appropriately chosen domains, is postponed to Theorem 5 of section 7. This will show that ($\mathfrak{F}$1) to ($\mathfrak{F}$4) are sufficient to completely characterize the dependency structures of relationships.

Example: Consider the family, for some fixed $K \subseteq \alpha$, $\mathfrak{F} = \{(A,B) | A \supseteq K \text{ or } A \supseteq B\}$. The meaning of this is that when A contains all the attributes in K, then there is no restriction on B; indeed the attributes in K determine all of the attributes in α; but if A does not contain all attributes of K, then the dependency (A,B) must be trivial, i.e. $A \supseteq B$. Showing that $\mathfrak{F}$ is a full family of dependencies is left to the reader.

The reader who believes at this point that the whole question of dependency structures might really be trivial is invited to construct an R with $\mathfrak{F}_R = \mathfrak{F}$ in the above example. (Two answers are obtainable from section 8 by considering K as the sole candidate key of a dependency structure.)

5. MAXIMAL ELEMENTS OF FULL FAMILIES

In any partially ordered set, the maximal elements are those for which there exists no greater element in the set. In the case of a full family $\mathfrak{F}$ with the partial order $\geqslant$, $(A,B) \in \mathfrak{F}$ is maximal iff for all $(A',B') \in \mathfrak{F}$ $(A',B') \geqslant (A,B)$ only if $A'=A, B'=B$. Intuitively this means that (A,B) is maximal iff A cannot be made smaller for the given B and B cannot be made larger for the given A without going outside the family $\mathfrak{F}$.

Let $\hat{\mathfrak{F}}$ denote the family of all maximal elements of the full family $\mathfrak{F}$. We shall usually use the symbol $\mathfrak{M}$ instead of $\hat{\mathfrak{F}}$ for simplicity. We introduce the convenient notation $A \nearrow B$ to mean $(A,B) \in \mathfrak{M}$.

Theorem 2: The family $\mathfrak{M}$ of maximal elements of a full family of dependencies $\mathfrak{F}$ satisfies:

($\mathfrak{M}$1) for all $A \subseteq \alpha$ there exists $(A',B') \geqslant (A,A)$ such that $A' \nearrow B'$,
($\mathfrak{M}$2) if $A \nearrow B$ and $A' \nearrow B'$ and $(A,B) \geqslant (A',B')$ then $A=A'$ and $B=B'$,
($\mathfrak{M}$3) if $A \nearrow B$ and $A' \nearrow B'$ and $A' \subseteq B$ then $B' \subseteq B$.

Conversely, any subset $\mathfrak{M}$ of $\mathcal{L}$ satisfying these conditions is the set of maximal elements of exactly one full family

$$\mathfrak{F} = \{(A,B) | \exists\, (A',B') \geqslant (A,B) \text{ such that } A' \nearrow B'\}$$

Proof: ($\mathfrak{M}$1) : Starting from $(A,A) \in \mathfrak{F}$ (by $\mathfrak{F}$1), we can climb up ($\geqslant$) in the finite set $\mathfrak{F}$ until we reach a maximal element (A',B') which is at least as informative as (A,A).
($\mathfrak{M}$2) : We can weaken $A \nearrow B$ to $A \to B$ and use the definition of (A',B') being a maximal element of $\mathfrak{F}$.
($\mathfrak{M}$3) : $(A',B') \geqslant (B,B')$ so $B \to B'$ by ($\mathfrak{F}$3). This, together with $A \to B$, gives $A \to B'$ by ($\mathfrak{F}$2). Hence by ($\mathfrak{F}$4) $A \to B \cup B'$. But (A,B) is maximal in $\mathfrak{F}$, so $B' \subseteq B$.

To prove the converse we note first that the only possible full family having $\mathfrak{M}$ as its set of maximal elements is the $\mathfrak{F}$ given in the statement of the theorem by ($\mathfrak{F}$3) and by the fact that in any finite set we can always climb up inside it from any (A,B) to a maximal element.

Furthermore, $\mathfrak{M}$ is the set of maximal elements of the given $\mathfrak{F}$ by ($\mathfrak{M}$2).

All that remains to be shown is that $\mathfrak{F}$ is a full family.
($\mathfrak{F}$1) is clear from ($\mathfrak{M}$1).
($\mathfrak{F}$2) : Let $A \to B$, $B \to C$. Then by the definition of $\mathfrak{F}$ we can find elements of $\mathfrak{M}$ $(A_1,B_1) \geqslant (A,B)$ and $(B_2,C_2) \geqslant (B,C)$ which satisfy $B_1 \supseteq B \supseteq B_2$, allowing the conclusion $C_2 \subseteq B_1$ by ($\mathfrak{M}$3). Hence $(A_1,B_1) \geqslant (A,C)$ and $A \to C$.
($\mathfrak{F}$3) is trivial.
($\mathfrak{F}$4) : It suffices by ($\mathfrak{F}$3) to consider only the case $A \nearrow B$, $A' \nearrow B'$.
Now for $A \cup A' \subseteq \alpha$ we apply ($\mathfrak{M}$1) to get $A'' \nearrow B''$ with $A'' \subseteq A \cup A' \subseteq B''$. Applying ($\mathfrak{M}$3) twice gives

$B \cup B' \subseteq B''$ and hence the result, completing the proof of the theorem.□

The maximal elements of a full family suffice to characterize the family completely. Unfortunately, the conditions ($\mathcal{M}$1) to ($\mathcal{M}$3) are still rather complex and we still cannot get a good picture of what the general full family looks like. The economy of the representation is shown by taking the maximal elements of the example given in the previous section

$$\mathcal{M} = \{(K,\alpha)\} \cup \{(A,A) \mid A \subseteq \alpha,\ A \not\supseteq K\}.$$

We leave it up to the reader to check ($\mathcal{M}$1) to ($\mathcal{M}$3).

6. SATURATED SUBSETS OF ATTRIBUTES

In this section we prove a result which may seem at first somewhat surprising, namely that the family

$$\mathcal{B} = \{B \mid (A,B) \in \mathcal{M}\}$$

is sufficient to characterize $\mathcal{F}$ and $\mathcal{M}$ completely! The elements of $\mathcal{B}$ will be termed saturated subsets according to $\mathcal{F}$ (or $\mathcal{M}$). Intuitively speaking, they are sets of attributes which already contain all attributes which they determine.

Theorem 3: Let $\mathcal{F}$ be a full family of dependencies over α and $\mathcal{M}$ its family of maximal elements. Let $\mathcal{B}$ be defined as above. Then $\mathcal{B}$ is an $\cap$-semilattice containing α [4], i.e. $\mathcal{B}$ satisfies
($\mathcal{B}$1) $\alpha \in \mathcal{B}$,
($\mathcal{B}$2) if $B_1 \in \mathcal{B}$ and $B_2 \in \mathcal{B}$ then $B_1 \cap B_2 \in \mathcal{B}$.
Conversely if $\mathcal{B}$ is any family of subsets of α satisfying ($\mathcal{B}$1) and ($\mathcal{B}$2) then there is exactly one full family $\mathcal{F}$ whose maximal elements $\mathcal{M}$ determine $\mathcal{B}$ as above, and indeed

$$\mathcal{F} = \{(A,B) \text{ for all } B' \in \mathcal{B}\ (A \subseteq B' \text{ implies } B \subseteq B')\}.$$

Proof:
($\mathcal{B}$1) follows from ($\mathcal{M}$1) by taking $A = \alpha$.
($\mathcal{B}$2) : Let $A_1 \nearrow B_1$, $A_2 \nearrow B_2$. By ($\mathcal{M}$1) $\exists$ $A' \nearrow B'$ such that $A' \subseteq B_1 \cap B_2 \subseteq B'$. But $A' \subseteq B_1$ so by ($\mathcal{M}$3) $B' \subseteq B_1$. Similarly $B' \subseteq B_2$, so that $B_1 \cap B_2 = B' \in \mathcal{B}$.

To show the converse, we first leave it up to the reader to show that the $\mathcal{F}$ defined above is indeed a full family. Properties ($\mathcal{B}$1) and ($\mathcal{B}$2) of $\mathcal{B}$ are not even needed for this part!

Any maximal element $(A,B) \in \mathcal{F}$ must be such that if $\mathcal{S} = \{B' \mid A \subseteq B'\}$ then $B = \cap\mathcal{S}$. Hence $B \in \mathcal{B}$ by ($\mathcal{B}$1) and ($\mathcal{B}$2) and the finiteness of $\mathcal{S}$. On the other hand if $B \in \mathcal{B}$ and we construct a maximal dependency of $\mathcal{F}$ $A_1 \nearrow B_1$ with $A_1 \subseteq B \subseteq B_1$ by ($\mathcal{M}$1), then certainly $A_1 \subseteq B$ and hence $B_1 \subseteq B$ by the definition of $\mathcal{F}$. So $B = B_1$ and we see that $\mathcal{F}$ does determine $\mathcal{B}$ via $\mathcal{M}$.

In order to show that $\mathcal{F}$ is the only full family determining the given $\mathcal{B}$ via $\mathcal{M}$, consider any $\mathcal{F}_1$ and its $\mathcal{M}_1$ which also determine $\mathcal{B}$. We show first that $\mathcal{F}_1 \subseteq \mathcal{F}$. For any $B' \in \mathcal{B}$ there is an $(A',B') \in \mathcal{M}_1$; and if $(A,B) \in \mathcal{F}_1$ with $A \subseteq B'$, we have for an $(A'',B'') \in \mathcal{M}_1$ with $(A'',B'') \geq (A,B)$ that $A'' \subseteq B'$. By ($\mathcal{M}$3) $B'' \subseteq B'$, so $B \subseteq B'$. Hence $(A,B) \in \mathcal{F}$ by definition. Finally we show $\mathcal{F} \subseteq \mathcal{F}_1$. For any $(A,B) \in \mathcal{F}$ we have by ($\mathcal{M}$1) some $(A',B') \in \mathcal{M}_1$ with $A' \subseteq A \subseteq B'$. By the definition of $\mathcal{F}$: $B \subseteq B'$. Hence $(A',B') \geq (A,B) \in \mathcal{F}_1$. This completes the proof of the theorem.□
Example: From the example in sections 4 and 5 we get

$$\mathcal{B} = \{\alpha\} \cup \{B \mid B \not\supseteq K\} .$$

This family clearly satisfies ($\mathcal{B}$1) and ($\mathcal{B}$2).

We note that the description of a full family $\mathcal{F}$ by means of $\mathcal{B}$ is more economical than by means of $\mathcal{M}$. But we can simplify the description still further if we consider a family $\mathcal{G}$ of generators of $\mathcal{B}$ under (finite) intersection, i.e. $\mathcal{G} \subseteq \mathcal{B}$ and

$$\mathcal{B} = \{\cap\mathcal{S} \mid \mathcal{S} \subseteq \mathcal{G}\} .$$

Here of course by convention $\cap\phi = \alpha$, so α is never required in $\mathcal{G}$. For any $\mathcal{B}$ a (unique!) smallest $\mathcal{G}$ exists, but $\mathcal{B}$ is completely described by giving any $\mathcal{G}$. It is this description which allows us to see just what full families are possible: we can pick an arbitrary family $\mathcal{G}$ of subsets of α and construct $\mathcal{B}$ and $\mathcal{F}$ from it. This process gives all possible $\mathcal{F}$.

We have the following, now trivial,

Theorem 4: Any family $\mathcal{G}$ of subsets of α gives rise to a full family

$$\mathcal{F} = \{(A,B) \mid \text{ for all } G \in \mathcal{G}\ (A \subseteq G \text{ implies } B \subseteq G)\}$$

such that $\mathcal{G}$ is a set of generators under intersection of its saturated sets $\mathcal{B}$. Any full family of dependencies over α can be obtained in this way.
Proof: $\mathcal{G}$ and $\mathcal{B}$, the family it generates under finite intersection, are easily shown to give rise to the same full family $\mathcal{F}$ (cf. Theorem 3).□

7. JUSTIFICATION OF THE AXIOMS

Theorem 5: Let $\mathcal{F}$ be any full family of dependencies over α. Then there exists a relationship R over α with integer domains such that $\mathcal{F}$ is equal to the dependency structure $\mathcal{F}_R$ of R.
Remark: We need to have the choice of sufficiently large domains to avoid unwanted dependencies.
Proof: Let $\mathcal{G}$ be a family of generators (of minimal cardinality, say) of the $\mathcal{B}$ defined from $\mathcal{F}$ via $\mathcal{M} = \hat{\mathcal{F}}$. Let $p_1, p_2, \ldots, p_k$ be the first k prime numbers where $\mathcal{G} = \{G_1, \ldots, G_k\}$. We consider a table with columns headed by the elements of α. The lines $f_i : \alpha \to N$ of R are defined for $i \in \{p_1^{n_1} \ldots p_k^{n_k} \mid n_j = 0 \text{ or } 1\}$ by $f_i(b) = \prod_{G_j \not\ni b} p_j^{n_j}$.
Intuitively speaking, we have removed from the columns in each G_j the information about the power of the prime p_j present in the prime decomposition of i. Thus the sets G_j are automatically "saturated": every column outside G_j contains information which the attributes of G_j cannot provide. We see that each $B \in \mathcal{B}$ is also "saturated". Hence $(A,B) \in \mathcal{F}_R$ iff the columns B contain no more primes than are present in A, i.e., for all $G_j \in \mathcal{G}$ $(A \subseteq G_j$ implies $B \subseteq G_j)$. Hence $\mathcal{F}_R = \mathcal{F}$.□
We leave the construction of an example to the reader.

This theorem shows that our axioms ($\mathcal{F}$1) to ($\mathcal{F}$4) characterize precisely all possible dependency structures of relationships. Now we can use the words "dependency structure of a relationship" and "full family of dependencies" interchangeably.

8. STRUCTURE OF THE FAMILY OF CANDIDATE KEYS

Definition: If $\mathcal{F}$ is a full family of dependencies, the elements of

$$\mathcal{C} = \{A \mid (A,\alpha) \in \hat{\mathcal{F}}\}$$

are called candidate keys of $\mathcal{F}$.

We know that a candidate key must exist by ($\mathcal{B}$1), and the example of sections 4, 5, and 6 shows that there may be only one candidate key: $\mathcal{C} = \{K\}$. In general there will be many as is shown by the

Theorem 6: The family of candidate keys $\mathcal{C}$ of a dependency structure satisfies

($\mathcal{C}$1) $\mathcal{C} \neq \phi$,

($\mathcal{C}$2) if $K_1, K_2 \in \mathcal{C}$ and $K_1 \subseteq K_2$ then $K_1 = K_2$.

Conversely any $\mathcal{C}$ satisfying ($\mathcal{C}$1) and ($\mathcal{C}$2) is the set of candidate keys of some full family $\mathcal{F}$.

Proof: ($\mathcal{C}$1) follows from ($\mathcal{B}$1), while ($\mathcal{C}$2) follows directly from ($\mathcal{M}$2). For the converse let $\mathcal{C}$ satisfy ($\mathcal{C}$1) and ($\mathcal{C}$2), and let

$$\mathcal{B} = \{B \mid \text{for all } K \in \mathcal{C} \quad B \not\supseteq K\} \cup \{\alpha\}.$$

$\mathcal{B}$ satisfies ($\mathcal{B}$1) and ($\mathcal{B}$2) and so gives rise to an $\mathcal{F}$ by theorem 3. $\mathcal{B}$ is generated by $\mathcal{G} = \mathcal{B} - \{\alpha\}$.

$$\begin{aligned}\mathcal{F} &= \{(A,B) \mid \text{if a set } G \subseteq \alpha \text{ not containing any } K \in \mathcal{C} \text{ contains } A, \text{ then } B \subseteq G\} \\ &= \{(A,B) \mid \text{if } A \text{ does not contain any } K \in \mathcal{C}, \text{ then } B \subseteq A\}.\end{aligned}$$

Hence those maximal $(A,B) \in \mathcal{F}$ for which $B = \alpha$ are the dependencies of the form (K,α) for $K \in \mathcal{C}$.□

Proof II: We give a second proof of the converse via a relationship R.

Let $\{L_1,\ldots,L_k\} = \{L \mid L \cap K \neq \emptyset \text{ for all } K \in \mathcal{C}\}$.

Let $L_0 = \emptyset$. Let $R = \{f_i \mid i = 0,\ldots,k\}$

where $f_i(a) = \begin{cases} i & \text{if } a \in L_i \\ 0 & \text{otherwise.}\end{cases}$

It is now easy to show that $\mathcal{C}$ is the set of all candidate keys of $\mathcal{F}_R$.□

9. APPLICATIONS

The dependency structure axioms ($\mathcal{F}$1) to ($\mathcal{F}$4) are not intended to be an optimal choice, however, they do provide a standard for verifying the correctness and completeness of other axiom systems. Consider, for example, the following one proposed in [5] by Delobel and Casey:

(DC 1) Transitivity: if $E \to F$ and $F \to G$ the $E \to G$.

(DC 2) Reflexivity: $E \to E$.

(DC 3) Projectivity: if $E \subseteq F$ then $F \to E$.

(DC 4) Additivity: if $E \to F$ and $E \to G$ then $E \to F \cup G$.

(DC 5) Pseudotransitivity: if $E \to F$ and $F \cup G \to H$ then $E \cup G \to H$.

(DC 6) Augmentation: if $E \to G$ then $E \cup F \to G$.

It is easy to verify that (DC 1), (DC 3), and (DC 4) together are equivalent to our axioms for full families of dependencies. The same is true of (DC 2), (DC 5), and (DC 6).

We shall now introduce boolean functions as in [5], but we shall use an interpretation of them which shows that their connection to dependency structures reaches deeper than the level of mere symbol manipulation.

Let α be a finite set of attributes. A boolean evaluation of these attributes is a function $e : \alpha \to \{0,1\}$. The set of all such e will be denoted by $\{0,1\}^\alpha$. Every e can be defined in terms of the set $A = \{b \mid b \in \alpha, e(b) = 1\}$, and $e = I_A$ is called the indicator function of A.

We consider boolean functions f of the variables α, that is, $f : \{0,1\}^\alpha \to \{0,1\}$. For each $b \in \alpha$ we define the boolean function f_b by: $f_b(I_A) = I_A(b)$ for all $A \subseteq \alpha$. As usual, we just write b instead of f_b. If $X \subseteq \alpha$, we write ΠX for the logical product of the boolean functions b for $b \in X$. We have: $\Pi X(I_A) = 1$ iff $X \subseteq A$.

If (X_i, Y_i) $i=1,\ldots,n$ is a set of dependencies over α, they define the boolean function

$$f = \sum_{i=1}^{n} \Pi X_i (\Pi Y_i)' \ .$$

Theorem 7: If f is defined by a set of dependencies then the smallest full family $\mathcal{F}$ containing those dependencies satisfies

$$X \to Y \text{ iff } \Pi X (\Pi Y)' \leqslant f \ .$$

Furthermore, $f(I_B) = 0$ iff B is saturated according to $\mathcal{F}$.

Proof: $f(I_B) = 0$ iff $X_i \subseteq B$ implies $Y_i \subseteq B$ for all $i=1,\ldots,n$. The set $\mathcal{B}$ of all such B satisfies ($\mathcal{B}$1) and ($\mathcal{B}$2), and so defines by theorem 3 a full family $\mathcal{F}$ whereby $(X,Y) \in \mathcal{F}$ iff for all $B \in \mathcal{B}$: $X \subseteq B$ implies $Y \subseteq B$. Clearly $(X_i,Y_i) \in \mathcal{F}$ for all i. If $\mathcal{F}'$ is any full family containing the (X_i,Y_i), it corresponds to a $\mathcal{B}'$ where for all $B' \in \mathcal{B}'$ and $i=1,\ldots,n$: $X_i \subseteq B'$ implies $Y_i \subseteq B'$. Hence $\mathcal{B}' \subseteq \mathcal{B}$ and so by theorem 3 $\mathcal{F}' \supseteq \mathcal{F}$. The condition $\Pi X(\Pi Y)' \leqslant f$ is equivalent to : for all $B \in \mathcal{B}$ ($X \subseteq B$ implies $Y \subseteq B$), i.e., $(X,Y) \in \mathcal{F}$.□

We immediately obtain the result of Appendix A of [5]:

Corollary A: If f is defined by a set of dependencies, then any representation of f as

$$f = \sum_{i=1}^{k} \Pi A_i (\Pi B_i)'$$

yields a set of dependencies (A_i,B_i) generating the same dependency structure.

We also have the result of Appendix B of [5]:

Corollary B: K is a candidate key of the dependency structure defining f iff ΠK is a prime implicant of $f + \Pi\alpha$ having no complemented variables.

Proof: $\Pi K(\Pi\alpha)' \leqslant f$ iff $\Pi K \leqslant f + \Pi\alpha$.□

10. CONCLUSIONS

The axiomatic formulation of full families of dependencies in section 4 is appropriate for the treatment of the most general type of dependency structure which can arise in a relational model of data. Although space does not permit giving details, it is clearly necessary to define normal forms not in terms of relationships R, but in terms of dependency structures $\mathcal{F}$ which are imposed *a priori* on ("time-varying") relationships R by the meanings of the attributes. At any time we must have $\mathcal{F}_R \supseteq \mathcal{F}$. The dependencies prescribed by $\mathcal{F}$ must be present in the actual data, perhaps along with other fortuitous dependencies which R is too small to exclude.

A precise analysis of the restrictions thus imposed on the "time-varying" relationships R should yield a better understanding of the decomposition of R into normal forms.

ACKNOWLEDGMENT

The author wishes to thank Gregor Bochmann, Vaclav Chvatal, and Neil Stewart for many helpful suggestions.

REFERENCES

[1] E.F. Codd, A Relational Model of Data for Large Shared Data Banks, Communications of the ACM, vol. 13, no. 6, June 1970, 377-387.

[2] E.F. Codd, A.L. Dean (eds.), 1971 SIGFIDET Workshop on Data Description, Access, and Control, Association for Computing Machinery, New York, 1971.

[3] E.F. Codd, Further Normalization of the Data Base Relational Model, Courant Inst. Comp. Sci. Symp. 6, Data Base Systems, Prentice-Hall, Englewood Cliffs, N.J., 1971, 33-64.

[4] G. Birkhoff, Lattice Theory, 3rd. ed., American Mathematical Society Colloquium Publ. XXV, Providence, R.I., 1967.

[5] C. Delobel, R.G. Casey, Decomposition of a data base and the theory of boolean switching functions, IBM Journal of Research and Development, vol. 17, no. 5, Sept. 1973, 374-386.