Characterizations of Multivalued Dependencies and Related Expressions
(Extended Abstract)

José Luis Balcázar and Jaume Baixeries
Dept. Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
c/ Jordi Girona, 1-3
08034 Barcelona
{balqui,jbaixer}@lsi.upc.es

Abstract. We study multivalued dependencies, as well as the propositional formulas whose deduction calculus parallels that of multivalued dependencies, and the variant known as degenerated multivalued dependen-
cies. For each of these sorts of expressions, we provide intrinsic character-
izations in purely semantic terms. They naturally generalize similar
properties of functional dependencies or Horn clauses.

1 Introduction

Multivalued dependencies (MVD) are a natural generalization of functional de-
pendencies, and an important notion in the design of relational databases. The
presence of functional dependencies that do not result from keys indicates the
possibility of decomposing a relation with no information loss. Multivalued de-
pendencies precisely characterize the relations in which a lossless-join decompo-
sition can be performed.

Early works on multivalued dependencies focused on providing a consistent
and complete calculus for entailment between dependencies ([6], [7], [8], [19],
[20]). In fact, the very same set of deduction rules is consistent and complete
both for logical entailment between Horn clauses and for semantic entailment
between functional dependencies. One way of explaining the connection is the
“comparison-based binarization” of a given relation $r$: a relation derived from $r$,
formed by binary tuples, each obtained from a pair of original tuples from $r$ by
attribute-wise comparison. Then it is easy to see that a functional dependency
holds in $r$ if and only if the comparison-based binarization, seen as a theory (i.e.
a set of propositional models), satisfies the corresponding Horn clause.

Similarly, in [16] a family of propositional formulas is identified, for which a
consistent and complete calculus not only exists but, additionally, corresponds to
a syntactically identical consistent and complete calculus for multivalued depen-
dencies. However, the connection is much less clear than simply considering the
comparison-based binarization; and, in particular, the database expressions that
most naturally correspond to the propositional formulas under a comparison-based binarization are so-called “degenerate multivalued dependencies”.

We prove here semantic, intrinsic characterizations of multivalued dependencies, degenerate multivalued dependencies, and multivalued dependency formulas. Our statements present alternative properties that hold for a relation $r$, or for a propositional theory $T$, exactly when a given multivalued dependency, respectively degenerate multivalued dependency, holds for $r$, or when a given multivalued dependency formula holds in $T$.

2 Multivalued Dependencies

Our definitions and notations from relational database theory are fully standard. We denote $R = \{A_1, \ldots, A_n\}$ the set of attributes, each with a domain $\text{Dom}(A_i)$; then a tuple $t$ is a mapping from $R$ into the union of the domains, such that $t[A_i] \in \text{Dom}(A_i)$ for all $i$. Alternatively, tuples can be seen as well as elements of $\text{Dom}(A_1) \times \cdots \times \text{Dom}(A_n)$. A relation $r$ over $R$ is a set of tuples. We will use capital letters from the end of the alphabet for sets of attributes, and do not distinguish single attributes from singleton sets of attributes. We denote by $XY$ the union of the sets of attributes $X$ and $Y$. Our binarization process is standard:

**Definition 1.** For tuples $t, t'$ of a relation $r$, $\text{ag}(t, t')$ (read: “agree”) is the set $X$ of attributes on which $t$ and $t'$ have coinciding values: $A \in \text{ag}(t, t') \iff t[A] = t'[A]$.

The following is the standard definition of multivalued dependency. Let $X, Y, \text{ and } Z$ be disjoint sets of attributes whose union is $R$. For tuples $t$ and $t'$, denote them as $xyz$ and $x'y'z'$ meaning that $t[X] = x, t[Y] = y$, etc.

**Definition 2.** A multivalued dependency $X \rightarrow Y | Z$ holds in $r$ if and only if for each two tuples $xyz$ and $x'y'z'$ in $r$, also $x'y'$ appears.

Our characterization mainly rests on the following definitions:

**Definition 3.** For a set of attributes $X$ of $R$, $\tau_r(X)$ (but when $r$ is clear from the context we drop the subscript) is the set of all pairs of tuples from $r$ whose agree set is $X$:

$$\tau_r(X) = \{(t, t') | \text{ag}(t, t') = X\}$$

**Definition 4.** For sets $T, T'$ of pairs of tuples, we denote $T \bowtie T'$ the set

$$T \bowtie T' = \{(t, t') | \exists t'' ((t, t'') \in T \wedge (t'', t') \in T') \wedge \exists t''' ((t', t''') \in T \wedge (t, t''') \in T')\}$$

Our main result about mutivalued dependencies is as follows:

**Theorem 1.** Let $X, Y, Z$ be pairwise disjoint sets of attributes of $R$, such that their union $XYZ$ includes all the attributes. Then the multivalued dependency $X \rightarrow Y | Z$ holds in $r$ if and only if, for each $X' \supseteq X$, $\tau(X') = \tau(X'Y') \bowtie \tau(X'Z')$, where $Y' = Y \setminus X'$ and likewise $Z' = Z \setminus X'$. 