On the Role of the Complementation Rule for Data Dependencies over Incomplete Relations

Flavio Ferrarotti\textsuperscript{1}, Sven Hartmann\textsuperscript{2}, and Sebastian Link\textsuperscript{1}

\textsuperscript{1} School of Information Management, Victoria University of Wellington, New Zealand
\textsuperscript{2} Institut für Informatik, Technische Universität Clausthal, Germany

Abstract. Recently, an axiomatization for functional dependencies (FDs) and multivalued dependencies (MVDs) has been established where arbitrary attributes can be specified as NOT NULL. That is, the information stored over such attributes must not be incomplete. The axiomatization subsumes previous axiomatizations of FDs and MVDs where every attribute is declared to be NOT NULL, and where no attribute is declared to be NOT NULL. We establish axiomatizations which underpin formally the intuition that the complementation rule is a mere means of database normalization. The results unburden the existing theory of the strong assumption that all attributes are known at the time when the dependencies are specified. The findings extend and unify previous results for the special cases above.

1 Introduction

A database system manages a collection of persistent information in a shared, reliable, effective and efficient way. Most commercial database systems are still founded on the relational model of data \cite{10}. Data administrators utilize various classes of data dependencies to restrict the relations in the database to those considered meaningful to the application at hand. According to \cite{12} functional dependencies (FDs) capture around two-thirds, and multivalued dependencies (MVDs) around one-quarter of all uni-relational dependencies (those defined over a single relation schema) that arise in practice. In particular, MVDs are frequently exhibited in database applications \cite{37}, e.g. after denormalization or in views \cite{1}. While research on this topic has been extensive, only very recently a theory has been established that can reason about FDs and MVDs exhibited by relations that satisfy arbitrary NOT NULL constraints \cite{21}.

Example 1. Consider a table SUPPLIES with column headers $A$(rticle), $S$(upplier), $L$(ocation) and $C$(ost). The table collects information about suppliers that deliver articles from a location at a certain cost.

\texttt{CREATE TABLE Supplies (Article CHAR[20],
Supplier VARCHAR NOT NULL,
Location VARCHAR NOT NULL,
Cost CHAR[8]);}
On the Role of the Complementation Rule for Data Dependencies 137

Suppose the database management system enforces the following constraints:
The FD $A \rightarrow S$ says that for every article there is a most one supplier, the FD $AL \rightarrow C$ says that the costs are determined by the article and the location, and the MVD $S \rightarrow AC$ says that the supplier determines the article and cost pairs independently of the location. Do the following meaningful constraints also need to be enforced explicitly, or are they already enforced implicitly: i) the MVD $A \rightarrow L$ and ii) the FD $A \rightarrow C$?

Indeed, the declaration of Supplier and Location as NOT NULL guarantees that both $A \rightarrow L$ and $A \rightarrow C$ are implied by $A \rightarrow S$; $AL \rightarrow C$ and $S \rightarrow AC$. However, reasoning about FDs and MVDs in the presence of an arbitrary null-free subschema (NFS), i.e. the set of attributes declared NOT NULL, is subtle. For example, if $S$ is not declared NOT NULL, then neither the FD nor the MVD is implied. Consequently, the opportunity to specify an arbitrary NFS provides the data administrator with a flexible mechanism to control the expressiveness of the consequence relation. Dedicated tools for reasoning about FDs and MVDs in the presence of arbitrary NFSs have been established [21]. The set

$$\mathcal{D} = \{R_F, D_F, U_F, U_M, T_M, C^R_M, I_{FM}, T_{FM}\}$$

of inference rules from Table I is a finite axiomatization [21].

**Example 2.** Let $R = ASLC$, $R_s = SL$, $\Sigma = \{A \rightarrow S; AL \rightarrow C; S \rightarrow AC\}$ as in Example 1. The inference

\[
\begin{array}{ccc}
I_{FM} : & A \rightarrow S & S \rightarrow AC \\
T_M : & A \rightarrow L & C^R_M : S \rightarrow L \\
T_{FM} : & A \rightarrow A & R_F : A \rightarrow A \\
U_M : & A \rightarrow AL & T_{FM} : A \rightarrow A \\
& A \rightarrow C & AL \rightarrow C
\end{array}
\]

shows that $A \rightarrow L$ and $A \rightarrow C$ can be inferred from $\Sigma$ by $\mathcal{D}$. Since $\mathcal{D}$ is sound, in particular, it follows that both dependencies are implied by $\Sigma$.

The inference in Example 2 can be criticized in two different aspects. When inferring MVDs, then applications of the $R$-complementation rule $C^R_M$ should be restricted to the very last step of the inference (if necessary at all). The ability to have inferences with this property for all implied MVDs would establish an axiomatization that appropriately reflects the database normalization process. Moreover, for every implied FD there should be an inference with no applications of the $R$-complementation rule $C^R_M$ at all. The desirability of these two features has already been motivated and axiomatizations with these features have been established for the special cases where every attributes is NOT NULL [8,9,29] and where every attribute is NULL [30]. In this paper, we will show that the axiomatization $\mathcal{D}$ has neither of these features. Subsequently, we will establish a finite axiomatization with both features. The results provide a unifying framework for all previous findings on these issues.