HORIZONTAL DECOMPOSITIONS BASED ON FUNCTIONAL-DEPENDENCY-SET-IMPLICATIONS

P. De Bra*

Department of Mathematics
University of Antwerp, U.I.A.
Universiteitsplein 1, 2610 Antwerp
Belgium

*Research assistant of the N.F.W.O.

Abstract

A new approach towards horizontal decompositions in the Relational Database Model is given. It is based on partial implications between sets of *functional dependencies.* This horizontal decomposition theory is especially useful for databases which must represent "real world" situations, in which there always are exceptions to rather severe constraints like *functional dependencies (fd's).*

The *functional-dependency-set-implication (fsi)* generalizes all previous work on horizontal decompositions using partial implications between (single) fd's.

The exceptions to a set of fd's are formalized using another new constraint, the *anti-functional-dependency-set (afs).* The membership problem is solved for mixed fsi's and afs', and a complete set of inference rules is given. The inheritance problem, i.e. which dependencies hold in the (two) subrelations generated by the horizontal decomposition) is shown to be solvable in polynomial time.

§1 Introduction

The *vertical decomposition* of relations into projections of these relations, based on *functional dependencies (fd's)*, was introduced with the *Relational Database Model* by Codd [Co], and has been exhaustively studied and generalized since. However, it relies on the assumption that the part of the real world, represented by the database, satisfies some rather severe constraints.

Because this assumption is highly unrealistic some mechanism for handling exceptions to these constraints is necessary. In [De1, De2, Pa, De3, De4, De5] a theory has been established that uses a *horizontal decomposition* of relations into restrictions (often called selections) of these relations, to put the "exceptions" in a well defined subrelation. This theory is based on "partial implications" between functional dependencies, given different names in each paper, as the class of these constraints became bigger and bigger.

In this paper we generalize the constraints of [De5] to include implications between sets of fd's instead of single fd's. The new constraint is called a *functional-dependency-set-implication (fsi).* It means that if a (part of a) relation satisfies a set of fd's, then it must also satisfy some other set of fd's, which is said to be implied by the first set. Note that this is an ad-hoc implication, based on the observation of the real world, not just a "logical" deduction (which has been studied exhaustively already [Ul]).

The exceptions to a set of fd's are formalized using another new constraint: the ***anti-functional-dependency-set (afs)***, a generalization of the ***anti-functional dependency (afd)*** of [De5].

In Section 2 we define the horizontal decomposition, based on fsi's. We also recall two theoretical tools, that are used throughout the horizontal decomposition theory: the ***Armstrong relation*** and the ***conflict concept***. In Section 3 the membership problem is solved for mixed fsi's and afs' and a complete set of inference rules is given. A complicated construction of a relation instance is recalled from [De5], which also leads to the solution of the inheritance problem in Section 4. Both the membership and the inheritance problem are shown to be solvable in polynomial time.

We suppose the reader is aware of the basic definitions and notations of the Relational Database theory [Ul].

§2 Horizontal Decompositions

The traditional vertical decomposition, based on ***functional dependencies (fd's)***, can only be applied to relations in which some fd's hold. Since fd's are rather restrictive they do not occur frequently in the "real world", at least not if no exceptions to any fd can be tolerated. For this reason a large number of weaker constraints have been defined in literature [Ul], which have a better chance of being satisfied, and which still lead to vertical decompositions. However, these constraints are less natural and do not all have the theoretical simplicity of fd's.

In [De1, De2, Pa] a method for handling exceptions to fd's is presented, using horizontal decompositions. The exceptions to an fd are put in a separate subrelation (obtained by taking a restriction of the relation), inducing the fd in the remaining (and main) part of the relation. Because the fd holds in this main part, it can be used to apply the classical vertical decomposition to this part.

Although an fd may not hold in the real world, the part that satisfies the fd may satisfy some other fd's too, which cannot be logically deduced from the first fd. Such "partial implications" between fd's have been studied in [De3, De4, De5], and have been shown to lead to the same horizontal decompositions as those of [De1, De2, Pa]. The drawback of all these horizontal decompositions is that they generate an awful lot of subrelations, exponential in the number of constraints.

In this paper we consider a more general class of constraints, using implications between sets of fd's. This enables the database designer to combine several implications between fd's into one implication between sets of fd's, reducing the number of constraints, and hence the number of generated subrelations.

We first illustrate the horizontal decomposition with the following example:

Example 2.1. Consider a large company with several divisions (e.g. factories) each having (several) departments (each) treating one or more jobs. Employees work in one or more departments (of one or more divisions). They have salaries and managers.

Although this may seem a rather unconstrained database, it may obviously obey the following constraint:

If in a *div*ision every *dep*artment treats only one *job*, every *emp*loyee has only one *job* and every *man*ager supervises only one *job* (for this division), then (the division is so large that) every *emp*loyee works in only one *dep*artment and has only one *sal*ary, and every *man*ager supervises (employees) in only one *dep*artment (for that division).

This constraint will be written as:

$$\{div, dep \rightarrow job; div, emp \rightarrow job; div, man \rightarrow job\} \overset{div}{\supset\!\!-} \{div, emp \rightarrow dep, sal; div, man \rightarrow dep\}$$

Note that none of the fd's of the second (or "implied") set are logical consequences of the first set of fd's. they are said to be implied by the first set of fd's by observing the real world.

□

We now define this constraint, and the horizontal decomposition induced by it, more formally.

Definition 2.2. Let X be a set of attributes.

A set of tuples S in a relation instance is called *X-complete* iff the tuples not belonging to S all have other X-projections than those belonging to S. Formally, if $t_1 \in S$, $t_2 \notin S$ then $t_1[X] \neq t_2[X]$.

A set of tuples S is called *X-unique* iff all the tuples of S have the same X-projection. Formally, if $t_1, t_2 \in S$ then $t_1[X] = t_2[X]$.

Definition 2.3. Let Z be a set of attributes, $\mathcal{F}_1$ and $\mathcal{F}_2$ be sets of fd's, such that $\forall X \rightarrow Y \in \mathcal{F}_1 \cup \mathcal{F}_2 : Z \subseteq X$.

The *functional-dependency-set-implication (fsi)* $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$, means that in every Z-complete set of tuples (in every instance) in which all the fd's of $\mathcal{F}_1$ hold, all the fd's $\mathcal{F}_2$ must hold too.

The sets of tuples which are both X-complete and X-unique play an important role in the horizontal decomposition theory. In the sequel we shall use the term *"X-value"* to refer to such a set of tuples, as well as for the X-projection of the tuples of that set.

The requirement that all "left hand sides" of the fd's of $\mathcal{F}_1$ and $\mathcal{F}_2$ must include Z is not a severe restriction. In Section 5 we shall show that eliminating this restriction does not necessarily lead to a bigger class of constraints.

The *functional-dependency-implications (fdi's)* of [De5] are special fsi's where $\mathcal{F}_1$ and $\mathcal{F}_2$ each contain only one fd. Since all previously defined constraints of [De1, De2, Pa, De3, De4] are special fdi's they are fsi's too.

In particular fd's can be expressed in many ways as fsi's, some of which are fdi's. $X \rightarrow Y$ is equivalent to $\{X \rightarrow X\} \overset{X}{\supset\!\!-} \{X \rightarrow Y\}$ for instance, which is an fdi (in fact even a "cfd" of [De3]). But $X \rightarrow Y$ is also equivalent to $\emptyset \overset{X}{\supset\!\!-} \{X \rightarrow Y\}$ for instance, which is not an fdi.

Example 2.2. Consider the relation of Example 2.1. The horizontal decomposition separates the divisions in which every employee, every department and every manager have only one job, from the other divisions. If one assumes that most large divisions have enough work to distribute the jobs in this way, a major part of the database may consist of information about such large divisions, hence the exceptions to these fd's only have a minor influence on the cost of solving queries or making updates. However, if no horizontal decomposition was applied to the database, the few exceptions would have a great influence on the efficiency of the database system, since they would prevent the classical vertical decomposition that speeds up the system and reduces redundancy.

The "user" need not know about this horizontal decomposition. If an update (in a "large" division) causes one of the fd's ($div, dep \rightarrow job$; $div, emp \rightarrow job$ or $div, man \rightarrow job$) to be violated, all tuples with that *div*-value are moved to the subrelation for the exceptions. If an update causes the three fd's to become satisfied the tuples with that *div*-value have to move to the other subrelation automatically. Hence the user need not know about the horizontal decomposition. However, it may also be useful to let some users access the subrelations and some other users only the "union". By doing this one can easily allow some users to create or remove exceptions, while preventing other users from doing so.

□

The restriction operator for separating the "large" divisions from the "small" ones is defined as follows:

Definition 2.4. Let $\mathcal{R}$ be a relation scheme, Z be a set of attributes, $\mathcal{F}$ a set of fd's such that $\forall X \rightarrow Y \in \mathcal{F} : Z \subseteq X$.

For every instance R of $\mathcal{R}$, the *restriction for $\mathcal{F}_Z$ of R*, $\sigma_{\mathcal{F}_Z}(R)$, is the largest Z-complete subset (of tuples) of R in which all fd's of $\mathcal{F}$ hold.

The *restriction for $\mathcal{F}_Z$ of $\mathcal{R}$*, $\sigma_{\mathcal{F}_Z}(\mathcal{R})$, is a scheme $\mathcal{R}_1$, (with the same attributes as $\mathcal{R}$,) of which the instances are exactly the restrictions for $\mathcal{F}_Z$ of the instances of $\mathcal{R}$.

We require $Z \subseteq X$ for all $X \rightarrow Y \in \mathcal{F}$ to make sure that X-values of R are not split up by taking a restriction for $\mathcal{F}_Z$.

Definition 2.5. The *horizontal decomposition of a scheme $\mathcal{R}$, according to the fsi $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$*, is the couple $(\mathcal{R}_1, \mathcal{R}_2)$, where $\mathcal{R}_1 = \sigma_{\mathcal{F}_Z}(\mathcal{R})$ and $\mathcal{R}_2 = \mathcal{R} - \mathcal{R}_1$.

Note from Definition 2.5 that the horizontal decomposition of a scheme, according to $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ does not depend on $\mathcal{F}_2$, but it induces the $\mathcal{F}_2$ in $\mathcal{R}_1$. Hence one can always perform a horizontal decomposition to generate a subrelation with a "desirable" set of fd's $\mathcal{F}_1$, by using the "trivial" fsi $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \emptyset$.

In $\mathcal{R}_2$, which contains the exceptions, for every X-value at least one of the fd's of $\mathcal{F}_1$ must not hold. In Example 2.1 this means that in such a division at least one department of employee or manager must have more than one job. (In these divisions nothing is known about the number of departments an employee or a manager works for, nor about an employee's salary).

The following constraint formalizes the notion of "exception".

Definition 2.6. Let $\mathcal{F}$ be a set of fd's, such that $\forall X \rightarrow Y \in \mathcal{F} : Z \subseteq X$.

The *anti-functional dependency set (afs)* $\not{\mathcal{F}}_Z$ means that in every nonempty Z-complete set of tuples, in every instance, at least one fd of $\mathcal{F}$ does not hold.

The *restriction for $\not{\mathcal{F}}_Z$ of R*, $\sigma_{\not{\mathcal{F}}_Z}(R)$, is the largest Z-complete set of tuples in which $\not{\mathcal{F}}_Z$ holds.

The *restriction for $\not{\mathcal{F}}_Z$ of a scheme $\mathcal{R}$*, $\sigma_{\not{\mathcal{F}}_Z}(\mathcal{R})$ is the scheme of which the instances are the restrictions for $\not{\mathcal{F}}_Z$ of the instances of $\mathcal{R}$.

One can easily see that $\mathcal{R} - \sigma_{\mathcal{F}_Z}(\mathcal{R}) = \sigma_{\not{\mathcal{F}}_Z}(\mathcal{R})$, hence the horizontal decomposition of $\mathcal{R}$ according to $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ is the couple of schemes $(\sigma_{\mathcal{F}_Z}(\mathcal{R}), \sigma_{\not{\mathcal{F}}_Z}(\mathcal{R}))$.

The *anti-functional dependency (afd)* introduced in [De5] is an afs $\not{\mathcal{F}}_Z$ for which $\mathcal{F}$ contains only one fd.

From now on we let a relation scheme $\mathcal{R}$ have a set $\mathcal{I}$ of fsi's and a set $\mathcal{A}$ of afs'. (Note that $\mathcal{I}$ also contains the fd's).

Having more than one fsi means that after the (horizontal) decomposition according to one fsi of $\mathcal{I}$ one may want to decompose the (two) subrelations again, using some other fsi of $\mathcal{I}$. Therefore one must (first) determine which fsi's hold in these subrelations. This is described in Section 4. Also, since decomposing a relation according to an fsi creates two subrelations which may sometimes both be decomposed further on, one obtains an exponential number of final subrelations. Therefore the designer must choose his fsi's carefully, putting many fd's in one fsi, to reduce the number of subrelations. This is the main advantage of fsi's over fdi's [De5], in which all the fd's are treated separately, leading to more fdi's, and hence exponentially more subrelations.

The presence of both fsi's and afs' in a relation scheme may induce a situation of "internal conflict" between the fsi's and the afs'. The easiest example of conflict is an fd $X \rightarrow Y$, which is a special fsi, and the afd $X \not\xrightarrow{X} Y$, which is a special afs.

Definition 2.7. A set $\mathcal{I} \cup \mathcal{A}$ of fsi's ($\mathcal{I}$) and afs' ($\mathcal{A}$) is *in conflict* iff the empty set of tuples is the only instance in which all dependencies of $\mathcal{I} \cup \mathcal{A}$ hold.

In Section 3 the membership problem for mixed fsi's and afs' is reduced to the conflict concept, which itself is reduced to the membership problem for fd's. So the conflict concept is an important theoretical tool.

In the proofs of Sections 3 and 4 a special instance is used, which is an *Armstrong relation for fd's* [Ar, De1]. It has a special property, also satisfied by the "direct product construction" of [Fa], but not by every (so called) Armstrong relation for fd's:

Theorem 2.8. *Let $Arm(\mathcal{F})$ denote the Armstrong relation for a set $\mathcal{F}$ of fd's [Ar, De1]. In $Arm(\mathcal{F})$ every fd, which is a consequence of $\mathcal{F}$, holds, and for every other fd $X \rightarrow Y$, the "corresponding" afd $X \not\xrightarrow{X} Y$ holds (which is an afs).*

□

§3 The Membership Problem for fsi's and afs'

In this section we reduce the membership problem for fsi's and afs' to a sequence of membership tests for fd's, for which many solutions are well known [Be, Ber].

We use the symbol $\models$ to denote the (logical) implication of a dependency by a set of dependencies, and the symbol $\vdash$ to denote the deduction of a dependency from a set of dependencies using the inference rules, given below. We shall prove the equivalence of $\models$ and $\vdash$, i.e. the completeness of the inference rules.

We denote the set of all the fd's which are consequences of a set $\mathcal{F}$ of fd's by $\mathcal{F}^*$. The set of all fd's $X \rightarrow Y$ of $\mathcal{F}^*$ for which $Z \subseteq X$ is denoted by $\mathcal{F}^{*Z}$.

$(F1)$: if $Y \subseteq X$ then $X \rightarrow Y$.
$(F2)$: if $X \rightarrow Y$ and $V \subseteq W$ then $XW \rightarrow YV$.
$(F3)$: if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$.
$(FS1)$: if $\mathcal{F}_2 \subseteq \mathcal{F}_1^{*Z}$ and $\forall X \rightarrow Y \in \mathcal{F}_1 : Z \subseteq X$ then $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$.
$(FS2)$: if $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ and $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_3$ then $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2 \cup \mathcal{F}_3$.
$(FS3)$: if $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ and $\mathcal{F}_2 \overset{Z}{\supset\!\!-} \mathcal{F}_3$ then $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_3$.
$(FS4)$: if $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ and $Z \rightarrow Z'$ then $\mathcal{F}_1^{*Z'} \overset{Z'}{\supset\!\!-} \mathcal{F}_2^{*Z'}$.

$(FS5)$: if $\mathcal{F}_1$ holds and $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ then $\mathcal{F}_2$ holds and if $\mathcal{F}_2$ holds and $\forall X \to Y \in \mathcal{F}_1 \cup \mathcal{F}_2 : Z \subseteq X$ then $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ holds.

$(FA1)$: if $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ and $\not\mathcal{F}_{2Z}$ then $\not\mathcal{F}_{1Z}$.

$(FA2)$: if $\not\mathcal{F}_{1Z}$ and $\forall X \to Y \in \mathcal{F}_2 : Z \subseteq X$ then $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$.

$(FA3)$: if $\not\mathcal{F}_Z$ and $Z \to Z'$ then $\not\mathcal{F}^{*Z'}_{Z'}$.

$(FA4)$: if $\not\mathcal{F}_{1Z}$ and $\mathcal{F}_1^{*Z} \subseteq \mathcal{F}_2^{*Z}$ and $\forall X \to Y \in \mathcal{F}_2 : Z \subseteq X$ then $\not\mathcal{F}_{2Z}$.

As fd's are special fsi's the use of fd's in these rules is allowed. In fact $FS5$ shows all representations of fd's as fsi's.

Theorem 3.1. *The rules $F1 \dots F3$, $FS1 \dots FS5$, $FA1 \dots FA4$ are sound.*

Proof. This is very similar to the proof for fdi's and afd's [De5]. We give the proof for $FA3$ as an example.

One can easily see that the fd $Z \to Z'$ means that every Z'-complete set of tuples is also Z-complete.

Suppose $\not\mathcal{F}^{*Z'}_{Z'}$ does not hold. Hence in some Z'-complete set of tuples S all fd's of $\mathcal{F}^{*Z'}$ hold. Since S is also Z-complete it remains to prove that all fd's of $\mathcal{F}$ hold in S.

Let $X \to Y \in \mathcal{F}$, then $XZ' \to Y \in \mathcal{F}^{*Z'}$. $Z \to Z'$ and $Z \subseteq X$ induce $X \to XZ'$ by augmentation $(F2)$. By transitivity $(F3)$ we infer that $X \to Y$ holds in S.

□

The proof of the completeness of the inference rules consists of the following steps: first we prove that $F1 \dots F3$, $FS1 \dots FS5$ are complete for the deduction of fd's from a set of fsi's. Then we show that they are also complete for the deduction of fsi's. From this proof one can easily derive a membership algorithm for fsi's, which is essentially a sequence of fd-membership tests. Finally we prove that $F1 \dots F3$, $FS1 \dots FS5$, $FA1 \dots FA4$ are complete for mixed fsi's and afs' by reducing this problem to the membership problem for fsi's only.

Throughout the proofs of this and the next section, we use the following set of fd's:

Definition 3.2. $FSAT_{\mathcal{I}}(\mathcal{F})$ is the smallest possible set of fd's, such that:

1. $\mathcal{F} \subseteq FSAT_{\mathcal{I}}(\mathcal{F})$.
2. If $\mathcal{F}_{i_1} \subseteq FSAT_{\mathcal{I}}(\mathcal{F})$ and $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!\!-} \mathcal{F}_{i_2} \in \mathcal{I}$ then $\mathcal{F}_{i_2} \subseteq FSAT(\mathcal{F})$.
3. If $FSAT_{\mathcal{I}}(\mathcal{F}) = (FSAT_{\mathcal{I}}(\mathcal{F}))^*$.

□

$FSAT_{\mathcal{I}}(\mathcal{F})$ can be constructed starting from $\mathcal{F}$ and by repeatedly trying to satisfy 2) and 3) of the definition. However, the construction will never be useful in a membership algorithm since taking the closure (step 3) of a set of fd's is very costly. Fortunately one can quite easily construct an efficient algorithm for verifying whether an fd is in $FSAT_{\mathcal{I}}(\mathcal{F})$. A similar algorithm is given for cfd's in the extended version of [De3].

Note that $FSAT_{\mathcal{I}}(\mathcal{F}) = FSAT_{\mathcal{I} \cup \mathcal{F}}(\emptyset)$. This equality will be used several times without further notice.

Lemma 3.3. $FSAT_{\mathcal{I}}(\mathcal{F}) = \{P \to Q : \mathcal{I} \cup \mathcal{F} \models P \to Q\}$.

Proof. Consider $Arm(FSAT_{\mathcal{I}}(\mathcal{F}))$. By Definition 3.2 and Theorem 2.8 it is clear that $\mathcal{I} \cup \mathcal{F}$ holds in $Arm(FSAT_{\mathcal{I}}(\mathcal{F}))$. Hence all the fd-consequences of $\mathcal{I} \cup \mathcal{F}$ also hold. By Theorem 2.8 this implies that all these fd's are in $(FSAT_I(\mathcal{F}))^*$. Step 3 of Definition 3.2 implies that these fd's are in $FSAT_I(\mathcal{F})$.

The opposite inclusion is obvious from Definition 3.2.

□

The above lemma shows how to detect whether an fd is a consequence of a set of fsi's.

Lemma 3.4. $FSAT_I(\mathcal{F}) = \{P \rightarrow Q : I \cup \mathcal{F} \vdash P \rightarrow Q\}$.

Proof. From Theorem 3.1 and Lemma 3.3 we know that: $\{P \rightarrow Q : I \cup \mathcal{F} \vdash P \rightarrow Q\} \subseteq FSAT_I(\mathcal{F})$.

For the opposite inclusion we show that the property, that all elements of $FSAT_I(\mathcal{F})$ can be deduced from $I \cup \mathcal{F}$, remains valid throughout the construction of $FSAT_I(\mathcal{F})$.

- If $P \rightarrow Q \in \mathcal{F}$ then the property is trivial.
- If $P \rightarrow Q$ is added to $FSAT_I(X,Y)$ in step 2 of Definition 3.2 then $P \rightarrow Q \in \mathcal{F}_{i_2}$ for some $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!-} \mathcal{F}_{i_2} \in I$. By induction all fd's of $\mathcal{F}_{i_1}$ can be inferred from $I \cup \mathcal{F}$. Hence by rule $FS5$ $I \cup \mathcal{F} \vdash \mathcal{F}_{i_2}$. Rule $FS5$ applied to $\mathcal{F}_{i_2}$ and $\mathcal{F}_{i_2} \overset{Z_i}{\supset\!-} \{P \rightarrow Q\}$ (holding by $FS1$) gives $I \cup \mathcal{F} \vdash P \rightarrow Q$.
- If $P \rightarrow Q$ is added in step 3 of Definition 3.2 then it can be deduced from fd's for which the property holds, by using $F1 \ldots F3$, which are the classical inference rules for fd's [Ul]. Hence $I \cup \mathcal{F} \vdash P \rightarrow Q$.

□

If one chooses $\mathcal{F} = \emptyset$ then the following result becomes obvious:

Corollary 3.5. $F1 \ldots F3$, $FS1 \ldots FS5$ *are complete for the inference of fd's from a set of fsi's.*

□

In the construction of $FSAT_I(\emptyset)$ only those fsi's $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!-} \mathcal{F}_{i_2}$ of I are used (in step 2) for which $I \models \mathcal{F}_{i_1}$ (and hence also $I \models \mathcal{F}_{i_2}$).

This leads to the following lemma:

Lemma 3.6. *Let* $I_Z = \{\mathcal{F}_{i_1} \overset{Z_i}{\supset\!-} \mathcal{F}_{i_2} \in I : I \models Z_i \rightarrow Z \text{ or } I \models \mathcal{F}_{i_1}\}$.
$I \models P \rightarrow Q$ *iff* $I_Z \models P \rightarrow Q$ *(for any* Z*).*

□

In the sequel we will also need the following remark, which can be easily deduced from the inference rules for fd's:

Remark 3.7 *If* $Z \rightarrow Z'$ *and* $\forall X \rightarrow Y \in \mathcal{F} : Z \subseteq X$ *then* $\mathcal{F}$ *and* $\mathcal{F}^{*Z'}$ *are equivalent (i. e.* $\mathcal{F}^* = (\mathcal{F}^{*Z'})^*$*). In general however, (if* $Z \not\rightarrow Z'$*)* $\mathcal{F}$ *is more powerful than* $\mathcal{F}^{*Z'}$.

□

The following lemma shows an important property of $FSAT_{I_Z}(\mathcal{F})$.

Lemma 3.8. *Let* I_Z *be as in Lemma 3.6. Let* $\mathcal{F}$ *be such that* $\forall X \rightarrow Y \in \mathcal{F} : Z \subseteq X$.
If $P \rightarrow Q \in FSAT_{I_Z}(\mathcal{F})$ *then* $I \models P \rightarrow Q$ *or* $I \models P \rightarrow Z$.

Proof. We prove that the property remains valid throughout the construction of $FSAT_{I_Z}(\mathcal{F})$.

- If $P \rightarrow Q \in \mathcal{F}$ then $P \rightarrow Z$ is trivial.
- If $P \rightarrow Q$ is added in step 2 of Definition 3.2 then $P \rightarrow Q \in \mathcal{F}_{i_2}$ for some $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!-} \mathcal{F}_{i_2} \in I_Z$. There are two possibilities (by the definition of I_Z): $I \models \mathcal{F}_{i_1}$ or $I \models Z_i \rightarrow Z$.
  - If $I \models \mathcal{F}_{i_1}$ then $I \models \mathcal{F}_2$ by $FS5$, hence obviously $I \models P \rightarrow Q \in \mathcal{F}_{i_2}$.

- If $I \models Z_i \rightarrow Z$ then $I \models P \rightarrow Z$ by augmentation (since $Z_i \subseteq P$ if $P \rightarrow Q \in \mathcal{F}_{i_2}$).

- If $P \rightarrow Q$ is added in step 3 then it is derived from other fd's (already in $FSAT_{I_Z}(\mathcal{F})$) by reflexivity, augmentation or transitivity [Ul].
  - If $Q \subseteq P$ then $P \rightarrow Q$ is trivial.
  - If $P = P'P''$, $Q = Q'Q''$, with $P' \rightarrow Q'$ already in $FSAT_{I_Z}(\mathcal{F})$ and $Q'' \subseteq P''$, then $P \rightarrow Q$ or $P \rightarrow Z$ is deduced from $P' \rightarrow Q'$ or $P' \rightarrow Z$ by augmentation.
  - If $P \rightarrow O$ and $O \rightarrow Q$ already are in $FSAT_{I_Z}(\mathcal{F})$ then $P \rightarrow Q$ or $P \rightarrow Z$ is deduced from $P \rightarrow O$ or $P \rightarrow Z$ and $O \rightarrow Q$ or $O \rightarrow Z$ by transitivity.

□

The following lemma partially solves the membership problem for fsi's:

Lemma 3.9. *Let I_Z be as in Lemma 3.6. Let $\mathcal{F}_1$ be such that $\forall X \rightarrow Y \in \mathcal{F}_1 : Z \subseteq X$, $\mathcal{F}_2$ such that $\forall X \rightarrow Y \in \mathcal{F}_2 : Z' \subseteq X$, and let $I \models Z' \rightarrow Z$.*

Then $I_Z \models \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2^{*Z}$ iff $\mathcal{F}_2 \subseteq FSAT_{I_Z}(\mathcal{F}_1)$.

Proof. If $I_Z \models \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2^{*Z}$ then obviously $\mathcal{F}_2^{*Z} \subseteq FSAT_{I_Z}(\mathcal{F}_1)$. Since $I \models Z' \rightarrow Z$ $\mathcal{F}_2^{*Z}$ is equivalent to $\mathcal{F}_2$ by Remark 3.7. Hence also $\mathcal{F}_2 \subseteq FSAT_{I_Z}(\mathcal{F}_1)$ by step 3 of Definition 3.2.

For the converse we proceed as in Lemma's 3.4 and 3.8, by proving that the property remains valid throughout the construction of $FSAT_{I_Z}(\mathcal{F}_1)$.

- If $\mathcal{F}_2 = \mathcal{F}_1$ then rule $FS1$ gives $\mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2^{*Z}$.
- If $\mathcal{F}_2$ is added to $FSAT_{I_Z}(\mathcal{F})$ in step 2 of Definition 3.2 then $\mathcal{F}_2 = \mathcal{F}_{i_2}$ for some $\mathcal{F}_{i_1} \overset{Z}{\supset\!-} \mathcal{F}_{i_2} \in I_Z$. There are 2 possibilities (by the definition of I_Z):
  - If $I \models \mathcal{F}_{i_1}$ then $I_Z \models \mathcal{F}_{i_1}$ by Lemma 3.6. Hence $I_Z \models \mathcal{F}_{i_2}$ by rule $FS5$. Hence also $I_Z \models \mathcal{F}_{i_2}^{*Z}$ since $\mathcal{F}_{i_2}$ induces $\mathcal{F}_{i_2}^{*Z}$. $\mathcal{F}_1 \overset{Z}{\supset\!-} \emptyset$ (holding by $FS1$) and $\emptyset \overset{Z}{\supset\!-} \mathcal{F}_{i_2}^{*Z}$ (a representation for fd's, by $FS5$) induce $\mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_{i_2}^{*Z} = \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2^{*Z}$ by rule $FS3$.
  - If $I \models Z_i \rightarrow Z$ then we have that $\mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_{i_1}^{*Z}$ by induction. Since $Z_i \rightarrow Z$, $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!-} \mathcal{F}_{i_2}$ induces $\mathcal{F}_{i_1}^{*Z} \overset{Z}{\supset\!-} \mathcal{F}_{i_2}^{*Z}$ by $FS4$. Hence $\mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_{i_2}^{*Z} = \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2^{*Z}$.
- If $\mathcal{F}_2$ is added to $FSAT_{I_Z}(\mathcal{F}_1)$ in step 3 of Definition 3.2 then $\mathcal{F}_2 \subseteq \mathcal{F}^*$ for some $\mathcal{F}$ that was a part of $FSAT_{I_Z}(\mathcal{F}_1)$ already.
  From Lemma 3.8 we know that for all $X \rightarrow Y \in \mathcal{F} : I_Z \vdash X \rightarrow Y$ or $I_Z \vdash X \rightarrow Z$.
  - If $I_Z \vdash X \rightarrow Y$ then $I_Z \vdash XZ \rightarrow Y$ (by $F2$), hence $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \{XZ \rightarrow Y\}$ by $FS3$ on $\mathcal{F}_1 \overset{Z}{\supset\!-} \emptyset$ and $\emptyset \overset{Z}{\supset\!-} XZ \rightarrow Y$ ($FS5$).
  - If $I_Z \vdash X \rightarrow Z$ then $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \{XZ \rightarrow Y\}$ holds by induction (since $XZ \rightarrow Y \in \{X \rightarrow Y\}^{*Z}$).

  Let $\mathcal{F}' = \{XZ \rightarrow Y \in \mathcal{F} : I_Z \vdash X \rightarrow Z \text{ or } I_Z \vdash X \rightarrow Y\}$, then by $FS2$ $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}'$. One can easily see that $\mathcal{F}^{*Z} = \mathcal{F}'^{*Z}$ (using $F1 \ldots F3$), hence $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}^{*Z}$ by $FS3$ on $\mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}'$ and $\mathcal{F}' \overset{Z}{\supset\!-} \mathcal{F}'^{*Z} = \mathcal{F}^{*Z}$ ($FS1$). Hence by $FS3$ and $FS1$ one infers $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2$.

□

From the proof of Lemma 3.9 one can see that:

Corollary 3.10. *Let I_Z be as in Lemma 3.6, let $\mathcal{F}_1$ and $\mathcal{F}_2$ be such that $\forall X \rightarrow Y \in \mathcal{F}_1 \cup \mathcal{F}_2$, $Z \subseteq X$, then $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!-} \mathcal{F}_2$ iff $\mathcal{F}_2 \subseteq FSAT_{I_Z}(\mathcal{F}_1)$.*

□

To complete the proof of the completeness of $F1 \ldots F3$, $FS1 \ldots FS5$ for fsi's it remains to show that the fsi's of $I - I_Z$ have no influence on $I \models \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$. To prove this we need a complicated construction of an instance, similar to that of [De5], which we shall also be needing to prove the completeness for mixed fsi's and afs'. Therefore we include the properties of this instance, related to afs', in the following lemma:

Lemma 3.11. *Let $I \cup \mathcal{A}$ be not in conflict. Let I_Z be as in Lemma 3.6. Let $\mathcal{A}_Z = \{\mathfrak{F}_{Z_i} \in \mathcal{A} : I \models Z_i \rightarrow Z\}$. Let $I_Z \cup \mathcal{A}_Z \not\models \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ (or let $I_Z \cup \mathcal{A}_Z \not\models \mathfrak{F}_{1Z}$).*

Then we can construct an instance in which $I \cup \mathcal{A}$ holds but in which $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ (resp. $\mathfrak{F}_{1Z}$) does not hold.

Proof. Suppose $I_Z \cup \mathcal{A}_Z \not\models \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$. We shall see later that in $R_1 = Arm(FSAT_{I_Z}(\mathcal{F}_1))$ $\mathcal{A}_Z$ holds. By Lemma 3.9 $\mathcal{F}_2 \not\subset FSAT_{I_Z}(\mathcal{F}_1)$, hence $\mathcal{F}_2$ does not hold in R_1. From Theorem 2.8 we can easily deduce that this means that for some $X \rightarrow Y \in \mathcal{F}_2$ $X \overset{X}{\not\rightarrow} Y$ holds in R_1, hence $X \overset{Z}{\not\rightarrow} Y$ holds (by $F1$ and $FA3$). We also know (from rule $FA2$) that $\mathfrak{F}_{1Z}$ cannot hold in R_1.

In R_1 a number of fsi's of $I - I_Z$ and a number of afs' of $\mathcal{A} - \mathcal{A}_Z$ may not hold. This will be solved by "adding" copies of $S = Arm(FSAT_I(\emptyset))$. In S $I \cup \mathcal{A}$ holds, as one can easily see.

Let some $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!\!-} \mathcal{F}_{i_2} \in I - I_Z$ not hold in R_1. Then (Theorem 2.8) all fd's of $\mathcal{F}_{i_1}$ hold and some fd's of $\mathcal{F}_{i_2}$ do not hold. Since $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!\!-} \mathcal{F}_{i_2} \notin I_Z$ for some fd $T \rightarrow U \in \mathcal{F}_{i_1} : I \not\models T \rightarrow U$.

Let the values that occur in S be renamed such that they all become different from the values of R_1, except that for some $t_1 \in R_1$, $t_2 \in S : t_1[\overline{Z_i}] = t_2[\overline{Z_i}]$, where $\overline{Z_i} = \{attribute\ A : I \models Z_i \rightarrow A\}$. The "modified" union of R_1 and S satisfies the following properties:

- In $R_1 \cup S$ I_Z still holds: let $\mathcal{F}_{j_1} \overset{Z_j}{\supset\!\!-} \mathcal{F}_{j_2} \in I_Z$ not hold. Then there are two cases:
  - either $I \models \mathcal{F}_{j_1}$, hence $I \models \mathcal{F}_{j_2}$ by $FS5$, and if $\mathcal{F}_{i_2}$ no longer holds then $\exists s_1 \in R_1, \exists s_2 \in S, \exists T_j \rightarrow U_j \in \mathcal{F}_{j_2}$: $s_1[T_j] = s_2[T_j]$ and $s_1[U_j] \neq s_2[U_j]$, and hence $T_j \subseteq \overline{Z_i}$ and $s_1[T_j] = s_2[t_j] = t_1[T_j]$, but then also $U_j \subseteq \overline{Z_i}$ since $I \models T_j \rightarrow U_j$, hence $s_1[U_j] = s_2[U_j] = t_1[U_j]$, a contradiction.
  - or $I \models Z_i \rightarrow Z$, but then $T_j \subseteq \overline{Z_i}$ (which holds for some $T_j \rightarrow U_j \in \mathcal{F}_{j_2}$ that does not hold in $R_1 \cup S$) and $Z_j \rightarrow Z$ would induce $Z \subseteq \overline{Z_i}$, a contradiction with $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!\!-} \mathcal{F}_{i_2} \notin I_Z$.
- In $R_1 \cup S$ $\mathcal{A}_Z$ still holds since it is impossible to violate an afs by taking a union of two instances in which that afs holds.
- In $R_1 \cup S$ every fsi of $I - I_Z$ and every afs of $\mathcal{A} - \mathcal{A}_Z$ which already holds in R_1 (and also in S of course) still holds. For the afs' the reason is the same as for those of $\mathcal{A}_Z$. For the fsi's we have that if $\mathcal{F}_{k_1} \overset{Z_k}{\supset\!\!-} \mathcal{F}_{k_2} \in I - I_Z$ holds in R_1 then $\mathfrak{F}_{k_1 Z_k}$ holds in R_1 and S, and such an afs is not violated by the union. $FA2$ shows that this afs implies $\mathcal{F}_{k_1} \overset{Z_k}{\supset\!\!-} \mathcal{F}_{k_2}$.
- In $R_1 \cup S$ $\mathfrak{F}_{1Z}$ still does not hold, since it does not hold in R_1 and since (as explained above) R_1 and R_2 do not "share" a common Z-value which could influence $\mathfrak{F}_{1Z}$ in R_1 (otherwise $Z \subseteq \overline{Z_i}$).
- In $R_1 \cup S$ $\mathfrak{F}_{2Z}$ may no longer hold, because it may not hold in S. But in the "R_1-part" of $R_1 \cup S$ $\mathcal{F}_2^Z$ still holds, since R_1 and S do not share a Z-value (and hence also no T-value for any $T \rightarrow U \in \mathcal{F}_2$).
- But in $R_1 \cup S$ the number of Z_i-values for which all $X_{i_n} \rightarrow Y_{i_n} \in \mathcal{F}_1$ hold (and for which some $T_{i_n} \rightarrow U_{i_n} \in \mathcal{F}_2$ does not hold) is decreased by one, since the Z_i-value containing t_1 collapses with the Z_i-value of S, containing t_2, (in which $\mathfrak{F}_{1Z_i}$ holds), and since S has no Z_i-values for which all $X_{i_n} \rightarrow Y_{i_n} \in \mathcal{F}_{i_1}$ hold.

By repeating the above construction for all other Z_i-values for which $\mathcal{F}_{i_1}$ holds one can generate a relation in which $\mathcal{F}_{i_1} \overset{Z_i}{\supset\!\!-} \mathcal{F}_{i_2}$ holds (since $\not{\mathcal{F}}_{i_1 Z_i}$ holds).

By then repeating the above construction for all fsi's of $I - I_Z$ which do not hold in $Arm(FSAT_{I_Z}(\mathcal{F}_1))$ one generates a relation in which I holds (and $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ still does not hold).

For the afs' of $\mathcal{A} - \mathcal{A}_Z$ the construction proceeds in a similar way.

If $I_Z \cup \mathcal{A}_Z \not\models \not{\mathcal{F}}_Z$ then we shall see later that in $Arm(FSAT_{I_Z}(\mathcal{F}))$ $\mathcal{A}_Z$ holds, as well as $I_Z \cup \{\mathcal{F}\}$.

The same construction as above leads to an instance in which $I \cup \mathcal{A}$ holds, and in which $\not{\mathcal{F}}_Z$ does not hold.

□

Theorem 3.12. *$F1 \ldots F3$, $FS1 \ldots FS5$ are complete for fdi's.*

Furthermore, let $\mathcal{F}_1$ and $\mathcal{F}_2$ be such that $\forall X \to Y \in \mathcal{F}_2 \cup \mathcal{F}_2 : Z \subseteq X$, then $I \vdash \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ iff $\mathcal{F}_2 \subseteq FSAT_{I_Z}(\mathcal{F}_1)$.

Proof. If $I \models \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ then by Lemma 3.11 $I_Z \models \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$. Lemma 3.9 yields $\mathcal{F}_2 \in FSAT_{I_Z}(\mathcal{F}_1)$, while Corollary 3.10 implies $I_Z \vdash \mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$.

The converse is trivial.

□

Before we prove the completeness of our rules for mixed fsi's and afs', we take a closer look at the conflict concept.

Lemma 3.13. *$I \cup \mathcal{A}$ is in conflict iff for some $\not{\mathcal{F}}_Z \in \mathcal{A}$ $I \models \mathcal{F}$ holds.*

Proof. The if-part is trivial.

For the only-if-part consider $Arm(FSAT_I(\emptyset))$. In $Arm(FSAT_I(\emptyset))$ I holds, hence if $I \cup \mathcal{A}$ is in conflict some $\not{\mathcal{F}}_Z \in \mathcal{A}$ does not hold. If for some fd $X \to Y \in \mathcal{F}$ $X \to Y$ does not hold in $Arm(FSAT_I(\emptyset))$ then by Theorem 2.8 $X \overset{X}{\not\rightarrow} Y$ must hold. With rules $FS1, FA3$ and $FA4$ one can easily deduce that $X \overset{X}{\not\rightarrow} Y$ induces $\not{\mathcal{F}}_Z$. Hence for no $X \to Y \in \mathcal{F}$ $X \overset{X}{\not\rightarrow} Y$ can hold. Theorem 2.8 then induces that for all $X \to Y \in \mathcal{F}$ $X \to Y$ holds, hence $\mathcal{F}$ holds and must be a consequence of $FSAT_I(\emptyset)$. Step 3 of Definition 3.2 implies that $\mathcal{F} \subseteq FSAT_I(\emptyset, \emptyset)$, hence $I \models \mathcal{F}$ by Lemma 3.3.

□

Lemma 3.14. *Let $I \cup \mathcal{A}$ be not in conflict.*

$I \cup \mathcal{A} \models \not{\mathcal{F}}_Z$ iff $I_Z \cup \mathcal{A}_Z \vdash \not{\mathcal{F}}_Z$.

Proof. The if-part is trivial.

For the only-if-part, suppose $I_Z \cup \mathcal{A}_Z \nvdash \not{\mathcal{F}}_Z$. We first show that $I_Z \cup \mathcal{A}_Z \not\models \not{\mathcal{F}}_Z$.

Suppose $I_Z \cup \mathcal{A}_Z \models \not{\mathcal{F}}_Z$, hence $I_Z \cup \mathcal{A}_Z \cup \mathcal{F}$ is clearly in conflict. Then in $Arm(FSAT_{I_Z}(\mathcal{F}))$ for some afs $\not{\mathcal{F}}'_V$ of $\mathcal{A}_Z$ $\mathcal{F}'$ holds (by Lemma 3.13 and Theorem 2.8). Hence (Theorem 2.8) $I_Z \cup \mathcal{F} \models \mathcal{F}'$. Since $I_Z \models V \to Z$ Lemma 3.9 yields $I_Z \models \mathcal{F} \overset{Z}{\supset\!\!-} \mathcal{F}'^{*Z}$.

$\not{\mathcal{F}}'_V$ induces $\not{\mathcal{F}}'^{*Z}_Z$ by rule $FA3$. Rule $FA1$ (with $\mathcal{F} \overset{Z}{\supset\!\!-} \mathcal{F}'^{*Z}$ induces $\not{\mathcal{F}}_Z$, a contradiction with $I_Z \cup \mathcal{A}_Z \nvdash \not{\mathcal{F}}_Z$.

Hence $I_Z \cup \mathcal{A}_Z \not\models \not{\mathcal{F}}_Z$ and $\mathcal{A}_Z$ holds in $Arm(FSAT_{I_Z}(\mathcal{F}))$ (this was used in Lemma 3.11, considering that $FA2$ deduces $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ from $\not{\mathcal{F}}_{1Z}$. Lemma 3.11 then says that there exists an instance in which $I \cup \mathcal{A}$ holds and in which $\not{\mathcal{F}}_Z$ does not hold. Hence $I \cup \mathcal{A} \not\models \not{\mathcal{F}}_Z$.

□

Lemma 3.15 *Let $I \cup A$ be not in conflict.*
$I \cup A \models F_1 \overset{Z}{\supset\!\!-} F_2$ *iff* $I_Z \vdash F_1 \overset{Z}{\supset\!\!-} F_2$ *or* $I_Z \cup A_Z \vdash \mathfrak{X}_{1Z}$.

Proof. The if part is trivial (using $FA2$ if $I_Z \cup A_Z \vdash \mathfrak{X}_{1Z}$).

For the only-if-part, assume that $I_Z \nvdash F_1 \overset{Z}{\supset\!\!-} F_2$ and $I_Z \cup A_Z \nvdash \mathfrak{X}_{1Z}$. By the proof of Lemma 3.14 $I_Z \cup A_Z \not\models \mathfrak{X}_{1Z}$. Hence by Lemma 3.11 there exists an instance in which $I \cup A$ holds and in which $F_1 \overset{Z}{\supset\!\!-} F_2$ does not hold. Hence $I \cup A \not\models F_1 \overset{Z}{\supset\!\!-} F_2$.

□

The above lemmas prove:

Theorem 3.16 *$F1 \ldots F3, FS1 \ldots FS5$ and $FA1 \ldots FA4$ are complete for mixed fsi's and afs'.*

□

The following property shows how to solve the membership problem (and is easy to prove):

Corollary 3.17 *Let $I \cup A$ be not in conflict.*
Let $I_Z = \{F_{i_1} \overset{Z_i}{\supset\!\!-} F_{i_2} \in I : F_{i_1} \subseteq FSAT_I(\emptyset)$ *or* $Z_i \rightarrow Z \in FSAT_I(\emptyset)\}$.
Let $A_Z = \{\mathfrak{X}'_{V_i} \in A : V_i \rightarrow Z \in FSAT_I(\emptyset)\}$.
$I \cup A \models \mathfrak{X}_Z$ *iff for some afs* $\mathfrak{X}'_{V_i}$ *of* A_Z *we have* $F' \subseteq FSAT_{I_Z}(F)$.
$I \cup A \models F_1 \overset{Z}{\supset\!\!-} F_2$ *iff* $I \cup A \models \mathfrak{X}_{1Z}$ *or* $F_2 \in FSAT_{I_Z}(F_1)$.

□

Using Corollary 3.17 and a membership test for *FSAT*, a polynomial time membership algorithm for mixed fsi's and afs' can be easily constructed. The major factor in the time complexity is the calculation of I_Z and A_Z.

§4 The Inheritance of Dependencies

In Example 2.1 the relation scheme contains only one fsi. In general however a scheme may have several fsi's and afs'. So after decomposing it according to one of its fsi's, we want to decompose the subschemes further on, using other fsi's. But not all fsi's (and afs') still hold in the subschemes. In this section we describe how to decide which dependencies hold in the subschemes. This is called the *inheritance problem.*

Notation 4.1. In the sequel we always treat the horizontal decomposition of a scheme R, with fsi's I and afs' A, according to $F_1 \overset{Z}{\supset\!\!-} F_2 \in I$, into the subschemes $R_1 = \sigma_{F_{1Z}}(R)$ with fsi's I_1 and afs' A_1, and $R_2 = \sigma_{\mathfrak{X}_{1Z}}(R)$ with fsi's I_2 and afs' A_2. We assume that $I \cup A$ is not in conflict, and also that $I \cup A \not\models F_1$ and $I \cup A \not\models \mathfrak{X}_{1Z}$ (otherwise R_1 resp. R_2 would always be empty). We only require $I_1 \cup A_1$ and $I_2 \cup A_2$ to be generating for the sets of all dependencies which hold in R_1 and R_2.

Remark 4.2. *Since fd's cannot be violated by taking a restriction, all the fd's which hold in R also hold in R_1 and R_2.*

□

The following inclusions are easy to prove:

Lemma 4.3.

$I_1 \subseteq \{\mathcal{F}_1' \overset{Z'}{\supset\!\!-} \mathcal{F}_2' : I \cup \mathcal{A} \cup \mathcal{F}_1 \models \mathcal{F}_1' \overset{Z'}{\supset\!\!-} \mathcal{F}_2'\}$.

$I_2 \subseteq \{\mathcal{F}_1' \overset{Z'}{\supset\!\!-} \mathcal{F}_2' : I \cup \mathcal{A} \cup \mathfrak{F}_{1Z} \models \mathcal{F}_1' \overset{Z'}{\supset\!\!-} \mathcal{F}_2'\}$.

$\mathcal{A}_1 \subseteq \{\mathfrak{F}_{Z'}' : I \cup \mathcal{A} \cup \mathcal{F}_1 \models \mathfrak{F}_{Z'}'\}$.

$\mathcal{A}_2 \subseteq \{\mathfrak{F}_{Z'}' : I \cup \mathcal{A} \cup \mathfrak{F}_{1Z} \models \mathfrak{F}_{Z'}'\}$.

Theorem 4.4. *An fsi or afs must hold in $\mathcal{R}_1$ (resp. $\mathcal{R}_2$) iff it is a consequence of $I_Z \cup \mathcal{A}_Z \cup \mathcal{F}_1$ (resp. $I_Z \cup \mathcal{A}_Z \cup \mathfrak{F}_{1Z}$).*

Proof. From Lemma 4.3 it follows that $(I_Z \cup \mathcal{A}_Z \cup \mathcal{F}_1)^* \subseteq (I_1 \cup \mathcal{A}_1)^* \subseteq (I \cup \mathcal{A} \cup \mathcal{F}_1)^*$ and also that $(I_Z \cup \mathcal{A}_Z \cup \mathfrak{F}_{1Z})^* \subseteq (I_2 \cup \mathcal{A}_2)^* \subseteq (I \cup \mathcal{A} \cup \mathfrak{F}_{1Z})^*$, where * means the "closure" operator, i. e. taking all the consequences of a set of dependencies (as we already did for fd's in Section 3). We prove that the first inclusions are equalities.

Let $I \cup \mathcal{A} \cup \mathcal{F}_1 \models \mathfrak{F}_{Z'}'$ but $I_Z \cup \mathcal{A}_Z \cup \mathcal{F}_1 \not\models \mathfrak{F}_{Z'}'$. We prove that $\mathfrak{F}_{Z'}'$ does not hold in $\mathcal{R}_1$.

$I_Z \cup \mathcal{A}_Z \cup \mathcal{F}_1 \not\models \mathfrak{F}_{Z'}'$ implies that $I_Z \cup \mathcal{A}_Z \cup \mathcal{F}_1 \cup \mathcal{F}_1'$ is not in conflict, by the proof of Lemma 3.14. By the proof of Lemma 3.11 one can construct an instance R in which $I \cup \mathcal{A}$ holds but in which $\mathfrak{F}_{Z'}'$ does not hold. One starts with $R_1 = Arm(FSAT_{I_Z \cup \mathcal{F}_1}(\mathcal{F}'))$ this time, and adds copies of $Arm(FSAT_I(\emptyset, \emptyset))$ to obtain R. From the construction, used in the proof of Lemma 3.11 one can easily see that $R_1 = \sigma_{\mathcal{F}_{1Z}}(R)$, since the copies of $Arm(FSAT_I(\emptyset, \emptyset))$ have $\mathfrak{F}_{1Z}$ and have different Z-values than those occurring in R_1. Hence we obtain an instance in which $I \cup \mathcal{A}$ holds, and such that $\mathfrak{F}_{Z'}'$ does not hold in R_1. Hence $\mathfrak{F}_{Z'}' \notin (I_1 \cup \mathcal{A}_1)^*$.

The proof of the other three cases (an fsi in R_1, an afs in R_2 and an fsi in R_2) is similar and therefore left to the reader.

□

From Theorem 4.4 one can easily deduce an algorithm which calculates the inherited dependencies in the same time as a membership algorithm.

A decomposition algorithm can be easily constructed for the following normal form:

Definition 4.8. A scheme $\mathcal{R}$ is said to be in *FSI-Normal Form, (FSINF)* iff for all fsi's $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ of I either $\mathcal{F}_1$ or $\mathfrak{F}_{1Z}$ holds in $\mathcal{R}$.

A decomposition $\{\mathcal{R}_1, \ldots, \mathcal{R}_n\}$ is in *FSINF* iff all the $\mathcal{R}_i, i = 1 \ldots n$ are in *FSINF*.

Note that in the "final" subschemes there are no "real" fsi's any more, only fd's and afs'.

Note also that if only fdi's are given, (i. e. $\mathcal{F}_1$ and $\mathcal{F}_2$ each contain only one fd) then the decomposition algorithm generates a decomposition into the *FDI-Normal Form* of [De5]. If the fdi is trivial (i. e. the "left" fd is trivial) then the decomposition algorithm generates a decomposition into the *Clean Normal Form* of [De2]. If all trivial fdi's are explicitly said to be inherited (from I to I_1 and I_2) then we obtain the decomposition into the *Inherited Normal Form* of [De2].

§5 Possible extensions

The fsi's $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ still have a rather severe restriction: $\forall X \rightarrow Y \in \mathcal{F}_1 \cup \mathcal{F}_2 : Z \subseteq X$. In this section we shall show how far this restriction can be removed, without seriously affecting the semantics of the constraint.

If one changes the restriction to $X \rightarrow Z$ instead of $Z \subseteq X$, one can easily prove the following remark:

Remark 5.1 *Let $\mathcal{F}_1'$ (resp. $\mathcal{F}_2$)= $\{X' \rightarrow Y : X' = XZ$ and $X \rightarrow Y \in \mathcal{F}_1$ (resp. $\mathcal{F}_2$)$\}$.*

The fsi $\mathcal{F}_1' \overset{Z}{\supset\!\!-} \mathcal{F}_2'$ plus the set of all fd's $XZ \rightarrow Y : X \rightarrow Y \in \mathcal{F}_1 \cup \mathcal{F}_2$ are equivalent to the "generalized" fsi $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$.

□

One cannot remove the restriction that $X \rightarrow Z$ must hold as well. The class of constraints would certainly become bigger, but in the subrelation $\mathcal{R}_1$, generated by the "generalized" fsi $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$ the set of fd's $\mathcal{F}_1$ does not hold any more, as one can easily see from the following example:

Suppose we have the "unconstrained" fsi $Emp \rightarrow Job \overset{Dep}{\supset\!\!-} Emp \rightarrow Sal, Man$, meaning that if an employee has only one job in some department, then he has only one salary and one manager for that department. There is something peculiar about this constraint: if for two departments the constraint holds, then (since the two departments together form a *Dep*-complete set of tuples) every employee who works in both departments must have the same job, the same salary and the same manager in both departments! To avoid this altered semantics one can change the definition of the constraint, replacing the phrase "X-complete set" by "X-value" (i.e. X-complete and X-unique set). The definition then becomes:

Definition 5.2 The *"generalized" functional-dependency-set-implication (gfsi)* $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$, means that in every Z-unique, Z-complete set of tuples (in every instance) in which all the fd's of $\mathcal{F}_1$ hold, all the fd's of $\mathcal{F}_2$ must hold too.

There is no restriction on X any more. However one can easily prove that such a gfsi is again equivalent to a normal fsi:

Remark 5.3 *Let $\mathcal{F}_1'$ (resp. $\mathcal{F}_2$)= $\{X' \rightarrow Y : X' = XZ$ and $X \rightarrow Y \in \mathcal{F}_1$ (resp. $\mathcal{F}_2$)$\}$.*

The fsi $\mathcal{F}_1' \overset{Z}{\supset\!\!-} \mathcal{F}_2'$ is equivalent to the gfsi $\mathcal{F}_1 \overset{Z}{\supset\!\!-} \mathcal{F}_2$.

□

The constraint from Example 2.1 now can be written as the gfsi

$$\{dep \rightarrow job; emp \rightarrow job; man \rightarrow job\} \overset{div}{\supset\!\!-} \{emp \rightarrow dep, sal; man \rightarrow dep\}$$

The meaning now reflects the formal description of the constraint more closely than in Example 2.1.

In future research a similar theory should be established for handling exceptions to other constraints, such as multivalued dependencies or inclusion dependencies.

References

[Ar] Armstrong W., Dependency structures of database relationships, *Proc. IFIP 74*, North Holland, pp. 580–583, 1974.

[Be] Beeri C., Bernstein P.A., Computational Problems related to the Design of Normal Form Relation Schemes, *ACM TODS*, vol. **4.1**, pp. 30–59, 1979.

[Ber] Bernstein P.A., Normalization and Functional Dependencies in the Relational Database Model, *CSRG-60*, 1975.

[Co] Codd E., Further normalizations of the database relational model, In *Data Base Systems* (R. Rustin, ed.) Prentice Hall, N.J., pp. 33–64, 1972.

[De1] De Bra P., Paredaens J., The membership and the inheritance of functional and afunctional dependencies, Proc. of the Colloquium on Algebra, Combinatorics and Logic in Computer Science, Gyor, Hungary.

[De2] De Bra P., Paredaens J., Horizontal Decompositions for Handling Exceptions to Functional Dependencies, in "Advances in Database Theory", Vol. II, pp. 123–144, 1983.

[De3] De Bra P., Paredaens J., Conditional Dependencies for Horizontal Decompositions, in "Lecture Notes in Computer Science", Vol. **154**, pp. 67–82, (10-th *ICALP*), Springer-Verlag, 1983.

[De4] De Bra P., Imposed-Functional Dependencies Inducing Horizontal Decompositions, in "Lecture Notes in Computer Science", Vol. **194**, pp. 158–170, (12-th *ICALP*), Springer-Verlag, 1985.

[De5] De Bra P., Functional Dependency Implications, Inducing Horizontal Decompositions, UIA-report 85-30, 1985.

[Fa] Fagin R., Armstrong Databases, *IBM RJ 3440*, 1982.

[Pa] Paredaens J., De Bra P., On Horizontal Decompositions, *XP2-Congress*, State Univ. of Pennsylvania, 1981.

[Ul] Ullman J., Principles of Database Systems, Pitman, 1980.