1998 International Computer Symposium
Workshop on Software Engineering and Database Systems
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

EXISTENCE CONDITIONS AND CLASSIFICATION OF THE INCONSISTENCIES IN THE DATABASE INTEGRATION *

Július Štuller

Institute of Computer Science — Academy of Sciences of the Czech Republic
182 07 Pod vodárenskou věží 2, Prague 8, Czech Republic
E-mail : stuller@uivt.cas.cz

ABSTRACT

The technological progress in the areas of the hardware and the software, together with the general expansion of the computers to almost all human activities, make it easier to realize the integration of many already existing databases.
Unfortunately the process of the databases integration can be accompanied by many various difficulties and problems. One of them is surely the possible occurrence of the inconsistencies appearing in this process of the integration.
These inconsistencies can occur at various levels and they can be of different types.

In the paper we study the existence conditions for these inconsistencies and we propose a classification of certain of these inconsistencies.

Keywords : database systems integration, inconsistency

I. Introduction

The technological progress in the areas of the hardware, specially in the field of the (secondary) memories where the ever increasing capacities are paradoxally in the last several years available at ever decreasing prices and smaller physical sizes, and the software, continuously more and more user friendly, efficient and cheaper, together with the general expansion of the computers to almost all human activities, make it easier to realize the integration of many already existing databases.

Every database can be seen, at least from the point of view of the logic, as a conjunction of different facts (and depending on the representation of these as data, information or knowledge, we can obtain either a classical database system, either an information system or even a kind of fashioned knowledge-base system) which leads naturally to the idea of representing such a database as a (formal) logic theory.
The states of such a database and the operations over such a database obey usually certain rules (so called integrity constraints in the database approach) which can again be expressed in the corresponding logic (for instance in the form of special axioms).

Unfortunately the process of the integration of (existing) databases (as an example see for instance the series of psycho-medical studies [Štuller, 1995a] & [Štuller, 1997a]) can be accompanied by many various difficulties and problems. One of them is surely the possible occurrence of the inconsistencies (in sense of the classical logic — see for instance [Štuller, 1995b] - [Štuller, 1995f]) appearing in this process of the integration of databases.
The inconsistencies in the integration of databases can occur at various levels and they can be of different types.

In this paper we will study the conditions which can lead to the existence of the inconsistencies in the databases integration and we will propose a basic classification of these inconsistencies.

II. Formulation of the Problem

First we will study the conditions for the existence of the inconsistencies in the integration of several databases under the following natural logical assumption :

A1 : *Each of the databases to be integrated has no inconsistencies (when taken alone) .*

*This work was supported by the Grant No. 201/97/1070 of the Grant Agency of the Czech Republic : *Inconsistency Resolution Methods in the Data/Knowledge Base Integration.*
The participation at the ICS '98 was supported by the Grant No. A1030601 of the Grant Agency of the Academy of Sciences of the Czech Republic : *Mathematical Foundations of Inference under Vagueness and Uncertainty*

Furthermore, for reasons of the simplification, and having in mind the current situation in the area of the database technologies where the Codd *relational data model* prevails, we will suppose that :

A2 : *All the databases to be integrated are relational ones :*

Let $\mathcal{B}_i$, $i \in \widehat{m}$, be m *relational databases* to be integrated ($m \geq 2$) , each consisting of k_i *relations* $R^i{}_j = \langle A^i{}_j , D^i{}_j , T^i{}_j \rangle$.
(See Appendix for notations and definitions)

We want to find the conditions which can lead to inconsistencies when trying to integrate some of the databases $\mathcal{B}_i$.

III. Relational Operations Leading to Inconsistencies

We will first suppose there exist (at least two) databases $\mathcal{B}_{i_1}$ and $\mathcal{B}_{i_2}$ each having (at least) one relation $R^{i_j}{}_{q_j}$ with one common, simple or compound, attribute, say C :

$$\textbf{A3 :} \ (\exists C)(\exists s \geq 2)(\forall j \in \widehat{s})(\exists \mathcal{B}_{i_j}) \\ (\exists R^{i_j}{}_{q_j} \in \mathcal{B}_{i_j})(C \subset A^{i_j}{}_{q_j})$$

(Otherwise it may be questionable to integrate the databases $\mathcal{B}_{i_j}$ ...)

From all the usual relational operations (operators) the only ones which can lead to possible inconsistencies are:

- the *unions* of the relations
- the *joins*
- and the corresponding *compositions.*

Let us start by the union (we will use its generalized form from [Štuller, 1997b]) :

Definition 1 Let $R_i = \langle A_i , D_i , T_i \rangle$ be m ($m \geq 2$) relations. The π **- union** of relations R_i is the relation noted $\bigcup_{\pi}{}_{i=1}^{m} R_i = \langle A , D , T \rangle$ such that :

1. $D(A) \cap (\bigcap_{i=1}^{m} D_i(\pi_i(A_i))) \neq \emptyset$
2. $T = \bigcup_{i=1}^{m} T_i(\pi_i(A_i))$

Convention 1 In the case of permutations π_i being *identities* we will omit the prefix π - and speak shortly only about the **union** and note it $\bigcup_{i=1}^{m} R_i$

A. Integration by the Union of the Relations

In order to be able to make the union of the relations $R^{i_j}{}_{q_j}$ we must first suppose they all have the same cardinality , say k
(greater or equal to two; if it was equal to one, the corresponding relations would be the lists, maybe ordered, which cannot lead to inconsistencies if the original relations had no inconsistencies ...) :

$$\textbf{A4 :} \ (\exists k \geq 2)(\exists s \geq 2)(\forall j \in \widehat{s})(\exists \mathcal{B}_{i_j}) \\ (\exists R^{i_j}{}_{q_j} \in \mathcal{B}_{i_j})(|A^{i_j}{}_{q_j}| = k)$$

Remark 1 If at least two of databases $\mathcal{B}_{i_j}$ do not consist of (simple) lists, we can always find, by successive projections, the corresponding (sub)relations $R^{i_j}{}_{q_j}$ with the required property.

First, for the simplification, we will suppose relations $R^{i_j}{}_{q_j}$ are defined over the same relational scheme: $S = \langle A , D \rangle$, that is : 1. $A^{i_j}{}_{q_j} = A$
2. $D^{i_j}{}_{q_j} = D$

$$\textbf{A5 :} \ (\forall j \in \widehat{s})(R^{i_j}{}_{q_j} \in S = \langle A , D \rangle)$$

Example 1

R_1	
Name	**Position**
Peter	researcher

R_2	
Name	**Position**
Peter	director

$R = R_1 \cup R_2$	
Name	**Position**
Peter	researcher
Peter	director

Even in this very simple example without any further supplementary information it is impossible to decide whether an inconsistency appeared in the process of the integration of databases. Such a supplementary information is in general expressed in one or several *integrity constraint(s).*

We will suppose that we have such an integrity constraint. Let it be the following:

Every value of the attribute Name *is associated with no more than one value of the attribute* Position.
(A particular case of a so called functional dependency)

More formally :

$$(\forall u , v \in T)((t(Name) = u(Name)) \\ \Rightarrow (t(Position) = u(Position)))$$

Let us denote by $\mathfrak{I}$ the set of all the possible *integrity constraints* over the given *universe of discourse* $\mathcal{U}$:

$\mathcal{U} = \mathcal{D}(A)$ where :

$$A = \bigcup_{i=1}^{m} \bigcup_{j=1}^{k_i} A^i{}_j$$

and

$$\mathcal{U} = \bigcup_{i=1}^{m} \bigcup_{j=1}^{k_i} D^i{}_j(A^i{}_j)$$

and by I a subset of the set $\Im$.

Let us further denote by $\mathcal{R}(I)$ the set of all the relations over given universe of discourse satisfying I .

It is obvious that the following holds in general :

Lemma 1

$$((\exists s \geq 2)(\forall j \in \widehat{s})(\exists \mathcal{B}_{i_j})(\exists R^{i_j}{}_{q_j} \in (\mathcal{B}_{i_j} \cap \mathcal{R}(I)))$$
$$\not\Rightarrow (\bigcup_{j=1}^{s} R^{i_j}{}_{q_j} \in \mathcal{R}(I))$$

Remark 2 The *union*, to be **meaningful**, should be done only after a thorough *semantical justification* and *verification* because *syntactical equality* of *attributes* and of the *corresponding domains* may be **misleading**, especially in the case of the *overloaded concepts* like *name*, *number*, *year* etc.

Returning to our simple example they are several possibilities :

- They are some **erroneous descriptions** of data:
  $(\exists i \in \widehat{m})(\exists j \in \widehat{k}_i)(\exists R^{i_j}{}_{q_j} = \langle A^{i_j}{}_{q_j}, D^{i_j}{}_{q_j}, T^{i_j}{}_{q_j}\rangle)$
  $(\exists\, C^{i_j}{}_{q_j} \subset A^{i_j}{}_{q_j})\,(C^{i_j}{}_{q_j}$ is "incorrect" $)$
  (that means some *attribute(s)* is (are) *wrong* : in our Example it could mean, for instance, that datum "director" is not a (value of the attribute) Position but it should be a (value from an another attribute) Function ...)

- They are some **erroneous data**
  (in at least one of the relations $R^{i_j}{}_{q_j}$:)
  $(\exists i \in \widehat{m})(\exists j \in \widehat{k}_i)(\exists R^{i_j}{}_{q_j} = \langle A^{i_j}{}_{q_j}, D^{i_j}{}_{q_j}, T^{i_j}{}_{q_j}\rangle)$
  $(\exists\, t \in T^{i_j}{}_{q_j})\,(t$ is "incorrect" $)$
  (i.e. this t *does not represent correctly a fact from the reality we are trying to capture in a database* - relation $R^{i_j}{}_{q_j}$; in our Example it could mean that either Peter is not a researcher or that he is not a director ...)

- They are some
  erroneous integrity constraints :
  $(\exists\, i \in I)\,(i$ is "incorrect" $)$
  (i. e. this i *does not correctly reflect the reality we are trying to model* ; in our Example it could mean that there may be more than one Position associated with one Name ...)

In all cases the *incorrect " items "* **must be removed.** Let us denote by :

- $\hat{A}$ all "incorrect" attributes in A :
  $\hat{A} = \{ a \in A \mid a$ is "incorrect" $\}$
- $\check{A}$ the subset of the set A containing no "incorrect" attributes :
  $\check{A} = A - \hat{A}$
- $\hat{R}$ the subrelation of the relation R containing all "incorrect" data :
  $\hat{R} = \langle A, D, \{ t : A \rightarrow D(A) \mid t$ "incorrect" $\}\rangle$
- $\check{R}$ the subrelation of the relation R containing no "incorrect" data :
  $\check{R} = R - \hat{R}$
- $\hat{I}$ the subset of the set I containing all "incorrect" integrity constraints :
  $\hat{I} = \{ i \in I : i$ is "incorrect" $\}$
- $\check{I}$ the subset of the set I containing no "incorrect" integrity constraints :
  $\check{I} = I - \hat{I}$

In the first case by the *" correction of attributes "* one usually means the replacement (*renaming*) of the incorrect attributes .
Such a replacement can be done only after a thorough (*semantical*) analysis of data corresponding to the appropriate incorrect attributes.
The incorrect attributes should be discovered and their replacement should be performed at the so called **schema integration** *stage* .

In the second case the result of the *"correction of data"* should be *new relations* (without incorrect data)

$\check{R}^{i_j}{}_{q_j}$ and the *new union* $\bigcup_{j=1}^{s} \check{R}^{i_j}{}_{q_j}$.

The incorrect data should be discovered and corrected at the so called **data integration** *stage* .

And finally, in the third case, as the result of the *" correction of the integrity constraints "* , one should obtain a *new set of the integrity constraints* $\check{I}$ (without incorrect constraints).
(At least some of) the incorrect constraints should be discovered and their correction should be performed again at the **schema integration** *stage* .

B. Integration by the π - Unions

Next we will suppose the relations $R^{i_j}{}_{q_j}$ are defined over such different relational schemata

$$S^{i_j}{}_{q_j} = \langle A^{i_j}{}_{q_j}, D^{i_j}{}_{q_j} \rangle$$

that there exist appropriate permutations $\pi^{i_j}{}_{q_j}$ in

$|\widehat{A^{i_j}{}_{q_j}}|$ that the following holds :

A6 : $\bigcap_{j=1}^{s} D^{i_j}{}_{q_j}(\pi^{i_j}{}_{q_j}(A^{i_j}{}_{q_j})) \neq \emptyset$

Example 2

R_1	
Name	**Position**
Peter	researcher

R_2	
Name	**Function**
Peter	director

$R = R_1 \cup_\pi R_2$	
Name	**Post**
Peter	researcher
Peter	director

The necessary prerequisite is the existence of the "appropriate" permutations $\pi^{i_j}{}_{q_j}$ in $|\widehat{A^{i_j}{}_{q_j}}|$ which must be semantically justifiable for the concrete databases – relations :
In our *Example 2* we presuppose that the (names of the) attributes **Position** and **Function** are synonyms (i.e. they are semantically equivalent).

If this is the case similar reasoning we used to the union of relations applies also to the π - union of the relations giving the same sources of the possible inconsistencies:
(Let us denote by :
$R^{m+1} = \langle A^{m+1}, D^{m+1}, T^{m+1} \rangle = \bigcup_{\pi}{}_{j=1}^{s} R^{i_j}{}_{q_j}$
$k_{m+1} = 1$)

- **erroneous descriptions** of data:
  $(\exists i \in \widehat{m+1})(\exists j \in \widehat{k_i})$
  $(\exists R^{i_j}{}_{q_j} = \langle A^{i_j}{}_{q_j}, D^{i_j}{}_{q_j}, T^{i_j}{}_{q_j} \rangle)$
  $(\exists C^{i_j}{}_{q_j} \subset A^{i_j}{}_{q_j})$ ($C^{i_j}{}_{q_j}$ is "incorrect")

  (in our Example 2 it could mean, for instance, that the attribute Post is not a *semantically equivalent* synonym of the attributes Position and Function ...)

- **erroneous data**
  (in at least one of the relations $R^{i_j}{}_{q_j}$:)
  $(\exists i \in \widehat{m})(\exists j \in \widehat{k_i})$
  $(\exists R^{i_j}{}_{q_j} = \langle A^{i_j}{}_{q_j}, D^{i_j}{}_{q_j}, T^{i_j}{}_{q_j} \rangle)$
  $(\exists t \in T^{i_j}{}_{q_j})$ (t is "incorrect")

  (in our Example 2 it could mean that either Peter is not a researcher or that he is not a director ...)

- **erroneous integrity constraints** :
  $(\exists i \in I)$ (i is "incorrect")

  (in our Example 2 it could mean that there may be more than one Post associated with one Name ...)

C. Generalization

Relaxing the condition **A5** (about the relations one wants to make an union over being defined over the same relational scheme) into weaker condition **A6** requiring the existence of permutations $\pi^{i_j}{}_{q_j}$ such that there exists the π - union of relations $R^{i_j}{}_{q_j}$, one can obtain the corresponding lemma for the π - union :

Lemma 2

$$((\exists s \geq 2)(\forall j \in \widehat{s})(\exists \mathcal{B}_{i_j})(\exists R^{i_j}{}_{q_j} \in (\mathcal{B}_{i_j} \cap \mathcal{R}(I))) \not\Rightarrow (\bigcup_{\pi}{}_{j=1}^{s} R^{i_j}{}_{q_j} \in \mathcal{R}(I))$$

Remark 3 In the case of the π - union, to obtain **meaningful** results, one should be even more careful to *semantically justify* the meaning of performing the operation of the π - union .

D. Integration by the (Equi-) Joins

In the following we will study the properties of the joins in the process of the integration of the (relational) databases .

Let us start by giving the definition of the simplest join (the so called *equi-join*) from [Štuller, 1997b] :

Definition 2 Let $R_i = \langle A_i, D_i, T_i \rangle$ be m ($m \geq 2$) relations and B_i be m sets of attributes such that :

$$((B_i \subset A_i)(\forall i \in \widehat{m})) \wedge (\bigcap_{i=1}^{m} D_i(\pi_i(B_i)) \neq \emptyset)$$

The **join** of the relations R_i, *according to the attributes sets* B_i, *with respect to the equality*, is the relation noted : $*_{\pi_1(B_1)=\pi_i(B_i)} R_i = \langle A, D, T \rangle$

where : 1. $A = \bigcup_{i=1}^{m} A_i$

2. $D = \bigcup_{i=1}^{m} D_i$

3. $T = *_{\pi_1(B_1)=\pi_i(B_i)} T_i$

Convention 2 In case of permutations π_i being the identities, the equality of B_i and such that they are *maximal* (in *set inclusion* sense) with such a property, we will omit the *index* $_{\pi_1(B_1)=\pi_i(B_i)}$ by the $*$ and call the join the *natural join* of R_i .

We will start by *illustrate* the **difference** between the *integration* by performing:
- one of the *joins* (the natural one) of the relations and
- one of the *unions* (the π - *union*) of the same relations :

Example 3

R_1	
Mother	**Son**
Eve	John

R_2	
Mother	**Daughter**
Eve	Anne

$R = R_1 * R_2$		
Mother	**Son**	**Daughter**
Eve	John	Anne

$R = R_1 \cup_\pi R_2$	
Mother	**Child**
Eve	John
Eve	Anne

Nevertheless, depending on the every concrete situation, one must choose the best appropriate operation to perform the integration of the databases.

We will illustrate the occurrence of the inconsistencies in the integration by joining the relations in the following example :

Example 4

R_1	
Husband	**Wife**
Joseph	Mary

R_2	
Mother	**Child**
Mary	Jesus

$R = R_1 *_{Wife=Mother} R_2$		
Husband	**Wife**	**Child**
Joseph	Mary	Jesus

Again, as in the case of the union, even in this very simple example without any further supplementary information it is impossible to decide whether an inconsistency appeared in the process of the integration of databases.

The comparison of this join with the π - union of the (same) relations :

Example 5

$R = R_1 \cup_\pi R_2$	
Man	**Woman**
Jesus	Mary
Joseph	Mary

shows that the integration by joins against the integration by unions :

- allows **new** *relationships between objects* (entities or their attributes) which
- can be the source of **new** *inconsistencies* (having for arguments some of such new relationships) in addition to the inconsistencies known from the unions .

Remark 4 What was said about the **importance** of the *semantical justification* for the π - union holds *even more* for the **joins** as the only condition on p relations ${R^{i_k}}_{q_k}$ to be joinable is :

$$\mathbf{A7:} \quad \bigcap_{k=1}^{p} {D^{i_k}}_{q_k}\,({\pi^{i_k}}_{q_k}\,({B^{i_k}}_{q_k})) \neq \emptyset \quad \text{where :}$$

$$(\forall k \in \widehat{p})\ ({B^{i_k}}_{q_k} \subset {A^{i_k}}_{q_k})$$

which is equal to the condition **A6** with the unique difference that ${B^{i_k}}_{q_k} \subset {A^{i_k}}_{q_k}$ and so one can have in principle up to :

$$\prod_{k=1}^{p} \left(\sum_{m=1}^{|{B^{i_k}}_{q_k}|} \binom{|{A^{i_k}}_{q_k}|}{m} \right)$$

possibilities of performing the join of p relations.

E. Other Operations of the Relational Algebra

All the other (usual) relational operations (and operators), except the *(equi-) compositions* , do not contribute to the process of the integrations of databases. We will not need to study the properties of the compositions as they are expressible in the corresponding joins (and projections).

IV. Inconsistencies classification

Definition 3 The following relational operations : *unions* , *(equi-) joins* and *(equi-) compositions* will be called the **integration operations**.

Definition 4 Let R_k be m $(m \geq 2)$ relations one wants to make an integration operation over, I_k be m corresponding sets of integrity constraints and I_{m+1} be the set of integrity constraints corresponding to the result of the integration operation such that : $I = \bigcup_{k=1}^{m+1} I_k$ is (logically) consistent .

We will call the inconsistencies in the result of the integration operation over the relations R_k the :

universe of discourse inconsistencies

$$\Leftrightarrow \ (\exists\, k \in \widehat{m})\ (\check{A}_k \neq A_k)$$

data inconsistencies

$$\Leftrightarrow \ (\exists\, k \in \widehat{m})\ (\check{R}_k \neq R_k)$$

integrity constraints inconsistencies

$$\Leftrightarrow \ (\exists\, k \in \widehat{m+1})\ (\check{I}_k \neq I_k)$$

semantical inconsistencies

$$\Leftrightarrow \ (\exists\, k \in \widehat{m})\ (\pi_k \neq \textit{Identity})$$

Definition 5 *Universe of discourse inconsistencies* and *integrity constraints inconsistencies* will be called the **conceptual inconsistencies.**

Remark 5 Every type of databases integration inconsistencies, according to *our* **classification**, originates from *different* **sources** and *can be best* **eliminated**, or *at least* **minimized**, at *various* **stages** of the integration of the concerned databases :

- the *conceptual inconsistencies* at the stage of the **schema integration**
- the *semantical inconsistencies* by **well-considered choice** of the *attribute(s) over which one wants to integrate the databases*
- the *data inconsistencies* by *thorough* **verification** and **validation**, at the *data entry* stage

Convention 3 In the following we will use the following notation :

u - **inconsistencies** : *universe of discourse inconsistencies*
d - **inconsistencies** : *data inconsistencies*
i - **inconsistencies** : *integrity constraints inconsistencies*
s - **inconsistencies** : *semantical inconsistencies*
c - **inconsistencies** : *conceptual inconsistencies* .

V. Existence Conditions for the Inconsistencies in the Database Integration

We have seen in the section **III** that the integration of (the relational) databases by integration operations may lead to inconsistencies .

In order to eliminate, as much as possible, the occurrences of these inconsistencies, one should try to , especially in the case of the validity of the conditions :

(**A1** : *Each of the considered databases to be integrated has no inconsistencies (when taken alone)* .
&
A2 : *All the databases to be integrated are relational ones* .)

- & **A5** : $((\forall j \in \widehat{s}) (R^{i_j}{}_{q_j} \in S = \langle A, D \rangle))$

  clear the databases to be integrated from :

  - **incorrect attributes** which *can lead to* the **u - nconsistencies**
  - **incorrect data** which *can lead to* the **d - inconsistencies**
  - **incorrect integrity constraints** which *can lead to* **i - inconsistencies**

- & **A6** : $(\bigcap_{j=1}^{s} D^{i_j}{}_{q_j} (\pi^{i_j}{}_{q_j} (A^{i_j}{}_{q_j})) \neq \emptyset)$

  semantically **deeply analyze** the *corresponding attributes* in the relations to be integrated by **π - unions** to eliminate the **s - inconsistencies**

- & **A7** : $\bigcap_{k=1}^{p} D^{i_k}{}_{q_k} (\pi^{i_k}{}_{q_k} (B^{i_k}{}_{q_k})) \neq \emptyset$

  $(B^{i_k}{}_{q_k} \subset A^{i_k}{}_{q_k}) (\forall k \in \widehat{p})$

  semantically **deeply analyze** the *corresponding attributes* in the relations to be integrated by **joins** to eliminate the **s - inconsistencies** .

VI. Related work

The process of the integration of databases has been studied from mid-eighties, with the emphasis on the schema integration (see e. i. [Batini et al., 1986] as one of the first papers and [Ramesh & Ram, 1997], [Santucci, 1998] and [Tseng et al., 1998] as ones of the last ones) and with less attention on the integration of data themselves (see for instance [Orlowska et al., 1997]) .

To our knowledge none of them proposed any kind of classification of the inconsistencies in the process of databases integration neither any existence conditions for these inconsistencies to appear .

VII. Conclusion

By **analyzing** some *simple examples* we have arrived at the **sources** of *possible inconsistencies* when integrating databases and we have **proposed** *certain* **classification** of these inconsistencies based on the their sources.

We have find *four* conditions **A4**, **A5**, **A6** and **A7** under which can occur different types of inconsistencies in the process of the integration of databases.

The conditions **A4**, **A5**, **A6** apply to the *integration by unions* while the condition **A7** applies to the *integrations by joins* .

In the future we would like to further develop our classification of the inconsistencies in databases integration with the focus on the possibilities of designing some kind of *support for the resolution* of these inconsistencies.

VIII. Appendix

Definition A **relation** in the RMD will be any triple $\langle A, D, T \rangle$ with :

1. A being a finite set of **attribute names** .
2. D being a mapping which maps every attribute name $a \in A$ to a **domain** , noted $D(a)$.
   Let us : *denote* by $D(A)$ the *union* of all $D(a)$ and *call* it the **universe of discourse** .
3. T being a finite set of **mappings** t from A to the universe of discourse $D(A)$ such that :
   $$t(a) \in D(a) \text{ for all } a \in A .$$

Notation 1 $\widehat{m} = \{1, 2, \cdots, m\}$ $\qquad (\widehat{0} = \emptyset)$

Notation 2 The *cardinality* of the set A will be denoted by $|\,\boldsymbol{A}\,|$.

Notation 3 Let $A_i = \{a_{ij} \mid j \in \widehat{|A_i|}\},\ i \in \{1,2\}$.

$$(\forall j \in \widehat{|A_i|})\ \ (D_1(a_{1j}) \cap D_2(a_{2\pi(j)}) \neq \emptyset)$$

(π being an appropriate permutation in $\widehat{|A_i|}$)

$$\Updownarrow$$

$$\boldsymbol{D}_1(\boldsymbol{A}_1) \cap \boldsymbol{D}_2(\pi(\boldsymbol{A}_2)) \neq \emptyset$$

Lemma $\quad (\boldsymbol{D}_1(\boldsymbol{A}_1) \cap \boldsymbol{D}_2(\pi(\boldsymbol{A}_2)) \neq \emptyset) \Rightarrow (|\,\boldsymbol{A}_1\,| = |\,\boldsymbol{A}_2\,|)$

Notation 4 $\displaystyle\bigcup_{i=1}^{m} \boldsymbol{T}_i(\pi_i(\boldsymbol{A}_i)) = \{t : A \rightarrow D(A) \mid (\exists i \in \widehat{m})(\exists u_i \in T_i)(u_i(\pi_i(A_i)) = t(A))\}$

Definition Let $R = \langle A, D, T \rangle$ be a relation and $A_1 \subset A$. The **projection** of the relation R over A_1 is the relation noted $\boldsymbol{R}[\boldsymbol{A}_1] = \langle A_1, D_1, T_1 \rangle$ such that :

1. $\boldsymbol{D}_1 = \boldsymbol{D}/\boldsymbol{A}_1$
   (the *restriction* of the mapping D on the subset A_1 of A)
2. $T_1 = \boldsymbol{T}[A_1]$

Notation 5 $\quad \boldsymbol{T}[A_1] = \{t : A_1 \rightarrow D_1(A_1) \mid (\exists u \in T)(t(A_1) = u(A_1))\}$

Definition Let $R_i = \langle A_i, D_i, T_i \rangle,\ i \in \{1, 2\}$, be two relations such that :

1. $(\exists \boldsymbol{A}_{21} \subset \boldsymbol{A}_2)\ \ |\,\boldsymbol{A}_1\,| = |\,\boldsymbol{A}_{21}\,|)$
2. $\boldsymbol{D}_1(\boldsymbol{A}_1) \cap (\boldsymbol{D}_2/\boldsymbol{A}_{21})(\pi(\boldsymbol{A}_{21})) \neq \emptyset)$
   (π being an appropriate permutation)
3. $\boldsymbol{T}_1(\boldsymbol{A}_1) \subset \boldsymbol{T}_2[\boldsymbol{A}_{21}](\pi(\boldsymbol{A}_{21}))$

Then we will say that the relation R_1 is a **subrelation** of the relation R_2 – what we will note by $\boldsymbol{R}_1 \subset \boldsymbol{R}_2$

Notation 6 $\quad \displaystyle \boldsymbol{D}(\boldsymbol{a}_j) = \bigcup_{i=1}^{m} D_i(a_j)\ ,\ \forall j \in \widehat{|A|} \qquad [\boldsymbol{A} = \bigcup_{i=1}^{m} \boldsymbol{A}_i]$

$$\Updownarrow$$

$$\boldsymbol{D} = \bigcup_{i=1}^{m} \boldsymbol{D}_i$$

Notation 7

$$*_{\pi_1(B_1)=\pi_i(B_i)}\ T_i = \{t : A \rightarrow D(A) \mid ((\forall i \in \widehat{m})(\exists u_i \in T_i))((t(A_j) = u_j(A_j)) \wedge (u_1(\pi_1(B_1)) = u_i(\pi_i(B_i)))))\}$$

Convention

We will call a set of attributes a **compound attribute** or even, shortly, only an **attribute** .
When such a set will have *exactly one* element, we will call it, whenever necessary, a **simple attribute** .

References

[Arens et al., 1993] ARENS Y., CHEE C. Y., HSU C., KNOBLOCH C. A.: *Retrieving and Integrating Data from Multiple Information Sources.* Journal of Intelligent and Cooperative Information Systems, 1993, 2, 2, 127-158

[Batini et al., 1986] BATINI C., LENZERINI M., NAVATHE S. B.: *A Comparative Analysis of Methodologies for Database Schema Integration.* ACM Computing Surveys, 1986, 18, 4, 323-364

[Blanco et al., 1994] BUILDING A FEDERATED RELATIONAL DATABASE SYSTEM : AN APPROACH USING A KNOWLEDGE-BASED SYSTEM. Journal of Intelligent and Cooperative Information Systems, 1994, 3, 4, 415-455

[Gupta et al., 1996] GUPTA A., JAGADISH H. V., MUMICK I. S.: *Data Integration using Self-Maintainable Views.* In: Proceedings of the 5th International Conference on Extending Database Technology (EDBT-96), Lecture Notes in Computer Science 1057, Springer-Verlag, 1996, 140-144

[Hull & Zhou, 1996] HULL R., ZHOU G.: *A Framework for Supporting Data Integration using the Materialized and Virtual Approaches.* In: Proceedings of the ACM SIGMOD International Conference on Management of data, 1996, 481-492

[Orlowska et al., 1997] ORLOWSKA M. E., LI H., LIU CH.: *On Integration of Relational and Object-Oriented Database Systems.* LNCS 1338 (Proceedings of SOFSEM'97), 1997, 294-312

[Ramesh & Ram, 1997] RAMESH V., RAM S.: *Integrity Constraint Integration in Heterogeneous Databases: An Enhanced Methodology for Schema Integration.* Information Systems, 1997, 22, 8, 423-446

[Santucci, 1998] SANTUCCI G.: *Semantic schema refinements for multilevel schema integration.* Data & Knowledge Engineering 1998, 25, 301-326

[Solaco et al, 1995] sc Garcia-Solaco M., Saltor F., Castellanos M.: A STRUCTURE BASED SCHEMA INTEGRATION METHODOLOGY. In: Proceedings of 11th IEEE International Conference on Data Engineering (ICDE-95), 1995, 505-512

[Štuller, 1995a] Provazníková H., ŠTULLER J., Štullerová N.: *Health and Prevention Care Possibilities for First Year Students of the Charles University 3rd Medical Faculty at Prague.* (In Czech: Zdraví a možnosti preventivní péče o studenty prvního ročníku 3. lékařské fakuty UK Praha.) In: Journal of Czechoslovak Psychology (*Československá psychologie*), 1995, 34/2, 159-169, Academia, ISSN 0009-062X

[Štuller, 1997a] Provazníková H., ŠTULLER J., Štullerová N., Berkovičová V.: *Life Style, Health and Achievements of Students of the Faculty of Medicine.* (In Czech: Životní styl, zdraví a prospěch studentů lékařské fakulty.) In: Journal of Czechoslovak Psychology (*Československá psychologie*), 1997, 41/3, 216-224, Academia, ISSN 0009-062X

[Štuller, 1995b] ŠTULLER J. : *Inconsistency Resolution in the Databases Integration.* In : Proceedings of the International Scientific Seminar DATABASE SYSTEMS ("DATABÁZOVÉ SYSTÉMY"), 1st-2nd June 1995, Bratislava, Slovakia, Centre of Edition, House of Technology of the Union of Slovak Scientific and Technical Societies, Bratislava, 102-107, ISBN 80-233-0348-1

[Štuller, 1995c] ŠTULLER J.: *Inconsistency Problems in the Information Systems Integration.* (Invited Lecture.) In: BIOMATH-95 : Proceedings of the International Symposium on Mathematical Modelling and Information Systems in the Biology, Ecology and Medicine BIOMATH-95, 23rd-27th August 1995, Sofia, Bulgaria, (Editors : Popova E. D., Markov S., Ullrich Ch.), DATECS Publishing Ltd. - Institute of Biophysics, BAN, Sofia, 1995, 77, ISBN 954-613-005-2

[Štuller, 1995d] ŠTULLER J.: *Inconsistency Conflict Resolution in the Integration of the Databases.* In: Proceedings of the International Conference "Computer Science", 5th-7th September 1995, Ostrava, Czech Republic, (Editors: Vondrák I., Štefan J.), Ostrava - MARQ, Ostrava Repronis, 283-289, ISBN 80-901751-7-1

[Štuller, 1995e] ŠTULLER J.: *Inconsistency Resolution in the Integration of Databases.* In: Proceedings of the 15th Database Conference DATASEM '95, 8th-10th October 1995, Brno, Czech Republic, M. Š. - PRINT, Letovice, 47-52, ISBN 80-900066-9-8

[Štuller, 1995f] ŠTULLER J.: *Inconsistency Conflict Resolution.* In: Proceedings of the XXII-th Winter School SOFSEM '95 - Seminar on Current Trends in Theory and Practice of Informatics, 25th November - 2nd December 1995, Milovy, Czech Republic, (Editors: Bartošek M., Staudek J., Wiedermann J.), Lecture Notes in Computer Sciences 1012, Springer-Verlag, Berlin, 1995, 469-474, ISBN 3-540-60609-2

[Štuller, 1997b] ŠTULLER J.: *Database Systems and Logic - I.* Technical Report No. 702, ICS AS CR, Prague, 1997, 35 pages

[Štuller, 1998a] ŠTULLER J.: *Classification of the Inconsistencies in the Databases Integration.* TR No. 746, ICS AS CR, Prague, 1998, 11 pages

[Štuller, 1998b] ŠTULLER J.: *Existence Conditions for the Inconsistencies in the Databases Integration.* Technical Report No. 760, ICS AS CR , Prague, 1998, 11 pages

[Tseng et al., 1998] TSENG F. S. C., CHIANG J.-J., YANG W.-P. : *Integration of Relations with Conflicting Schema Structures in Hererogeneous Database Systems.* Data & Knowledge Engineering 27, 1998, 231-248