Database Systems and Logic - I

Štuller, Július

1997

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

DATABASE SYSTEMS AND LOGIC - I

Július ŠTULLER

Technical report No. 702

January 1997

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+42 2) 6605 3200 fax: (+42 2) 85 85 789
e-mail: stuller@uivt.cas.cz
http://www.uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

DATABASE SYSTEMS AND LOGIC - I

Július ŠTULLER [1]

Technical report No. 702
January 1997

Abstract

Every database can be seen, at least from the point of view of 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).
In order to enlarge the expressiveness and the possibilities of the existing database systems by allowing them to process the uncertainty (probalistic, possibilistic, degree of belief) and the fuzziness (vagueness, degree of truth) it is possible to try to extend the underlying logic from the classical one to one of the fuzzy logics.

Keywords
database systems, logic, incomplete information

[1]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* .

Contents

Chapter 1

Introduction

The close relation between the databases and the logic was recognized very early. For instance, in [Schwartz, 1971] the author wrote:

> The problems of data-base system design are separated into two categories : *abstract* or hierarchy-independent problems, and *concrete* , or hierarchy-dependent problems.
> ... **abstract** problems, ... could still be defined and studied even if data sets, however large, were always held in a large single-level memory big enough to hold whatever files, permanent or temporary, that a data system required.
> ... core memory containing 2^{60} 300-bit words; even a fast computer would take a century to access all these words, so that any data base which cannot be stored in such a memory is in real sense too large to be handled by our present data processing technology. (Approximately 4.10^{10} GB ...)
> In contrast to the abstract theory ... the **concrete** theory ... problems specific to the storage of large amounts of data on " block addressable " or " serially addressable " and generally electromechanical memories; specifically, drums, discs, and tapes.
> ... even if all problems arising from block addressability are ignored, a large number of design problems remain.
> Many of these are optimization problems of various kinds, generally having to do with methods for reducing the size of the otherwise very lengthy searches necessary to locate particular items to be retrieved.
> ... the retrieval processes to be carried out are easily described in set-theoretical terms, so that the problems of the data base area are problems of *efficiency* rather than problems of description.
>
> Data base problems are generally quite simple from the *logical* point of view, and easily formulated in set-theoretical terms.
> ... From an abstract point of view, a data base can be regarded as an encoded representation of certain **sets** $S_1, S_2, \ldots, S_N$ (the *files* of the data base) together with a certain collection of **mapping** $f_1, \ldots, f_m$.

Certain of these mappings will define *value* or *attribute* **functions**, i. e. will assign to one or another of their sets S_i attributes whose meaning is external to the data base itself. ...
A mapping of this kind may be indicated symbolically by writing
$f \ : \ S_i \rightarrow V$, where V is the range of values of f .
Other mappings will be *cross-reference mappings* which assign elements of one set S_j to elements of another set S_i . ...
Such a map may be indicated by writing $f \ : \ S_i \rightarrow S_j$.

Within a data base one characteristically finds :

a. Relatively few, but often quite large sets S_i .
(These are the main *files* of the data base.)

b. Items may be added to and substracted from sets, and particular values of maps may be changed with fair frequency as a data base is updated ...

... the operations associated with data base processing are from the abstract set theoretical point of view extremely simple. They generally only require that certain straightforward combinations of the basic operations : *subset extraction, union, intersection, counting, totalling* and *maximization* be carried out.

EXAMPLE 1

How many employees belonging to organization central staff speak Chinese?

$$\underline{\text{print}}\ \#\ \{\ \text{x} \in \text{employees} \mid \text{department (x)}\ \ \underline{\text{eq}}\ \ \text{centralstaff}\ \ \underline{\text{and}}\ \ \text{Chinese} \in \text{languagesspoken (x)}\ \}$$

We can rewrite the last two lines as follows :

$$\text{Card}\ (\ \{\ \text{x} \in \text{E} \mid \text{d (x)} \ = \ c_1 \ \ \underline{\text{and}}\ \ c_2 \in \text{l (x)}\ \}\ )$$

and we see that the nucleus of a general query is simply the following one :

$\{\ \text{x} \in \text{S} \mid \ \text{P (x)}\ \}$, where P is an appropriate predicate .

Chapter 2

Codd relational data model

Its appearance [Codd, 1970] in the early seventies influenced almost all the areas of database research and technology.
Let us remind just a few phrases from the abstract of this famous paper :

> Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). . . . Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed.

2.1 Relations

Definition 1a
(Codd)

Given sets $S_1, S_2, \ldots, S_M$ (not necessarily distinct) , R is a **relation** on these m sets if it is a *subset of the cartesian product* $S_1 \times S_2 \times \ldots \times S_M$ (set of m-tuples each of which has its first element from S_1 , its second element from S_2 , and so on) .
S_j is the jth **domain** of R.

Note 1

One can use the " array representation " of a relation :

Morocco	Rabat
Libya	Tripolis
Tunisia	Tunis

with the following properties :

P1 : Each row represents an m-tuple of R .
P2 : The ordering of rows is immaterial.
P3 : All rows are distinct.

P4 : The ordering of columns is significant — it corresponds to the ordering S_1 , S_2 , . . . , S_M of the domains on which R is defined.

Cain	Abel
Brutus	Ceasar

P5 : The significance of each column is partially conveyed by labeling it with the name of the corresponding domain.

US–President :	**Vice–President :**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

Remark 1

Even if the columns are labeled by the name of the corresponding domains, the ordering of columns should matter :
one can have a relation with two (or more) identical domains
— see the following example .

EXAMPLE 2

Part :	**Part :**	**Quantity :**
Computer	System board	1
System board	I/O Support	1
I/O Support	8-bit IAS Slot	1
I/O Support	16-bit IAS Slot	6
I/O Support	32-bit VESA Slot	3
I/O Support	Keybord attachement	1
I/O Support	Speaker attachement	1

Codd proposed in his original paper in such a case that the ambiguous domains names " be qualified by a distinctive *role name* , which serves to identify the role played by that domains in the given relation " .
Instead we will present a modification of his original **Definition 1a** :

Definition 1b

A **relation** in the RELATIONAL DATA MODEL (RMD) will be any (ordered) triple $\langle A, D, T \rangle$ where

1. A is a *finite set* of **attribute names** (distinct *words of finite length over an alphabet*) .
2. D is a *mapping* which maps every attribute name $a \in A$ to a **domain** , noted $D(a)$.
   (Domains are *nonempty sets* — need not be distinct !)
3. T is a *finite subset* of the cartesian product of all the attribute names domains $D(a)$.

The previous example then gets the following form :

EXAMPLE 3

Parts explosion problem $= \langle A, D, Components \rangle$ where :

$A = \{ Assembly, Subassembly, Quantity \}$

$D : D(Assembly) = Parts$

$D(Subassembly) = Parts$

$D(Quantity) = Natural\ Numbers$

(with the domains *Parts* and *Natural Numbers*)

Components : instead of an array representation we will in the next utilize a " *tabular representation* " :

Assembly	**Subassembly**	**Quantity**
Computer	System board	1
System board	I/O Support	1
I/O Support	8-bit IAS Slot	1
I/O Support	16-bit IAS Slot	6
I/O Support	32-bit VESA Slot	3
I/O Support	Keybord attachement	1
I/O Support	Speaker attachement	1

Remark 2

By permuting the columns of such a **table** or (equivalently) permuting the order of the attribute names domains in the cartesian product we obtain the same information .
So Codd had to use the term **relationship** as an *equivalence class* of relations that are "*equivalent* " *under permutation of domains* (relationships as " domain-unordered counterparts " of relations) .
Instead we will again present yet another modification of his original definition of a relation :

Definition 1

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 $D\,(\,a\,)$.
   (We will 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 $t\,(\,a\,) \in D\,(\,a\,)$ for all $a \in A$.

Note 2

We will utilize the same tabular representation as before, but the table representing a relation will now have the following properties :

P1 : Each row represents a mapping t from T .
P2 : The ordering of rows is immaterial.
P3 : All rows are distinct.
P4 : The ordering of columns is immaterial.

Convention 1

Instead of the " attribute names " we will speak shortly only about the " **attributes** " .

Convention 2

We will still utilize the name " tuple " for the elements of T .

Having the right definition of the relation we can return to the Codd's vision of a data bank :

> The totality of data in a data bank may be viewed as a collection of time-varying relations. These relations are of assorted degrees. As time progresses, each m - ary relation may be subject to insertion of additional m - tuples, deletion of existing ones, and alteration of components of any of its existing m - tuples .

To be able to study in more details the relations we will start by giving the notion of the *equality* of relations .

2.1.1 Equality of Relations

Definition 2

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

1. $A_1 = A_2$
2. $D_1 = D_2$
3. $T_1 = T_2$

Then we will say that the two relations are **equal**
(for what we will use the usual notation : $R_1 = R_2$) .

EXAMPLE 4

R_1	
President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

R_2	
Vice–President	**President**
Johnson	Keneddy
Humprey	Johnson
Ford	Nixon
Rockefeller	Ford
Mondale	Carter
Bush	Reaggan
Quale	Bush
Gore	Clinton

$$\boldsymbol{R_1 = R_2}$$

Remark 3

The notion of the equality is a particular case of a more general notion, namely of *equivalence* which we will introduce next.

2.1.2 Equivalence of Relations

Notation 1

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

Definition 3a

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

1. $| \boldsymbol{A}_1 | = | \boldsymbol{A}_2 | \qquad (= m)$

$$(A_i = \{ a_{ij} \mid j \in \widehat{m} \}, i \in \{1,2\})$$

2. $(\forall j \in \widehat{m}) (D_1 (a_{1j}) = D_2 (a_{2j}))$

$$\Updownarrow$$

Notation 2 $\qquad D_1 (A_1) \simeq D_2 (A_2)$

$$\Updownarrow$$

Notation 3 $\qquad \boldsymbol{D}_1 \simeq \boldsymbol{D}_2$

3. $| T_1 | = | T_2 | \qquad (= n)$

$$(T_i = \{ t_{ik} \mid k \in \widehat{n} \}, i \in \{1,2\})$$

$$(\forall k \in \widehat{n}) (\forall j \in \widehat{m}) (t_{1k} (a_{1j}) = t_{2\pi(k)} (a_{2j}))$$

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

$$\Updownarrow$$

Notation 4 $\qquad (\forall k \in \widehat{n}) (t_{1k} (A_1) = t_{2\pi(k)} (A_2))$

$$\Updownarrow$$

Notation 5 $\qquad T_1 (A_1) = \pi (T_2 (A_2))$

$$\Updownarrow$$

Notation 6 $\qquad T_1 (A_1) \simeq T_2 (A_2)$

$$\Updownarrow$$

Notation 7 $\qquad \boldsymbol{T}_1 \simeq \boldsymbol{T}_2$

Then we will say that the two relations are **equivalent** .

Notation 8 $\qquad \boldsymbol{R}_1 \simeq \boldsymbol{R}_2$

EXAMPLE 5

R_1	
US–President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

R_2	
President	**Vice–Pres.**
Keneddy	Johnson
Johnson	Humprey
Nixon	Ford
Ford	Rockefeller
Carter	Mondale
Reaggan	Bush
Bush	Quale
Clinton	Gore

Remark 4

But we can go even further.

In the **Definition 3a** we can replace the point 2. by the following one :

2. $(\forall j \in \widehat{m})\ (D_1(a_{1j}) = D_2(a_{2\pi(j)}))$

$(\pi$ being an appropriate permutation in $\widehat{m})$

$\Updownarrow$

Notation 9 $D_1(A_1) \simeq D_2(\pi(A_2))$

$\Updownarrow$

Notation 10 $\boldsymbol{D}_1 \sim \boldsymbol{D}_2$

This leads us naturally to the following definition of the equivalence of relations :

Definition 3b

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

1. $|\boldsymbol{A}_1| = |\boldsymbol{A}_2|$ $(= m)$

$(A_i = \{a_{ij} \mid j \in \widehat{m}\}, i \in \{1,2\})$

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

$(\pi$ being an appropriate permutation in $\widehat{m})$

$\Updownarrow$

Notation 11 $\boldsymbol{D}_1(\boldsymbol{A}_1) \cap \boldsymbol{D}_2(\boldsymbol{\pi}(\boldsymbol{A}_2)) \neq \boldsymbol{\emptyset}$

3. $|T_1| = |T_2|$ $(= n)$

$(T_i = \{t_{ik} \mid k \in \hat{n}\}, i \in \{1,2\})$

$(\forall k \in \hat{n})(\forall j \in \widehat{m})\ (t_{1k}(a_{1j}) = t_{2\rho(k)}(a_{2\pi(j)}))$

$(\rho$ being an appropriate permutation in $\hat{n})$

$\Updownarrow$

Notation 12 $(\forall k \in \hat{n})(t_{1k}(A_1) = t_{2\rho(k)}(\pi(A_2)))$

$\Updownarrow$

Notation 13 $T_1(A_1) = \rho(T_2(\pi(A_2)))$

$\Updownarrow$

Notation 14 $T_1(A_1) \simeq T_2(\pi(A_2))$

$\Updownarrow$

Notation 15 $\boldsymbol{T}_1 \sim \boldsymbol{T}_2$

Then we will say that the two relations are **equivalent** .

Notation 16 $\boldsymbol{R}_1 \sim \boldsymbol{R}_2$.

Lemma 1

$\simeq$ is a special case of $\sim$.

Proof : Simply take for the permutation π the identity .

EXAMPLE 6

R_1	
US–President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

R_2	
Vice–Pres.	**President**
Johnson	Keneddy
Humprey	Johnson
Ford	Nixon
Rockefeller	Ford
Mondale	Carter
Bush	Reaggan
Quale	Bush
Gore	Clinton

$$\boldsymbol{R}_1 \sim \boldsymbol{R}_2$$

Lemma 2

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

Lemma 3

$$((\boldsymbol{T}_1 \sim \boldsymbol{T}_2) \wedge (| \boldsymbol{T}_i | \neq 0)) \Rightarrow$$
$$((\boldsymbol{D}_1 (\boldsymbol{A}_1) \cap \boldsymbol{D}_2 (\boldsymbol{\pi} (\boldsymbol{A}_2))) \neq \boldsymbol{\emptyset})$$

Corollary 1

$$((\boldsymbol{T}_1 \sim \boldsymbol{T}_2) \wedge (| \boldsymbol{T}_i | \neq 0)) \Rightarrow (| \boldsymbol{A}_1 | = | \boldsymbol{A}_2 |)$$

Corollary 2

In case $| \boldsymbol{T}_i | \neq 0$, the first and the second condition in the **Definition 3b** are redundant.

Now we can give the following definition of the equivalence of relations .

Definition 3c

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

$$(\boldsymbol{T}_1 \sim \boldsymbol{T}_2) \wedge (| \boldsymbol{T}_i | \neq 0) .$$

Then we will say that the two relations are **equivalent** .

If we admit that all **empty relations** (relations with $| \boldsymbol{T} | = 0$) are equivalent, we obtain the *final* definition of equivalence of relations :

Definition 3

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

$$\boldsymbol{T}_1 \sim \boldsymbol{T}_2 .$$

Then we will say that the two relations are **equivalent** .

2.2 Set of Relations

Thank to the previous definitions we can decide whether two relations — from (certain) set of relations , noted $\Re$ — are equivalent or even equal. In the following we will often not distinguish between equivalent relations. (In fact, in such a case, we will operate on the *factorized* set $\Re / \sim$) .

2.2.1 Ordering

We can define an *ordering* between relations :

Definition 4a

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

1. $A_1 = A_2$
2. $D_1 = D_2$
3. $T_1 \subset T_2$

Then we will say that the relation R_1 is a **subrelation** of the relation R_2 — what we will note : $R_1 \subset R_2$.

EXAMPLE 7

MURDERED US PRESIDENTS	
President	**Vice–President**
Keneddy	Johnson

US PRESIDENTS	
President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

MURDERED US PRESIDENTS $\subset$ US PRESIDENTS

Remark 5

Again we can generalize the notion of a subrelation in several directions :

In the **Definition 4a** we can replace the points 1. — 3. by the following ones :

1. $\mid \boldsymbol{A}_1 \mid \; = \; \mid \boldsymbol{A}_2 \mid \qquad (= m)$

$$(\; A_i \; = \; \{ \, a_{ij} \mid j \in \widehat{m} \, \} \, , \; i \in \{ \, 1, 2 \, \} \;)$$

2. $(\; \forall \, i \in \widehat{m} \;) \; (\; D_1 \, (\, a_{1i} \,) \subset D_2 \, (\, a_{2\pi(i)} \,) \;)$

$$\Updownarrow$$

Notation 17 $\quad \boldsymbol{D}_1 \, (\boldsymbol{A}_1) \; \subset \; \boldsymbol{D}_2 \, (\, \boldsymbol{\pi} \, (\boldsymbol{A}_2) \,)$

$(\; \pi$ being a permutation in $\widehat{m} \;)$

or even this new point 2. by the following one :

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

3. $(\; \forall \, t \in T_1 \;) \; (\; \exists \, u \in T_2 \;) \; (\, t \, (A_1) \; = \; u \, (\, \pi \, (A_2) \,) \;)$

$$\Updownarrow$$

Notation 18 $\quad \boldsymbol{T}_1 \, (\boldsymbol{A}_1) \; \subset \; \boldsymbol{T}_2 \, (\, \boldsymbol{\pi} \, (\boldsymbol{A}_2) \,)$

According to **Lemma 2** the first condition is redundant and so we obtain the following definition :

Definition 4b

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

1. $\boldsymbol{D}_1 \, (\boldsymbol{A}_1) \; \cap \; \boldsymbol{D}_2 \, (\, \boldsymbol{\pi} \, (\boldsymbol{A}_2) \,) \neq \boldsymbol{\emptyset}$
2. $\boldsymbol{T}_1 \, (\boldsymbol{A}_1) \; \subset \; \boldsymbol{T}_2 \, (\, \boldsymbol{\pi} \, (\boldsymbol{A}_2) \,)$

Then we will say that the relation R_1 is a **subrelation** of the relation R_2 .

We will later once more return to the generalization of the definition of subrelation after introducing the operation *projection* .

In any case the notion of subrelation give us already now the possibility to define a *partial* **ordering** on the set of relations .

2.3 Set Operations on Relations

Since relations are, roughly speaking, sets (of mappings), we can apply all the usual set operations on them.

2.3.1 Unary operations

Definition 5

Let $R = \langle A, D, T \rangle$ be a relation.

The **active domain** of the attribute a $(\in A)$, *with respect to* the relation R, is the following subset of the domain $D(a)$:

$$\alpha D(a, R) = \{ d \in D(a) \mid (\exists t \in T)\ (t(a) = d) \}$$

The **active complement** of the relation R is the relation $\tilde{R} = \langle A, D, \tilde{T} \rangle$ where :

$$\tilde{T} = \{ t : A \rightarrow D(A) \mid ((\forall a \in A)\ (t(a) \in \alpha D(a, R))) \bigwedge (t \notin T) \}$$

The **complement** of the relation R is the (ordered) triple : $\bar{R} = \langle A, D, \bar{T} \rangle$ where :

$$\bar{T} = \{ t : A \rightarrow D(A) \mid ((\forall a \in A)(t(a) \in D(a))) \bigwedge (t \notin T) \}$$

Lemma 4

In case of an infinite universe of discourse the complement of a relation is **not** a relation.

Corollary 3

The complement of a relation is a *partial unary operation* on the set of relations .

Corollary 4

The active complement of a relation is a *total unary operation* on the set of relations .

Note 3

In the following (if not noted explicitly otherwise) we will use only the active complement which we will call shortly *the complement* .

2.3.2 Binary operations

Definition 6

Let $R_i = \langle\, A_i\,,\, D_i\,,\, T_i\,\rangle$ be relations with equal cardinalities of A_i such that : $\boldsymbol{D}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \;\cap\; \boldsymbol{D}_2(\boldsymbol{\pi}_2(\boldsymbol{A}_2)) \neq \boldsymbol{\emptyset}$
(π_i being appropriate permutations) .

A $\boldsymbol{\pi}$ **- intersection** of relations R_i is the relation noted $\boldsymbol{R}_1 \cap_\pi \boldsymbol{R}_2 = \langle\, A\,,\, D\,,\, T\,\rangle$ such that :

1. $|\,\boldsymbol{A}\,| = |\,\boldsymbol{A}_1\,| = |\,\boldsymbol{A}_2\,|$
2. $\boldsymbol{D}(\boldsymbol{A}) \;\cap\; \boldsymbol{D}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \;\cap\; \boldsymbol{D}_2(\boldsymbol{\pi}_2(\boldsymbol{A}_2)) \neq \boldsymbol{\emptyset}$
3. $T = \{\, t : A \rightarrow D(A) \mid ((\exists u \in T_1) \wedge (\exists v \in T_2))$ $(t(A) = u(\pi_1(A_1)) = v(\pi_2(A_2)))\}$

$$\Updownarrow$$

Notation 19 $\quad \boldsymbol{T} = \boldsymbol{T}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \;\cap\; \boldsymbol{T}_2(\boldsymbol{\pi}_2(\boldsymbol{A}_2))$

A $\boldsymbol{\pi}$ **- difference** of relations R_1 and R_2 is the relation noted $\boldsymbol{R}_1 -_\pi \boldsymbol{R}_2 = \langle\, A\,,\, D\,,\, T\,\rangle$ such that :

1. $|\,\boldsymbol{A}\,| = |\,\boldsymbol{A}_1\,|$
2. $\boldsymbol{D}(\boldsymbol{A}) \;\cap\; \boldsymbol{D}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \neq \boldsymbol{\emptyset}$
3. $T = \{\, t : A \rightarrow D(A) \mid$ $((\exists u \in T_1)(t(A) = u(\pi_1(A_1)))) \wedge$ $((\forall v \in T_2)(t(A) \neq v(\pi_2(A_2))))\}$

$$\Updownarrow$$

Notation 20 $\quad \boldsymbol{T} = \boldsymbol{T}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \;-\; \boldsymbol{T}_2(\boldsymbol{\pi}_2(\boldsymbol{A}_2))$

A $\boldsymbol{\pi}$ **- union** of relations R_i is the relation noted $\boldsymbol{R}_1 \cup_\pi \boldsymbol{R}_2 = \langle\, A\,,\, D\,,\, T\,\rangle$ such that :

1. $|\,\boldsymbol{A}\,| = |\,\boldsymbol{A}_1\,| = |\,\boldsymbol{A}_2\,|$
2. $\boldsymbol{D}(\boldsymbol{A}) \;\cap\; \boldsymbol{D}_1(\boldsymbol{\pi}_1(\boldsymbol{A}_1)) \;\cap\; \boldsymbol{D}_2(\boldsymbol{\pi}_2(\boldsymbol{A}_2)) \neq \boldsymbol{\emptyset}$
3. $T = \{\, t : A \rightarrow D(A) \mid ((\exists u \in T_1) \vee (\exists v \in T_2))$ $((t(A) = u(\pi_1(A_1)))$ $\vee (t(A) = v(\pi_2(A_2))))\}$

$$\Updownarrow$$

Notation 21 $\quad \boldsymbol{T} = \boldsymbol{T}_1(\boldsymbol{A}_1) \;\cup\; \boldsymbol{T}_2(\boldsymbol{A}_2)$

Convention 3

In the case of permutations π_i being *identities* we will omit the prefix $\boldsymbol{\pi}$ **-** and speak shortly only about (respectively) : the **intersection** ,
the **difference** ,
the **union** ,

and note them respectively : $\boldsymbol{R}_1 \cap \boldsymbol{R}_2$, $\boldsymbol{R}_1 - \boldsymbol{R}_2$ and $\boldsymbol{R}_1 \cup \boldsymbol{R}_2$.

The *insertion of additional m - tuples* then corresponds to the **union** of the appropriate relations, the *deletion of existing m - tuples* corresponds to the **difference** of the appropriate relations, and the *alteration of components of any of existing m - tuples* can be expressed as a *deletion* followed by an *insertion* .

2.3.3 Algebraic properties of relational set operations

From the point of view of the *general algebra* it can be easily shown that :

The **unary** operation : the *complement* (no active) and
the **binary** operations : the *intersection*
the *difference*
the *union*
are **partial** *operations* .

The *intersection* and the *union* are : **commutative** and **associative** .

Thanks to the *associativity* we can generalize these two operations to arbitrary *higher* **arity** $\boldsymbol{n}$.

Let us denote by $R_\emptyset$ the following *empty relation* : $R_\emptyset = \langle\, \emptyset\, ,\, \emptyset\, ,\, \emptyset\, \rangle$.

Now we can make the *intersection* and the *difference* **total** operations (like the *active complement*) by posing for the cases not covered by the **Definition 6** :

$$\boldsymbol{R}_1 \cap \boldsymbol{R}_2 = \boldsymbol{R}_\emptyset$$
$$\boldsymbol{R}_1 - \boldsymbol{R}_2 = \boldsymbol{R}_1$$

Especially :

$$\boldsymbol{R}_1 \cap \boldsymbol{R}_\emptyset = \boldsymbol{R}_\emptyset$$
$$\boldsymbol{R}_1 - \boldsymbol{R}_\emptyset = \boldsymbol{R}_1$$
$$\boldsymbol{R}_\emptyset - \boldsymbol{R}_1 = \boldsymbol{R}_\emptyset \ .$$

Adding :

$$\boldsymbol{R}_1 \cup \boldsymbol{R}_\emptyset = \boldsymbol{R}_1$$

$R_\emptyset$ acts as :

- the **zero** (null) for the *intersection*
- the *right* **unity** for the *difference*
- the *left* **zero** for the *difference*
- the **unity** for the *union* .

Resume Thanks to the previous generalizations we can sum up :

The set of relations is :

- a partly ordered **groupoid** with respect to the *difference*
- a partly ordered associative *Abelian* **groupoid** with respect to the *union*
- a partly ordered *Abelian* **semigroup** with respect to the *intersection* .

Convention 4

In the following we will call the set operations over relations the **basic** operations .

What was really new in Codd RMD were the other kinds of operations over the relations which we will introduce in the next and which we will call :

2.4 Higher operations

2.4.1 Projections

Definition 7

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. $T_1 = \{\, t : A_1 \rightarrow D_1(\,A_1\,) \mid (\,\exists\, u \in T\,)\,(\,t\,(\,A_1\,) = u\,(\,A_1\,)\,)\,\}$
$\Updownarrow$

Notation 22 $\boldsymbol{T}_1 = \boldsymbol{T}\,[\,\boldsymbol{A}_1\,]$

EXAMPLE 8

US PRESIDENTS	
President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

R
President
Clinton
Bush
Reaggan
Carter
Ford
Nixon
Johnson
Keneddy

Note 4

R = US PRESIDENTS [**President**]

Now we can return to the generalization of the **inclusion** of relations :

Definition 4

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

Then we will say that the relation R_1 is a **subrelation** of the relation R_2 — what we will note : $R_1 \subset R_2$.

Example 9

MUSP
President
Keneddy

US Presidents	
President	**Vice–President**
Clinton	Gore
Bush	Quale
Reaggan	Bush
Carter	Mondale
Ford	Rockefeller
Nixon	Ford
Johnson	Humprey
Keneddy	Johnson

Note 5

MUSP = Murdered US Presidents [**President**]

MUSP $\subset$ US Presidents

Lemma 5

There exists a **minimal element** in the set of relations, namely the the relation $R_\emptyset$.

Definition 8

Let $R = \langle A , D , T \rangle$ be a relation and $A_1 \subset A$.
The **antiprojection** 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 :

$$(\boldsymbol{A}_2 = \boldsymbol{A} - \boldsymbol{A}_1)$$

1. $\boldsymbol{D}_1 = \boldsymbol{D} / \boldsymbol{A}_1$
2. $T_1 = \{ t : A_1 \to D_1 (A_1) \mid (\forall u \in T [A_2]) ((t , u) \in T) \}$

$$\Updownarrow$$

Notation 23 $\quad \boldsymbol{T}_1 = \boldsymbol{T} \,]\, \boldsymbol{A}_1 [$

Notation 24

(t, u) denotes mapping $w : A \rightarrow D(A)$ such that :

$$\boldsymbol{w}(\boldsymbol{A}_1) = \boldsymbol{t}(\boldsymbol{A}_1) \text{ and}$$
$$\boldsymbol{w}(\boldsymbol{A}_2) = \boldsymbol{u}(\boldsymbol{A}_2) \ .$$

Remark 6

The set $T[A_1]$ can be expressed also as follows :

$$(A_2 = A - A_1)$$

$$T[A_1] = \{ t : A_1 \rightarrow D_1(A_1) \mid (\exists u \in T[A_2])((t, u) \in T) \}$$

Comparing it with the expression of the set $T\,]A_1[$:

$$T\,]A_1[= \{ t : A_1 \rightarrow D_1(A_1) \mid (\forall u \in T[A_2])((t, u) \in T) \}$$

we can state the following :

Lemma 6

The antiprojection differs from the projection only by replacing the *existential* **quantifier** by the *general* one .

Corollary 5

In general case the following *inclusion* holds : $\boldsymbol{R}\,]\boldsymbol{A}_1[\subset \boldsymbol{R}[\boldsymbol{A}_1]$

Example 10

Languages spoken	
Name	**Language**
Boris	Russian
François	French
John	English
Peter	English
Peter	French
Peter	German
Peter	Russian
Wolfgan	German

Languages spoken [**Name**]
Name
Boris
François
John
Peter
Wolfgan

Languages spoken] **Name** [
Name
Peter

Definition 9

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{ 1 , 2 \}$, be two relations with disjunctive sets of attribute names : $A_1 \cap A_2 = \emptyset$.
(This can be always fulfilled by renaming the attribute names.)
The **cartesian product** of the relations R_i is the relation noted $\boldsymbol{R}_1 \times \boldsymbol{R}_2 = \langle \boldsymbol{A} , \boldsymbol{D} , \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{A} = \boldsymbol{A}_1 \cup \boldsymbol{A}_2$
2. $D / A_i = D_i / A_i \quad , \quad i \in \{ 1 , 2 \}$

$$\Updownarrow$$

Notation 25 $\quad \boldsymbol{D} = \boldsymbol{D}_1 \cup \boldsymbol{D}_2$

3. $T = \{ t : A \to D (A) \mid (t (A_i) = u_i (A_i)) (u_i \in T_i) \}$

$$\Updownarrow$$

Notation 26 $\quad \boldsymbol{T} = \boldsymbol{T}_1 \times \boldsymbol{T}_2$

Remark 7

The *cartesian product* is in certain sense an **inverse** operation to the operation *projection* .
But in general case we have only the following lemma :

Lemma 7

Let $R = \langle A , D , T \rangle$ be a relation and $A_i \subset A$, $i \in \{ 1 , 2 \}$, such that : $A_2 = A - A_1$.
Then : $\quad \boldsymbol{R} \subset (\boldsymbol{R} [\boldsymbol{A}_1] \times \boldsymbol{R} [\boldsymbol{A}_2])$

EXAMPLE 11

R	
Name	**Language**
John	English
Paul	French

R [**Name**]
Name
John
Paul

R [**Language**]
Language
English
French

R [**Name**] × R [**Language**]	
Name	**Language**
John	English
John	French
Paul	English
Paul	French

2.4.2 Joins

Definition 10

Let $R_i = \langle A_i, D_i, T_i \rangle$, $i \in \{1, 2\}$, be two relations and $B_i \subset A_i$, $i \in \{1, 2\}$, be two sets of attributes such that $D_1(B_1) \cap D_2(\pi(B_2)) \neq \emptyset$.

The **join** of the relations R_1 and R_2 , *according to the attributes (sets)* B_1 *and* B_2 , *with respect to the* **equality** , is the relation noted $\boldsymbol{R}_1 *_{B_1 = \pi(B_2)} \boldsymbol{R}_2 = \langle \boldsymbol{A}, \boldsymbol{D}, \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{A} = \boldsymbol{A}_1 \cup \boldsymbol{A}_2$

2. $D(a_j) = D_1(a_j) \cup D_2(a_j) \quad , \quad \forall j \in |\widehat{A}|$

$$\Updownarrow$$

Notation 27 $$\boldsymbol{D} = \boldsymbol{D}_1 \cup \boldsymbol{D}_2$$

3. $\boldsymbol{T} = \{ t : A \to D(A) \mid ((\forall i \in \{1, 2\})(\exists u_i \in T_i)) \; ((t(A_j) = u_j(A_j)) \wedge (u_1(B_1) = u_2(\pi(B_2))))\}$

$$\Updownarrow$$

Notation 28 $$\boldsymbol{T} = \boldsymbol{T}_1 *_{B_1 = \pi(B_2)} \boldsymbol{T}_2$$

Convention 5

In case of π being the identity, equality of B_i and such that they are **maximal** (in *set inclusion* sense) with such a property, we will omit the *index* $_{B_1 = \pi(B_2)}$ by the $*$ and call the join shortly the **natural join** of $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

EXAMPLE 12

$\boldsymbol{R}_1$	
Head *Name*	**Department** *Number*
Ladislav	21
Marcel	23
Emil	24
Zdeněk	25
Václav	27

$\boldsymbol{R}_2$	
Department	
Number	*Name*
21	Numerical optimization
22	Knowledge based systems
23	Neural networks
24	Non-linear modelling
25	Applied Linear Algebra

$\boldsymbol{R}_1 * \boldsymbol{R}_2$		
Head	**Department**	
Name	*Name*	*Number*
Ladislav	Numerical optimization	21
Marcel	Neural networks	23
Emil	Non-linear modelling	24
Zdeněk	Applied Linear Algebra	25

Convention 6

In the next 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** .

We can generalize the previous definition of join by replacing the equality by an arbitrary (binary) relation …

Definition 11

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{1, 2\}$, be two relations and $B_i \subset A_i$, $i \in \{1, 2\}$, be two attributes such that there exists a *binary relation* (in *classical mathematical* sense …) Θ defined on the *cartesian product* of cartesian products of corresponding domains $D_i ({a^i}_j)$, ${a^i}_j \in B_i$, which we will denote by : $\boldsymbol{D}_1{}^X (\boldsymbol{B}_1) \times \boldsymbol{D}_2{}^X (\boldsymbol{B}_2)$ where :

Notation 29 $$\boldsymbol{D}_i{}^X (\boldsymbol{B}_i) = \times_{{a^i}_j \in B_i} \boldsymbol{D}_i ({\boldsymbol{a}^i}_j)$$

The **join** of the relations R_1 and R_2 , *according to the attributes* B_1 *and* B_2 , *with respect to the relation* Θ , is the relation noted $\boldsymbol{R}_1 *_{\Theta (B_1 , B_2)} \boldsymbol{R}_2 = \langle \boldsymbol{A} , \boldsymbol{D} , \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{A} = \boldsymbol{A}_1 \cup \boldsymbol{A}_2$
2. $\boldsymbol{D} = \boldsymbol{D}_1 \cup \boldsymbol{D}_2$
3. $\boldsymbol{T} = \{ \, t : A \rightarrow D (A) \mid ((\forall i \in \{1, 2\}) (\exists u_i \in T_i)) \; ((t (A_j) = u_j (A_j)) \wedge ((u_1 (B_1) , u_2 (B_2)) \in \Theta)) \}$

$$\Updownarrow$$

Notation 30 $$\boldsymbol{T} = \boldsymbol{T}_1 *_{\Theta (B_1 , B_2)} \boldsymbol{T}_2$$

Convention 7

The join *with respect to the relation* Θ will be called the Θ**-join** .

EXAMPLE 13

$\boldsymbol{R}_1$		
Name	**Age** *(years)*	**Annual Salary** *(thousands USD)*
Peter	33	50
John	40	15
Július	42	9

$\boldsymbol{R}_1 *_{\textbf{Age} < \textbf{Annual Salary}} \boldsymbol{R}_1$		
Name	**Age** *(years)*	**Annual Salary** *(thousands USD)*
Peter	33	50

Finally we can even generalize the notions of the *intersection* — **Definition 6** , of the *cartesian product* — **Definition 9** and of the *join with respect to the equality* (*natural join*) — **Definition 10** into the following one :

Definition 12

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{ 1 , 2 \}$, be two relations .
The **join** of the relations R_1 and R_2 is the relation noted
$\boldsymbol{R}_1 \circledast \boldsymbol{R}_2 = \langle \boldsymbol{A}, \boldsymbol{D}, \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{A} = \boldsymbol{A}_1 \cup \boldsymbol{A}_2$
2. $\boldsymbol{D} = \boldsymbol{D}_1 \cup \boldsymbol{D}_2$
3. $\boldsymbol{T} = \{ t : A \rightarrow D(A) \mid ((\forall i \in \{1, 2\}) (\exists u_i \in T_i)) \; (t(A_j) = u_j(A_j))\}$

$$\Updownarrow$$

Notation 31 $\boldsymbol{T} = \boldsymbol{T}_1 \circledast \boldsymbol{T}_2$

Lemma 8

In case we have in the **Definition 12** :

1. $\boldsymbol{A}_1 = \boldsymbol{A}_2$ *then* $\boldsymbol{R}_1 \circledast \boldsymbol{R}_2 = \boldsymbol{R}_1 \cap \boldsymbol{R}_2$
2. $\boldsymbol{A}_1 \cap \boldsymbol{A}_2 = \boldsymbol{\emptyset}$ *then* $\boldsymbol{R}_1 \circledast \boldsymbol{R}_2 = \boldsymbol{R}_1 \times \boldsymbol{R}_2$
3. $(\exists \boldsymbol{B}_i \subset \boldsymbol{A}_i)$, $i \in \{1, 2\}$,
   such that $\boldsymbol{B}_1 = \boldsymbol{B}_2$ and
   $\boldsymbol{B}_i$ are *maximal* (set inclusion sense)
   then $\boldsymbol{R}_1 \circledast \boldsymbol{R}_2 = \boldsymbol{R}_1 * \boldsymbol{R}_2$

Example 14

Languages spoken	
Name	**Language**
Boris	Russian
François	French
John	English
Peter	English
Peter	French
Peter	German
Peter	Russian
Wolfgan	German

Languages required
Language
English

(Languages spoken $\circledast$ Languages required) [**Name**]
Name
John
Peter

Remark 8

It is also possible to define so called (*operator* of the) **selection** :

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

$A_2 \subset A_1$ and

$| T_2 | = 1$.

Let us denote by y the *image* of the (unique) element of T_2 :

$$y = u (A_2) , \quad u \in T_2 .$$

The **selection** , noted $\boldsymbol{\sigma}_{A_2 = y}$, is defined as :

$$\boldsymbol{\sigma}_{A_2 = y} (\boldsymbol{R}_1) = \boldsymbol{R}_1 \circledast \boldsymbol{R}_2$$

or *equivalently* :

$$\boldsymbol{\sigma}_{A_2 = y} (\boldsymbol{R}_1) = \langle \boldsymbol{A}_1 , \boldsymbol{D}_1 , \boldsymbol{T} \rangle \text{ such that :}$$

$$\boldsymbol{T} = \{ t : A_1 \rightarrow D_1 (A_1) \mid t (A_2) = y \}$$

Example 15

R	
Name	**Language**
Boris	Russian
François	French
John	English
Peter	English
Peter	French
Peter	Russian
Peter	German
Wolfgan	German

$\boldsymbol{\sigma}_{\textbf{Language} = \text{English}}$ (R)	
Name	**Language**
John	English
Peter	English

Definition 13

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

The **sum** of the relations R_1 and R_2 is the relation noted

$\boldsymbol{R}_1 + \boldsymbol{R}_2 = \langle \boldsymbol{A} , \boldsymbol{D} , \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{A} = \boldsymbol{A}_1 \cup \boldsymbol{A}_2$
2. $\boldsymbol{D} = \boldsymbol{D}_1 \cup \boldsymbol{D}_2$
3. $\boldsymbol{T} = \{ t : A \rightarrow D (A) \mid (((\forall i \in \{1, 2\}) (\exists u_i \in T_i)) (\bigvee_{j=1}^{2} t (A_j) = u_j (A_j))) \wedge ((\forall a \in A) (\exists k \in \{1, 2\}) (t (a) \in \alpha D (a , R_k))) \}$

$$\Updownarrow$$

Notation 32 $\quad \boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2$

Remark 9

Comparing the expression for the $\boldsymbol{T}_1 \circledast \boldsymbol{T}_2$:

$$\boldsymbol{T}_1 \circledast \boldsymbol{T}_2 = \{ t : A \rightarrow D(A) \mid ((\forall i \in \{1, 2\})(\exists u_i \in T_i)) \\ (t(A_j) = u_j(A_j))\}$$

which can be also expressed as follows :

$$\boldsymbol{T}_1 \circledast \boldsymbol{T}_2 = \{ t : A \rightarrow D(A) \mid ((\forall i \in \{1, 2\})(\exists u_i \in T_i)) \\ (\textstyle\bigwedge_{j=1}^{2} t(A_j) = u_j(A_j))\}$$

or even as :

$$\boldsymbol{T}_1 \circledast \boldsymbol{T}_2 = \{ t : A \rightarrow D(A) \mid (((\forall i \in \{1, 2\})(\exists u_i \in T_i)) \\ (\textstyle\bigwedge_{j=1}^{2} t(A_j) = u_j(A_j))) \wedge \\ ((\forall a \in A)(\exists k \in \{1, 2\})(t(a) \in \alpha D(a, R_k)))\}$$

and the expression for $\boldsymbol{T}_1 + \boldsymbol{T}_2$:

$$\boldsymbol{T}_1 + \boldsymbol{T}_2 = \{ t : A \rightarrow D(A) \mid (((\forall i \in \{1, 2\})(\exists u_i \in T_i)) \\ (\textstyle\bigvee_{j=1}^{2} t(A_j) = u_j(A_j))) \wedge \\ ((\forall a \in A)(\exists k \in \{1, 2\})(t(a) \in \alpha D(a, R_k)))\}$$

we can state the following :

Lemma 9

The sum of the relations differs from the join of the relations only by replacing the **conjunction** by the corresponding **disjunction** .

Corollary 6

In general case the following *inclusion* holds :

$$\boldsymbol{R}_1 \circledast \boldsymbol{R}_2 \subset \boldsymbol{R}_1 + \boldsymbol{R}_2$$

Definition 14

Let $R_i = \langle A_i, D_i, T_i \rangle$, $i \in \{1, 2\}$, be two relations such that there exists a join $R_1 *_{\Theta(B_1, B_2)} R_2$.

Denote : $A_i' = A_i - B_i$, $i \in \{1, 2\}$.

The **composition** of the relations R_1 and R_2 , *according to the attributes* B_1 *and* B_2 , *with respect to the relation* Θ , is the relation noted $\boldsymbol{R}_1 \cdot_{\Theta(B_1, B_2)} \boldsymbol{R}_2$ and defined as :

$$(\boldsymbol{R}_1 *_{\Theta(B_1, B_2)} \boldsymbol{R}_2)[\boldsymbol{A}_1' \cup \boldsymbol{A}_2']$$

Convention 8

The composition *with respect to the relation* Θ will be called the Θ**- composition** .

Convention 9

In case the corresponding join will be the *natural one*, we will speak about the **natural composition** and we will note it simply as :

$$\boldsymbol{R}_1 \cdot \boldsymbol{R}_2 \quad .$$

Now we can return to the the situation of the EXAMPLE 12 :

EXAMPLE 16

R_1	
Head *Name*	**Department** *Number*
Ladislav	21
Marcel	23
Emil	24
Zdeněk	25
Václav	27

R_2	
Department	
Number	*Name*
21	Numerical optimization
22	Knowledge based systems
23	Neural networks
24	Non-linear modelling
25	Applied Linear Algebra

$R_1 \cdot R_2$	
Head	**Department**
Name	*Name*
Ladislav	Numerical optimization
Marcel	Neural networks
Emil	Non-linear modelling
Zdeněk	Applied Linear Algebra

2.4.3 Restrictions

Definition 15

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{ 1 , 2 \}$, be two relations and $B_i \subset A_i$, $i \in \{ 1 , 2 \}$, be two attributes of equal *cardinalities* $(\mid B_1 \mid = \mid B_2 \mid)$.

The **restriction** of the relation R_1 *by* the relation R_2 , *according to the attributes* B_1 *and* B_2 , is the relation R noted $\boldsymbol{R}_1 \ {}_{B_1} \mid {}_{B_2} \ \boldsymbol{R}_2$ such that :

1. $\boldsymbol{R} \subset \boldsymbol{R}_1$
2. $\boldsymbol{R} [\boldsymbol{B}_1] \subset \boldsymbol{R}_2 [\boldsymbol{B}_2]$
3. $\boldsymbol{R}$ is the *maximal* relation satisfying the previous two conditions .

Convention 10

In the next we will use alternatively also the name B_1 , B_2 – **restriction** of the relation R_1 *by* (the relation) R_2 .

Convention 11

When $B_1 = A_2$ we will speak shortly only about the **restriction** of the relation R_1 *by* (the relation) R_2 .

And again we can give a generalization of the restriction with respect to a general (binary) relation ...

Definition 16

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{ 1 , 2 \}$, be two relations and $B_i \subset A_i$, $i \in \{ 1 , 2 \}$, be two attributes such that there exists a *binary relation* (in *classical mathematical* sense ...) Θ defined on the *cartesian product* of cartesian products of corresponding domains $D_i ({a^i}_j)$, that is :

$$\Theta \subset \boldsymbol{D}_1{}^X (\boldsymbol{B}_1) \times \boldsymbol{D}_2{}^X (\boldsymbol{B}_2)$$

The **restriction** of the relation R_1 *by* the relation R_2 , *according to the attributes* B_1 *and* B_2 , *with respect to the relation* Θ , is the relation noted $\boldsymbol{R}_1 \mid_{\Theta (B_1 , B_2)} \boldsymbol{R}_2 = \langle \boldsymbol{A}_1 , \boldsymbol{D}_1 , \boldsymbol{T} \rangle$ such that :

1. $\boldsymbol{T} = \{ t \in T_1 \mid ((\exists u \in T_2) \ (t (B_1) , u (B_2)) \in \Theta)) \}$
2. It is the *maximal* relation satisfying the previous condition .

Convention 12

In the next we will use alternatively also the name
B_1, B_2 – Θ –**restriction** of the relation R_1 *by* (the relation) R_2 .

EXAMPLE 17

$\boldsymbol{R}_1$		
Name	**Language**	**Level**
Boris	Russian	high
François	French	excellent
John	English	superior
Peter	English	superior
Peter	French	medium
Peter	German	medium
Peter	Russian	medium
Wolfgan	German	excelent

$\boldsymbol{R}_2$	
Language	**Level**
English	superior
French	excellent
German	excellent

$\boldsymbol{R}_1 \ _{\textbf{Language , Level}} \mid {}_{\textbf{Language , Level}} \ \boldsymbol{R}_2$		
Name	**Language**	**Level**
François	French	excellent
John	English	superior
Peter	English	superior
Wolfgan	German	excelent

2.4.4 Division

Definition 16

Let $R_i = \langle A_i , D_i , T_i \rangle$, $i \in \{1 , 2\}$, be two relations such that : $((A_2 \subset A_1) \wedge (A_2 \neq A_1))$ and

$D_1(A_2) \cap D_2(\pi(A_2)) \neq \emptyset$

(π being an appropriate permutation)

Denote : $A_3 = A_1 - A_2$.

The **division** of the relation R_1 *by* the relation R_2 is the relation noted $\boldsymbol{R}_1 \div \boldsymbol{R}_2 = \langle \boldsymbol{A}_3 , \boldsymbol{D}_3 , \boldsymbol{T}_3 \rangle$ such that :

1. $\boldsymbol{D}_3 = \boldsymbol{D}_1 / \boldsymbol{A}_3$
2. $\boldsymbol{T}_3 = \{ t : A_3 \rightarrow D_3(A_3) \mid ((\forall v \in T_2)(\exists u \in T_1))$
$((t(A_3) = u(A_3)) \wedge (u(A_2) = v(A_2)))\}$

EXAMPLE 18

$\boldsymbol{R}_1$

Name	Language
Boris	Russian
François	French
John	English
Peter	English
Peter	French
Peter	Russian
Peter	German
Wolfgan	German

$\boldsymbol{R}_2$

Language
English
French
German
Russian

$\boldsymbol{R}_1 \div \boldsymbol{R}_2$

Name
Peter

$\boldsymbol{R}_3$

Language
English
French
German

$\boldsymbol{R}_1 \div \boldsymbol{R}_3$

Name
Peter

2.4.5 Algebraic properties of higher relational operations

From the point of view of the *general algebra* it can be shown that :

The *(pseudo -)* **unary** operations (over set of relations $\Re$) : the *projection*
the *antiprojection*

(in fact *partial* mappings from the cartesian product :
$\Re \times \aleph$ into $\Re$ where $\aleph$ is the *set of attribute names*)

can be made **total** by *leaving the condition* : $\boldsymbol{A}_1 \subset \boldsymbol{A}$
and *requiring* that for the *projection* we have :

$$\boldsymbol{R}\,[\,\boldsymbol{A}_1\,] = \langle\ \boldsymbol{A}_1 \cap \boldsymbol{A}\,,\ \boldsymbol{D}\,/\,(\,\boldsymbol{A}_1 \cap \boldsymbol{A}\,)\,,\ \boldsymbol{T}\,[\,\boldsymbol{A}_1 \cap \boldsymbol{A}\,]\ \rangle$$

respectively that for the *antiprojection* we have :

$$\boldsymbol{R}\,]\,\boldsymbol{A}_1\,[\ = \langle\ \boldsymbol{A}_1 \cap \boldsymbol{A}\,,\ \boldsymbol{D}\,/\,(\,\boldsymbol{A}_1 \cap \boldsymbol{A}\,)\,,\ \boldsymbol{T}\,]\,\boldsymbol{A}_1 \cap \boldsymbol{A}\,[\ \rangle$$

The **binary** operations :

the *cartesian product* (by leaving the condition :
$A_1 \cap A_2 = \emptyset$
we obtain the definition of *join*)
the *join*
the *natural join*
the *natural composition*
the *sum*

and the *(pseudo -)* **binary** operation : the *join with respect to the equality*
(in fact partial mappings from the cartesian product :
$\Re^2 \times \aleph^2$ into $\Re$)

are **commutative** and **associative** (thanks to the *associativity* we can generalize them to an arbitrary *higher* **arity** $\boldsymbol{n}$) .

The *(pseudo -)* **binary** operations :

the Θ - *join*
the Θ - *composition*
the $B_1\,, B_2 - \Theta -$ *restriction*

(in fact partial mappings from the cartesian product :
$\Re^2 \times \aleph^2 \times$ *Mathematical Relations* into $\Re$)

the $B_1\,, B_2 -$ *restriction*
the *restriction*
the *division*

(in fact partial mappings from the cartesian product :
$\Re^2 \times \aleph^2$ into $\Re$)

are NOT **commutative** .

$\boldsymbol{R}_{\emptyset}$ acts as :

- the **unity** for the :

  cartesian product
  join with respect to the equality
  natural join
  join
  natural composition
  sum
  division
  restriction

Resume Thanks to the previous generalizations we can sum up :

The set of relations is :

- a partly ordered **groupoid** with respect to the :

  Θ - *join*
  Θ - *composition*
  $B_1 , B_2 \ - \Theta - restriction$
  $B_1 , B_2 \ - \ restriction$
  restriction
  division

- a partly ordered associative *Abelian* **groupoid** with respect to the :

  join with respect to the equality

- a partly ordered *Abelian* **semigroup** with respect to the :

  cartesian product
  join
  natural join
  natural composition
  sum

Bibliography

[Anvari - Rose, 87] ANVARI M, ROSE G. F.: *Fuzzy relational databases.* In: Analysis of Fuzzy Information, Volume II (Editor: Bezdek), CRC Press, Boca Raton, 1987

[Buckles - Petry, 82] BUCKLES B. P., PETRY F. E.: *A fuzzy representation of data for relational databases.* Fuzzy Sets and Systems 7 (1982), 213-226

[Buckles - Petry, 84] BUCKLES B. P., PETRY F. E.: *Extending the fuzzy database with fuzzy numbers.* Information Sciences 34 (1984), 145-155

[Buckles - Petry - Sachar, 89] BUCKLES B. P., PETRY F. E., SACHAR H. S.: *A domain calculus for fuzzy relational databases.* Fuzzy Sets and Systems 29 (1989), 327-340

[Codd, 1970] CODD E. F.: *A Relational Model of Data for Large Shared Data Banks.* Communications of the ACM 13 (1970), 6, 377-387

[Codd, 1979] CODD E. F.: *Extending the database relational model to capture more meaning.* ACM Transactions on Database Systems 4 (1979), 4, 397-434

[Dubois - Lang - Prade, 1991] DUBOIS D., LANG J., PRADE H.: *Timed Possibilistic Logic.* Fundamenta Informaticae XV, 1991, 3-4, 211-234

[Gallaire - Minker - Nicolas, 1978] GALLAIRE H., MINKER J., NICOLAS J.: *Logic and Databases.* New York, Plenum Press, 1978

[Grant, 1979] GRANT J.: *Partial value in a tabular database model.* Information Processing Letters 9 (1979), 2, 97-99

[Grant, 1980] GRANT J.: *Incomplete information in a relational database.* Fundamenta Informaticae 3 (1980), 3, 363-378

[Grant - Minker, 1986] GRANT J., MINKER J.: *Answering queries in indefinite databases and the null value problem.* Advances in Computing Research (Editor: Kanellakis P.) The Theory of Databases, Volume 3, 247-267, JAI Press, 1986

[Hájek - Harmanec, 1992] HÁJEK P., HARMANEC D: *On Belief Functions (The present state of Dempster-Shafer theory).* In: Advances in Artificial Intelligence (Editor: Mařík V.), Berlin, Springer Verlag, 1992, 286-307

[Hájek, 1993] HÁJEK P.: *On Logics of Approximate Reasoning.* Neural Network Word 6 (1993), 733–744.

[Hájek - Havránek - Jiroušek, 1992] HÁJEK P., HAVRÁNEK T., JIROUŠEK R.: *Uncertain Information Processing in Expert Systems.* Boca Raton, CRC Press, 1992, 285 pages, ISBN 0-8493-6368-3

[Imielinsky - Lipsky, 1984] IMIELINSKY T., LIPSKY W. (JR.): *Incomplete information in relational databases.* Journal of ACM 31 (1984), 4, 761-791

[Imielinsky - van der Meyden - Vadaparty, 1995] IMIELINSKY T., VAN DER MEYDEN R., VADAPARTY K.: *Complexity tailored design - a new methodology for databases with incomplete information.* Journal of Computer and System Sciences 53 (1995), 3, 405-432

[Ioannidis - Ramakrishnan, 1995] IOANNIDIS Y. E, RAMAKRISHNAN R.: *Containment of Conjuctive Queries: Beyond Relations as Sets.* ACM Transactions on Database Systems 20 (1995), 3, 288-324

[Ioannidis - Wong, 1991] IOANNIDIS Y. E., WONG E.: *Toward an algebraic theory of recursion.* Journal of ACM 38 (April 1991), 2, 329-381

[Jagadish, 1995] JAGADISH H.V.: *The INCINERATE Data Model.* ACM Transactions on Database Systems 20 (March 1995), 1, 71-110

[Koubarakis, 1994] KOUBARAKIS M.: *Database models for infinite and indefinite temporal information.* Information Systems 19 (1994), 2, 141-173

[Lim - Srivastava - Shekhar, 1993] LIM E.-P., SRIVASTAVA J., SHEKHAR S. : *Resolving Attribute Incompatibility in Database Integration - An Evidential Reasoning Approach.* Technical Report TR 93-14, University of Minnesota, Department of Computer Science, July 1993, 36 pages

[Lipsky, 1979] LIPSKY W.: *On semantic issue connected with incomplete information systems.* ACM Transactions on Database Systems 4 (1979), 3, 262-296

[Liu - Sunderraman, 1990] LIU K. C., SUNDERRAMAN R.: *Indefinite and maybe information in relational databases.* ACM Transactions on Database Systems 15 (1990), 1, 1-39

[Liu - Sunderraman, 1991] LIU K. C., SUNDERRAMAN R.: *A generalized relational model for indefinite and maybe information.* IEEE Transactions on Knowledge and Data Engineering 3 (1991), 1, 65-77

[Medina - Pons - Vila, 1994] MEDINA J. M., PONS O., VILA M. A.: *GEFRED - A generalized model of fuzzy relational databases.* Information Sciences 76 (1994), 87-109

[Morrissey, 1990] MORRISSEY J. M.: *Imprecise information and uncertainty in information systems.* ACM Transactions on Information Systems 8 (1990), 2, 159-180

[Ola, 1992] Ola A.: *Relational databases with exclusive disjunctions.* In: Proceedings of IEEE Data Engineering Conference 1992, 328-336

[Ola, Ozsoyoglu, 1993] Ola A., Ozsoyoglu G.: *Incomplete relational database models based on intervals.* IEEE Transactions on Knowledge and Data Engineering 5 (1993), 2, 293-308

[Prade - Testemale, 1984] Prade H., Testemale C.: *Generalizing database relational algebra for the treatment of incomplete/uncertain information and vague queries.* Information Sciences 34 (1984), 115-143

[Reiter, 1984] Reiter R.: *Toward a Logical Reconstruction of Relational Database Theory.* In: On Conceptual Modelling - Perspective from Artificial Intelligence, Databases and Programming Languages (Editors: Brodie M. L., Mylopoulos J., Schmidt J. W.), New York, Springer-Verlag, 1984, 191-238

[Reiter, 1986] Reiter R.: *A sound and sometimes complete query evaluation algorithm for relational databases with null values.* Journal of ACM 33 (1986), 2, 349-370

[Rundensteiner - Hawkes - Bandler, 89] Rundensteiner E. A., Hawkes L. W., Bandler W.: *On nearness measures in fuzzy relational data models.* International Journal of Approximate Reasoning 3 (1989), 267-298

[Schwartz, 1971] Schwartz J. T.: *Abstract and concrete problems in the theory of files.* In: Courant Computer Science Symposium - Data Base Systems (May 24th-25th 1971), Englewood Cliffs - New Jersey, Prentice-Hall 1972

[Štuller, 1995a] Štuller J. : *Inconsistency Resolution in the Databases Integration.* In : Proceedings of the International Scientific Seminar DATABASE SYSTEMS ("DATABÁZOVÉ SYSTÉMY"), an accompanying program of the exposition COFAX '95, 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, 1995b] 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, 1995c] 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, 1995d] 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, 1995f] 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

[Tseng - Chen - Yang, 1993a] Tseng F. S. C., Chen A. L. P., Yang W. P.: *Answering heterogeneous database queries with degrees of uncertainty.* International Journal on Distributed and Parallel Databases (1993), 281-302

[Tseng - Chen - Yang, 1993b] Tseng F. S. C., Chen A. L. P., Yang W. P.: *Searching a minimal semantically-equivalent subset of a set of partial values.* Journal on VLDB 2 (1993), 4, 489-512

[Vadaparty - Naqvi, 1995] Vadaparty K., Naqvi S.: *Using Constraints for Efficient Query Processing in Nondeterministic Databases.* IEEE Transactions on Knowledge and Data Engineering 7 (1995), 6, 850-864

[Zaniolo, 1984] Zaniolo C.: *Database systems with null values.* Journal of Computer and System Sciences 28 (1984), 142-166

[Zemankova - Kandel, 1984] Zemankova M., Kandel, A.: *Fuzzy Relational Data Bases - A Key to Expert Systems.* Verlag TUV Rheinland, 1984

[Zemankova, 1985] Zemankova M.: *Implementing imprecision in information systems.* Information Science 37 (1985), 107-141

[Chiu - Chen, 1995] Chiu J-S., A. L. P.: *An Exploration of Relationships Among Exclusive Disjunctive Data.* IEEE Transactions on Knowledge and Data Engineering 7 (1995), 6, 928-940