DEMONSTRATIO MATHEMATICA
Vol. XXXV No 4 2002

Janusz A. Pomykała

SIMILARITY IN RELATIONAL DATABASES AND IN INFORMATION SYSTEMS THEORY

Abstract. The motivation of the paper comes form two sources—theory of relational databases (RDB) and the information systems theory (IST). On the one hand functional and multivalued dependencies in RDB capture a large portion of the semantics of real world relations, but it has proved useful to consider also other classes of dependencies eg. join or template dependencies. It is known that there is an equivalence between functional dependencies in a relational database and a certain fragment of propositional logic. This was extended by many authors to include both functional and multivalued dependencies, and complete axiomatizations were given. Also for fully join and for template dependencies complete axioms are known.

Dependencies of attributes in information systems theory (IST) can be expressed in terms of indiscernibility relations derived from the system, in particular data constraints are modeled by them. A generalization of this theory to dependencies in other information frames is an open problem. We propose here an attempt to solve it for frames based on similarity relations. We define dependencies for weak and strong similarity relations with parameters and develop logical formalism for reasoning about them.

In RDB theory we propose the notion of "similarity of records", giving the motivations from medicine (eg. similar symptoms should imply similar diagnosis or treatment) and from economy (similar market informations should be followed by similar economic movement or decisions). In consequence we introduced the notions of similarity dependency between sets of attributes in RDB. Examples are shown that the notion introduced is different from functional, multivalued, join and template dependencies in RDB. We analyse Armstrong axioms and Fagin axioms in this context, finding sound (but as yet not necessarily complete) axiomatization of similarity dependency in RDB.

1. Information systems

We recall again that the aim of the following paper is twofold: first give a broad motivation for studying similarity of systems and similarity relations (or equivalently tolerance relations) in geometry, and logic, and second give a new definition of dependency of attributes in RDB and IST.

Any collection of data specified as a structure (O, A, V, f) *such that* O *is a nonempty set of objects,* A *is a nonempty set of attributes,* V *is a nonempty* *set of values of information function f, is referred to as an information system. In the sequel we assume that $f : O \times A \to V - \{\emptyset\}$.*

A family of information systems indexed by a set I will be denoted by $(\mathbb{O}, \mathbb{A}, \mathbb{V}, \mathbb{F})^I$, i.e. for every $i \in I$

$$f_i : O_i \times A_i \to V_i.$$

Usually we assume that

$$V = \bigcup \{V_a : a \in A\},$$

V_a is also called a domain of the attribute a.

In this paper we assume that with every attribute $a \in A$ is related a tolerance relation (i.e. reflexive and symmetric relation) $\tau(a)$. In most cases this relation shall be defined in the following way:

$$\mathrm{Sim}(a)xy \quad \text{iff} \quad f(x,a) \cap f(y,a) \neq \emptyset$$
$$\mathrm{sim}(a)xy = \mathrm{Sim}(a)xy.$$

For $B \subseteq A$ we define

$$\mathrm{Sim}(B)xy \quad \text{iff} \quad \forall b \in B \ \mathrm{Sim}(b)xy$$
$$\mathrm{sim}(B)xy \quad \text{iff} \quad \exists b \in B \ \mathrm{sim}(b)xy.$$

$\mathrm{Sim}(B)$ is called (strong) similarity relation and $\mathrm{sim}(B)$ is called weak similarity with respect to the set of attributes $B \subseteq A$. Some authors use the notation sim, $w$sim, $s$sim, respectively (cf. [3], [21]).

The set $\{f(x,a) : a \in A\}$ shall be called an information about the object x, in short a record of x or a row determined by x. We shall say that two records determined by x, y are strongly τ-similar iff $\forall a \in A \ f(x,a)\tau(a)f(y,a)$. We will also consider the case when the above notion is restricted to a set $B \subseteq A$ i.e. two records $\{f(x,a) : a \in B\}$ and $\{f(y,b) : b \in B\}$ are similar with respect to the set $B \subseteq A$ iff

$$\forall b \in B \ f(x,b)\tau(b)f(y,b).$$

In other words we can say that two records are strongly τ-similar if for every attribute the respective values of attributes (i.e. the values of information function $f(x,a)$, $f(y,a)$) are similar with respect to the family of tolerances $\{\tau(a) : a \in A\}$.

We shall say that two objects (records) x, y are weakly τ-similar if for some attribute $a \in A$, values $f(x,a)$, $f(y,a)$ are similar with respect to $\tau(a)$.

In symbols: $\exists b \in B f(x,b)\tau(b)f(y,b)$. We denote strong relation by $\tau(B^\wedge)$ and weak one by $\tau(B^\vee)$, respectively.

Now let us assume that one information system (O, A, V, f) is established.

With the above definitions of similarity of objects and records we can examine the following questions:

1. find a record r' similar but different to a given record r (if it exists),
2. produce a record similar to a given one,
3. find all records similar to a given one,
4. check for a given two records r, r' if they are similar,
5. determine the level of similarity for a given record r and a family of records $w_1, \ldots, w_m$,
6. determine the properties of the dependency of attributes with respect to similarity relations.

We can express special kind of similarity of records by formulating the proper query in the system. It is however strongly determined by the possibilities of a given RDB system. By analogy to indiscernibility matrices (see Skowron, Rauszer [39]) we propose to use similarity matrices. As regards similarity queries based on distance access method compare for [4].
Let me examine first the question 6. We begin with the definition:

The set of attributes $Y \subseteq A$ depends on the set $X \subseteq A$ with respect to the similarity relation Sim *if and only if*

$$\mathrm{Sim}(X) \leq \mathrm{Sim}(Y).$$

We shall write in symbols

$$X \xrightarrow{S} Y \quad \text{or} \quad X \xrightarrow{\mathrm{Sim}} Y.$$

In the same way we can define dependency of attributes with respect to weak similarity relation sim:

$$X \xrightarrow{s} Y \quad \text{iff} \quad \mathrm{sim}(X) \leq \mathrm{sim}(Y)$$

(here $\leq$ is usual inclusion relation).

In other words, $X \xrightarrow{S} Y$ if strong similarity of objects with respect to the set of attributes X implies strong similarity of objects with respect to the set of attributes Y.

REMARK. Dependency of attributes with respect to indiscernibility relations is formulated eg. in [3], [32]. The dependency with respect to similarity was formulated in 1994.

Let me recall now Armstrong axioms for functional dependency (Let me recall that XY abbreviates $X \cup Y$.):

B1 If $Y \subseteq X \subseteq A$ then $X \to Y$.
B2 If $X \to Y$ and $Z \subseteq A$ then $XZ \to YZ$.
B3 If $X \to Y \to Z$ then $X \to Z$.

The axioms hold for strong similarity relation Sim. They hold is some other classes of strong relations also, cf. MacCaull [18]. Let me show eg. B1: $\mathrm{Sim}(X)xy$ iff $\forall a \in X\ \mathrm{Sim}(a)xy$ but $Y \subseteq X$ so all the more $\forall a \in Y\ \mathrm{Sim}(a)xy$, i.e. $\mathrm{Sim}(X) \leq \mathrm{Sim}(Y)$ which means $X \xrightarrow{S} Y$.

REMARK. For weak similarity we have just the opposite:

$$\text{if} \quad Y \subseteq X \subseteq A \quad \text{then} \quad Y \xrightarrow{\mathrm{sim}} X.$$

Now let me prove B2 for Sim:
By hypothesis we have $\mathrm{Sim}(X) \leq \mathrm{Sim}(Y)$, $\mathrm{Sim}(XZ) = \mathrm{Sim}(X \cup Z)$ by notation, so

$$\begin{aligned} \mathrm{Sim}(X \cup Z)xy \text{ iff} \forall a \in (X \cup Z)\ \mathrm{Sim}(a)xy \text{ iff} \\ \forall a \in X\ \mathrm{Sim}(a)xy \wedge \forall a \in Z\ \mathrm{Sim}(a)xy \\ \mathrm{Sim}(Y \cup Z)xy \text{ iff} \forall (a \in Y)\ \mathrm{Sim}(a)xy \wedge \forall a \in Z\ \mathrm{Sim}(a)xy \end{aligned}$$

hence $\mathrm{Sim}(X \cup Z) \leq \mathrm{Sim}(Y \cup Z)$.

Ad. B3: It follows by transitivity of inclusion $\leq$. ∎

Also for weak similarity relation the axioms B2, B3 holds easily:

Since $\mathrm{sim}\, XZ = \mathrm{sim}\, X \cup \mathrm{sim}\, Z$ and $X \to Y$ means that $\mathrm{sim}\, X \leq \mathrm{sim}\, Y$ and this gives that

$$\mathrm{sim}\, XZ \leq \mathrm{sim}\, Y \cup \mathrm{sim}\, Z = \mathrm{sim}\, YZ.$$

B3 holds by transitivity of inclusion.

Let me finally introduce the mixed similarity dependency in the following way:

$$X \xrightarrow{\mathrm{S\,s}} Y \text{ iff } \mathrm{Sim}(X) \leq \mathrm{sim}(Y) \text{ and}$$
$$X \xrightarrow{\mathrm{s\,S}} Y \text{ iff } \mathrm{sim}(X) \leq \mathrm{Sim}(Y).$$

In Words: $X \xrightarrow{\mathrm{S\,s}} Y$ if for all objects $x, y \in O$ strong similarity of x, y with respect to the set of attributes X implies weak similarity of x, y with respect to the set Y. ∎

REMARK. The reason for a difference between Sim and sim relations in the axiom B1 is the following:
If $A \leq B$ and $\exists x \in A\varphi$ then $\exists x \in B\varphi$.
If $A \leq B$ and $\forall x \in B\varphi$ then $\forall x \in A\varphi$.

In other words the pair $(\leq, \forall)$ has slightly different properties then the pair $(\leq, \exists)$.

2. Tolerance relation

"Any model of perception must take account of the fact that we cannot distinguish between points that are sufficiently close" (Zeeman [46]). Similar statement was formulated for choice behaviour by Luce. In consequence the notions of tolerance, threshold, just noticable difference has been formulated. We can say that the above notions of similarity and tolerance, threshold, just noticable difference have the same physical and philosophical fundation and the same role to play.

If we substitute closeness for identity then we can define tolerance geometry. In the Approach of Roberts it is important to "Study finite sets and axioms necessary and sufficient for isomorphism (or homomorphism) into certain kinds of spaces".

We recall now a simple axiom system of Tarski for two-dimensional elementary geometry. The system consist of twelve individual axioms A1–A12, and the infinite collection of all elementary continuity axioms A13. We shall use the original notation of the axioms.

A1 [Identity axiom for betweennes] $\bigwedge_{xy}[\beta(xyx) \to (x = y)]$,
A2 [transitivity axiom for betweennes] $\bigwedge_{xyzn}[\beta(xyn) \wedge \beta(yzn) \to \beta(xyz)]$,
A3 [connectivity axiom for betweennes]
$\bigwedge_{xyzn}[\beta(xyz) \wedge \beta(xyn) \wedge \beta(x \neq y) \to \beta(xzn) \vee \beta(xnz)]$,
A4 [reflexivity axiom for equidistance] $\bigwedge_{xy}[\delta(xyyx)]$,
A5 [identity axiom for equidistance] $\bigwedge_{xyz}[\delta(xyzz) \to (x = z)]$,
A6 [transitivity axiom for equidistance]
$\bigwedge_{xyzuvw}[\delta(xyzn) \wedge \delta(xyvw) \to \delta(zuvw)]$,
A7 [Pasch's axiom] $\bigwedge_{txyzu} \bigvee_{v}[\beta(xtu) \wedge \beta(yuz) \to \beta(xvy) \wedge \beta(ztr)]$,
A8 [Euclid's axiom]
$\bigwedge_{txyzn} \bigvee_{vw}[\beta(xut) \wedge \beta(yuz) \wedge (x \neq u) \to \beta(xzv) \wedge \beta(xyw) \wedge \beta(vty)]$,
A9 [five segment axiom]

$$\bigwedge_{xx'yy'zz'uu'} [\delta(xyx'y') \wedge \delta(yzy'z') \wedge \delta(xux'u') \wedge \delta(yuy'u') \\ \wedge\beta(xyz) \wedge (x'y'z') \wedge (x \neq y) \to \beta(zuz'u')],$$

A10 [axiom of segment construction] $\bigwedge_{xyuv} \bigvee_{z}[\beta(xyz) \wedge \beta(yzuv)]$,
A11 [lower dimension axiom] $\bigwedge_{xyz}[\neq \beta(xyz) \wedge \neg\beta(yzx) \wedge \neg\beta(zxy)]$,
A12 [upper dimension axiom]

$$\bigwedge_{xyzuv} [\delta(xuxv) \wedge \delta(yuyv) \wedge \delta(zuzv) \wedge (u \neq v) \\ \to \beta(xyz) \vee \beta(yzx) \vee \beta(zxy)],$$

A13 [elementary continuity axioms]
All sentences of the form
$\bigvee_{vw} \cdots \Big\{\bigwedge_{z} \bigvee_{xy}[\varphi \wedge \psi \to \beta(zxy)] \to \bigvee_{xy}[\varphi \wedge \psi \to \beta(xuy)]\Big\}$,

where φ stands for any formula in which the variables $x, v, w, \dots$ but neither y nor z nor u, occur free and similarly for ψ, with x and y interchanged.

We recall that if $B(x, y, z)$ is ternary relation of betweennes B on a set A, Tarski-type axiomatization for tolerance geometry, one-dimension case is the following:

C2. $B(x, y, z) \to B(z, y, x)$
C3. $B(x, y, z)$ or $B(x, z, y)$ or $B(y, x, z)$
C4. $B(x, y, u)$ and $B(y, z, u) \to B(x, y, z)$
C5. If $u \neq v$ then $B(x, u, v)$ and $B(u, v, y) \to B(x, u, y)$
C6. $B(x, y, z)$ and $B(y, x, z) \to x = y$
C7. $x = y \to B(x, y, z)$.

THEOREM (Roberts). *Suppose B is a ternary relation on a finite set A. Then Axioms* C2 - C7 *are necessary and sufficient for the existence of a* $1-1$ *function $f : A \to R$ so that for all $x, y, z \in A$*

$$B(x, y, z) \Leftrightarrow [f(x) \leq f(y) \leq f(z) \text{ or } f(z) \leq f(y) \leq f(x)].$$

Let us now recall the definition of classical betweenes: there is $f : A \to R$ $\forall_{x,y,z \in A}\ B(x, y, z) \Leftrightarrow |f(x) - f(y)| + |f(y) - f(z)| = |f(x) - f(z)|$, and for ε-betweenes we have:

$$B(x, y, z) \Leftrightarrow |f(x) - f(y)| + |f(y) - f(z)| < |f(x) - f(z)| + \varepsilon.$$

The tolerance axioms are stated in terms of B and a relation I on A defined from B by:

$$xIy \Leftrightarrow B(x, y, x).$$

Let R be a simple (i.e. total) order on A. We say that R is compatible with I if for all $x, y, u, v \in A$

$$xRuRvRy \ \&\ xIy \to uIv.$$

Axioms for ε-betweeness on the line:

T1. There is a simple order R on A compatible with I
T2. $B(x, y, z) \to B(z, y, x)$
T3. $B(x, y, z)$ or $B(x, z, y)$ or $B(y, x, z)$
T4. $B(x, y, u)$ and $B(y, z, u)$ and $B(x, y, z) \to uIy$ and uIz
T5. If uIv, then $B(x, u, v)$ and $B(u, v, y) \to B(x, u, y)$
T6. $B(x, y, z)$ and $B(y, x, z) \to xIy$ or $(zIx$ and $zIy)$
T7. $xIy \to B(x, y, z)$.

THEOREM (Roberts). *Suppose B is a ternary relation on a finite set A and $\varepsilon > 0$ is given. Then Axioms* T1 - T7 *are necessary and sufficient for the* *existence of a function $f : A \to \mathcal{R}$ satisfying*

$$B(x,y,z) \Leftrightarrow |f(x) - f(y)| + |f(y) - f(z)| < |f(x) - f(z)| + \varepsilon.$$

The above two theorems of Roberts show the importance of similarity in geometry.

We come back to them later, now let me define several relations considered in information systems:

a) $x\tau y$ iff $|x - y| < \varepsilon$,
b) $x\tau y$ iff $\rho(x,y) < \varepsilon$,
c) $(x_1 \dots x_n)\,\tau\,(y_1 \dots y_n)$ iff $\rho\Big((x_1 \dots x_n)\,,\,(y_1 \dots y_n)\Big) < \varepsilon$, where ρ is a distance function,
d) $(x_1 \dots x_n)\,\tau\,(y_1 \dots y_n)$ iff $\exists_{i,j}(|x_i - y_j| < \varepsilon$,
e) Assume that $x \in O_1$, $y \in O_2$, $(O_1 A_1 V_1 f_1)$, $(O_2 A_2 V_2 f_2)$ are information systems and we define:

$$x\tau y \text{ iff } \exists_{a_1,a_2}\ f\,(x,a_1) \cap f\,(y,a_2) \neq \emptyset$$

and

$$x\tau\,(B_1,B_2)\,y \text{ iff } \exists b_1 \in B_1,\ b_2 \in B_2, \quad f\,(x,b_1) \cap f\,(y,b_2) \neq \emptyset.$$

Consider also the modification given by the following:

$$|f(x,b_1) - f(y,b_2)| < \varepsilon,$$

where f is real valued.

REMARK. It is a general question how to relate different objects in different information systems if for some reason we are obliged to do this.

f) Assume that $\tau_1\ \tau_2\ \tau_3\ \dots\ \tau_n$ are tolerances on U, we define the relation τ^n

$$x\tau^n y \text{ iff } \left(\exists x_1 \dots x_n\ x_1 = x\ x_n = y\ \&\ \forall_i\ x_i\tau_i\ x_{i+1}\right).$$

We say that τ^n is generated by $\tau_1 \dots \tau_n$. It is clear that in most interesting cases, there is n_0 s.t. for all $n > n_0$ we can not say reasonably that objects x, y related by τ^n are similar. In other words there is a threshold of similarity for a given information system. Finally let us consider.

g) Similarity of texts (We define this relation for simple texts only):

assume that $T_1 = a_1 a_2 \dots a_n$ and $T_2 = b_1 b_2 \dots b_n$
where $a_i, b_i \in$ char,

We define

$$f\,(T_1 T_2) \overset{\text{df}}{=} \sum_i \{a_i \neq b_i\}\,,$$

and in more generalized setting:

$$\mu(T_1T_2) = \frac{1}{2}\sum_i \{a_i \neq b_i\} + \frac{1}{4}\sum_i \{a_i \neq c_i\} + \\ + \ldots + \frac{1}{2^n}\sum \{a_i \neq \gamma_i\},$$

where $c_1 = b_2, \ldots, c_{n-1} = b_n$, $c_n = \emptyset$, $d_1 = b_3$, $d_2 = b_n, \ldots, d_{n-2} = b_n$, $d_{n-1} = \emptyset$ etc. are the original texts shifted to the right. It is easy to formulate analogical function μ for texts shifted to the left or to both sides.

Let us finally mention that many properties of tolerance relation are examined by Chajda, Zelinka [5].

3. Similarity of systems

The basic notion which expresses similarity of systems or algebras is the notion of homomorphism. We propose here slightly different notion, relating similarity of systems to the similarity structures which can be defined in (or on) the system. More exactly, we shall say that two systems are similar, if they have the same (i.e. isomorphic) similarity structures. In this context by similarity structures we mean:

a) family of tolerance relations,
b) topology or several topologies,
c) hypergraph or connecting net,
d) family of approximations operations.

Every time we fix only one similarity structure, just for simplicity and convenience.

EXAMPLE: SIMILARITY STRUCTURE ON INFORMATION SYSTEM

By similarity structure on the system (O, A, V, f) we mean the following structure:

$$(B, C, D, \ldots, T)\Big(\mathrm{Sim}(B), \mathrm{Sim}(C)\ldots\Big)\Big(U_1^B \ldots U_{B(i)}^B\Big)$$
$$\Big(U_1^C \ldots U_{C(i)}^C\Big) \ldots \Big(U_1^T \ldots U_{T(i)}^T\Big),$$

where $B, C, D, \ldots, T$ are subsets of $A, \mathrm{Sim}(B), \ldots$ are similarity relations,

$$\Big(U_1^B \ldots U_{B(i)}^B\Big) \ldots \Big(U_1^T \ldots U_{T(i)}^T\Big)$$

are partitions of the universe of objects O, satisfying conditions:

$$\forall_{E \in \{B,C,D,\ldots,T\}} \forall_{x,y \in U_i^E} \ \mathrm{Sim}(E)xy$$

and

$$\forall x \in U_i^E,\ y \in U_j^E \text{ for } i \neq j \quad \text{non} \quad (x\,\mathrm{Sim}(E)y).$$

Of course the above definition can be formulated also for weak similarity relations sim(B),... or other kind of tolerances related to attributes. The proposed notion of similarity of information systems can be especially useful in case when we have fixed set of objects O and dynamically changing sets A, V, f. In other words we can compare fixed set of objects from a different perspectives (in this case expressed by sets of attributes A, sets of values V and functions f) (cf. [30], [33]).

EXAMPLE. The following two systems are similar with respect to a similarity relation Sim:

	a_1	a_2
x_1	1	0
x_2	1	0
x_3	1	0
x_4	7	0
x_5	7	2
x_6	7	2

	a_3	a_4
x_1	$\{7,8\}$	1
x_2	$\{7\}$	1
x_3	$\{8\}$	1
x_4	$\{0\}$	1
x_5	$\{0\}$	$\{3,5\}$
x_6	$\{0\}$	$\{5,10,12\}$

Relaxing the condition that similarity structures for similar systems have to be isomorphic, we can obtain more general notion.

4. Examples and further motivation

In psychological investigations we often deal with experiments in which one has to estimate the values of a stimulus (for instance the light or sound-stimulus) on a given measurement scale. This estimation is given by the interval in which we expect to find the actual value of the stimulus; in other words the postulated value has to be considered taking into account a certain error.

If the results of such an experiment are presented in terms of information systems (see Pawlak [24]), then the values of the information function should be identified with the subsets of attribute values (instead of single values). In such circumstances the associated indiscernibility relation (for the objects) does not have to be the equivalence relation but it is the tolerance relation and consequently the family of the elementary sets forms a cover of U (but not necessarily a partition).

EXAMPLE. Let a_i, $i = 1, \ldots, n$ be the natural numbers. We define the tolerance relation $\xi \subset N \times N$, (which is not the equivalence one if $(a_i, a_j) = 1$ for some $i \neq j$, $1 \leq i, j \leq n$) as follows

$$x \xi y \Leftrightarrow \exists_{1 \leq i \leq n} x \equiv y \,(\operatorname{mod} a_i)\,.$$

Let us describe what are the sets O_k, I_k and E in case when $n = 2$, $a_1 = 2$, $a_2 = 3$. We have five elemntary sets, three of them are arithmetical progressions mod 3 and two of them are arithmetical progressions mod 2. I_k is the arithmetical progressions, namely. $O_k = \{n : n \equiv k(\bmod 2)\} \cup \{n : n \equiv k(\bmod 3)\}$.

EXAMPLE. Let us consider the following information system (U, A, V, F) where $U = \{p, q, r, s, t, u, v, w, x, y, z\}$, $A = \{a\}$, $V = [0, 10)$ and the values of F are intervals contained in V:

	$F(x,a)$
p	$\{0\}$
q	$(0,2)$
r	$(1,2)$
s	$(2,4)$
t	$(3,5)$
u	$(4,5)$
v	$(5,7)$
w	$(6,8)$
x	$(7,9)$
y	$(8,10)$
z	$(9,10)$

It is easy to check that the relation ξ_A is the following:

$$\begin{aligned}\xi_A = \{&\langle q,r\rangle, \langle r,q\rangle, \langle s,t\rangle, \langle t,s\rangle, \langle t,u\rangle, \langle u,t\rangle, \langle u,w\rangle,\\ &\langle w,v\rangle, \langle w,x\rangle, \langle x,w\rangle, \langle x,y\rangle, \langle y,x\rangle, \langle y,z\rangle, \langle z,y\rangle\}\\ &\cup\{\langle m,m\rangle, : m \in U\}.\end{aligned}$$

The family of all elementary sets contains the following sets:

$$E = \{\{p\}, \{q,r\}, \{s,t\}, \{t,u\}, \{v,w\}, \{w,x\}, \{x,y\}, \{y,z\}\}.$$

As regards the family of the kernels, we have one two-element kernel $I_q = I_r = \{q, r\}$. The other ones are one-element sets.

We have the following indiscernibility neighbourhoods: $O_p = \{p\}$, $O_q = O_r = \{q,r\}$, $O_s = \{s,t\}$, $O_t = \{s,t,u\}$, $O_u = \{t,u\}$, $O_v = \{v,w\}$,$O_w = \{v,w,x\}$,$O_x = \{w,x,y\}$, $O_y = \{x,y,z\}$, $O_z = \{y,z\}$.

EXAMPLE. Suppose one has to select a group X of persons of the set U, which are to participate in a polar expedition, in a way that assures the best possible collaboration (i.e. without conflicts). For any person x belonging to U, two experts (the psychologist and the physician for instance) are to indicate independently the suitable "good working" groups with person x.

From data obtained by the experts we form the information system and we analyse it. The experts $Ex\,1$, $Ex\,2$ will play the role of the "attributes".

	$Ex\,1$	$Ex\,2$
x_1	$f(x_i, Exj)$	
x_2		
x_3		
$\vdots$		
x_n	$i = 1, \ldots, n,$	$j = 1, 2.$

The value of the information $f(x, Ex\,1)$ is the set of persons that expert 1 regards as good working group containing the person x. Identically we define $f(x, Ex\,2)$. Let us observe that the families D^1, D^2 defined in the following way:

$$D^1 = \{f(x, Ex\,1) : x \in U\}$$

$$D^2 = \{f(x, Ex\,2) : x \in U\}$$

are the covers of the set U. We define the tolerance relation on U as follows: for any $x, y \in U$

$$x\xi y \Leftrightarrow \exists D \in D^1 \ \& \ \exists D' \in D^2 \left(\{x, y\} \subset G \ \& \ \{x, y\} \subset D'\right).$$

Now we give the interpretation for the notions of elementary (with respect to the opinion of both experts) if it is the greatest set of which all the members can stay together in the polar station without conflicts.

Indiscernibility neighbourhood O_x is the set of all the people, who will no be in conflict with person x (pairly).

The kernel I_x is the set of all persons (in agreement with each other) having the same "relation" as x to all the remaining members of U.

We shall illustrate our approach on the simple numerical example:

	$Ex\,1$	$Ex\,2$
x_1	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$
x_2	$\{x_2, x_3, x_4\}$	$\{x_1, x_2, x_3\}$
x_3	$\{x_2, x_3, x_4\}$	$\{x_2, x_3, x_4\}$
x_4	$\{x_4, x_5\}$	$\{x_4, x_5\}$
x_n	$\{x_1, x_5\}$	$\{x_1, x_4, x_5\}.$

We have the following elementary sets and indiscernibility neighbourhoods:

$$E = \{\{x_1, x_2, x_3\}, \{x_1, x_5\}, \{x_2, x_3, x_4\}, \{x_4, x_5\}\}$$

$$O_{x_1} = \{x_1, x_2, x_3, x_5\}, \quad O_{x_2} = \{x_1, x_2, x_3, x_4\} = O_{x_3},$$

$$O_{x_3} = \{x_1, x_2, x_3, x_4\}, \quad O_{x_4} = \{x_2, x_3, x_4, x_5\}, \quad O_{x_5} = \{x_1, x_4, x_5\}.$$

At last there are the following kernels:

$$I_{x_1} = \{x_1\},\ I_{x_2} = I_{x_3} = \{x_2, x_3\},\ I_{x_4} = \{x_4\},\ I_{x_5} = \{x_5\}$$

on the universe U.

The example suggests that in order to find the needed group of people it is convenient to proceed as follows: we take the greatest kernel and the (possibly) missing persons should be chosen from the biggest elementary set including this kernel. In our example the kernel is I_{x_2} and the third element to be added is x_1 or x_4.

5. Rough regions and Kuratowski lemma

It is well known that if we apply to a set A two operations

- the closure and the complement,
- in a fixed topological space $(U, ^{-})$ then the number of sets that can be obtained from A in this way is less or equal to 14 (Kuratowski[15]).

This means that if we apply the closure and the complement operations to the sets A and we form an equality, then the number of relations defined with respect to these equalities in the family of all subsets of U has to be finite also. The equivalencies of the similar kind are sometimes applied in computer science and data analysis, for example so called rough, bottom and top equality (see Nowotny, Pawlak [19]). In this paper we construct 18 relations (including rough top and bottom equalities) (see Pomykała [29]) obtaining as a special case also topological rough sets (see Wiweger [45]).

Basic definitions

Assume that $(U, ^{-})$ is a topological space. If w is a finite sequence of and $^{-}$ (where $'$ means the complement operation) then we shall write $w \in$ Word$(' , ^{-})$. In other words w is a word over the alphabet $\{' \ ^{-}\}$. For $A \subseteq U$ we define:

$$A^{\emptyset} = A$$

$$A^{w} = A' \text{ for } w = '$$

$$A^{w} = A^{-} \text{ for } w = ^{-}$$

and inductively for arbitrary $w \in$ Word$(', ^{-})$.

If R is an equivalence relation in U we denote by U and L, Pawlak's closure and interior operations on U, i.e. for $A \subseteq U$

$$U\,A = \overline{A} = \bigcup\{[x] : x \in A\}$$

where $[x]$ is an equivalence class of x w.r.t.R.

L is defined to be conjugated to U, i.e. for each $A \subseteq U$, $LA = -U - A$.

In the sequel the operation $^{-}$ will be equal to U for some relation R. Usually we write $\overline{A}$, $\underline{A}$ instead of $U\,A$, $L\,A$, respectively. $(U, ^{-}) = (U, \overline{R})$ is called approximation space.

Now let us define the following relations on $P(U)$, let A, $B \subseteq U$:

$$
\begin{aligned}
&A \cong_1 B \text{ iff } A = B\\
&A \cong_2 B \text{ iff } A = -B\\
&A \cong_3 B \text{ iff } A = \underline{B}\\
&A \cong_4 B \text{ iff } A = -\underline{B}\\
&A \cong_5 B \text{ iff } A = \overline{B}\\
&A \cong_6 B \text{ iff } A = -\overline{B}\\
&A \cong_7 B \text{ iff } \underline{A} = B\\
&A \cong_8 B \text{ iff } \underline{A} = -B\\
&A \cong_9 B \text{ iff } \underline{A} = \underline{B}\\
&A \cong_{10} B \text{ iff } \underline{A} = -\underline{B}\\
&A \cong_{11} B \text{ iff } \underline{A} = \overline{B}\\
&A \cong_{12} B \text{ iff } \underline{A} = -\overline{B}\\
&A \cong_{13} B \text{ iff } \overline{A} = B\\
&A \cong_{14} B \text{ iff } \overline{A} = -B\\
&A \cong_{15} B \text{ iff } \overline{A} = \underline{B}\\
&A \cong_{16} B \text{ iff } \overline{A} = -\underline{B}\\
&A \cong_{17} B \text{ iff } \overline{A} = \overline{B}\\
&A \cong_{18} B \text{ iff } \overline{A} = -\overline{B}
\end{aligned}
$$

Let us observe that the first relation is the equivalence relation, the second is symmetric, the third is transitive, and the fifth is transitive and similarly 13-th, finally 7-th, 9-th and 17-th are equivalences, 18-th is symmetric.

Main result

Our main theorem is the following

THEOREM. *If* $(U, ^{-}) = (U, U)$ *is the approximation space for some equivalence relation* R *and* $w1, w2 \in$ Word$(^{-}, ')$ *then the relation* $\equiv$ *(on* $P(U)$*)*

defined by the condition

$$A \equiv B \textit{ iff } A^{w1} = B^{w2}$$

is equal to one of the above 18 relations.

Proof. We need 3 Lemmas

LEMMA 1 (Kuratowski 1922). *Suppose that we apply to a set A the operations $^-$ and $'$. The number of sets that we obtain is less or equal to* 14.

LEMMA 2. *If (U, U) is the approximation space and A' denotes the complement of A then there exist no more then 6 sets obtained by applying to the set A the operations of closure and of the complement. The following inclusions are generally valid among them:*

$$A'^{-\,'} \subseteq A \subseteq A^{-}$$
$$A^{-\,'} \subseteq A' \subseteq A'^{-}$$

LEMMA 3. *The following equalities holds:*

$$A'^{-\,'} = A'^{-\,'\,-} = A'^{-\,'\,-\,'\,-\,'}$$
$$A^{-\,'\,-\,'} = A^{-\,'\,-\,'\,-} = A^{-}$$
$$A^{-\,'} = A^{-\,'\,-} = A^{-\,'\,-\,'\,-\,'}$$
$$A'^{-} = A'^{-\,'\,-\,'} = A'^{-\,'\,-\,'\,-}$$

Now, in view of Kuratowski lemma we infer that the words $w1$, $w2$, in the equality $A^{w1} = B^{w2}$ may be reduced to words over the set

$$W = \{{}'\,{}^{-}\,{}', \emptyset, {}^{-}, {}^{-}\,{}', {}', {}'\,{}^{-}\}.$$

Considering every pair of words $w1$, $w2$ belonging to W and examining the equalities $A^{w1} = B^{w2}$ it is easy to check that $A^{w1} = B^{w2}$ iff $A \cong_i B$ for some $i \in \{1, \dots 18\}$. Finally let us observe that all equivalences $\cong_1, \dots, \cong_{18}$ are different. The proof of the theorem is completed.

REMARK. The general construction of this paper for arbitrary Kuratowski closure operation will be given in a forthcoming paper.

We leave open the problem of the description of all rough constructions defined with respect to the operations introduced in the paper [27].

It seems to be worth studying the structure of the algebras created from the family of all pairs $(\operatorname{int} A, \operatorname{cl} A)$ in a fixed topological space. ∎

6. The completeness theorem

The concepts of functional and multivalued dependencies in RDB are important tools for database design. The complete and sound axiomatizations are known (see [1], [10], [2]). In this place we first recall axioms in the

RDB context, and we show the modification in information systems theory case (cf. [18]).

Let us recall that a functional dependency (see [6]) is a statement of the form $X \to Y$, where X, Y are sets of attributes. Dependency $X \to Y$ holds in a relation R if for every pair r_1, r_2 of tuples, if $r_1[X] = r_2[X]$, then $r_1[Y] = r_2[Y]$.

A multivalued dependency (see [10]) is a statement of the form

$$X \twoheadrightarrow Y, \quad \text{where} \quad X, Y \subseteq A.$$

Let $Z = A - X - Y$. The dependency $X \twoheadrightarrow Y$ holds in R if for all r_1, r_2 in R, if $r_1[X] = r_2[X]$, then there are r_3 and r_4 in R such that
1) $r_3[X] = r_1[X]$, $r_3[Y] = r_1[Y]$, $r_3[Z] = r_2[Z]$;
2) $r_4[X] = r_2[X]$, $r_4[Y] = r_2[Y]$, $r_4[Z] = r_1[Z]$.

We say that a dependency d is a consequence of a set of dependencies D if for all relations R, d holds in R if all the dependencies D hold in R.

We recall now axioms for multivalued dependencies:

D0: Let X, Y, Z be sets of attributes such that $X \cup Y \cup Z = A$ and $Y \cap Z \subseteq X$. Then $X \twoheadrightarrow Y$ iff $X \twoheadrightarrow Z$.
D1: If $Y \subseteq X$ then $X \twoheadrightarrow Y$.
D2: If $Z \subseteq W$ and $X \twoheadrightarrow Y$, then $XW \twoheadrightarrow YZ$.
D3: If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$ then $X \twoheadrightarrow Z - Y$.
Let D be a set of multivalued dependencies, and let $d_1 d_2 \dots d_n$ be dependences. We say that $d_1 d_2 \dots d_n$ is a derivation from D if the following holds:
for all i, either $d_i \in D \cup \{d_1 \dots d_{i-1}\}$ or d_i can be inferved from $D \cup \{d_1 \dots d_{i-1}\}$ by an application of one of the axioms (inference rules).

The inference rules are sound if every dependency d that can be derived from D is also a logical consequence of D. The inference rules (i.e. here D0–D3) are complete if every dependency d that is a consequence of D can also be derived from D. In ([1], [10], [2], cf. [38] [43]), it is proved that the axioms (inference rules) for functional dependency (Armstrong axioms) and for multivalued dependency are sound and complete, with respect to relational semantics.

Let D be a set of similarity dependencies and $X \subseteq A$. By X^* we denote the closure of X with respect to D i.e. it is the set of attributes B such that $X \xrightarrow{S} B$ can be deduced from D by Armstrong's axioms. We shall follow the proof of completeness given by Ullman [20] (see also Armstrong [1], Fagin [10]).

LEMMA. *$X \xrightarrow{S} Y$ follows from Armstrong's axioms iff $Y \subseteq X^*$.*

Proof. Let $Y = A_1 \dots A_n$ and suppose $Y \subseteq X^*$. By definition of closure, $X \xrightarrow{S} A_i$ is implied by axioms for all i. By the property

1) If $X \xrightarrow{S} Y$ and $X \xrightarrow{S} Z$ then $X \xrightarrow{S} YZ$

we obtain that $X \xrightarrow{S} Y$ follows. Conversely, suppose $X \xrightarrow{S} Y$ follows from axioms. For every i, $X \xrightarrow{S} A_i$ holds by the property of decomposition:

2) If $X \xrightarrow{S} Y$ and $Z \subseteq Y$ then $X \xrightarrow{S} Z$.

By application of the lemma it holds:

THEOREM. *Armstrong's axioms are sound and complete in relational semantics.*

Modifying the original proof it is possible to obtain:

THEOREM. *Armstrong's axioms are sound and complete with respect to information systems semantics.*

REMARK. It is possible to analyse properties of knowledge representation systems using the relations of indiscernibility, similarity and informational inclusion, together (cf. Vakarełow [44]). It is very interesting research topic, how to apply Vakarełow method to dependency theory unifying different relations with parameters.

Proof. We checked above that axioms are sound i.e. respective properties of similarity relation Sim hold.

Let D be a set of dependencies and suppose $X \xrightarrow{S} Y$ cannot be inferred from axioms. Consider the following system:
$O = \{x, y\}$, A – the set of attributes over which D is defined, $V = \{\{1\}, \{\emptyset\}\}$, and $f(x,a) = 1$ for every $a \in A$, $f(y,a) = 1$ for $a \in X^*$, $f(y,a) = \emptyset$ for $a \in A - X^*$.

First let me show that all dependencies in D are satisfied in the system (O, A, V, f).

Suppose $V' \xrightarrow{S} W$ is in D but is not satisfied in the system. Then $V' \subseteq X^*$ otherwise two records x, y are not similar for some attribute in V. And therefore can not destroy the dependency $V' \to W$. On the other hand W can't be a subset of X^*, or $V \to W$ is satisfied by our system.

Let A_1 be an attribute of W which does not belong to X^*. Since $V' \subseteq X^+$, $X \xrightarrow{S} V'$ follows from the axioms by Lemma. Dependency $V' \xrightarrow{S} W$ is in D, so by transitivity we have $X \xrightarrow{S} W$. By reflexivity condition $W \xrightarrow{S} A_1$, so $X \xrightarrow{S} A_1$ follows from the axioms. But then A_1 is in X^*. By contradiction, each $V' \to W$ in D is satisfied in the system (O, A, V, f).

Now let us show that $X \xrightarrow{S} Y$ is not satisfied. Assume, that it is satisfied in the system. As $X \subseteq X^*$, it follows that $Y' \subseteq X^*$, else the two tuples of system are similar on X but are not similar on Y'. But then Lemma tells that $X \to Y$ can be inferred from axioms, a contradiction. Therefore $X \to Y$ is not satisfied by (O, A, V, f) even though each dependency of D holds. ∎

Finally we try to express multivalued dependency related to similarity relation Sim. Let us take $Z = A - X - Y$. There are many possibilities to define the notion analogical to multivalued dependency, let me propose the following:

a) $X \twoheadrightarrow Y$ iff $X \to YZ$,
b) $X \twoheadrightarrow Y$ iff $\forall_{xy} \left(X \xrightarrow{S} Y \land \exists b \in Z \ \mathrm{Sim}(b)\right) xy$,
c) $X \twoheadrightarrow Y$ iff $\forall_{xy} (\mathrm{Sim}(X)xy \Rightarrow (\mathrm{Sim}(Y)xy \lor \mathrm{Sim}(Z)xy))$,
d) $X \twoheadrightarrow Y$ iff $\mathrm{Sim}\, X \leq \mathrm{Sim}(X \cup Y) \circ \mathrm{Sim}\, Z$.

Of course we can also consider the case of embedded dependencies i.e. under assumption $Z \subseteq A - X - Y$. Due to space limitations we consider multivalued case in a different article.

In this paper we consider only the definition c), which is interesting for us from the point of view of the applications.

Functional and strong similarity dependency

Now let me explain the relation between functional and similarity dependency of attributes.

Let us consider the following systems:

	a	b
x	1	2
y	1	2

and

	a	b
x	$\{1\}$	$\{2\}$
y	$\{1\}$	$\{2\}$.

In the first one we can define indiscernibility relation but not similarity, at least in a natural way. In the second system we can define both relations - indiscernibility and similarity. Therefore, in some cases it is technically useful to transform a system putting

$$F(x, a) = \{f(x, a)\} \quad \text{for all} \quad x, a.$$

Under such assumption we can say that the similarity dependency is a generalization of the functional dependency. More exactly:

THEOREM. *If the functional dependency $X \to Y$ holds in a given information system $(OAVf)$ then the similarity dependency $X \xrightarrow{S} Y$ holds in the transformed system $OAVF$, where $F(x, a) = \{f(x, a)\}$.*

In view of this theorem, functional dependency can be interpreted as a similarity dependency.

On the other hand, similarity dependency is in an essential way different from multivalued dependency. This together means that similarity dependency is a generalization of functional one in a different direction then the multivalued dependency.

Let me recall the example from Fagin [10]. It is shown there the multivalued dependency Employee $\twoheadrightarrow$ {Salary, Year} and it is argued that it does not hold

$$\text{Employee} \twoheadrightarrow \text{Salary} \quad \text{nor} \quad \text{Employee} \twoheadrightarrow \text{Year}.$$

In case of similarity dependency we have:

Sim (Employee) $\leq$ Sim (Salary, Year) means that:

$$\forall_{xy} \text{ Sim (Employee) } xy \Rightarrow$$
$$\Rightarrow \text{Sim (Salary) } xy \cap \text{ Sim (Year) } xy$$

which implies that Sim (Employee) $\xrightarrow{S}$ Sim (Salary) and Sim (Employee) $\xrightarrow{S}$ Sim (Year).

Therefore: strong similarity dependency can not be equivalent to multivalued dependency.

1. Similarity dependency is a generalisation of functional dependency.
2. It differs from multivalued dependency since its definition is independent from the context i.e. $X \xrightarrow{S} Y$ depends only on X, Y and not on the attributes from $A - (X \cup Y)$.
3. It differs from MVD since the rule of complementation does not hold for similarity dependency i.e. the rule. If $X \cup Y \cup Z = A$ and $Y \cap Z \subseteq X$ then $X \twoheadrightarrow Y$ iff $X \twoheadrightarrow Z$, is not true for $\xrightarrow{s}$.
   On the other hand reflexivity, augmentation and transitivity rules hold. Also the rules of Pseudo-transitivity, union and decomposition holds for similarity dependency.

Now it is natural question if weak similarity dependency sim is a better candidate to be equivalent to (or to emulate) multivalued dependency?

Weak similarity dependency shall be examined in the second part of this article. In this place we only formulate the axioms:

1. $X \xrightarrow{s} X$,
2. $X \xrightarrow{s} Y \xrightarrow{s} Z$ implies $X \xrightarrow{s} Z$,
3. if $Y \cap Z \neq \emptyset$ then $X \xrightarrow{s} (Y \cap Z)$ implies $(X \xrightarrow{s} Y \ \& \ X \xrightarrow{s} Z)$,
4. it is not true in the full class of information systems, that $(X \cap Y) \xrightarrow{s} Z$ implies $(X \xrightarrow{s} Z) \ \& \ (Y \xrightarrow{s} Z)$.
5. $X \xrightarrow{s} Y \cap Z$ implies $X \xrightarrow{s} Y$ and $X \xrightarrow{s} Z$.
6. $X \cap Y \xrightarrow{s} Z$ implies $X \xrightarrow{s} Z$ and $Y \xrightarrow{s} Z$.

They are called reflexivity axiom, transitivity axiom, decomposition of intersection in second coordinate axiom and decomposition of intersection in the first coordinate, respectively.

We propose also the following axioms for mixed dependencies with respect to weak and strong similarity relations together:

1. Armstrong axioms for strong similarity dependency,
2. Axioms 1–6.

It is not in general true that:

1. $X \xrightarrow{S} Y$ implies $X \xrightarrow{s} Y$,
2. $X \xrightarrow{s} Y \xrightarrow{S} Z$ implies $X \xrightarrow{s} Z$,
3. $X \xrightarrow{S} Y \xrightarrow{s} Z$ implies $X \xrightarrow{s} Z$,
4. $X \xrightarrow{S} Y \cap Z$ implies $X \xrightarrow{s} Y$ and $X \xrightarrow{s} Z$ under condition that $Y \cap Z \neq \emptyset$. ∎

REMARK. Axioms are not sound in the full class of similarity structures.

7. Similarity on the family of objects

Now let me define some relations on the set of objects $\mathbb{O}$. Let us denote now sets of objects by M, N, P, Q, X, Y, Z. For $a, b \in \mathbb{A}$ we define

$$\mathrm{Sim}(m)ab \text{ iff } f(m,a) \cap f(m,b) \neq \emptyset$$
$$\mathrm{sim}(m)ab = \mathrm{Sim}(m)ab.$$

Next, for $P \subseteq \mathbb{O}$ we define

$$\mathrm{Sim}(P)ab \quad \text{iff} \quad \forall_{p \in P} \quad \mathrm{Sim}(p)ab$$
$$\mathrm{sim}(P)ab \quad \text{iff} \quad \exists_{p \in P} \quad \mathrm{sim}(p)ab.$$

These relations are called strong similarity of attributes with respect to the set of objects P and weak similarity of attributes with respect to P, respectively.

The set $\{f(p,a) : p \in P\}$ is called the information about the attribute a w.r.t. the family of objects P.

Now let us assume that we have also the set of tolerances on the set of attributes:

$$\tau(x), \tau(y), \tau(z), \ldots \quad \text{for} \quad x, y, z, \ldots \in \mathbb{O}.$$

We shall say that two informations about attributes a, b w.r.t. $P \subseteq \mathbb{O}$ are strongly similar if

$$\forall_{p \in P} \ f(p,a)\tau(p)f(p,b).$$

On the other hand, two attributes a, b are weakly similar w.r.t. P iff

$$\exists_{p \in P} \ f(p,a)\tau(p)f(p,b).$$

In other words we can say that two columns (attributes a, b) are strongly similar if for every object $p \in P$ the respective values of information function f, $f(p,a)$, $f(p,b)$ are similar with respect of the family of tolerances $\tau(p)$, $p \in P$. By analogy we can describe weak similarity of attributes.

Now let us consider definition of dependency between sets of objects.

We begin with the definition:

the set of objects $Y \subseteq \mathbb{O}$ *depends on the set* $X \subseteq \mathbb{O}$ *with respect to the similarity relation* Sim *if and only if*

$$\mathrm{Sim}(X) \leq \mathrm{Sim}(Y).$$

We shall write in symbols

$$X \xrightarrow{S} Y \quad \text{or} \quad X \xrightarrow{\mathrm{Sim}} Y.$$

In the same way we can define dependency of objects with respect to weak similarity relation sim:

$$X \xrightarrow{s} Y \quad \text{iff} \quad \mathrm{sim}(X) \leq \mathrm{sim}(Y)$$

(here $\leq$ is usual inclusion relation).

In other words, $X \xrightarrow{S} Y$ if strong similarity of attributes with respect to the set of objects X implies strong similarity of attributes with respect to the set of objects Y.

Let us define functional dependency for objects as a statement of the form $X \to Y$, where X, Y are sets of objects. Dependency $X \to Y$ holds in a relation R (or in information system) if for every pair r_1, r_2 of columns in the relation (in the system), if $r_1[X] = r_2[X]$, then $r_1[Y] = r_2[Y]$.

A multivalued dependency is a statement of the form

$$X \twoheadrightarrow Y, \quad \text{where} \quad X, Y \subseteq \mathbb{O}.$$

Let $Z = A - X - Y$. The dependency $X \twoheadrightarrow Y$ holds in R if for all r_1, r_2 in R, if $r_1[X] = r_2[X]$, then there are r_3 and r_4 in R such that

1) $r_3[X] = r_1[X]$, $r_3[Y] = r_1[Y]$, $r_3[Z] = r_2[Z]$;
2) $r_4[X] = r_2[X]$, $r_4[Y] = r_2[Y]$, $r_4[Z] = r_1[Z]$.

We say that a dependency d is a consequence with respect to relational semantics of a set of dependencies D if for all relations R, d holds in R if all the dependencies D hold in R. We will say that d is a consequence of D if for all systems, d holds in (OAVF) if all dependencies D hold in it.

We recall now axioms for multivalued dependencies:

D0: Let X, Y, Z be sets of attributes such that $X \cup Y \cup Z = A$ and $Y \cap Z \subseteq X$. Then $X \twoheadrightarrow Y$ iff $X \twoheadrightarrow Z$.
D1: If $Y \subseteq X$ then $X \twoheadrightarrow Y$.
D2: If $Z \subseteq W$ and $X \twoheadrightarrow Y$, then $XW \twoheadrightarrow YZ$.

D3: If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$ then $X \twoheadrightarrow Z - Y$.
Let D be a set of multivalued dependencies, and let $d_1 d_2 \ldots d_n$ be dependiences. We say that $d_1 d_2 \ldots d_n$ is a derivation from D if the following holds:
for all i, either $d_i \in D \cup \{d_1 \ldots d_{i-1}\}$ or d_i can be inferred from $D \cup \{d_1 \ldots d_{i-1}\}$ by an application of one of the axioms (inference values).

The inference rules are sound if every dependency d that can be derived from D is also a logical consequence of D. The inference rules (i.e. here D0–D3) are complete if every dependency d that is a consequence of D can also be derived from D. In [22] it is proved that the axioms (inference rules) for functional dependency (Armstrong axioms) and for multivalued dependency are sound and complete.

Let D be a set of similarity dependencies for sets of objects and $X \subseteq A$. By X^* we denote the closure of X with respect to D i.e. it is the set of objects B such that $X \xrightarrow{S} B$ can be deduced from D by Armstrong's axioms.

Modyfying proof of completeness for attributes we can obtain the completeness theorem for objects also:

THEOREM. *Armstrong's axioms expressing functional dependency for sets of objects are sound and complete.*

At the end of this section let me state the following language:

<u>Language:</u> $O_1, O_2, \ldots$ propositional symbols
$A_1, A_2, \ldots$ attributes symbols
s, S weak and strong similarity symbols

<u>Axioms:</u>

$$S(A_i)O_j \to s(A_1)O_j,$$

$$s(A_i)(O_1 \wedge O_2) \to s(A_i)O_1 \wedge s(A_i)O_2,$$

$$s(A_i)(O_1 \vee O_2) \leftrightarrow s(A_i)O_1 \vee s(A_i)O_2,$$

$$S(A_i)(O_1 \wedge O_2) \to S(A_i)O_1 \wedge S(A_i)O_2,$$

$$S(A_i)(O_1 \vee O_2) \leftrightarrow S(A_i)O_1 \vee S(A_i)O_2,$$

$$s(A_i \cup A_j)O_1 \to S(A_i)O_1 \vee S(A_j)O_1,$$

$$s(A_i \cap A_j)O_1 \to S(A_i)O_1 \wedge S(A_j)O_1,$$

under condition $A_i \cap A_j \neq \emptyset$.

If $A_i \cap A_j = \emptyset$ then we assume by convention:

$$S(\emptyset)O_1 = \emptyset,$$

$$S(A_i \cup A_j)O_1 \to S(A_i)O_1 \wedge S(A_j)O_1,$$

$$S(A_i)O_1 \wedge S(A_j)O_1 \to S(A_i \cap A_j)O_1.$$

Standard interpretation:

O_j – sets of objects,

A_i – sets of attributes,

s – weak similarity relation,

S – strong similarity relation.

8. Normal forms

Let me recall now the definitions of normal forms in the formulations given by Fagin and Date.

A relation schema is a pair (A, D) where A is a set of attributes and D is a set of dependences involving only these attributes.

An attribute is a key attribute if it is contained in some key. Otherwise it is a nonkey attribute.

A relation schema is in third normal form ($3NF$) if whenever $X \rightarrow A$ is a nontrivial FD of the schema, where A is a single attribute, then either X is a superkey or A is a key attribute.

A relation schema is in BCNF if whenever $X \rightarrow Y$ is a nontrivial FD of the scheme, necessarily X is a superkey.

A relation schema is in fourth normal form (4NF) if whenever $X \twoheadrightarrow Y$ is a nontrivial MVD of the schema, necessarily X is a superkey.

We shall consider normal forms in view of our similarity dependency in the second part of this article.

9. Further examples

EXAMPLE.

Real time systems with the hierarchy [RTS], [Petri net].

Input = digital data for software system
Output = digital data that control external hardware

The time between the presentation of a set of inputs and the appearance of all the associated outputs is called response time

Hard r.t.s. = resp. time is explicity bounded
Soft r.t.s. = those in which performance is degraded
but not destroyed when response
time constraints are not mat.

R.t.s. which are reactive or embedded have ongoing interactions with their environment.

Event : any occurence that results in a change
in the sequential flow or program execution.

Synchronous events : occur at predictable times
such as execution of a branch
instruction or hardware trop.

Asynchronous events occur at unpredictable points in the flow-of-control and are usually caused by external sources such as a clock signal.

Both types of events can be signaled to the CPU by hardware interrupts.

There is an inherent delay between when an interrupt occurs and when the CPU begins reacting to it, called the interrupt latency.

Task driven by interrupt, that occur aperiodically are called sporadic tasks. Systems in which interrupts occur only at fixed frequencies are called fixed-rate-systems and those with interrupts occuring sporadically are called sporadic systems.

A higher-priority task is said to preempt a lower-priority task if it interrupts the lower priority task; that is a lower priority task is running when the higher priority task signal that is about to begin.

Systems instead of round-robin or first-come-first-served scheduling are called preemptive priority systems. The priorities assigned to each of the task associated with that Interrupt.

The above description of some features of real time systems is given in the literature, but I was not able to recall the proper reference. Any way we propose the following definitions: input data D_1, D_2 are similar if the corresponding output data H_1, H_2 are also similar.

Next, we shall say that two events Ev_1 Ev_2 are similar if the change in sequential flow or in program execution will not destroy the main aim of the program or the main functions of the external hardware.

Finally, two interrupt events ie_1, ie_2 are ε-similar if the difference between corresponding interrupt latencies is smaller than $\varepsilon > 0$.

EXAMPLE. Let us assume that we have the following ROI - let it be a medical image showing the flow of contrast or the flow of the blood (cf. Goszczyńska [11]).

Vertical lines show where approximately is the contrast in the blood after $t_1, t_2, \ldots, t_n$ moments of time. Let us denote the region between t_i, t_j and the anatomical border observed on the image by $R(t_i, t_j)$. We have that

$$\operatorname{Apr}(R(t_i, t_j)) = \operatorname{Apr}(R(t_0, t_j)) - \operatorname{Apr}(R(t_0, t_i)).$$

So in this case with the flow of contrast in the blood we can observe which anatomical rejon is filled by contrast after the given period of time.

On the other hand we can also devide the ROI on equal parts along the X-axis and estimate moments of time in which every point is obtained. Here we have the function $F(p_i) = \operatorname{Apr}(R(t_0, t_j))$ and more precisely speaking

$p_i \rightarrow t_j$ i.e. given a point p_i we find t_j such that after t_j-th moment of time we obtained the point p_i.

In the examinations like the above we can compare the series of time with the similar space properties. In other words definition of similarity of time series can be used to distinquish space properties of the flow, and vice versa – by considering similar space properties we can reason about "proper" time series in a conducted experiment.

We plan the second part of the article in the following way:

1. Applications in medical imaging and in genomic databases.
2. Normal forms and decomposition algorithms in view of similarity relations.
3. Mixed FD and MVD axioms in view of Sim and sim relations.
4. Relation of similarity dependency to join and template dependency.
5. Mixed dependency with respect to ind, Ind, sim and Sim relations.
6. Application of Orłowska-Mac Caull tableaux procedure for the implication problem for association rules.
7. The role of symmetry in biocybernetics.
8. Why RDB and IST are different theories?
9. Quants, atoms and similarity – searching for a new laws of theoretical physics.
10. In the paper of Togawa and Otsuka a model of cortical neural structure consisting of threshold elements is proposed in which the single cell representation hypothesis is introduced. We suggest that it is possible to apply in this model the ideas related to tolerance and approximation, and in this way to obtain the better understanding of mental processes such as consciousness and cognition.

Final remarks

We list some problems and ideas which can be further developed:

1. Formulate and examine definitions more throughly of similarity of systems, algebras and logics.
2. Formulate dependency theory on a lattice (cf. Lee [17]).
3. Examine approximation operations on a lattice (cf. Iwiński [12]).
4. Express algebraic properties of algebras defined by operations R^* for special classes of R.
5. Find applications to medical imaging.
6. Develop similarity of many sorted algebra (Bidirectional morphisms) (cf. A. J. Pomykała [31]).
7. Make comparison between the notions of approximation in Partition space, in Cover Space and in the algebra of Images.

8. Examine the elementary axioms of geometry in view of similarity neighbourhood-formulations using different systems of axioms (cf. Tarski [41], Roberts [37]).
9. Finally, we suggest to consider also questions:
- when two systems are similar (and not homomorphic), and describe more exactly a relation between homomorphism and similarity of systems.
10. Karen Kwast [16] considered the definition of reduct and dependency in the following general setting: take any relation R satisfying only a single requivement – distribution over the attributes. Formally:
(*) $\forall_{r,s \in R} : r\langle X \rangle s$ iff $\forall_A \in X : r\langle A \rangle s$. As a consequence, $\langle \emptyset \rangle = R \times R$.
Then she formulated definitions of independent set of attributes, dispensable element, the core of X and the reduct of the set of attributes. She axiomatixed dispensable subsets and considered reduced reducts and showed the connection to normal forms. It is possible to use some result of her to relativise reducts and dependency to both Ind and Sim relations, generally speaking every relation satisfying (*) belongs to similar formalisation.

REMARK. The paper was presented on the Relmics 6 conference in Holland. Extended and orthogonal version of it shall be submitted in the Proceedings of the conference.

Finally let me recall axioms for mixed functional and multivalued dependencies cf. [2]. Axioms 1, 2, 3 are equivalent to Armstrong's axioms.

4. $X \to Y$, $YW \to Z$ implies $XW \to Z$.
5. $X \to Y$, $X \to Z$ implies $X \to YZ$.
6. $X \to YZ$, implies $X \to Y$ and $X \to Z$.
7. If $X \cup Y \cup Z = A$ and $Y \cap Z \subseteq X$ then $X \twoheadrightarrow Y$ iff $X \twoheadrightarrow Z$.
8. If $Y \subseteq X$ then $X \twoheadrightarrow Y$.
9. If $Z \subseteq W$ and $X \twoheadrightarrow Y$ then $XW \twoheadrightarrow YZ$.
10. If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$ then $X \twoheadrightarrow Z - Y$.
11. If $X \to Y$ then $X \twoheadrightarrow Y$.
As regards to mixed inference rules, the following holds:
If $X \to Y$ then $X \xrightarrow{S} Y$.
On the other hand the rules:
12. If $X \to Z$ and $Y \to Z'$ where $Z' \subseteq Z$ and Y and Z are disjoint, then $X \to Z'$.
13. If $X \twoheadrightarrow Y$ and $XY \to Z$ then $X \to Z - Y$,
do not hold. (Here $\to$ means functional dependency.)

End of part I.

References

[1] W. W. Armstrong, *Dependency structures of data base relationships*, Information Processing 74.

[2] C. Beeri, R. Fagin, J. Howard, *A complete axiomatization for functional and multivalued dependencies in database relations*, Proc. ACM SIG MOD, D. C. P. Smith, ed. New York, NY, 1977, p. 47.

[3] W. Buszkowski, E. Orłowska, *Indiscernibility-Based Formalization of Dependencies in Information Systems*, (chapter 9 in [21]).

[4] P. Ciaccia, M. Patella, P. Zezula, *Processing Complex Similarity Queries with Distance - Based Access Methods*, in: LNCS 1377, Schek, Saltor, Ramos, Alonso (Eds.), Advances in Database Technology – EDBT'98.

[5] I. Chajda, B. Zelinka, *Tolerance relation on lattices*, Casopis pro pestovani matematiky, roc. 99 (1974), Praha.

[6] E. F. Codd, *Further normalisation of the data base relational model*, IBM Research Laboratory.

[7] E. F. Codd, *A Relational Model of Data for Large Shared Data Banks*, Communications of the ACM, vol. 13, No 6, June 1970.

[8] Date, *Introduction to database systems*, (polish translation 1999).

[9] I. Dűntsch, G. Gediga, *Rough set data analysis.*

[10] R. Fagin, *Multivalued Dependencies and a New Normal Form for Relational Databases*, ACM Transactions on Database Systems, Sept. 1977, vol. 2, No 3.

[11] H. Goszczyńska, personal communication.

[12] T. B. Iwiński, *Algebraic approach to rough sets*, Bull. Polish Acad. Sci. Math. 35 (1987).

[13] J. Jărrinen, *Knowledge Representation and Rough Sets*, Tucs Dissertations, No 14, March 1999.

[14] B. Konikowska, *A logic for Reasoning about Similarity*, (chapter 15 in [8]).

[15] K. Kuratowski, *Sur l'operation $\overline{A}$ de l'Analysis Situs*, Fund. Math. 3 (1922), 182–199.

[16] K. Kwast, *Reducts on partial database relations*, manuscript 13.08.1993.

[17] T. T. Lee, *An Algebraic Theory of Relational Databases*, The Bell System Technical Journal vol. 62, No. 10, December 1983.

[18] W. MacCaull, *A proof System for Dependencies for Information Relations*, Fundamenta Informaticae, vol. 42, No 1, April 2000.

[19] M. Nowotny, Z. Pawlak, *Independence of attributes*, Bull. Polish Acad. Sci Math. 36, 1988.

[20] E. Orłowska, *Semantics of vague concepts*, Foundations of Logic and Linguistics Plenum Press, 1983.

[21] E. Orłowska, (ed.), *Incomplete Information: Rough Set Analysis*, Physica-Verlag 2000.

[22] E. Orłowska, Z. Pawlak, *Logical foundations of knowledge representation*, ICS PAS Rep. 537, Warszawa 1984.

[23] K. Otsuka and T. Togawa, *A model of cortical neural network structure*, 20th Int. Conf. of IEEE Engineering in Medicine and Biology Society, vol. 20, No 4, 1998.

[24] Z. Pawlak, *Information Systems, Theoretical Foundations*, Inform. Syst. 6 (1981), 205–218.

[25] Z. Pawlak, *Rough classification*, Internat. J. Man-Machine Studies 20 (1984).

[26] J. A. Pomykała, *Approximation Operations in Approximation Space*, Bull. Polish Acad. Sci. Math. vol. 35, No. 9–10, 1987.

[27] J. A. Pomykała, *On definability in the nondeterministic information system*, Bull. Polish Acad. Sci. Math. vol. 36, No. 3–4, 1988.
[28] J. A. Pomykała, *A remark about the paper of H. Rasiowa and W. Marek*, Bull. Pol. Ac. Math., vol. 36, No. 7–8, 1988.
[29] J. A. Pomykała, *Approximation, Similarity and Rough Constructions*, ILLC Preparation Series for Computation and Complexity Theory CT-93-07, University of Amsterdam.
[30] J. A. Pomykała, *On similarity based approximation of information*, Demonstratio Math.. 27, No. 3–4 (1994), 663–671.
[31] J. A. Pomykała, *On the notion of homomorphism in many-sorted algebras*, Demonstratio Math. 36 (2003).
[32] J. A. Pomykała, E. de Haas, *Note on categories of information systems*, Demonstratio Math. 27, No. 3–4 (1994), 641–650.
[33] J. A. Pomykała, J. M. Pomykała, *On regularity of hypergraph sequences*, Demonstratio Math. 27, No. 3–4 (1994), 651–662.
[34] H. Rasiowa, W. Marek, *Gradual Approximating Sets by Means of Equivalence Relations*, Bull. Pol. Ac. Math. vol. 35, 1987.
[35] C. M. Rauszer, *Algebraic Properties of Functional Dependencies*, Bull. Pol. Ac. Sc., vol. 33, No 9–10, 1985.
[36] C. M. Rauszer, *An Equivalence between Theory of Functional Dependencies and a Fragment of Intuitionistic logic*, Bull. Pol. Ac. Sc., vol. 33, No 9–10, 1985.
[37] F. S. Roberts, *Tolerance Geometry*, JSL, vol. XIV, No 1, January 1973.
[38] Y. Sagiv, C. Delobel, D. Stott Parker Jr., R. Fagin, *An Equivalence Between Relational Database Dependencies and a Fragment of Propositional Logic*, Journal of the ACM, vol. 28, No. 3, July 1981.
[39] A. Skowron, C. Rauszer, *The Discernibility matrices and functions in Information Systems*, (in [40]).
[40] R. Słowiński (ed.), *Intelligent Decision Support*, Handbook of Applications and Advances of the Rough Set Theory, Kluwer, 1991.
[41] A. Tarski, *What is elementary geometry*, Symposium on the Axiomatic Method.
[42] T. Traczyk, *Common extension of Boolean informational systems*, Ann. Soc. Math. Polon. Ser. IV, Fund. Inform. 2 (1978), 63–70.
[43] J. D. Ullman, *Principles of database systems*, Computer Science Press, 1980.
[44] D. Vakarelov, *Modal logics for knowledge representation systems*, TCS 90, Elsevier (1991), 433–456.
[45] A. Wiweger, *On topological rough sets*, Bull. Pol. Ac. Sc. Math., vol. 37, No 1–6, 1989.
[46] E. C. Zeeman, *The Topology of the Brain and Visual Perception*, in M. K. Fort, J. R. Ed., Prentice Hall, 1962.
[47] W. Żakowski, *Approximations in the space* (U, π), Demonstratio Math. 16 (1983).

MANAGER ACADEMY
SOCIETY FOR TRADE INITIATIVES
ul. Kawęczyńska 36
03-772 WARSAW, POLAND
pomykala_andrzej@mac.edu.pl

Received November 26, 2001.