Intelligence Integration in Distributed Knowledge Management

Dariusz Król
Wroclaw University of Technology, Poland

Ngoc Thanh Nguyen
Wroclaw University of Technology, Poland

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content:	Kristin Klinger
Managing Development Editor:	Kristin M. Roth
Senior Managing Editor:	Jennifer Neidig
Managing Editor:	Jamie Snavely
Assistant Managing Editor:	Carole Coulson
Copy Editor:	Lanette Ehrhardt
Typesetter:	Jeff Ash
Cover Design:	Lisa Tosheff
Printed at:	Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Intelligence integration in distributed knowledge management / Dariusz Krol and Ngoc Thanh Nguyen, editors.

p. cm.

Includes bibliographical references and index.

Summary: "This book covers a broad range of intelligence integration approaches in distributed knowledge systems, from Web-based systems through multi-agent and grid systems, ontology management to fuzzy approaches"--Provided by publisher.

ISBN 978-1-59904-576-4 (hardcover) -- ISBN 978-1-59904-578-8 (ebook)

1. Expert systems (Computer science) 2. Intelligent agents (Computer software) 3. Electronic data processing--Distributed processing. I. Krol, Dariusz. II. Nguyên, Ngoc Thanh.

QA76.76.E95I53475 2009

006.3--dc22

2008016377

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter I
Logical Inference Based on Incomplete and/or Fuzzy Ontologies

Juliusz L. Kulikowski
Polish Academy of Sciences, Poland

ABSTRACT

In this chapter, a concept of using incomplete or fuzzy ontologies in decision making is presented. A definition of ontology and of ontological models is given, as well as their formal representation by taxonomic trees, bi-partite graphs, multigraphs, relations, super-relations and hyper-relations. The definitions of the corresponding mathematical notions are also given. Then, the concept of ontologies representing incomplete or uncertain domain knowledge is presented. This concept is illustrated by an example of decision making in medicine. The aim of this chapter is to give an outlook on the possibility of ontological models extension in order to use them as an effective and universal form of domain knowledge representation in computer systems supporting decision making in various application areas.

INTRODUCTION

The concept of *ontology* co-opted by computer specialists from ancient philosophy means organization of concepts in domains which might encompass selected application areas: management, law, engineering, medicine, and so forth (Chute, 2005; Pisanelli, 2004). As such, ontology of a domain is a form of computer-acceptable representation of knowledge about a part of an abstract or real world being an object of consideration or decision making. In general, an ontology Ω can be presented in the form of a set C of concepts and a finite family of *ontological models* M_k, $k = 1,2,\ldots,K$, defined as relationships described on selected subsets of C. The relationships may be of various kinds; however, taxonomies T_i, $i=1,2,\ldots,I$, of the concepts are mandatory elements of the ontology. The aim of this chapter is contributing to this concept in the particular cases when ontologies are being used in computer-based decision supporting systems have not been enough finely described. The chapter is organized as follows. In the beginning a concept of ontological models and their application to decision making are presented. Here, the models based on taxonomic trees, graphs, multigraphs, relations and hyperrelations are shortly described. Nondeterministic ontological models, including fuzzy models and models based on a concept of semi-ordering of syndromes of relations, are described next. Short conclusions are collected in the last section of the chapter. Our aim in this chapter is the presentation of intuitive aspects of the proposed approach to decision making, rather than revealing its strong theoretical backgrounds.

ONTOLOGIES AND ONTOLOGICAL MODELS

Taxonomies

In the simplest cases, the idea of ontology can be reduced to a *taxonomy of concepts* assigned to objects, phenomena or processes appearing in an examined part of abstract or of real world and being analyzed from some fixed points of view. For instance, in sociological investigations a concept of *People living in the town* can be specified by a structure called a *rooted tree,* as shown in Figure 1.

Figure 1. Example of a taxonomic tree based on the attribute "Status"

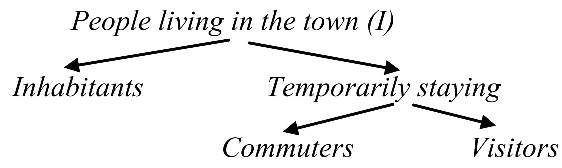
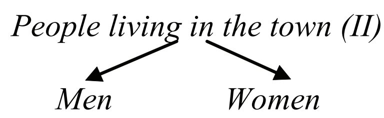
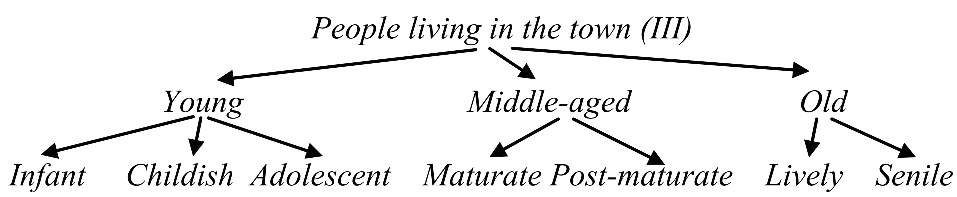


Figure 2. A taxonomic tree based on the attribute "Gender"



However, the same concept may be presented in several other ways (Figures 2 and 3) and so forth. The roots of the trees have been assigned above to the basic concept *People living in the town,* while the subjected nodes correspond to some subordered concepts. It is also assumed that on each level of any tree the subordered concepts totally cover the corresponding higher-level concept. So-interpreted rooted trees are called *taxonomic trees.* The fact that even in this simple case the part of real world under examination is represented by an ontology consisting not of a single but of several semantically linked taxonomic trees is worthy of being remarked. In general, formal structures constituting ontologies (in the above-defined, narrow sense) will be called *ontological models.* This given ontology

Figure 3. A taxonomic tree based on the attribute "Age"



thus consists of three ontological models having the form of taxonomic trees, linked semantically because their roots have been assigned to the same top-level concept.

And still, the class of problems whose solution might be supported by this ontology is rather poor. It might contain, for example, designing a database of inhabitants of the town, planning some social activities or investments in the town, or it might be used in any deliberations concerning the population of the town. However, more advanced applications of this ontology are limited by its evident deficiencies:

1. The taxonomic trees contain no information about the statistical structure of the world as a composition of designates (real entities) represented by a given tree;
2. No relationships between the concepts belonging to different taxonomic trees have been described by the ontology; and
3. Taxonomic tree do not define concepts, but only characterize hierarchical relationships between higher level and lower level concepts.

Graphs

Ontologies reduced to taxonomic trees only are thus rather ineffective in real world description and as tools supporting decision making. Let us also remark that trees in their graphical form are suitable to be analyzed by a man in the case of low numbers of nodes, and for computer-aided analysis they should be represented in the form of digital data structures.

However, trees are a sort of *graph*, the last being formally described by a triple (Tutte, 1984):

$$G = [C, \Lambda, \varphi] \qquad (1)$$

where C denotes a set of *nodes*, Λ stands for a set of *edges* and φ is a function (*incidence function*) assigning edges to some ordered pairs of nodes so that any edge can be assigned to at most one pair of nodes. An edge l_{ij} assigned to the pair $[c_i, c_j]$ of nodes is called *outgoing from* c_i and *incoming to* c_j.

There are several possibilities of defining a *tree* as a sort of graph. The simplest one is based on a statement that a graph becomes a tree if the number of its nodes is 1 larger than the number linking those edges. A tree is called a *rooted tree* if: 1) it contains exactly one node, called a *root*, to which no incoming edge is assigned and 2) to each other node exactly one in-coming edge is assigned. The nodes of a rooted tree to which no outgoing edges have been assigned are called *leafs* of the tree.

The taxonomies of an ontology are represented by rooted trees whose roots have been assigned to the top-level concepts, while other nodes correspond to the subordered concepts. Any concept in a taxonomic tree is characterized by its *level-number*, that is, the number of edges connecting the corresponding node with the root. For example, in the given taxonomy of *People living in the town (I)* the concept *Inhabitants* is a first-level, while *Visitors* is a second-level one. The top-level concepts are 0-level ones.

On the basis of graph algebra operations (Kulikowski, 1986) simple ontological models represented by graphs can be used to create more sophisticated ontological models. For instance, several taxonomic trees corresponding to the same top-level concept can be represented in the form of a unified taxonomic tree. This can be illustrated in the case of two taxonomic trees. For this purpose, a Cartesian product of the trees (in general, of the graphs) can be used. Let $G^{(1)} = [C^{(1)}, \Lambda^{(1)}, \varphi^{(1)}]$, $G^{(2)} = [C^{(2)}, \Lambda^{(2)}, \varphi^{(2)}]$ be two graphs. Their Cartesian product $G = G^{(1)} \times G^{(2)}$ is defined as a graph such that:

1. The set of its nodes $C = C^{(1)} \times C^{(2)}$, which means that each node of G is an ordered pair of some nodes of $G^{(1)}$ and $G^{(2)}$;

2. The set of its edges $\Lambda = \Lambda^{(1)} \times \Lambda^{(2)}$; and
3. Its incidence function φ assigns an edge $l_{prqs} = [l^{(1)}_{pr}, l^{(2)}_{qs}]$ to the pair of nodes $c_{pr} = [c^{(1)}_{p}, c^{(2)}_{r}]$, $c_{qs} = [c^{(1)}_{q}, c^{(2)}_{s}]$, if and only if $l^{(1)}_{pr}$ is assigned by $\varphi^{(1)}$ to the pair of nodes $[c^{(1)}_{p}, c^{(1)}_{r}]$ and $l^{(2)}_{qs}$ is assigned by $\varphi^{(2)}$ to the pair of nodes $[c^{(2)}_{q}, c^{(2)}_{s}]$.

For example, a Cartesian product of the first two taxonomic trees based on the attributes "*Status"* and *"Gender"* takes the form in Figure 4.

For the sake of formal accuracy, it has been assumed that the graphs $G^{(1)}$and $G^{(2)}$ admit existence of edges of the form $l^{(1)}_{ii}$, $l^{(2)}_{jj}$ (*loops*) linking each node with itself.

Using graphs (instead of trees only) in ontologies provides some additional possibilities to describe relationships between concepts. For example, let us take once more into consideration the first two taxonomic trees canceled to their upper two levels. Let $A = \{a_1, a_2, \ldots, a_K\}$ be a set of persons living in the given town. They can be classified according to the above-given ontology, that is, assigned to the leaves of the taxonomic trees. However, we would like to represent the persons and the assigned to them attributes: *Gender* = {*M – Man, W – Woman*}*, Status* = {*I – Inhabitant, TS – Temporarily staying*} in the form of a more concise structure. For this purpose, a graph *G'* will be constructed whose set of nodes *C'* consists of three disjoint subsets: $C' = A \cup \{M,W\} \cup \{I, ST\}$ and the incidence function φ admits edges connecting only persons with their attributes so that each person is connected with exactly two attributes: first, belonging to the subset {*M, W*} and second belonging to {*I, ST*}, as illustrated in Figure 5.

This bipartite graph represents a distribution of the attributes *Gender* and *Status* in a subset *A* of *persons living in the town*. However, it is not a tree, as it can be proved by counting and comparing the numbers of its nodes and edges. Using graphs as ontological models makes it possible using typical algebraic operations on graphs to construct more sophisticated models as compositions of some simpler ones. This can be illustrated by the following example.

Let *G'* be the bipartite graph illustrated in Figure 5 and *G''* be a bipartite graph representing the distribution of the attribute *Age* = {*Y – Young, MA – Middle aged, O – Old*} in the defined set *A* of elements (persons), as shown in Figure 6.

The *sum of graphs* $G' \cup G''$ can be defined as a graph $G^* = [C^*, \Lambda^*, \varphi^*]$ such that $C^* = C' \cup C''$, $\Lambda^* = \Lambda' \cup \Lambda''$, and φ* is an incidence function such that to a pair of nodes an edge is assigned if it is assigned by at least one of the incidence functions, φ' or φ'' (if two different edges to the given pair of nodes have been assigned by both incidence functions, then the problem, whose edges should be finally assigned to it, can be arbitrarily solved).

Using the definition of a sum of graphs to the graphs shown in Figure 5 and Figure 6, one

Figure 4. Cartesian product of two taxonomic trees

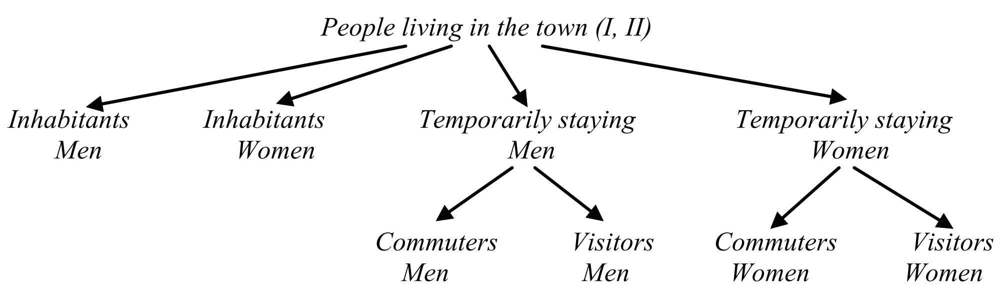


Figure 5. A bipartite graph representing distribution of two attributes

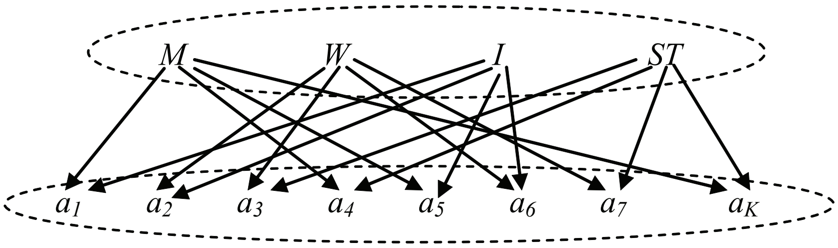


Figure 6. A bipartite graph representing distribution of the attribute Age

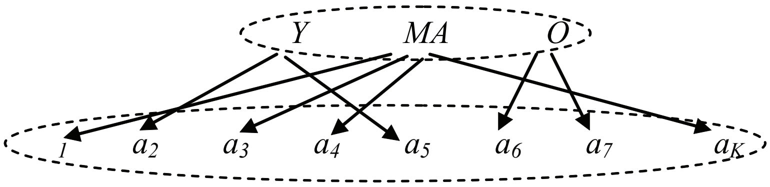
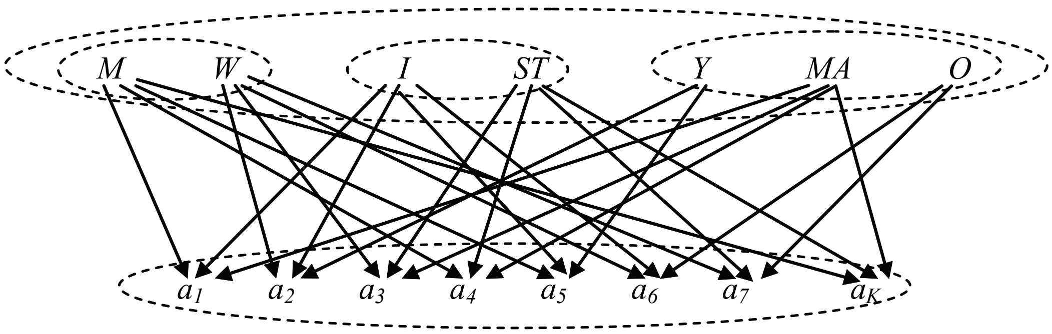


Figure 7. Sum of graphs: A bipartite graph representing a distribution of three attributes



obtains a graph G^* illustrating the distribution of three attributes, shown in Figure 7.

In similar way, ontological models describing distribution of higher numbers of attributes over fixed sets of elements using the algebra of graphs operations can be constructed. Such models may have the form of bipartite graphs, the subset of nodes representing the attributes being subdivided into mutually disjoint lower-level subsets of values of the given attributes. For effective calculations, the graphs should be stored in computer memory in the form of the corresponding connection matrices or incidence matrices (Kulikowski, 1986). However, the graphs presented in Figures 5, 6 and 7 as ontological models are rather untypical, because the idea of ontological models consists in knowledge presentation in aggregated form rather than by individual listing of instances. The algebra of graphs provides us with more universal and flexible tools for ontological models construction than taxonomic trees. Alas, in certain cases this tool is not quite suitable to a presentation of knowledge about the real world, as it can be shown by the following example.

Let us assume that a problem consists in description of the impact of papers published in a scientific journal on distribution of scientific results in the world. For this purpose, a corresponding ontological model should be constructed. Let us try to construct it in the form of a bipartite graph:

$$G = [V \cup T, \Lambda' \cup \Lambda'', \varphi], \quad (2)$$

where V is a subset of nodes assigned to affiliations of authors, T is a subset of nodes assigned to the topics covering the profile of the journal, Λ' is a subset of oriented edges (arcs) connecting nodes belonging to V with these belonging to T, Λ'' is a complementary subset of oriented edges connecting nodes belonging to T with these belonging to V and φ is an incidence function such that:

1. An edge (arc) l'_{ip} is connecting a node v_i, $v_i \in V$, with a node t_p, $t_p \in T$, if and only if at least one paper has been published in the journal such that affiliation of (at least one) its authors was v_i and the topic of the paper can be classified as belonging to t_p; and
2. An edge (arc) l''_{qj} is connecting a node t_q, $t_q \in T$, with a node v_j, $v_j \in V$, if and only if at least one paper published in the journal, whose topic can be classified as belonging to t_q, $t_q \in T$, has been cited somewhere by an author whose affiliation was v_j, $v_j \in V$.

A hypothetical part of such a graph is shown in Figure 8.

Let us take into consideration a partial graph consisting of the nodes v_2, v_5, v_8 and t_2 shown in Figure 9.

Several interpretations of these partial graphs are possible:

1. An author from v_2 has published in the journal a paper on t_2;
2. An author from v_5 has published in the journal a paper on t_2;
3. An author from v_8 has published in the journal a paper on t_2;
4. authors from v_2 and v_5 have commonly published in the journal a paper on t_2;
5. authors from v_2 and v_8 have commonly published in the journal a paper on t_2;
6. authors from v_5 and v_8 have commonly published in the journal a paper on t_2;
7. authors from v_2, v_5 and v_8 have commonly published in the journal a paper on t_2; or
8. an author from v_5 has cited at least one of the above-mentioned papers.

However, in the last case, it is not clear: was it a self-citation (four possibilities) or a citation of papers written by other authors (three possibilities)? Therefore, the ontological model presented in Figure 8 does not reflect all types of scientific information distribution caused by papers published in the given journal.

Figure 8. A bipartite graph representing affiliation of authorship and citations of papers

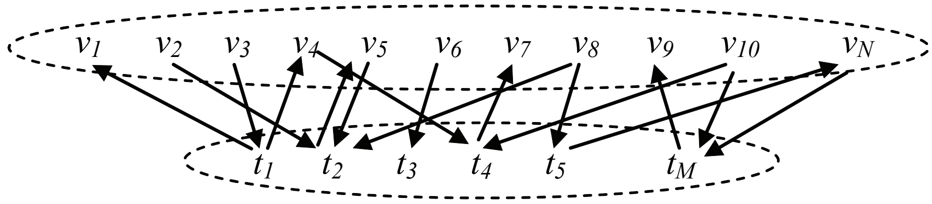
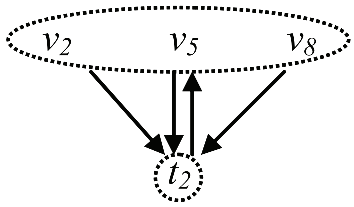


Figure 9. A partial graph of the graph shown in Figure 8



Multigraphs

In general, many relationships existing in the real world cannot be adequately presented by ontological models given in the form of graphs. Some larger possibilities are offered using *multigraphs*, that is, graphs whose incidence function admits more than one edge to any given pair of nodes. This can be illustrated by the following example.

Let us take into consideration a problem of young population flow and migration between the schools in a certain region. For analysis of the problem, an ontology consisting of several ontological models should be created, such as:

a. Taxonomic models of a regional population of pupils and students (sexuality, social background, etc.);
b. Taxonomic model of regional schools of any types and levels; or
c. Ontological model describing the flow of young population between the schools.

Our attention here will be focused on the last ontological model. For this purpose, it will be defined as a *weighted multigraph*:

$$M = [\Sigma , F, R^{+}, \varphi] \quad (3)$$

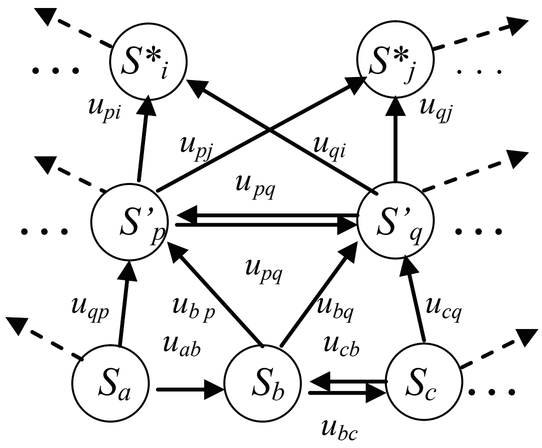
where:

- Σ is a set of nodes assigned to the regional schools (extended by adding the category *"Other"* for the schools outside the region);
- F is a set of oriented edges (arcs) assigned to the flows of pupils and students between the schools within the region, as well as coming from outside or going away from the region; the edges should also indicate a subclassification of flows based on the taxonomies following from the type a) models;
- R^{+} is a non-negative real half-axis used as a scale of flow intensity; and
- φ is a multigraph (vector) incidence function assigning to any pair of nodes $[S_i, S_j]$ an arc $f^{(r)}_{ij} \in F$ and a value $u^{(r)}_{ij} \in R^{+}$ if and only if between the corresponding pair of schools a flow of intensity $u^{(r)}_{ij}$ of the r-th category pupils (students) takes place.

A part of a multigraph of this type is illustrated in Figure 10. For the sake of simplicity multiple arcs have been replaced by the single ones and the denotations of arcs have been reduced to the weights of arcs (flow intensities in *persons/year*) presented in a concise symbolic form (in fact, they are numerical vectors whose components correspond to different sorts of pupils, for example, to *Boys* and *Girls*).

This model makes it possible to show, for example, which universities in the region are directly supplied with former pupils by given secondary schools or what is a social background of pupils or students entering the given schools. However, on the basis of this model it is not possible give a reply to a question such as, which elementary schools educate the highest percentage of pupils who graduate from the universities? The inadequacy of the above-described ontological model to answer these kinds of questions consists in the fact that graphs as well as multigraphs describe relationships between pairs of objects only, while our question concerns relationships among (in the simplest case) triples of elements:

Figure 10. A simplified partial multigraph representing the flow (migration) of pupils (students) between regional schools



[*elementary school*, *secondary school*, *university*]. The information about a detailed structure of flows entering a given node is lost as a result of aggregation of flow components, and it is not possible to reconstruct it by any outgoing flows' components analysis.

Relations

If $Q_1, Q_2, \ldots, Q_n$ are some nonempty sets taken in the given linear order and $C = Q_1 \times Q_2 \times \ldots \times Q_n$ is their Cartesian product, then any subset:

$$r \subseteq C \tag{4}$$

is called a relation described on the (linearly ordered) family of sets $[Q_1, Q_2, \ldots, Q_n]$. According to the definition, r is a set of n-tuples of the form $[a, b, \ldots, h]$ such that $a \in Q_1$, $b \in Q_2, \ldots,$ and $h \in Q_n$, called sometimes *syndromes* of the relation.

For a fixed linearly ordered family of sets and the corresponding Cartesian product C it is possible to take into consideration a family Φ of all possible subsets of C including C itself and an empty subset $\varnothing$. Φ is thus a family of all possible relations that can be defined on the given family of sets. On the other hand, it is possible to apply to it the general set-algebraic rules (Rasiowa & Sikorski, 1968) which in this case becomes an algebra of relations described on the family of sets $[Q_1, Q_2, \ldots, Q_n]$. Moreover, this algebra can also be extended on all relations described on any subsets of this family assuming that the original linear order has been preserved (Kulikowski, 1972). The *extended algebra of relations*, being in fact a sort of Boolean algebra, becomes a flexible tool not only for description of relations between any final number of arguments, but also for the creation of more sophisticated relations as algebraic compositions of some simpler ones, as well as for the creation of higher-order relations (*superrelations*) defined as relations between relations (Kulikowski, 1992).

Multi-argument relations cannot be easily presented in graphical form. However, there are several methods of description of a new relation:

- By listing the syndromes of the relation;
- Be presenting it as an algebraic composition of some other, known relations; or
- By presenting a testing function making it possible to decide whether the relation is satisfied by any given syndrome.

The first method can be illustrated by the following example. The problem of young population flow and migration between the schools will be considered again. We would like to create an ontological model making possible the investigation of contribution of elementary schools in the region to the educational productivity of universities, taking into account the sex of the graduate students. For this purpose five sets will be taken into consideration:

- $Q_1 = \{B, G\}$ describing sex (*Boys, Girls*};
- $Q_2 = \{S_a, S_b,\ldots,S_h\}$ describing elementary schools in the given region;
- $Q_3 = \{S'_p, S'_q,\ldots, S'_t\}$ describing secondary schools;
- $Q_4 = \{S^*_i, S^*_j,\ldots,S^*_k\}$ describing universities; and
- $Q_5 \equiv R^+$ a non-negative real half-axis representing flow intensities.

On the basis of the Cartesian product $C = Q_1 \times Q_2 \times Q_3 \times Q_4 \times Q_5$ it can be defined a relation r given in the form of a list of syndromes of the form

$$\boldsymbol{v} = [x, S_\alpha, S'_\beta, S^*_\gamma, w], \qquad (5)$$

where $x \in Q_1$, $S_\alpha \in Q_2$, $S'_\beta \in Q_3$, $S^*_\gamma \in Q_4$, $w \in Q_5$. Each syndrome represents a component of the flow with the additional characterizing it parameters. The relation can easily be represented in computer, however, it cannot be so easily plotted on a plane. Answering the former question: what is the contribution of a given elementary school, let it be S_α, to supplying a given university, let it be S^*_γ, with, say, girl students (G) is then reduced to selection from r, a subrelation $r' \subseteq r$ consisting of all syndromes of the form

$$\boldsymbol{v'} = [G, S_\alpha, \mathrm{F}, S^*_\gamma, w], \qquad (6)$$

where F denotes an undefined data value (here denoting any secondary school). The final answer can be reached by summing over F all values w of the syndromes of r'.

As mentioned before, the algebra of relations makes possible the combining of ontological models in order to get more suitable forms of reality description. For example, if $r^{(\kappa)}, r^{(\lambda)}$ are two relations of similar structure described on the same family of sets $[Q_1, Q_2,\ldots, Q_n]$, then a sum of relations

$$r = r^{(\kappa)} \cup r^{(\lambda)} \qquad (7)$$

is a relation consisting of all syndromes satisfying $r^{(\kappa)}$ or $r^{(\lambda)}$. In the above-described example, if $r^{(\kappa)}$ and $r^{(\lambda)}$ describe the flow of pupils (students) in two consecutive school years, then r describes it in the two school years taken together.

Another situation arises if the relations $r^{(\kappa)}$ and $r^{(\lambda)}$ are described on different families of sets, say, respectively, on $[Q^{(\kappa)}_1, Q^{(\kappa)}_2,\ldots, Q^{(\kappa)}_n]$ and $[Q^{(\lambda)}_1, Q^{(\lambda)}_2,\ldots, Q^{(\lambda)}_m]$. In this case, assuming that both families are conformably ordered, the algebraic operations can be defined according to the extended relations algebra rules (Kulikowski, 1992). In particular, a sum of relations can be defined as a relation described on the sum of families of sets $[Q^{(\kappa)}_1, Q^{(\kappa)}_2,\ldots, Q^{(\kappa)}_n] \cup [Q^{(\lambda)}_1, Q^{(\lambda)}_2,\ldots, Q^{(\lambda)}_m]$ consisting of syndromes such that each syndrome even 1) in its part belonging to $[Q^{(\kappa)}_1, Q^{(\kappa)}_2,\ldots, Q^{(\kappa)}_n]$ satisfies $r^{(\kappa)}$, or 2) in its part belonging to $[Q^{(\lambda)}_1, Q^{(\lambda)}_2,\ldots, Q^{(\lambda)}_m]$ it satisfies $r^{(\lambda)}$.

As an example, let a problem of air-passengers flow intensity in selected airports be considered. For this purpose, a set $A = \{a_1, a_2, \ldots, a_K\}$ of international airports will be considered. It will be multiplied in three versions: as departure airports A', as transit airports A^* and destination airports A''. In addition, a set V of flow intensity values, $V \equiv R^+$, where R^+ is a non-negative half-axis, will be taken into account. Then two Cartesian products will be constructed: $C = A' \times A'' \times R^+$, $C^* = A' \times A^* \times A'' \times R^+$. Let us also select a subset $D \subset A$ of particular interest, say, of international airports in a certain country. On the basis of C

two relations can be described: 1) r' describing direct flights starting from any airport of D, $D \subset A'$, and terminating in any airport of A', and 2) r'' describing direct flights starting from any airport of A' and terminating in any airport of D, $D \subset A''$. In addition, on the basis of C^* a relation r^* describing transit flights from any airport of A' through any airport of D to any airport of A'' will be described. The syndromes of the above-mentioned relations thus indicate the names of starting, transit or terminating airports between which the flights took place within a certain time-period, as well as intensity of the corresponding flow of passengers. Let us assume that a total flow of passengers through the airports of D are of interest. Then, an extended algebraic sum of relations: $r = r' \cup r^* \cup r''$ should be taken into account and the corresponding arithmetic sum of intensities should be calculated. The syndromes of r are quadruples of a general form: *starting airport*, *transit airport*, *terminate airport*, *intensity of the flow of passengers*, such that exactly one starting, transit or terminate airport belongs to D, and the other airports within the set A are unlimited.

In a similar way, extended intersection of relations can be used in ontological models creation. For example, let us take into consideration a family $F = [Q_1, Q_2, Q_3, Q_4, Q_5, Q_6]$ of sets where Q_1 denotes a set of names of *teachers*, Q_2 a set of *subjects*, Q_3 a set of *scholar classes*, Q_4 a set of *classrooms*, Q_5 a set of *weekdays* and Q_6 a set of *scholar hours*. On the Cartesian product $C' = Q_1 \times Q_2 \times Q_3$ it may be defined as a relation r' between *teachers, subjects* and *scholar classes.* On the Cartesian product $C'' = Q_2 \times Q_3 \times Q_5$ a relation r'' between *subjects, scholar classes* and *weekdays* in a similar way can be defined. At last, a relation r''' between *classrooms*, *weekdays* and *scholar hours* on the Cartesian product $Q_4 \times Q_5 \times Q_6$ can be established. The relations r', r'' and r''' can be established independently of each on each other by taking into account some constraints imposed on the corresponding syndromes. Then, a problem arises of the construction of a relation r containing all syndromes consisting of *teachers, subjects, scholar classes, classrooms, weekdays* and *scholar hours* satisfying the constraints. This relation can be defined as an extended algebraic intersection of relations:

$$r = r' \cap r'' \cap r''' \quad (8)$$

whose syndromes, by definition, projected on C' satisfy the relation r', projected on C'' satisfy r'', and projected on C''' satisfy r'''. Then, finally, on the basis of the relation r, an optimized timetable can be constructed.

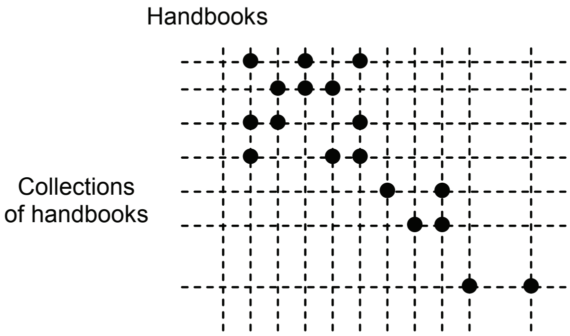
It might seem that the extended algebra of relations is a tool sufficient enough to construct a large class of ontological models. The following examples show that it is not quite so.

Hyper-Graphs

Let C be a set of scholar handbooks offered at a book market. A problem of recommending collections of handbooks for teaching given subjects during a multiyear education process will be considered. For this purpose, an ontology describing the regional educational subsystem should be constructed. However, our attention will be focused on ontological models describing the admissible collections of handbooks satisfying some educational requirements. Otherwise speaking, it is necessary to select according to some educational criteria a family of subsets of C assuming that the subsets are not obviously mutually disjoint. The first possibility is to construct a hyper-graph (Berge, 1973) whose nodes are assigned to the elements of C and any subset of nodes assigned to the handbooks satisfying the educational criteria constitutes a hyper-edge of the hyper-graph. Such hyper-graphs can be represented by a diagram, shown in Figure 11.

In this diagram vertical lines represent nodes, while dots lying on horizontal lines represent hyper-edges. A serious shortcoming of hyper-

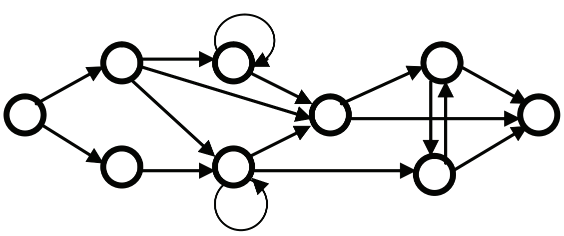
Figure 11. Diagram of a hyper-graph



graphs used as ontological models exists in their inability to describe an order (if any exists) of the nodes belonging to the same hyper-edge, and of belonging to several hyper-edges where the orders are different. This will be illustrated by the next example.

It will be taken into consideration a medical clinic specialized in a certain sort of diagnostic and therapeutic services. It is desired to create an ontology describing admissible processes of individual medical treatment of patients. For example, each process of this type should start by registration of the patient, then a series of diagnostic procedures should be followed by the proper medical treatment, and at last the process is finished by discharging the patient from the clinic. It is possible to construct a directed graph representing admissible logical sequences of operations of which any instance of medical

Figure 12. A directed graph representing admissible logical sequences of operations



treatment process consists. An example of such a graph is given in Figure 12. The nodes represent here operations, while directed edges (arcs) link pairs of operations which can be performed just after the former one.

Any instance of the process can be embedded in the graph. Each process starts at the extreme-left node and is finished in the extreme-right node. The loops existing at two nodes show that the given operations can be repeated. However, it is not possible to separate from the graph the admissible instances of the process only. For example, it is not evident whether some operations can be repeated one, two or more times and whether or not they can be repeated independently on the preceding subsequences of operations. A more complete ontological model of medical treatment processes should thus represent all medically or organizationally admissible sequences of operations of various lengths, which can be embedded in the above-presented graph.

Hyper-Relations

Let $F = [Q_1, Q_2, \ldots, Q_n]$ be a finite family of sets. A family K_F of all subfamilies of F including F itself and an empty family $\varnothing$ will be considered. Then, each subfamily $H_g \subseteq F$, $H_g \in K_F$, g denoting the subfamilies, creates a family U_g of all linearly ordered subfamilies of sets obtained as a result of all possible permutations of H_g. Each element of U_g is thus a linearly ordered subfamily of F and, as such, a Cartesian product of its elements can be constructed. Next, on the basis of this Cartesian product some relations can be created. The syndromes of each such relation are thus some finite strings of elements belonging to and taken in the order of sets constituting the given Cartesian product. A *first-type hyper-relation* (a *h*-relation) is then defined as any sum (in the set algebra sense) of relations defined on any subfamilies H_g created in the above-defined way (Kulikowski, 2006). The following example should make it clearer. There will be taken into consideration:

- A family of sets $F = \{A, B, D\}$;
- A family of its subfamilies $K_F = \{\emptyset, \{A\},\{B\},\{D\},\{A,B\},\{A,D\},\{B, D\},\{A, B, D\}\}$;
- Selected subfamilies of sets $H_6 = \{A, D\}$, $H_7 = \{B, D\}$;
- Families of permutations of the subfamilies H_6 and H_7:
  $U_6 = \{[A, D], [D, A]\}$, $U_7 = \{[B, D], [D, B]\}$;
- Cartesian products based on U_6 and U_7:
  $C_{6,1} = A \times D$, $C_{6,2} = D \times A$, $C_{7,1} = B \times D$, $C_{7,2} = D \times B$;
- Selected relations described on the above-given Cartesian products:
  $r' \subseteq C_{6,1}$, $r'' \subseteq C_{6,2}$, $r''' \subseteq C_{7,2}$;
- *h*-relations:
  $H_1 = A \cup D \cup r' \cup r''$, $H_2 = r' \cup r'' \cup r'''$, and so forth.

The syndromes of H_1 are linearly ordered strings consisting of one or two element while all syndromes of H_2 are strings consisting of two elements.

On the basis of any given family *F* of sets, a universe U_F of all possible *h*-relations created on the basis of *F* can be considered. The elements of U_F (i.e., *h*-relations) being defined as some sets are subjected to the set algebra rules, which in this case can be interpreted as an algebra of *h*-relations. This makes it possible to create more sophisticated *h*-relations as algebraic compositions of some simpler ones. Hyper-relations, as well as the algebra of hyper-relations, are thus a flexible tool for the creation of ontological models, more powerful than graphs or relations.

NONDETERMINISTIC ONTOLOGIES

Until now, ontologies consisting of deterministic models were considered. We tried to show that decision making based on ontologies may be improved by using ontological models suitable to the description of the area of interest with a required accuracy. "Suitable" means here the covering of the area of interest without making it too large, based not on aggregated concepts, nor going too deeply into the details. However, ontologies being a form of presentation of our knowledge about the world, they may be also based on ambiguous concepts and nondeterministic relations. Decision making based on uncertain information is one of basic problems in artificial intelligence investigations (Bubnicki, 2002; Grzegorzewski, Hryniewicz, & Gil, 2002; Rutkowski, Tadeusiewicz, Zadeh, & Zurada, 2006). Only certain aspects of this problem, strongly connected with using nondeterministic ontological models in decision making, will be considered here.

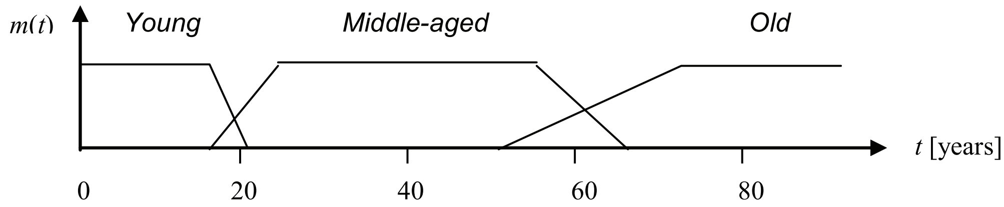
Fuzzy Ontological Models

Let us go back to the taxonomic trees shown in Figure 2 and Figure 3. In the first case, the concepts *Men* and *Women* are strongly defined and the respective ontological model is no doubt deterministic. On the other hand, the concepts *Young, Middle-aged* and *Old* used in the second ontological model can be interpreted:

a. Deterministically, as:
   Young ≡ [aged not more than 18 years],
   Middle-aged ≡ [aged more than 18 and not more than 60 years],
   Old ≡ [aged more than 60 years]; or
b. Nondeterministically, say, using a fuzzy sets approach (Zadeh, 1975a, 1975b, 1975c) and the membership functions shown in Figure 13.

In the second case, a particular case of nondeterministic ontologies, a *fuzzy ontology,* is presented. It might seem that no essential difference between the deterministic and the above-mentioned nondeterministic specification of concepts exists, because the strongly-defined membership functions make strong fixing between the *"Young"* and *"Middle-aged"* as well as

Figure 13. Fuzzy specification of the subconcepts of Age

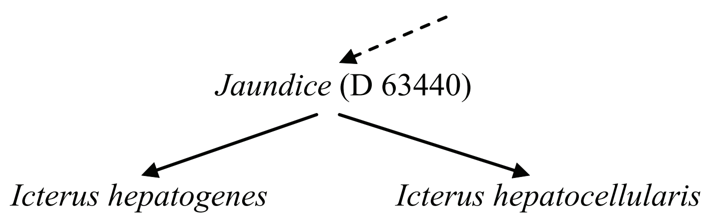


between the *"Middle-aged"* and *"Old"* concepts possibly due to a *defuzzyfication,* that is, to an operation consisting in fixing strong limits between the concepts. However, a difference between the deterministic and fuzzy ontology becomes evident if a practical decision based on fuzzy ontology is to be made. For example, if building of a network of sport fields for young people in the town is considered, then the fuzzy concept of "*young*" better suits to a characterization of the expected users of the fields than the deterministic one. The problem is that even if a border between *Young* and *Middle-aged* at the age of 18 years was fixed, not all people younger than 18 years would like to attend the sport fields and, on the other hand, some people older than 18 years would like to attend them. This example shows that the *fuzzy* or *nondeterministic* ontological model does not mean *worse* than a *deterministic* one in the case if it better describes the state of our knowledge about the area of interest and more exact knowledge is not available.

One should distinguish between decision making based on incomplete and on fuzzy ontology. Let us consider a taxonomic subtree of liver diseases (Coté, 1975).

Let it be known that a) a certain drug *D* is effective and recommended in *icterus hepatogenes* and rather ineffective in the case of *icterus hepatocellularis* therapy, and b) in a given population *p%* of patients diagnosed as affected with *liver jaundice* are in fact suffering from *icterus hepatogenes* and (100-*p*)% are suffering from *icterus hepatogenes*. Then, if a patient has been roughly diagnosed as affected with *liver jaundice* without indication of the type of jaundice and he is recommended to take the drug *D*, the decision is based on an incomplete model, canceled to its higher level ontological model. The expected effectiveness of the therapy in this case is at most *p*%. On the other hand, if diagnostic methods used to discriminate between the *hepatogenes* and *hepatocellularis icterus* work with *q% specifity* (i.e., the percentage of patients diagnosed as affected by a given disease really suffering from it), the given patient has been diagnosed as affected by *icterus hepatogenes* and, consequently, he has been recommended to take *D,* then the expected effectiveness of this therapy will be at most *q%*. The ontological model on which this decision is based is complete; however, if it is interpreted

Figure 14. A selected taxonomic subtree of liver diseases



as a taxonomy of possible diseases in patients diagnosed as affected by one or another type of liver jaundice, it is fuzzy. Therefore, incompleteness and fuzziness of ontological models lead to deterioration of decisions based on them. However, numerical values of this deterioration should be differently evaluated.

It is easy to take into account fuzziness in ontological models based on relations or hyper-relations. For this purpose, a set M defining an linearly ordered numerical scale of *weights* of syndromes will be defined. If $C = Q_1 \times Q_2 \times ... \times Q_n$ is a Cartesian product of n given sets on which a relation r has been defined, then an extended Cartesian product $C^* = C \times M$ and a relation $r^* \subseteq C^*$ can be taken in consideration. The syndromes of r^* have the form:

$$\sigma^* = [\sigma, \mu], \qquad (9)$$

where $\sigma \in C$, $\mu \in M$. The component μ in the simplest case may describe a *membership level* of σ as a syndrome of the fuzzy relation r^*. The membership level, in general, is not subjected to any additional constraints: it is used only to a relative assessment of the syndromes of r^*. If, for example $\sigma^*_1 = [\sigma_1, \mu_1]$ and $\sigma^*_2 = [\sigma_2, \mu_2]$ such that $\mu_1 < \mu_2$ are given, then this means that σ_1 in a certain sense is *"*less credible*"* than σ_2 as a syndrome of the relation. According to the context, "less credible" may mean: "less frequently," "with lower probability," "guaranteed by less known experts," and so forth. In many cases, such fuzzification of ontological models is sufficient as a basis of decision making. Let us remark that in the above-described example exact numerical membership levels are unimportant for decision making, because $\mu_1 < \mu_2$ holds for 0.1 < 0.15, 2 < 3, 46% < 58%, and so forth. This leads to a conclusion that the membership scale M can be defined up to any increasing continuous functional transformation preserving the sign of values.

Semi-Ordered Ontological Models

In certain cases using fuzzy ontological models of the above-presented type does not satisfy the requirements of effective decision making. Let us consider a case of choosing effective drugs for therapy of a certain class of diseases. For this purpose, two relation-based ontological models will be taken into account. First, there will be considered the following sets:

- Q_1 a set of available drugs;
- Q_2 a set of medical indications (diseases) for applying the drugs;
- Q_3 medical contraindications for applying the drugs; and
- Q_4 cost of the drug.

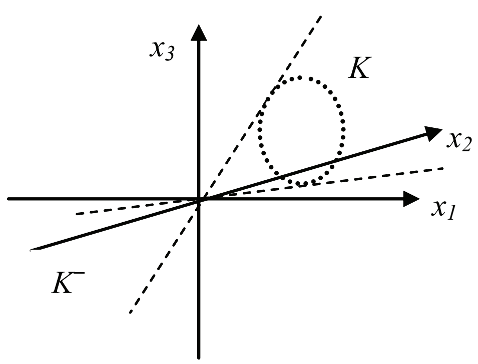
On the basis of these sets, a relation r' can be defined assuming that if there are more than one medical indication of contraindication for applying a given drug, then they should be presented by several relation syndromes. In addition, the following sets will be considered:

- Q_1 a set of drugs (as before);
- Q'_2 a set of diseases;
- Q'_3 a set of additional patients' health state characteristics; and
- M a scale of *credibility values.*

On the basis of these sets a fuzzy relation r'' can be defined, the component μ, $\mu \in M$, expresses the credibility (a positive real value between 0 and 1) that the drug is effective if the patient has been properly diagnosed and his additional health state characteristics have been correctly described. In order to combine information contained in r' and r'' an intersection of the relations $r = r' \cap r''$ defined, according to the extended relation algebra rule (Kulikowski, 1972, 1992), as a relation described on the Cartesian product: $C = Q_1 \times Q_2 \times Q_3 \times Q_4 \times Q'_2 \times Q'_3 \times M,$

such that its syndromes projected on $Q_1 \times Q_2 \times Q_3 \times Q_4$ satisfy *r'* and projected on $Q_1 \times Q'_2 \times Q'_3 \times M$ satisfy *r"* will be constructed. The relation *r* is fuzzy due to the credibility component $\mu \in M$ of its syndromes. However, it is not quite suitable to the requirements of making decisions about recommendation of a drug for the given patient. This is because it is not quite sure that: 1) the real patient's disease is identical to the result of diagnosis (the element of Q'_2) and, as a consequence, whether there is a full consistency in the syndromes between the elements of Q_2 and Q'_2, and 2) for similar reasons, whether there is a full consistency between the elements of Q_3 and Q'_3. Therefore, the relation *r* should be extended by adjoining to it two components: *a*) a measure ν of logical consistency between the syndrome's components of Q_2 and Q'_2, and *b*) a measure ρ of logical consistency between the syndrome's components of Q_3 and Q'_3. The way of defining the *logical consistency* is not substantial here. We would like only to show that the extended relation *r** contains three parameters, μ, ν and ρ, causing its fuzziness. At last, it becomes necessary to establish a method of relative assessment of the syndromes according to the values of the weight vectors $\boldsymbol{w} = [\mu, \nu, \rho]$. For this purpose, an additional ontological model can be created: a linear 3-dimensional semi-ordered real vector space $K^{(3)}$. One possibility of doing this exists in defining $K^{(3)}$ as a Kantorovich space (Kantorovich, Vulich, & Pinsker, 1950). As a component of ontology, $K^{(3)}$ represents the preferences established by the decision-makers (medical doctors, in the above-described example) for choosing the best decision(-s) from those, indicated by the fuzzy relational ontological model *r**. From a formal point of view, the principle of semi-ordering of vectors in a *K*-space consists in defining in a given linear vector space a *non-negative cone* K^+, its mirror-reflection being denoted by K^-, as illustrated in Figure 15.

Figure 15. Illustration of a 3-dimensional Kantorovich space



If $\boldsymbol{v}^{(1)}$ and $\boldsymbol{v}^{(2)}$ are two vectors belonging to the *K*-space and their difference satisfies the condition:

$$\boldsymbol{v}^{(1)} - \boldsymbol{v}^{(2)} \in K^+ \qquad (10)$$

then it is said that $\boldsymbol{v}^{(1)}$ *is preceded by* $\boldsymbol{v}^{(2)}$ ($\boldsymbol{v}^{(1)}$ *is preferred with respect to* $\boldsymbol{v}^{(2)}$, what can be shortly denoted as $\boldsymbol{v}^{(1)} \prec \boldsymbol{v}^{(2)}$. If neither $\boldsymbol{v}^{(1)} \prec \boldsymbol{v}^{(2)}$ nor $\boldsymbol{v}^{(1)} \prec \boldsymbol{v}^{(2)}$, then it is said that $\boldsymbol{v}^{(1)}$ and $\boldsymbol{v}^{(2)}$ are *mutually incomparable*. In the last case, some additional criteria should be used in order to select the best decision from a subset of mutually incomparable ones.

CONCLUSION

It was shown in the chapter that ontologies used as a support in computer-aided decision making usually consist of several ontological models being a form of presentation of knowledge about a given area of interest. Ontological models can be constructed on the basis of various formal models: taxonomic trees, graphs, multigraphs, relations, hyper-relations, and so forth. However, deterministic models not always describe adequately the state of our knowledge about the area of interest. That is why in certain cases canceled or otherwise incomplete ontological models as well as nondeterministic models should be used. The nondeterministic ontological models, in the simplest case, may be presented as fuzzy models, that is, models based on fuzzy concepts in the Zadeh sense. In a more general case, nondeterministic models can be presented in the form of nondeterministic relations, that is, relations whose syndromes have been semi-ordered. In particular, the concept of a semi-ordered linear vector space to construction of nondeterministic ontological models can be used.

REFERENCES

Berge, C. (1973). *Graphs and hypergraphs*. Amsterdam: North-Holland.

Bubnicki, Z. (2002). *Uncertain logics, variables and systems*. Springer-Verlag. LNICS No. 276.

Chute, C.G. (2005). Medical concept representation. In H. Chen et al. (Eds.), *Medical informatics: Knowledge management and data mining in biomedicine* (pp. 163-182). Springer-Verlag.

Coté, R.A. (Ed.). (1975). SNOMED: Systematized nomenclature of medicine. *Diseases*. ACP.

Grzegorzewski, P., Hryniewicz, O., & Gil, M. A. (Eds.). (2002). *Soft methods in probability, statistics and data analysis*. Heidelberg, Germany: Physica Verlag.

Kantorovich, L.B., Vulich, B.Z., & Pinsker, A.G. (1950). *Functional analysis in semi-ordered spaces* (in Russian). Moscow: GITTL.

Kulikowski, J.L. (1972). *An algebraic approach to the recognition of patterns*. CISM Lecture Notes No. 85. Wien: Springer-Verlag.

Kulikowski, J.L. (1986). *Outline of the theory of graphs (*in Polish). Warsaw, Poland: PWN.

Kulikowski, J.L. (1992). Relational approach to structural analysis of images. *Machine Graphics and Vision, 1*(1/2), 299-309.

Kulikowski, J.L. (2006). Description of irregular composite objects by hyper-relations. In K. Wojciechowski et al. (Eds.), *Computer vision and graphics* (pp. 141-146). Springer-Verlag.

Pisanelli, D.M. (Ed.). (2004). *Ontologies in medicine.* Amsterdam: IOS Press.

Rasiowa, H., & Sikorski, R. (1968). *The mathematics of metamathematics.* Warsaw: PWN.

Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., & Zurada, J. (Eds.). (2006). *Artificial intelligence and soft computing–ICAISC 2006.* Berlin: Springer-Verlag.

Tutte, W.T. (1984). *Graph theory.* Menlo Park, CA: Addison-Wesley.

Zadeh, L.A. (1975a). The concept of a linguistic variable and its application to approximate reasoning. Part I. *Information Science, 8*, 199-249.

Zadeh, L.A. (1975b). The concept of a linguistic variable and its application to approximate reasoning. Part II. *Information Science, 8*, 301-357.

Zadeh, L.A. (1975c). The concept of a linguistic variable and its application to approximate reasoning. Part III. *Information Science, 9*, 43-80.