DOI 10.2478/ama-2018-0025

THE CLOSEST INCOMPLETE DISTRIBUTED INFORMATION SYSTEM FOR MEDICAL QUERY ANSWERING SYSTEM

Katarzyna IGNATIUK*, Agnieszka DARDZIŃSKA*

*Department of Biomechanics and Biomedical Engineering, Bialystok University of Technology, ul. Wiejska 45c, 15-351 Białystok, Poland

ignatiukkatarzyna@gmail.com, a.dardzinska@pb.edu.pl

received 29 September 2017, revised 28 June 2018, accepted 30 June 2018

Abstract: The common issue for medical information systems are missing values. Generally, gaps are filled by statistically suggested values or rule-based methods. Another approach is to use the knowledge of information systems working under the same ontology. The medical incomplete system receives a query unable to answer, because of some unknown patient attributes. So, it has to communicate with other medical systems. The result of the collaboration is collective knowledgebase. In this paper, we propose a measure supporting choice of closest pair of systems. It determines the distance between the two systems. We use ERID algorithm to extract rules from incomplete, distributed information systems. Each constructed rule has confidence and support. They allowed to determine the distance between a pair of medical information systems. The proposed solution was verified on the basis of several "manipulated" medical information systems. Next, the solution was verified in systems with randomly selected data. The satisfying results were obtained and based on them, the proposed measure can be successfully used in medical systems to support the work of doctors and the treatment of patients.

Key words: Query Answering System, Knowledge Extraction, Incomplete Medical Information System, Distributed Medical Information System, The Closest Information System

1. INTRODUCTION

Millions of decisions shape human life. Some of them are minor. Other seriously affect the further course of life. That is why in the last few decades, methods supporting the decision process in areas important for the quality of human life have been improved, including medicine (Yoo et al., 2012). Medical databases have a huge amount of information about patients and provide great opportunities for data mining algorithms development and putting new challenges ahead of them, for example how to treat information gaps or the variety of information types, how to process the enormity of accumulated knowledge or assess its significance.

In this paper, special attention will be paid to the incompleteness of medical information systems. This is an important issue affecting doctors decision and patient treatment. The aim of authors was to propose the method supporting problem solution, based on the rules, their confidences and their supports.

2. THE INFORMATION SYSTEM AND RULES

The concept of an information system is commonly used to define a combination of components that collaborate to collect, process, store and disseminate information to support the decision, control, analysis and visualization of an organization (Laudon and Laudon, 2012). In the theory of classification rules and action rules, to describe interactions between data, the notion of the information system represented formally below is also considered correct. It says that the collected data can create the information system. It is enough that the database objects are described by a finite number of attributes whose values are also defined and there is an interaction between values that produce information. More formally, the information system (S) we mean a triplet $S = (X, A, V)$, where (Dardzinska, 2013):

- X is a nonempty, finite set of objects,
- A is a nonempty, finite set of attributes,
- $V = \cup \{V_a : a \in A\}$ is a set of attribute values, where V_a is a set of attribute values a for any $a \in A$.

We assume, that (Dardzinska, 2013):

- $V_a \cap V_b = \emptyset$ for any $a, b \in A$ such that $a \neq b$,
- $a : X \rightarrow V_a$ is a partial function for every $a \in A$.

When some attribute a does not return for object X any value V_a then the information system is incomplete. Tab. 1 presents example of this information system which is incomplete.

Tab. 1. Part of the incomplete information system S based on blood database

Patient	RBC [10^6/µl]	HGB [g/dl]	WBC [10^3/mm^3]	MCH [pg]	PLT [10^3/µl]	Blood disease
x_1	4.1	9.0	6.2	23.2		anemia
x_2	2.8		41.8	23.4	87.1	leukemia
x_3	7.4	19.2		36.1	350.7	polycythemia vera
x_4	3.2	9.5	3.2		90.3	pancytopenia

Information system objects from Tab. 1 are patients. The attributes of these objects are the characteristics of patients: {RBC, HGB, WBC, MCH, PLT, Blood disease} and the values of attrib-

utes create values of these features {4.1, 9.0,…, 90.3}. Attributes can be divided into stable, semi-stable and flexible. For example stable is attribute gender, semi-stable is patient's age and flexible are blood parameters RBC, HGB, WBC, MCH and PLT from Tab.1. One of the flexible attributes is decision attribute against which the patient can be classified, for example blood disease.

Single patient data informs about his or her health condition. To change this state, the doctor indicates treatment based on his knowledge and many years of practice or he can use knowledge in the form of patterns extracted from the medical databases. One of the methods of presenting patterns are classification rules or action rules, whose general form is represented by the dependence (Dardzinska, 2013): "antecedent→consequent". The antecedent is created by the values of the classification attributes (stable, semi-stable, flexible). The consequent is created by the values of the decision attribute. Example rule for the object x_4 from Tab. 1 has the form (RBC = 3.2) * (HGB = 9.5) * (WBC = 3.2) * (MCH = 22.3) * (PLT = 90.3) → (blood disease = pancytopenia). The value of the blood disease is a consequent and the rest of the rule is the antecedent.

To assess each extracted rule, two statistical measures are generally used: confidence (conf) and support (sup). The support determines by what percentage of all rules, the specific rule is supported. More formally, the support of the rule "antecedent → consequent" is the ratio of the observation number which fulfill the condition "antecedent ∩ consequent" to the number of all observations. The confidence determines how much we can trust the rule. More formally, the confidence of the rule "antecedent → consequent" it is the ratio of the observation number that fulfill the condition "antecedent ∩ consequent" to the number of observation that fulfill the condition "antecedent".

2.1. Types of Incomplete Information Systems

One of the main problems of medical information systems are information gaps, which hinder the process of extracting patterns and affect information about the extracted rules (conf and sup).

By incompleteness we mean empty spaces in information system because of different reasons. In medical information system, the cause of these incompleteness may be unreliable supplemented patient documentation, loss of part of documentation, errors during transferring information from paper to electronic documentation or intentional encrypting of sensitive patient data. Tab. 1 shows an example of an incomplete information system based on a medical database.

If the all attributes in S are defined and described, then S is complete. We say then, that all attributes in S are total functions. Otherwise, the system is incomplete. Information system presented in Tab. 1 is incomplete, because values PLT(x_1), HGB(x_2), WBC(x_3) and MCH(x_4) are not defined.

In the information systems may occur 4 types of incompleteness (Dardzinska, 2013).

Type 1

Incompleteness of the first type is defined by the assumption that at least one attribute $a \in A$ is a partial function (Dardzinska, 2013):

$$(\exists x \in X)(\exists a \in A)[a(x) = Null] \quad (1)$$

Tab. 1 illustrates the Type 1 of Incomplete Information System. Null value is interpreted as „undefined" value. In the system of this type, an undefined value can take different values, not necessarily the value which already exists in the system.

Type 2

By the incompleteness of this type we understand the situation where all attributes in $S = (X, A, V)$ are functions of the type $a: X \rightarrow 2^{V_a} - \{\emptyset\}$ (Dardzinska, 2013).

If $a(x) = \{a_1, a_2, \dots, a_n\} \subseteq V_a$ then we can say that the value of attribute a is one from $a_1, a_2, \dots, a_n$.

If $a(x) = V_a$, then all values of the attribute a are equally probable and $a(x) = \text{null}$ corresponds to "blank".

An example of incomplete information system of Type 2 is shown in Tab. 2.

Tab. 2. Incomplete information system of Type 2

Patient	Name	Last name	Blood disease
x_1	Ann, Lily, John	Lake	anemia, leukemia
x_2	Emily	Green	
x_3	John, Emily		polycythemia vera
x_4	Lucy, John	Taylor, Kelly	anemia, pancytopenia

$S = (X, A, V)$ is an incomplete information system of Type 2 and is represented by Tab. 2. We assume that: $X = \{x_1, x_2, x_3, x_4\}$, A = {Name, Last name, Blood disease} and $V = V_N \cup V_S \cup V_{Bd}$ where: V_N = {Ann, Lily, John, Emily, Lucy}, V_S = {Lake, Green, Taylor, Kelly}, V_{Bd} = {anemia, leukemia, polycythemia vera, pancytopenia}. Each value from the V_S is just as likely for Last name (x_3), that means Last name (x_3) = Lake OR Last name (x_3) = Green OR Last name (x_3) = Taylor OR Last name (x_3) = Kelly.

Each value from the V_{Bd} is just as likely for Blood disease (x_2), that means Blood disease (x_2) = anemia OR Blood disease (x_2) = leukemia OR Blood disease (x_2) = polycythemia vera OR Blood disease (x_2) = pancytopenia.

Type 3

For the incompleteness of type 3 we assume, that all attributes in $S = (X, A, V)$ are functions of type $a: X \rightarrow 2^{V_a}$ (Dardzinska, 2013). This type differs from the previous one, because we allow having the empty set as the value of some attributes in S. When $a(x) = \emptyset$, then the value of attribute a for the object x does not exist.

An example of incomplete information system of Type 3 is shown in Tab. 3.

Tab. 3. Incomplete information system of Type 3

Patient	Name	Last name	Blood disease
x_1	Ann, Lily, John	Lake	anemia, leukemia
x_2	Kate	Smith, Green	
x_3	John, Joseph	∅	polycythemia vera
x_4	Lucy, Emily	Taylor	∅

$S = (X, A, V)$ is an incomplete information system of Type 3 and is represented by Tab.3. We assume that: $X = \{x_1, x_2, x_3, x_4\}$, A = {Name, Last name, Blood disease} and $V = V_N \cup V_S \cup V_{Bd}$ where: V_N = {Ann, Lily, John, Kate, Joseph, Lucy, Emily}, V_S = {Lake, Smith, Green, Taylor}, V_{Bd} = {anemia, leukemia, polycythemia vera, pancytopenia}.

DOI 10.2478/ama-2018-0025

Last name (x_3) = Ø which is interpreted as „x_3 doesn't have last name". It means that the value is not missing because we know that this value doesn't exist for x_3. The same for blood disease(x_4) = Ø. The object x_4 doesn't have any blood disease.

Type 4

For this type of incompleteness, we assume that all attributes in $S = (X, A, V)$ are functions of the type: $a: X \rightarrow 2^{V_a \times R}$. When we assume that $a(x) = \{(a_1, p_1), (a_2, p_2), \ldots (a_n, p_n)\}$ and p_i is a confidence for a_i, then (Dardzinska, 2013):

$$\sum_{i=1}^{n} p_i = 1 \quad (2)$$

An example of incomplete information system of Type 4 is shown in Tab. 4.

Tab. 4. Incomplete information system of Type 4

Patient	Name	Last name	Blood disease
x1	Ann, Lily, John	Lake	(anemia, 1/3), (leukemia, 2/3)
x2	Emily	Green	Ø
x3	John, Emily		(polycythemia vera, 1)
x4	Lucy, John	Taylor, Kelly	(anemia, 1/2), (pancytopenia, 1/2)

$S = (X, A, V)$ is an incomplete information system of Type 4 and is represented by Tab.4. We assume that: X ={x_1, x_2, x_3, x_4}, A = {Name, Last name, Blood disease} and $V = V_N \cup V_S \cup V_{Bd}$ where: V_N = {Ann, Lily, John, Emily, Lucy}, V_S = {Lake, Green, Taylor, Kelly}, V_{Bd} ={anemia, leukemia, polycythemia vera, pancytopenia}. Blood disease (x_1) = {(anemia, $1/3$), (leukemia, $2/3$)} will be interpreted as „the confidence that x_1 has an anemia is $1/3$ or that he has a leukemia is $2/3$. The object x_2 has not blood disease. The object x_3 has polycythemia vera with the confidence equal to 1. For the object x_4, the anemia and pancytopeniaa are equally likely. The confidence in this case is $1/2$.

2.2. Distributed Information Systems

Let us assume that incomplete information system S is given, and the query q is submitted to this system. The syntax of the query q contains values unknown to S. For example the value of MCH for object x4 from Tab. 1 when the query is q(RBC, HGB, MCH) = (RBC = 3.2)*(HGB = 9.5) * (MCH = ?). Missing values should be replaced by statistical or rule-based methods suggested values, for example, by the rules extracted in Chase algorithm system (Dardzinska and Ras, 2003). Another approach is to create Query Answer System (QAS) (Ras and Dardzinska, 2006; Ras and Joshi, 1997) that uses the knowledge collected from several information systems working under the same ontology (Mizoguchi, 2003). Ontologies (Guarino, 1998; Guarino and Giaretta, 1995; Van Heijst, 1997) are widely used to build a semantical bridge between independent systems that can collaborate and understand each other. This is particularly important for semantical inconsistencies caused by different interpretation of attributes and their values by different systems. For instance, one medical system can interpret the concept illness differently than other one. QAS can be built by Information systems from different locations, independently built and collecting and storing data at a single location. In this case we talk about distributed information systems. The notion of a distributed information system was introduced in (Ras and Joshi, 1997) and next applied in (Ras, 1997; Ras, 2001; Ras, 2002; Dardzinska, 2004).

By an incomplete distributed information system we mean a pair $DIS = (\{S_i\}_{i \epsilon I}, L)$ where (Dardzinska, 2013):

- $S_i = (X_i, A_i, V_i)$ is an information system for any $i \in I$, and $V_i = \cup \{V_{ia} : a \in A_i\}$,
- $\exists i \in I\ S_i$ is incomplete,
- L is a symmetric, binary relations on the set I,
- I is a set of sites.

Two systems S_i, S_j are called neighbors in distributed information system if $(i, j) \in L$.

A distributed information system is object-consistent if the following condition holds (Dardzinska, 2013):

$$(\forall i)(\forall j)\big(\forall x \in X_i \cap X_j\big)\big(\forall a \in A_i \cap A_j\big)$$

$$[\big(a_{[S_i]}(x) \subseteq a_{[S_j]}(x)\big)\ or\ (a_{[S_j]}(x) \subseteq a_{[S_i]}(x))],$$

where a_s denotes that a is an attribute in S.

The inclusion $((a_{[S_i]}(x) \subseteq a_{[S_j]}(x))$ means that the system S_i has more precise information about the attribute a in object x than system S_j.

Object-consistency means that information about objects in one of the systems is either the same or more general than in the other. Saying other words, two consistent systems cannot have conflicting information about any object x which is stored in both of them. System in which the above condition does not hold is called object-inconsistent.

The result of collaboration between the systems is creation of the knowledgebase which collects rules defined as expressions written in predicate calculus and originates from various information systems

In this paper, we will present the method which helps to decide whether the selected information system is the closest one (in a semantical sense) to the system which has to answer the query q. We assume that all information systems work under the same ontology. Our proposal is to use the ERID algorithm (Dardzinska and Ras, 2006) without the minimum confidence to extract rules from each distributed incomplete information system. Confidences and supports of rules are used to construct the measure of the distance between pair of systems. In this paper we propose measure which is the modification of the work from (Dardzinska et al., 2017).

3. SEARCHING THE CLOSEST INFORMATION SYSTEM

Assume, we have a set of collaborating distributed information system (DIS) working under the same ontology. The user asks a query $q(Q)$ for an information system (S, K) from DIS, where $S = (X, A, V)$, K – knowledgebase (empty at the beginning, $K = \emptyset$), Q are the attributes used in $q(Q)$, and $A \cap B \neq \emptyset$ (Dardzisnka et al., 2017). All attributes in $Q \backslash [A \cap Q]$ are called foreign for (S, K). Since (S, K) can collaborate with other information systems in DIS, values of hidden or missed attributes for

DOI 10.2478/ama-2018-0025 acta mechanica et automatica, vol.12 no.2 (2018)

(S, K) can be extracted from their information systems in DIS.

Assume now that we have three, object-consistent and incomplete collaborating information systems with knowledgebase connected with them: $(S, K), (S_1, K_1), (S_2, K_2)$ where $S = (X, A, V)$, $S_1 = (X_1, A_1, V_1)$, $S_2 = (X_2, A_2, V_2)$ and $K = K_1 = K_2 = \emptyset$ (Dardzisnka et al., 2017). If the consensus between (S, K) and (S_1, K_1), based on the knowledge extracted from $S(A \cap A_1)$ and $S_1(A \cap A_1)$, is chosen by (S, K) as closer information system than consensus (S, K) and (S_2, K_2), it becomes more helpful in solving user given query. Rules defining hidden attribute values for S are then extracted at S_1 and stored in K.

Assuming that systems $S_1$and S_2 store the same sets of objects and use the same attributes describing them, system S_1 is more complete than system S_2. So, how to choose the system which is closer to the system that is unable to answer the question alone?

First, the attributes common to the two systems should be indicated. Next, the ERID algorithm creates rules for each system and respectively each attribute is decisive. We choose rules that exist in both systems and calculate for each confidence and support.

On the basis of these measures and distance between two the same rules in different systems, the factor of fitting two systems is calculated:

$$d(S_i, S_j) = \frac{\sum_r d_r(S_i \to S_j)}{\max(\sum sup_r S_i, \sum sup_r S_j)} \quad (3)$$

where:

$$d_r(S_i \to S_j) = \sqrt{(conf_r S_i \cdot sup_r S_i)^2 + (conf_r S_j \cdot sup_r S_j)^2} \quad (4)$$

S_i is closer to S_j than S_k when $d(S_i, S_j)$ is closer to 1 than $d(S_i, S_k)$.

From all the distributed systems we choose the one with maximum value of $d(S_i, S_j)$, which corresponds to the closest information system to the client. The existence of the knowledgebase K will guarantee to the client that the Query Answering System has maximum precision in answering questions asked to the incomplete system.

Example:

Let us assume we have three medical information systems: S_1, S_2 and S_3. They are presented in Tab. 5, Tab. 6 and Tab. 7. The systems are incomplete, object-consistent, created in different locations and they create Query Answering System.

Tab. 5. Information system S_1

X	a	b	c	g
x_1	1	2		3
x_2	2	2		3
x_3	3	3		1
x_4	1	1		2
x_5	2	3		1
x_6	3	2		2

Information system S_1 received a query $q(a, c, g) = a_3 * c_? * g_2$ and has no information about hidden attribute c, which appears in other system such as S_2 and S_3. Our goal is to choose one of them, from which we will be able to predict the values of attribute *c* in system S_1 and to answer query $q(a, c, g)$. Because attributes a, b, g are common for all the systems, first we extract the rules describing them in terms of other attributes. If the system is incomplete, we use ERID algorithm. For each rule we calculate support and confidence in a standard way (Dardzinska, 2004). Next, we pair the systems: S_1 with S_2 and S_1 with S_3. For each pair we select rules the same way for the two systems. Tab. 8 and Tab. 9 present joint rules for paired systems with the confidence and support for each rule and in each system.

Tab. 6. Information system S_2

X	a	b	c	d	e	g
x_7	1	1	3		1	3
x_8	2	1	2		2	2
x_9	1	2	2		1	3
x_{10}	3	2	1		2	2
x_{11}	1	3	3		1	1
x_{12}	3	3	1		2	1

Tab. 7. Information system S_3

X	a	b	c	d	e	g
x_{13}	2	3	1	2		1
x_{14}	1	2	2	2		3
x_{15}	2	1	1	3		3
x_{16}	3	2	2	3		2
x_{17}	1	2	3	1		3
x_{18}	2	3	1	2		1
x_{19}	3	2	3	3		2
x_{20}	1	1	3	1		3

Tab. 8. The common rules for systems S_1 and S_2 with their confidence and support

	S_1		S_2	
	$conf$	sup	$conf$	sup
$b_2 \to a_1$	0.(3)	1	0.5	1
$b_2 \to a_3$	0.(3)	1	0.5	1
$b_1 \to a_1$	1	1	0.5	1
$b_3 \to a_3$	0.5	1	0.5	1
$g_3 \to a_1$	0.5	1	0.5	2
$g_1 \to a_3$	0.5	1	0.5	1
$g_2 \to a_3$	0.5	1	0.5	1
$b_2 * g_3 \to a_1$	0.5	1	0.5	1
$b_2 * g_2 \to a_3$	0.5	1	0.5	1
$a_1 \to g_3$	1	1	0.(6)	2
$a_3 \to g_1$	0.5	1	0.5	1
$a_3 \to g_2$	0.5	1	0.5	1
$b_1 \to g_2$	1	1	0.5	1
$b_2 \to g_3$	0.(6)	2	0.5	1
$b_2 \to g_2$	0.(3)	1	0.5	1
$b_3 \to g_1$	1	2	1	2
$a_1 * b_2 \to g_3$	1	1	1	1
$a_3 * b_2 \to g_2$	1	1	1	1
$a_3 * b_3 \to g_1$	1	1	1	1

DOI 10.2478/ama-2018-0025

Tab. 9. The common rules for systems S_1 and S_3 with their confidence and support

	S_1		S_3	
	$conf$	sup	$conf$	sup
$b_1 \rightarrow a_1$	0.5	1	1	1
$b_1 \rightarrow a_2$	0.5	2	0.(3)	1
$b_2 \rightarrow a_3$	0.5	2	0.(3)	1
$b_3 \rightarrow a_2$	1	2	0.(3)	1
$g_1 \rightarrow a_2$	1	2	0.5	1
$g_2 \rightarrow a_3$	1	2	0.5	1
$g_3 \rightarrow a_1$	0.75	3	0.5	1
$g_3 \rightarrow a_2$	0.25	1	0.5	1
$b_2 * g_2 \rightarrow a_3$	1	2	1	1
$b_2 * g_3 \rightarrow a_1$	1	2	0.5	1
$b_3 * g_1 \rightarrow a_2$	1	2	0.5	1
$a_1 \rightarrow g_3$	1	3	0.5	1
$a_2 \rightarrow g_1$	0.(6)	2	0.5	1
$a_3 \rightarrow g_2$	1	2	0.5	1
$a_2 \rightarrow g_3$	0.(3)	1	0.5	1
$b_2 \rightarrow g_3$	0.5	2	0.(6)	2
$b_2 \rightarrow g_2$	0.5	2	0.(3)	1
$b_3 \rightarrow g_1$	1	2	1	2
$a_1 * b_2 \rightarrow g_3$	1	2	1	1
$a_2 * b_3 \rightarrow g_1$	1	1	1	1
$a_3 * b_2 \rightarrow g_2$	1	2	1	1

Next the factor of fitting two systems: S_1 and S_2 is calculated:

$$d(S_1, S_2) = \frac{21.21}{\max(21,22)} = \frac{21.21}{22} = 0.964.$$

And the same for systems S_1 and S_3:

$$d(S_1, S_3) = \frac{37.197}{\max(40,23)} = \frac{37.197}{40} = 0.929.$$

Since the factor between S_1 and S_2 is closer to 1 than the factor between S_1 and S_3, we choose S_2 as the closer information system to the communication with the incomplete system S_1. From all rules describing attribute c in terms of a, b and g, we choose the rules by which the system S_1 can answer the query $q(a, c, g) = a_3 * c_1 * g_2$. Based on the system S_1, attribute c has a value equal to 1. The rules from knowledgebase K allow us to answer other questions in the system S_1.

4. CONCLUSION

One of the main problems of medical information systems is incompleteness. It has a significant impact on the discovered knowledge from medical databases. To help the decision process in the incomplete system, a method of discovering rules based on knowledge gathered in distributed information systems was proposed.

In this study, we proposed the factor of fitting two systems which can help to find the closest information systems. On the basis of this measure, it was possible to build more precise knowledgebase about patients and answer the query asked for system without valuable information.

Our method has been analyzed based on the medical information systems with missing data and allowed to ascertain which system integration gives better results.

We plan to investigate how our measure will behave in the systems with rules extracted by ERID algorithm with minimum confidence and minimum support.

REFERENCES

1. **Dardzinska A.** (2004), Null Values and Chase in Distributed Information System, *Knowledge-Based Intelligent Information and Engineering Systems,* Springer, 1143–1149.
2. **Dardzinska A.** (2013), *Action Rules Mining*, Springer-Verlag, Berlin, 5–19.
3. **Dardzinska A., Ignatiuk K., Zdrodowska M.** (2017), Query Answering System as a Tool in Incomplete Distributed Information System Optimization Process, *Proceedings of FDSE 2017*, HoChi Minh City, Vietnam, 101–109.
4. **Dardzinska A., Ras Z.** (2003), $CHASE_2$ - Rules Based Chase Algorithm for Information Systems of Type λ, *Active Mining*, Springer-Verlag, 255–267.
5. **Dardzinska A., Ras Z.** (2006), Extracting Rules from Incomplete Decision Systems: *System ERID, Foundations and Novel Approaches in Data Mining,* Springer, 143–153.
6. **Guarino N.** (1998), Formal Ontology in Information Systems, *Proceedings of FOIS'98, Trento, Italy*, 3–15.
7. **Guarino N., Giaretta P.** (1995), Ontologies and knowledge bases, towards a terminological clarification, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing,* 25–32.
8. **Laudon K., Laudon J.** (2012), Management Information System: Managing the Digital Firm, *Prentice Hall*, New Jersey, 14–16.
9. **Mizoguchi R.** (2003), Tutorial on ontological engineering - Part 1: Introduction to Ontological Engineering, *New Generation Computing*, 21 (4), 365–384.
10. **Ras Z.** (1997), Collaboration control in distributed knowledge-based system, *Information Sciences*, 96 (3), 193–205.
11. **Ras Z.** (2001), Query answering based on distributed knowledge mining, *Proceedings of the 2nd Asia-Pacific Conference on Intelligent Agent Technology: Research and Development*, Maebashi City, Japan, 17–27.
12. **Ras Z.** (2002), Reducts-driven query answering for distributed knowledge systems, *International Journal of Intelligent Systems*, 17 (2), 113–124.
13. **Ras Z., Dardzinska A.** (2006), Solving Failing Queries through Cooperation and Collaboration, *World Wide Web Journal,* 9 (2), 173–186.
14. **Ras Z., Dardzinska A.** (2009), Cooperative Multi-hierarchical Query Answering Systems, *Encyclopedia of Complexity and Systems Science*, Springer, New York, 1532–1537.
15. **Ras Z., Joshi S.** (1997), Query approximate answering system for an incomplete DKBS, *Foundamenta Informaticae Journal*, 30 (3), 313–324.
16. **Van Heijst G., Schreiber A., Wielinga B.** (1997), Using explicit ontologies in KBS development, *International Journal of Human and Computer Studies*. 46 (2), 183–292.
17. **Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J., Hua L.** (2012), Data mining in healthcare and biomedicine: A survey of the literature, *Journal of Medical Systems*, 36 (4), 2431–2448.

Research was performed as a part of project no. MB/WM/6/2017 and financed with use of funds for science of MNiSW.