Proceedings of the International Conference of Control, Dynamic Systems, and Robotics
Ottawa, Ontario, Canada, May 15-16 2014
Paper No. 56

Model of Structurization and Transformation of Knowledge in Non-formalizable Domain

Alexey Tsvetkov
Program Systems Institute of the Russian Academy of Sciences
Medical Informatics Research Center (INTERIN)
Program Systems Institute of RAS, Pereslavl-Zalessky, 152020, Yaroslavl Region Russia,
cio@interin.com

Abstract - This paper discusses some theoretical and practical issues of creation of information and expert systems, namely: structuring and transformation of knowledge for non-formalizable domains; dynamic interaction between the subject domains. It is shown that the proposed model structure and transformation of knowledge in non-formalizable domains carries a large innovation potential and allows several orders of magnitude lower cost of ownership and the development of information and expert systems.

***Keywords*:** Non-formalizable domains, structuring and transformation of knowledge.

1. Introduction

Since in this work one of the main concepts is considered "non-formalizability" that initially as an antipode to consider the concept of "formalizability".

As shown by N.N. Nepeivoda (1985), "... under the formalization means the totality of a strict set of rules for certain activities and verifiable".

Basic principles of the theory of non-formalizable concepts considered N.N. Nepeivoda (1996).

For this research we select Principles 1, 7, 8, 9:

- Principle 1. Concepts can be described only in their relationship. A set of interrelated concepts can be described as a signature σ (called humanitarian studies thesaurus);
- Principle 7. At any given moment for the specific purpose of the relationship concepts are described as classical theory Th_{α}. This theory is called non-formalizable hypostasis system concepts;
- Principle 8. Among these theories is the theory Th_0, which is a sub-theory of any Th_{α};
- Principle 9. There is a computable function φ, is building for each pair of theories $Th_{\alpha} \subset Th_{\beta}$, theory $Th_{\varphi(\alpha,\beta)}$, expanding Th_{α}, but incompatible with Th_{β}.

That is, it is assumed that any subject area is a reflection of the real world in which over time there is a change. And, therefore, for each of the subject areas is true that can appear and disappear new elements, as well as change between them. That is, some formalizable subject area over time may become partially formalized, and in the extreme case - non-formalizable. If, based on the formalization of a domain built an information system or an expert system, then going to the domain of formal type to the other two - an information system or an expert system will give inadequate information. Consider this subject area formalized or not - depends on the time of the life cycle of specific information or expert system. Thus, the decision on the type of domain - is always a compromise. For some subject areas during the transition from formal to non-formalized type or partially formalized takes several years, but for others - for years or even months (short lifetime formalization domain).

In this paper, the subject area, which formalizability lifetime is short, will be defined as non-formalized subject area

In connection with the above, the following problems are relevant:

1. Timely diagnosis irrelevance domain description in the existing information or expert system;
2. Rapid modification of existing information or expert system in accordance with changes in the subject area;
3. The timing of the need for full reengineering of existing information or expert system.

In connection with item 1., item 2. and item 3. arise following subtasks that are discussed in this paper.

For item 2. it:

dynamic structuring and transformation of knowledge for non-formalizable domains.

For item 1 and item 3 it:

as a change in each (only one at the same time in some or all at once, as an extreme case) of structures describing several domains, of which at least one is a non-formalizable subject area, affects each of the structures or generated from their structure yet to be described domain.

Additionally, you can mention another aspect related to the fact that there are formalized subject areas. This economic aspect.

Really at the first sign of degeneration formalized domain in non-formalized subject area, which will be shown that information or expert system constructed in accordance with a given subject area, begin issuing inadequate information, owners and developers of information or expert systems need to reflect on the need to modernize or new development of their systems.

Addressing these challenges the traditional approach takes a lot of time and requires specialists with high qualifications and a large financial investment.

2. A Brief Overview of Models of Knowledge Representation

To represent knowledge, there are many models. Their classification in terms of mathematical validity is shown in Table 1.

Table 1. Model of knowledge representation.

Empirical models	**Theoretical models**
Production models	Logic Models
Network models	Formal grammars
Frame-based model	Combinatorial models
Lenems	Algebraic models
Neural networks, Genetic algorithms	

Empirical models based on the approach that examines principles of human memory and problem solving mechanism simulates human. As shown, for example, R. Davis et al. (1993), Ronald J. Brachman et al. (2004), Stuart Russell and Peter Norvig (2003).

Theoretical models based on an approach that involves theoretical basis based on formal logic, formal grammars, combinatorial models, graph theory, etc. As shown, for example, J-L Ermine (2005), Louay Karadsheh (2009).

The most popular are the following empirical model:

- Production models – this modrls is based on the rules in the form of sentences of the form "IF <condition> THEN <action> <postcondition>." Production models have significant drawbacks: the large number of productions are becoming complex consistency check, when a large number of productions (according to various sources from several hundred to a thousand) the proper functioning of the expert system is not guaranteed;
- Network model (or semantic nets) - it is actually a graph in which nodes correspond to concepts and objects, and the arcs - relationships between objects. Lack of models: an unambiguous definition of a semantic network currently does not exist;

- Frame model - based on frames, ie data structures for some conceptual object. The frame structure comprises slots, which may be either terminal, or are themselves frames. The disadvantages of frame models are: the complexity of frame systems reduces performance inference mechanism and changes in family hierarchy, exception handling difficult, there is no possibility to build a chain of statements, there is no mechanism output control;
- Lenems - combines existing paradigms of knowledge representation: logical, structural (network and frame-based models), procedural, and visual features are more advanced than other empirical models. But in cases where knowledge is created on the basis of subject area, which is quite dynamic and interacts with other subject domains, based on the lenems become cumbersome and often unacceptable.

Consider in more detail the most popular approach - production models. This can be explained by the fact that production models are evident, have a simple inference engine, make it easy to make changes and additions.

To the above definition production model may be added that part of the expression "IF <condition> THEN <action>" product called the kernel, and the nuclei themselves can be divided into deterministic and non-deterministic (see Table 2).

Table 2. Production classification.

Class core of production	**Feasibility of actions (factor α={0,…,1})**	**Examples of productions**
Deterministic	Performed with a degree of confidence	“**IF** ' Give a bouquet of flowers is a lovely lady ', **THEN** ' Lady be gracious to you ' **WITH CONFIDENCE LEVEL α** = 0,75 “
Nondeterministic	Holds with probability	'**IF**' fired a gun in the Moon ', **THEN** ' bullet hits in the moon ' **WITH PROBABILITY α** = 0 "

In addition, the productions may be unique or alternative. Unambiguous examples of productions are shown in Table 2. Alternative productions have a form as shown in Expressions (1) и (2).

$$IF \ < \text{condition} >, THEN \begin{cases} < \text{action} >_1 \ with\ confidence\ level\ \propto_1 \\ < \text{action} >_2 \ with\ confidence\ level\ \propto_2 \\ < \text{action} >_n \ with\ confidence\ level\ \propto_n \end{cases} \quad (1)$$

$$\text{IF} < \text{condition} >, THEN \begin{cases} < \text{condition} >_1 \text{ with probability} \propto_1 \\ < \text{condition} >_2 \text{ with probability} \propto_2 \\ < \text{condition} >_n \text{ with probability} \propto_n \end{cases} \quad (2)$$

Conceptually production model is typically implemented as a knowledge base that consists of a set of facts and rules of inference. In addition, the use of the machine output, which controls the enumeration rules. Can be both direct (from data to find a solution), and inverse (on solutions to the data) conclusions. Data - a set of facts which uses the inference engine. On the basis of such a concept can provide some abstract expert system (see Fig. 1).

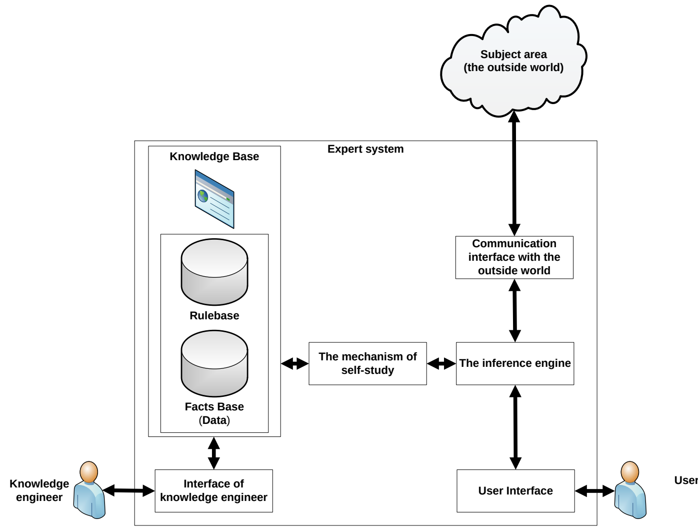

Fig. 1. Abstract expert system.

Furthermore, as seen from Fig. 1 there are a couple of problems mentioned above:

- Work of the expert system is not only difficult to formalize subject areas, but also in non-formalizable domains;
- Job expert system using multiple domains, of which at least one is a non-formalizable.

3. Model of Data Structuring

In (Tsvetkov A.A., 2013) proposed a synthetic model of representing the semantics of the domain, which can adequately describe a given subject area, as well as easy to upgrade model without loss of information from a previous version. In addition, this model allows you to generate a model for related subject areas in which there is general concepts (entities). There is some element of training - available to add new entities, which are characteristic for the other domain.

In accordance with the synthetic model representing the semantics described in (Tsvetkov A.A., 2011), consider a superset that includes descriptions of the structure of all domains

$$S = \{S_1, S_2, \dots, S_n | \, S_i \subset S\} \tag{3}$$

Boolean *P(S)* includes all subsets that describe the structures for all subject areas.

Let a subset $S_i = S_h$ is a description of a subject area, which is denoted by "h", and includes a subsets E_h, R_h, A_h, which are sets of entities, relationships and attributes, respectively. Then a subset S_h can be expressed as the following expression

$$S_h = \{E_h, R_h, A_h | \, E_h, R_h, A_h \subset S_h\} \tag{4}$$

It is assumed that all the sets belonging to S_h, can be performed algebraic operations that are valid in the theory of sets, and sets itself in the form

$$E_h = \{e_{h1}, e_{h2}, \dots, e_{hl} | \, e_{hj} \subset E_h\} \tag{5}$$
$$R_h = \{r_{h1}, r_{h2}, \dots, r_{hm} | \, r_{hs} \subset R_h\} \tag{6}$$
$$A_h = \{a_{h1}, a_{h2}, \dots, a_{hn} | \, a_{ht} \subset A_h\} \tag{7}$$
$$S_h = (E_h \times A_h) \times R_h \tag{8}$$

each member of which can be represented as valid combinations of threes e_{hj}, r_{hs}, a_{ht}, and the set is not empty S_h.

In the process of forming the set $\mathbb{S}_{hi+1}$ is compared with the knowledge that entities, attributes, and relationships with those members of the set $\mathbb{S}_{hi}$. If detected the presence of common members of the sets $\mathbb{S}_{hi}$ и $\mathbb{S}_{hi+1}$, then we can say that there is their intersection $\mathbb{S}_{hi} \cap \mathbb{S}_{hi+1}$(in case of full equality sets is not considered), which in this case can be called the "core" formed set $\mathbb{S}_{hi+1}$. That is, the task of further filling of the set $\mathbb{S}_{hi+1}$ is reduced to the formation of the difference between "core" and the set $\mathbb{S}_{hi+1}$.

$$\mathbb{C}_{hi+1} \setminus \mathbb{S}_{hi+1} = \{\mathbb{E}_{hi+1}, \mathbb{R}_{hi+1}, \mathbb{A}_{hi+1} \mid \mathbb{E}_{hi+1}, \mathbb{R}_{hi+1}, \mathbb{A}_{hi+1} \in \mathbb{C}_{hi+1} \mid \mathbb{E}_{hi+1}, \mathbb{R}_{hi+1}, \mathbb{A}_{hi+1} \notin \mathbb{S}_{hi+1}\} \quad (9)$$

4. Using the Model of Structuring Data

A graphical representation of the described process (using the example of the description of the subject area $\mathbb{S}_{hi+1}$ from the description of the subject area $\mathbb{S}_{hi}$) is shown in Fig. 2.

Suppose that there is some pro deemed formalized at this time interval, and described nonempty $\mathbb{S}_{hi}$, including a subset of attributes, entities, relationships, as well as filled with content (knowledge). And suppose you want to generate a description about each, which will be described in a variety of $\mathbb{S}_{hi+1}$, and for which the knowledge engineer and expert in the data domain formed reflection, i.e. a formal description of the data domain.

For this process can be intuitive an example of the following. Let us assume that the general practitioner wants to deepen their knowledge ($\mathbb{S}_{hi}$), for example, in the field of endocrinology ($\mathbb{S}_{hi+1}$). What is he doing? Takes books on endocrinology and begins to get knowledge in this area. But some of the knowledge they already have, i.e. there is a "core" of the new knowledge ($\mathbb{C}_{hi+1}$). If the doctor sees in the text that already knows what he will miss this stuff, if the text contains something new for him ($\mathbb{C}_{hi+1} \setminus \mathbb{S}_{hi+1}$), this material necessary to study. I.e. in the memory of a doctor formed a certain image domain "Endocrinology" ($\mathbb{S}_{hi+1}$).

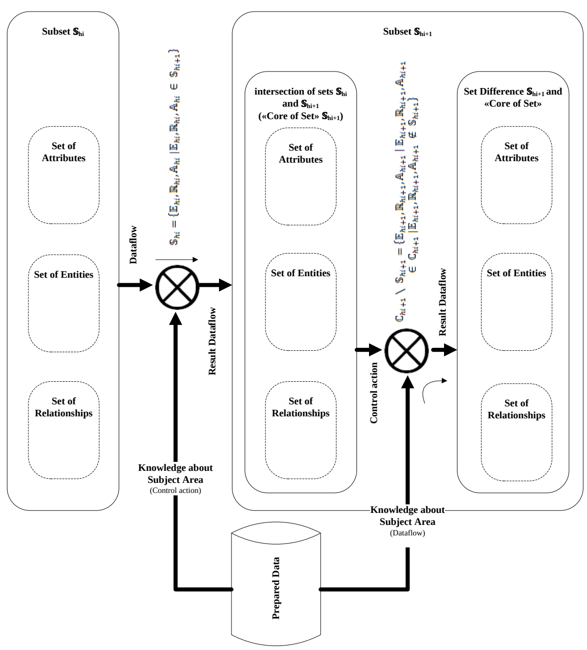

Fig. 2. The model of generation for describing of the subject area $\mathbb{S}_{hi+1}$ from the description of the subject area $\mathbb{S}_{hi}$.

On Fig. 3 notation BPMN 2.0 shows the sequence of processes in the implementation of the model of structuring data.

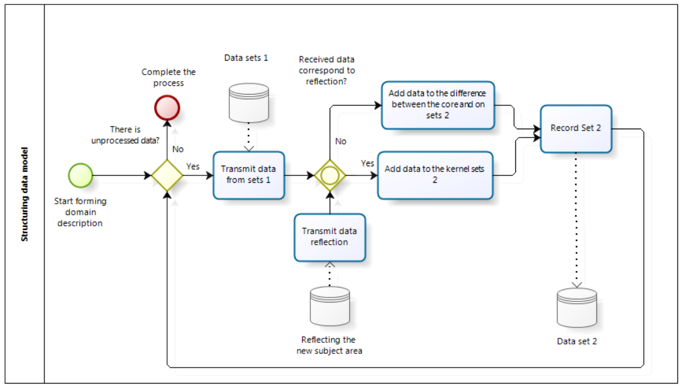

Fig. 3. Sequence of processes for structuring data model.

Consider a case where there are two non-empty sets $\mathbb{S}_{hi}$ and $\mathbb{S}_{hi+1}$, describing some of the subject areas. Let these sets have empty intersection $\mathbb{S}_{hi} \cap \mathbb{S}_{hi+1}$, which corresponds to the model described in the section « Model of Data Structuring » the present work. But in contrast to the example given in the preceding section, each of the sets is reacted with the actual subject area and to each other (as shown in Fig. 4).

Suppose that at some point in time, the system of two sets describes the formal subject areas. If at any point in time one of the real subject areas will be changes, then begin a process similar to that described in this section of the present work. That is, intersection $\mathbb{S}_{hi} \cap \mathbb{S}_{hi+1}$ will be constantly changing as the part $\mathbb{S}_{hi}$, and from the $\mathbb{S}_{hi+1}$.

This process can be called "dynamic formalization of non-formalizable subject areas", because with any change in the real subject area is a change in the structure the domain model, which, in turn, extends to other subject areas related through the intersection with the model given subject area.

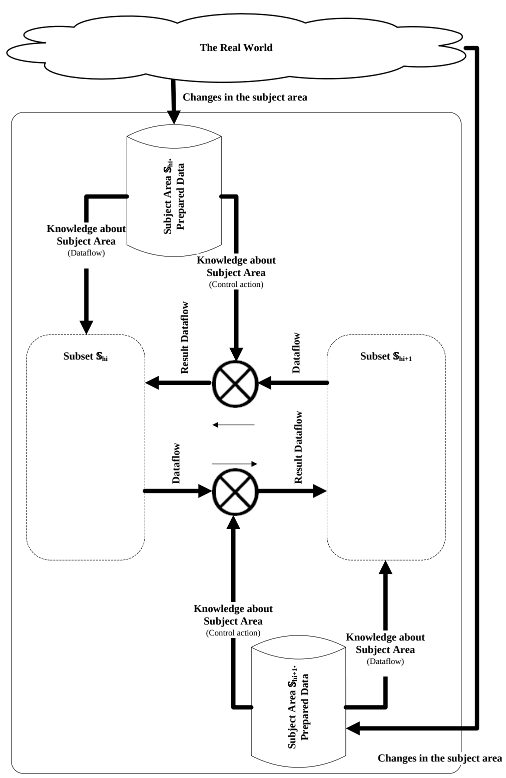

Fig. 4. A dynamic model of interaction of 2 domains

6. Conclusion

In this paper we proposed definition non-formalizable domain based on the lifetime of a given domain in the "formalizable" and in accordance with this definition shows that it is possible to structure and transform data and knowledge in the limited time points for use within the boundaries of some of the information and the expert system.

But, on the other hand, based on a dynamic model of the interaction of two or more subject areas may increase time formalizability each of these domains, which allows:

- The development of information and expert systems use approaches that are used to formalize domains;
- Extend the life of existing information and expert systems between upgrades or complete development of new systems that will reduce the cost of ownership of information assets, and hence increase their value.

Acknowledgements

I express my gratitude for the valuable advice, comments and encouragement in the preparation of articles for Professor N.N. Nepeivoda. Thank the Director of the Research Center for Medical Informatics PhD Yadulla Guliev for financial assistance in organizing a trip to the conference CDSR-14. Thanks also to the organizers of the conference CDSR-14.

References

Davis, R., Shrobe, H. and Szolovits P. (1993) What is a Knowledge Representation? *AI Magazine*, 14(1):17-33, 1993.

Ermine J-L (2005) A Theoretical and formal model for Knowledge Management Systems. Proceedings of ICICKM’2005 (2nd International Conference on Intellectual Capital and Knowledge Management), Dubai, 21-22 november 2005, Dan Remenyi ed., ISBN 1-905305-14-1, pp 187-199

Louay Karadsheh, Ebrahim Mansour, Samer Alhawari, Ghassan Azar, Naser El-Bathy (2009) A Theoretical Framework for Knowledge Management Process: Towards Improving Knowledge Performance. Communications of the IBIMA Volume 7, 2009 ISSN:1943-7765

Nepejvoda N.N. (1985) About Formalization of Non-formalizable Concepts: Auto-production systems of theories. – Semiotics and Informatics. VOL. 25, pages. 46-93. The periodical issue of All-union Institute of Scientific and Technical Information (VINITI), Moscow, 1985.

Nepejvoda N.N. (1996) Applied Logic. Tutorial. – Novosibirsk Centre on Mathematical Sciences, 1996.

Peter Jackson (1998) Introduction to Expert Systems, Third Edition, Addison-Wesley, 1998.

Ronald J. Brachman, Hector J. Levesque (2004) *Knowledge Representation and Reasoning*, Morgan Kaufmann, 2004 ISBN 978-1-55860-932-7.

Stuart Russell and Peter Norvig (2003) Artificial Intelligence: A Modern Approach, Second Edition, Prentice-Hall, 2003.

Tsvetkov A.A. (2011) Synthetic Model of Representing the Semantics of Different Subject Areas. – Proceedings of Conference “Technologies Of Informatisation In Professional Activity". VOL.1, IZHEVSK, NOVEMBER 8-12, 2011.

Tsvetkov A.A. (2013) Model of the structuring of knowledge on the example subject area "Health and Medicine". – 6-th National Multiconference on Control Problems (September 30 – October 5, 2013). VOL. 1, Rostov-on-Don University, 2013.