KNOWLEDGE REPRESENTATION IN THE AUTOMATED LEARNING SYSTEMS

Vitalii Lytvynov, Iryna Posadska

***Abstract**: The role of knowledge representation in IIS is considered in current article. Features of interaction of different categories of users in the learning process are presented. Methods of knowledge formalization have been analyzed. Features and advantages of usage of tools based on UML for knowledge representation are identified.*

***Keywords**: knowledge representation, intelligent learning system, subject domain, reference model, UML, formalized form.*

***ACM Classification Keywords**: I.2.4 Knowledge Representation Formalisms and Methods*

Introduction

To archive the high level in modern cultural, social, economic, scientific and technological development Ukraine requires comprehensively developed, professionally trained specialists. That is the reason why the problem of development and practical implementation of the most effective information technologies of studying, which are corresponding to the general concept of education, is becoming increasingly relevant. The essence of this concept is to prepare the individual, who has not only basic knowledge, sufficient for the requirements of the current level of production, science, culture and the state, but who also has the ability to actively creative professional and social activities. An important role is given to the new software and technological tools of individualization of learning process [Kazymyr, 2013].

Each of the directions of the development of information technologies for learning is based on the methodological and theoretical principles and has inherent advantages and disadvantages in the solution of different problems in the organization of the knowledge acquiring process.

From the beginning the development and usage of learning computer technologies in terms of methodology has been developed in two directions.

The first direction is based on the idea of programmed learning [Джордж, 1984].

Automated learning systems (ALS) for various academic disciplines are developed and operated as the key element of the current direction. The core of the ALS is the author's systems that allow teacher (developer) to enter his course materials in the database and to program algorithms of its study by using of special author's languages or other means.

The second direction of implementation of computer technology in learning is closely associated with the processes of informatization of various branches of human activity.

Learning computer programs, software packages, elements of automated systems designed to automate the time-consuming calculations, optimization, investigation of the properties of objects and processes based on mathematical models etc. has been developed in this direction.

Features of the functioning of IIS

Since the beginning of the 1980s new trend in the computerization of education called intelligent learning system (IIS) was rapidly developed. IIS is based on works in the area of artificial intelligence [Холдинг, 1991].

Among the variety of learning software only one class of such systems – IIS - provides the flexibility of learning. Knowledge models are used as the basic element of IIS. Knowledge models can be divided to the following varieties:

- knowledge about the subject domain of the course, considered as a reference model;
- knowledge about the subject domain of the course, formed in the learner representation;
- knowledge about the learning process.

The characteristic feature of IIS is the presence of specific user category - knowledge engineer. Knowledge engineer is a specialist, which main task is formalization of knowledge, design of knowledge base and filling it with knowledge about the subject domain of the course.

In the IIS it is necessary to represent the knowledge not only in text and graphic form, but also in a formalized form - that is the responsibility of knowledge engineer. Together with the standard categories of users - the learner, tutor, expert - knowledge engineer forms a complete set of IIS users.

The interaction of different types of users in the learning process is presented below (Fig. 1.).

The main function of the expert in the subject domain is to prepare the course about the subject domain i.e.:

- selection of materials necessary to study;
- representation of knowledge in text form;
- development of accompanying illustrative materials;
- design of tables, graphs, and other supporting materials.

Structured text and graphic materials of course (table of contents, alphabetical terminological pointers, etc.) are stored in a database. The knowledge engineer extracts them from database and formalizes the knowledge about the subject domain, namely he prepares reference model of subject domain. Actually, the knowledge engineer performs following functions:

- designs the model of knowledge base fragment based on structured text and graphic materials of the course;
- determines matches of text and graphic material with a formalized kind of knowledge;
- generates a set of formalized questions for the course section based on the subject domain of the section;
- creates a reference model of subject domain;
- verification of the subject domain.

After this tutor determines the strategy of learning and sequence of work with partitions, based on the text and graphic information obtained from the database. Also he provides the learner a portion of text and graphic material of learning courses and conducts follow-up testing after studying of learning material.

It is assumed that the learning course is divided into a number of fragments - portions.

The main task of the learner is to read the portion of text and graphic material of the course, which was given by tutor, and answer a set of questions on it. Portion of the course, which were presented to learner to study, can be restructured using correct answers on the testing.

After reading the portions learner gets a set of testing questions, which he can answer selecting one of the proposed options. That is the way to form subject domain of knowledge about the course, formed in representations of learner.

Next, the reference subject domain model and subject domain model formed in learner representations is compared, and the mismatch of Δ of these subject domains is determined. On the basis of Δ a set of links to re-study of the text of learning material is constructed, and a set of questions to be asked a learner after re-study of the course is formed. These steps are repeated while condition $\Delta \leq \rho$ on the fragment is reached, where ρ- threshold value of number of unstudied elements of subject domain.

To the tasks of knowledge engineer is added next: generation of questions to test knowledge about unstudied material of fragment, generation of links to the re-learning of the text fragment.

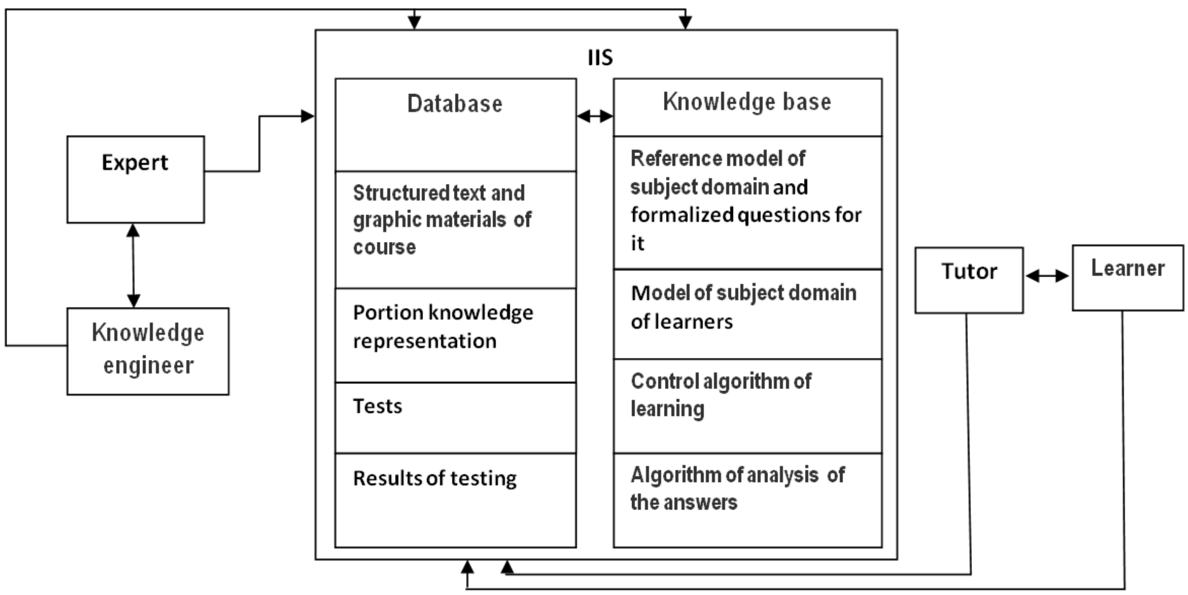

Fig. 1. The interaction of different types of users in the learning process

Considering the interaction of all categories of users of IIS, it becomes clear that a formalized subject domain model of the course is stored in the knowledge base, and if it is necessary, is used to create flexible learning scenarios, depending on the level of learner knowledge in current section. Also there is formalized correspondence between text and graphics representation and formalized representation of the course, the reference domain model, and formalized learner models, created based on the test results about fragment of subject domain of the course in the knowledge base. Also there are control algorithm of learning, algorithm of analysis of the answers, which contains the formalized questions about a fragment (portion) of subject domain, generated tests of the unstudied material of fragment and generated link to the re-learning of the text fragment.

The algorithms, presented below, are not considered in this article. At the initial stage of development of IIS the main task is to develop a domain model, so the main question is to select form of representation of formalized knowledge.

Formalized domain model gives own structure of course, independent of the subjectivity of domain experts. On the basis of these elements there is a possibility to construct the system of tests and to re-create a formalized subject domain model for learner, which can be different from the reference subject domain.

Standard methods of knowledge representation

As it can be seen from Fig.1, almost all blocks of learning system work with knowledge. Therefore it is necessary to analyze the object of research more detailed.

There is a wide range of methods of knowledge representation [Уэно, 1989]:

- logical methods;
- semantic networks;
- frames;
- production systems.

Logical methods. The main idea of the approach of the construction of logical models of knowledge representation is that all the information required for solution of applied tasks, is considered as a collection of facts and statements, which are presented as formulas of some logic. Knowledge is represented as a collection of such formulas, and obtaining of new knowledge is reduced to the implementation of procedures of logical inference. The concept of a formal theory is the basis of logical models of knowledge representation and is determined by cortege,: $S = \langle B, F, A, R \rangle$, where B - a countable set of basic symbols (alphabet), F - set, which is called the formula, A - selected subset of a priori true formulas (axioms), R - a finite set of relations between formulas, called rules of inference.

Semantic networks. Model of knowledge representation using semantic networks consists of vertices, called nodes, corresponding objects, concepts or events, and arcs that link them and describe relationships between objects that are considered [Скрэгг, 1983]. Arcs can be determined by different methods. Usually arcs type «IS-A» (relationship «is») and «HAS-PART» (relationship «has part of») are used to represent the hierarchy. They also create an inheritance hierarchy in the network. It means that the elements of the lower layer in network can inherit the properties of elements of higher level. It saves memory, because information about inheritable properties does not need to be repeated in each network node.

Frames. This model of knowledge representation using frames where the frame is a data structure representing a stereotypical situation like being inside some sort of a living room or gathering for a party. Several kinds of information attached to each frame. Part of this information describes the way of frame usage. Other part contains information about what we can expect hereinafter. Other part is dedicated to the information what should be done if these expectations are not confirmed [Минский, 1979]. The organization of frame model is in many ways similar to the semantic network. It is a network of nodes and relationships that are organized hierarchically: the upper nodes represent general concepts, and their subordinate units are the special cases of these concepts. In the system based on frames, the concept of each node is determined by a set of attributes-slots and values of these attributes. Each slot can be associated with specific procedures that are performed when the information in the slots (attribute values) changes. Each slot can be associated with any number of procedures.

Production systems. In such model knowledge is represented as a set of rules such as «If - then». Knowledge processing systems, which use such representation, are called production systems. The structure of a production system includes the rule (knowledge) base, working memory and the interpreter of the rules (solver), which implements a specific inference engine. The facts are presented in the form of pairs: attribute - value, and in cases, where there are several objects with identical attributes - in the form of triples: object - attribute - value. Operational knowledge is represented like the rules of type: condition – action or condition - goal. The rule is interpreted as: to achieve the rule goal it is necessary to achieve the goals of the condition. To solve the problem the control program selects a goal and tries to achieve it.

The main approaches to choose of formalized knowledge representation

However, we tend to follow the views of Y. Klykov about knowledge representation. According to the views of Y. Klykov knowledge is a set of data, facts and inference rules about the world which include the information about the properties of objects, regularity of the processes, phenomena and rules that are used by this information to make decisions [Клыков,1980].

Since information about world is concerned, knowledge is a set of models describing certain phenomena and features of the world, its structure or behavior. In this sense, systems theory, theory of similarity, models theory is the ideal foundation for a formal knowledge representation.

In general form, as well as system, models of phenomena, that form the knowledge, can be considered as $\langle X, R \rangle$, where X - set of concepts, $R \subseteq X \times X$ - set of relationships between concepts. However this form can be represented as a description of object models that are specific representatives of the world, and a description of models of classes of specific objects - $S = \{s_i\}, i \in N_+$.

Typically, in the theory of artificial intelligence knowledge is determined as description of the model classes, at the same time the specific representatives of classes are the facts.

This distribution of roles for models allows formulating a number of tasks of knowledge:

- tasks of operations on sets for knowledge;

- tasks of determination of fact belonging to the selected class (type of knowledge)
- tasks of construction of knowledge model on a given set of facts;
- tasks of generation of facts for a given class (knowledge models);
- tasks of optimal choice of fact for a given class (knowledge models);
- tasks of identifying of incompletely determined fact to fact of a given class;
- tasks of identifying of relationships between knowledge, etc.

Knowledge representation as a set of concepts and relations between them is well combined with structural models, which tend to be static generally, and in which the time factor does not play a significant role. There are also more specialized means of description of static structural systems, for example, formal grammar, logical calculus, neural networks, etc.

To describe such models many specialized formalisms to set behavior dynamics of systems was developed. There are a system of differential equations, finite automata [Глушков, 1961], Markov and semi-Markov processes [Королюк, 1982], aggregative model [Бусленко, 1973], Petri nets [Питерсон, 1984], logic-dynamic models [Жук, 1975], etc. All these means of dynamics description in explicit or implicit form deal with set of system states and transitions between states. They are used to describe both the facts and the knowledge of a given class. Models, reflecting the dynamics of behavior, are often called dynamic models.

Dynamic models are characterized by the fact that tasks to be solved for the structural model are included in the list of classes of solving tasks, but they are complemented by the prediction tasks or reachability of states, tasks of admissible trajectory of behavior, tasks of determining the moments of the crash of the trajectory, etc.

Conducting the analysis of learning system **as a system of storing, actualizing, performing the transformation of knowledge and providing them to users,** it is necessary to specify the main features of the above functions.

1. Forms of knowledge representation, which are used for storage:

- Natural-linguistic textual-graphical form of representation is a form, which is used in almost every system, including modern learning system. This form took an important part in the interaction the learner with learning system, playing the role of a friendly user interface. However, it has serious disadvantages in the organization of the structuring of knowledge, assessing of the level of received knowledge and adaptation of learning scenarios for a particular user.
- Formalized form is a form, where the elements of storing are programs and structures of knowledge bases, where features of formalisms are reflected and used to express the structure and dynamics of the models that form the knowledge and facts. Since these formalisms are very strongly linked to the subject domain of learning, it is impossible to claim that there is a better formalism for knowledge representation. That would be equivalent to the assertion about the existence of the best mathematical theory in the simulation of systems.

Selection of formalized form of knowledge representation depends on many factors: the type of the considered systems (discrete, uninterrupted, discrete-uninterrupted etc.), tasks solved in the courses of learning system, level of difficulty of obtaining secondary knowledge from primary knowledge, for example, obtaining values of the characteristics of the studied system, complexity of maintaining of knowledge up to date, and others.

2. Methods of maintaining knowledge up to date

It is known that the external world is changing. Quality learning should be carried out on the up to date models. That's why there is the undeniable importance of this function in the learning system. Its complexity depends heavily on knowledge representation forms.

Most simply, it is implemented with natural-linguistic textual-graphical form by editing old text and entering of new one.

Formalized storage form of knowledge has two ways to update:

- knowledge engineer works directly with formalized representation by editing old formal representations and adding new ones;
- training of system.

The first method is suitable in almost all forms of formalized knowledge representation, the second one is suitable only in case of knowledge representation in the form of a neural network.

Regarding the second method, there were identified methods of updating the models with teacher and without teacher, and special training methods, which are specific to different types of neural networks.

3. Methods of knowledge transformation

Actually, solution of tasks using models is a process of transformation of primary knowledge and facts into secondary (derivatives) knowledge. An example is the calculation of the characteristics of quality the system model. Transformation in the learning system can be maintained by creation of individual calculation procedures of derived indicators, or by using of standardized procedures for these purposes. The last one is usually associated with the method of knowledge representation, more precisely, to standard procedures that are typical for one or another formalism.

The standard procedures used for obtaining the secondary knowledge for some formalism are considered below.

Formal grammar is grammatical parsing procedures and procedures for generating of sentences of a language using grammar.

Logical calculus is procedures of logical inference of theorems, for example, the method of resolutions or algorithms of formulary transformation.

Petri nets are algorithms of testing reachability of the network from the initial.

Neural networks are calculation procedures of output values of network for given input.

Finite state machines are procedures of testing reachability of a given state from the initial.

ER models are model validation procedures or procedures for assessing of quality parameters of the model.

Returning to the interpretation of the concept of knowledge, there is still no answer to the question: what is the best representation of knowledge. Especially considering the fact that we are often faced with a situation where during teaching of discipline it is sometimes necessary to use models based on different formalisms.

Selection of the most appropriate knowledge representation transforms to choosing the most appropriate tool.

And there is the most obvious decision to perform the choice of such tool. These tools should be based on a set of formalisms. The most well-known tools of similar class are the tools of Rational Rose [IBM official website, 2014], which used Unified Modelling Language (UML) as input language.

Well known [Douglass, 1999], UML is a multi-language system with the possibility of assignment of structural models (class diagrams), scenarios of interaction of models elements (sequence diagrams and interaction of objects diagrams), setting of dynamics of the system (state diagram), the interaction of parallel processes (activity diagrams), etc.

Such a wide range of languages within a single tool allows effective usage of such tools with knowledge systems, particularly in learning systems.

Using system tools based on UML it is possible to perform the transformation from one type of formalized knowledge representation in the other and return. An example is a transformation from representation of procedural knowledge represented by object oriented algorithmic language to the structural representation in the form of class diagrams. This transformation is often called the reverse-engineering [Samuelson, 2002].

4. Knowledge representation for learners should be conducted within natural language constructions. This assumes that in the learning scenarios feedback references from natural language elements to formalized elements and backward should be represented. Refinement of scenario should be conducted on the basis of learner’s test results. The reason for this can be the automatic creation of formalized domain model of learner and its comparison with reference model.

Conclusions

Knowledge is used almost on every phase of functioning of learning systems and is stored in systems knowledge bases. The importance of the questions about the knowledge representation and knowledge determination increases.

It can be concluded from the article that knowledge can be presented in different forms: from text and graphics representation to formalized representations. Since knowledge can be considered as complex of models that reflect the world around us, the formalisms, which are used to represent them, are quite varied.

And the problem of choosing the correct formalism for learning systems, oriented on a variety of learning courses, includes not only the choice of the optimal formalism, but also the choice of tools, allowing the usage of a variety of formalisms.

In our opinion, such tools of representation of subject domain are the systems, which use the UML.

Bibliography

[Kazymyr, 2013] V.V. Kazymyr, M.V. Tevkun, I.S. Posadska, O.P. Drozd. Methods of collection data to evaluate the quality of distance learning system. Вісник Чернігівського державного технологічного університету: зб. наук. праць. – Чернігів: ЧНТУ, – 2013.– № 69. – С. 144.

[Джордж, 1984] Джордж Ф. Основы кибернетики: Пер. с англ. / Под ред. А.Л. Горелика. – М.: Радио и связь, 1984. – 272 с.

[Холдинг, 1991] Человеческий фактор: В 6 т. / Пер. с англ. / Д. Холдинг, Н. Голдстейн, Р. Эбертс и др. – Т. 3: Моделирование деятельности, профессиональное обучение и отбор операторов. – Ч. 2: Профессиональное обучение и отбор операторов). – М.: Мир, 1991. – 302 с.

[Уэно, 1989] Уэно Х., Исудзука М. Представление и использование знаний. – М.: Мир, 1989. – 220 с.

[Скрэгг, 1983] Скрэгг Г. Семантические сети как модели памяти // Новое в зарубежной лингвистике. – М.: Радуга, 1983. Вып. 12. – С. 228 – 271.

[Минский, 1979] Минский М. Фреймы для представления знаний. – М.: Мир, 1979. – 152 с.

[Клыков, 1980] Клыков Ю.И. Банки данных для принятия решений / Ю.И. Клыков, Л.Н. Горьков. – М.: Сов. радио, 1980. – 208 с.

[Глушков, 1961] Глушков В. М. Абстрактная теория автоматов. «Успехи математических наук», 1961, т. 16, в. 5.

[Королюк, 1982] Королюк В.С., Турбин А.Ф. Процессы марковского восстановления в задачах надежности систем. - Киев : Наук, думка, 1982.— 236 с.

[Бусленко, 1973] Бусленко Н. П., Калашников В. В., Коваленко И. Н. Лекции по теории сложных систем. М.: Сов.радио, 1973.

[Питерсон, 1984] Питерсон Дж. Теория сетей Петри и моделирование систем. - М.: Мир, 1984. - 264 с.

[Жук, 1975] Жук К. Д. Исследование структур и моделирование логико-динамических систем: моногр. / К. Д. Жук, А. А. Тимченко, Т. И. Доленко. — Киев : Наукова думка. — 1975. — 199 с.

[IBM official website, 2014] Rational Rose. [Online]. – Available: http://www-03.ibm.com/software/products/en/ratirosefami

[Douglass, 1999] Bruce Powel Douglass. Real – Time UML Second Edition. Developing Efficient Objects for Embedded Systems. Addison-Wesley, 1999, 238 p.

[Samuelson, 2002] Samuelson P., Scotchmer S. The Law and Economics of Reverse Engineering // The Yale Law Journal, 2002, Vol. 111, No. 7, pp. 1575-1663.

Authors' Information

Vitalii Lytvynov – Dr. Sc., Prof. Chernihiv, National University of Technology, 95, Shevchenko street, Chernihiv-27, Ukraine, 14027; vlitvin@ukrsoft.ua

Major Fields of Scientific Research: modeling of complicated systems, computer-aided management systems, decision support systems

Iryna Posadska – Ph.D. Student, Chernihiv National University of Technology, 95, Shevchenko street, Chernihiv-27, Ukraine, 14027; irrkin@gmail.com

Major Fields of Scientific Research: automated learning systems, information technologies and methods of artificial intelligence