Knowledge Bases Construction of Systems for Solving Intellectual Problems

Victor Martynov, Igor Boyko, Alexander Guminsky
Academy of Science of Belarus

There are some Knowledge Representation (KR) models that have become classical: the logical, production, frame models, the semantic network. The logical model is used to represent knowledge in calculus of first-order predicates and drawing conclusions by means of syllogism construction. In the production models the knowledge is represented by the totality of rules like "if... then..." (the Phenomenon-Reaction). A frame is a structure of data (image) for representing a stereotyped situation. The information, belonging to the frame is contained in a slot (constituent of a frame) [1].

The concept of the semantic network is not completely defined in the current works. However, it may be noted that the KR by the semantic networks is no more than representation of the data structure in the form of nodes and arcs (relations).

Within this system the nodes of the network represent facts in the form of complete sentences or either complete sentences or separate concepts determine separate concepts and the relations between these. For instance: a fact in the frame -- this is a relation with several objects linked by this relation in an intelligent way [2]. In addition to this the fact itself can be represented in the form of a semantic network.

As far as the productions are concerned, here the representation of concrete data with the aid of a triplet is used: object-attribute-meaning, which can be represented in the form of a frame that in its turn can be represented in the form of production and vice-versa. In fact, all the listed KR models can be reduced to semantic networks. In other words, a semantic network cannot have a one-meaning definition because it acts through logical, production and frame models of KR.

The fact that all classical KR models can be reduced to the semantic networks is caused by their basic elementary units like "object-relation-object", where the object is

understood both as a subject and relation. The totally of these structures is organized into bigger blocks.

In the logical and frame models this structure represented in the explicit way. In the production models it is present in the form of **"If... then..."**. It should be noted that actually all knowledge bases built on the classical models have the relation like structure.

The likeness of the elementary units is evidence that classical KR models are actually one-type models from the point of view of their ideology, which means that these models can be reduced to one basic model. It also proves that knowledge presented by means of one model can be used in such a way as if it were represented by means of another model. For example, a frame model with sets of rules can be used as a production one for the strategy of conclusion. The semantic network and predicate calculus (the logical model), as Scragg showed in his work [3] are equivalent both in power and in means of KR.

Based on classical KR models, applied systems of problem solution have been developed. Auxiliary means of the domain description are also used along with the latter. These systems proved efficient but attempts to enlarge the domain to enforce the systems abilities bring about the necessity to adopt the system to the new knowledge that makes it inflexible for development. Combinations of KR have become increasingly popular when applied systems are developed. But combinations of ideologically similar models cannot bring to fundamental solution of the problem.

Therefore, a new approach to KR needed. Let us agree that the structural unit "object-relation-object" has been exhausted and there must be a more profound structuring of the basic unit.

Suggestion of the atom concept produced a major breakthrough in the molecular physics and gave vise to the nuclear physics. Similarly, having compared the unit "object-relation-object" to a molecule, we suggest that "relation" should be something like an atom. It brings about the necessity of its structuring.

For this purpose, we will use the grammar of the Universal Semantic Code (USC) transformation formalized as axiomatic basis of the USC algebra. USC is a language of interpretation and calculus of meanings. It has the following peculiarities. It is a language of complete explication of meaning, i.e. each combinatorial type of a string of elements has one and only one meaning.

It is a language of universal canonization, i.e. restrictions imposed on its system do not depend on the fragment of the universe which the language signifies. Universal canonization is based on the so called absolute language universals which are defined as regularities necessarily characteristic of any system of elements capable of performing the role of the language.

It is a language of a non-conventional representation of semantics, i.e. the strings of the USC are not ascribed any definite meaning, the latter being not preset for them. The semantic interpretation of all strings generated by the USC is deduced from universal axioms. It is a system capable of interpreting the universe, i.e. of forming new concepts and making up hypotheses of causes and consequences situations.

It is a language of representation and transformation of data. The representation of data is performed by the data transformation. This is a unique feature of the USC that enables us to get some values from other values. All the transformations on the USC are syntactical (position-caused).

If **a, b, c** are positions of the variables **X, Y, Z** accordingly, then the transformation **(X*Y)*Z → X*(Y*Z)** (where * is a binary operation of matching on the set of variables) does not mean an exchange of positions, because thus their order is preserved -- **(a*b)*c → a*(b*c)** although the succession of operational steps is altered.

At the same time the transformation **(X*Y)*Z → (Y*Y)*Z** means an exchange of positions -- **(a*b)*c → (b*b)*c** while the succession of the operations is preserved. The syntactic character of the string variables allows to formalize them by means of a certain algebra.

The computer language PROLOG is based on formalized logics of first-order predicates, similarly the USC is based on the axioms of certain algebra. These are examples of formal reading:

(X*Y)*Z -- 'X by means of Y affects Z' or X*(Y*Z) -- 'X holds Y in Z'.

The brackets put in the strings determine the correctness of their reading.

The Knowledge Base (KB) in our system of intellectual problems is based on the axioms of the USC algebra and is formed as an oriented graph (presentation in the form of a matrix or a list is also possible). The nodes of the graph are represented in the form of USC strings. The arcs are in the form of USC axioms or theorems of the given algebra. It is obvious that the solution of the intellectual problem can be realized in the form of a route set by the succession of arcs.

The algorithm of the problem solution is based on the successive drawing of a route from the target situation to the initial one. If it is not feasible, the procedure of this kind can be repeated as many times as possible. But when the following procedure proves not feasible, then it is to be admitted that the means forming the system are insufficient to solve the problem.

Apart from the axioms the KB of our system contains the semantic vocabulary of most commonly used verbs of scientific and technology literature. Each of the verbs is either defined in the USC or has a reference to the synonym having a similar definition. The user forms the KB in respect to the given domain. His utterances are limited by only one verb that is picked up by the user in the form of an infinitive and input into the computer. There it is referred to the verb vocabulary and by means of definition of a USC string a necessary set of positions is given. The user fills the positions in with relevant names and again input into the computer. If the verb is absent in the vocabulary, the system suggests to the user to input a synonym or to simplify the whole utterance.

Here is an example of a questionnaire and the relevant vocabulary item. The vocabulary item "to design" is described by a complex string:

((X*Y)*W)*(Z*Z) -- 'X by means of Y affects W so that Z exists'.

The questionnaire: Who X by means what Y affects what Z designs what W?

The strings input this way and denoting the initial and target situations are placed on the relevant nodes of the graph drawing a route between them. The transfer algorithm reduced to correlation of the initial and the target situations can be defined as one-step inference, which has a form of a complex string.

Besides USC there are actually only two projects to make up a language with formalized semantics: the model of conceptual dependence proposed by R.C.Schank [4] and the sense <==> text model of I.A.Mel'chuk [5]. The basis of both are some primitives (semantic elements): primitive actions in Schank's model and lexical functions in Mel'chuk's, which form semantic notation of utterances. The primitives of the given models do not claim to be complete, independent and consistent in the strict sense of the word because of their empirical elaboration. The deductive theory of knowledge representing language has been embodied in USC for the first time. Nowadays it becomes clear that none of the variants of artificial intelligence can be effective without formal representation and transformation of sense since only under these conditions the computer modeling of mind processes is ensured.

A third version of USC (USC-3) [6] is entirely in agreement with the given principles. That is why it was realized as computer assisted. A fourth version of USC (USC-4) [7] differs from USC-3 in two fundamental characteristics: in exclusion of special means for representation of information and modality (these categories are represented by traditional ternary strings of elementary symbols) and in the explication of USC-4 as certain algebra.

We accept a set of axioms within the scope of this algebra. Each axiom represents a regular transformation of sense in explicit form [8]. It is a pity that majority of investigators do not understand so far that no kind of artificial intelligence systems can exist without semantic explication in the sense of substitution of the strict concept for the

intuitional one. The admission of the fact compels us to return to the main questions of knowledge representing languages, bases of knowledge, knowledge as such.

We shall demonstrate our understanding of the problem by displaying some examples. It is natural enough for a human to come to the following conclusion: *The engineer has seen the device before that is why he would recognize it* or in a more general form: *X has seen Y → X would recognize Y.*

If our system were intellectual enough it would know how to draw this immediate conclusion. In other words, the creator of artificial intelligence systems has to know the way of teaching a computer to draw such a kind of conclusion. Regretfully he does not know how to do it.

Moreover, he cannot perceive how the human does it. Let us try to assist the system giving for comparison another instance of deduction: *He has already played Rossini's "Tarantella" that is why he would play it* or in a more general form: *X has played Y → X would play (X can possibly play Y).*

A human identifies the verb to play in spite of grammatical differences. It is evident that the former deduction can be reduced to the postulate of modal logic P ==> <> P. Though we guess the first deduction is reducible to the same postulate, we do not know how to explain it if only for a human. The human easily uses such deduction but by intuition purely.

Let us begin from the right part of the first utterance *(X would recognize Y)*. This sentence signifies a possible result of action represented in the left part *(X can possibly recognize Y)*. It should be emphasized that after transformation the right part has coincided with the same part of the former utterance and the above-mentioned postulate.

Concerning the left part, the verb *to see* may be interpreted as *to receive information, to get to know*. Then we note the whole as follows: *X has known Y → X can possibly know (recognize) Y.* It became clear that in this case we are concerned with the same postulate of modal logic: P → <> P.

In order to identify P in left and right parts of our conclusion we must put semantics into formalisms of representation. By USC-4 means we write down this deduction in the following way:

$$((X*X)*X)*((\overline{Y*Z})*((\overline{Z*W})*W))) ==> (X*(X*X))*((\overline{Y*Z})*((\overline{Z*W})*W)))$$

In the canonized English it should be read: *X has known that Y is identical with Z which be longs quality W consequently X can possibly (get to) know that Y is identical with Z which be longs quality W.*

It is easy to understand that the formal representation and the canonized language as well demonstrate obviously the identity of P in left and right parts of the conclusion and distinguish the modal operator in the right one.

The number of such instances can be multiplied, but it is quite enough to make clear that the only means of human conclusion formalization is a powerful semantic code. Its potentialities can be realized if the working system of artificial intelligence would be supplied with specialized dictionary for translating natural language phrases into semantic notation (USC notation) or if a human would use USC type of language in the process of his intercourse with computer. We see no other solution.

Currently the principles of the KB construction of our intellectual system are applied by means of the computer language LPA PROLOG. The study is being performed within the project 'Invention Machine' and is meant to be sold at the commercial market.

References

1. Minsky M., Framework for Representation Knowledge. In: The Psychology of Computer Vision. New York, 1975.

2. Winston P.H., Artificial Intelligence. London-Amsterdam, 1977.

3. Skragg G., Semantic Nets as Memory Models. In: Computational Semantics. Amsterdam, 1978.

4. Schank R.C., Conceptual Information Processing. Amsterdam-Oxford-New-York, 1975.

5. Melchuk I.A., Experience of theory of linguistic models "Sense <==> Text". Moskva, 1974. (Russian).

6. Martynov V.V., USC-3: New Variant of a Language for Representing Knowledge and Effecting Calculations. In: Artificial Intelligence. Proceedings of the IFAC Symposium, Oxford-New-York-Toronto-Sydney-Francfurt, 1983.

7. Martynov V.V., USC-4 - The Special Language of Knowledge representation and transformation. In: News of Academy of Science of USSR, Technical Cybernetics, 5, 1989. (Russian).

8. Martynov V.V., USC Definitions, Axioms and Immediate Inferences. In: Knowledge-Dialog-Solving. Leningrad, 1991.