Computer Semantic Search of Inventive Solutions

Igor Boyko, Ph.D.
HP Labs, Palo Alto

ABSTRACT

The article covers problems of knowledge bases (KB) development and search automation of inventive solutions with intellectual computer systems. The following problems of the intellectual computer systems for inventive solution search are considered:

- correct inventive problem formulating;
- expanding a field of solution search by analogy on an abstract level for different domains;
- inventive solutions computing.

The article presents a design method of inventive solutions semantic search on a basis of the fifth version of Universal Semantic Code (USC). The article develops of Genrich Altshuller's problem formulating method in ARIZ and shows new search possibilities of the inventive solutions with semantic tools [1].

INTRODUCTION

The problems solved by means of heuristic methods are considered as inventive problems. For solving of such type of the problems the expert systems (ES) has been developed. The KB of any ES consists of facts and processing rules. The inventive ES has to help a user to make an invention starting with a problem formulation and finishing with a problem solution. The knowledge in the ES must be represented with some knowledge representation (KR) model. Practically, the expert system is an intellectual system, in which KB quality depends on a chosen knowledge representation model.

KR is oriented on a possibility of computer solving of intellectual problems [2]. The choice of the KR model allows avoiding the complications in development of the intellectual systems. However modern KR models can not satisfy of developer demands to KB of intellectual systems and particularly for inventive systems since they are built disregarding of natural language semantics [4].

About a real intellectuality of the computer system is possible to speak only if and when its KB is built on a semantic basis. But semantic building of KB is unenforceable without using a corresponding tool. For that we use Universal Semantic Code [6], which is more powerful in its semantic possibilities then well-known KR models.

SEQUENCES OF TECHNICAL OPERATIONS

A solution of an inventive problem usually is claimed as a patent for a method, a device or a substance.

2009 Edition
The TRIZ Journal, USA, March, 2001.

Where, we define following:

- the method is a collection of technical operations (TO) connected in a united technological process for achieving a determined purpose;

- the device is a collection of elements connected in a united design for implementing a determined method.

After analysis of the patents for methods the order of verbs, representing technical operations, was formed as a following structure [7]:

'... → verb 1 → verb 2 → verb 3 → ...', it is equal to:

'... → TO1 → TO2 → TO3 → ...'.

There is a more complex structure:

'... → verb 1 → verb 2 and verb 3 → verb 4 → ...', it is equal to:

'... → TO 1 → TO 2 and TO 3 → TO 4 → ...'.

Obviously that each 'TO' is located before or after some 'TO', i.e. all 'TOs' can be simultaneously of two types:

1. Previous
2. Subsequent

The type of TO depends on a point of view it is analyzed.

PATENT ANALYSIS

Each invention solves some problem to achieve a certain goal. The problem formulation is a target situation represented as a one or several TOs.

For example, the US patent 4711854: "Method of measuring moisture in a burnable absorber." The goal of the invention is 'measuring moisture in a burnable absorber'. The description of the invention is:

"A source of an inert carrier gas is provided that sends an inert carrier gas through line into enclosed chamber 1, which contains graphite crucible. A sample is placed within graphite crucible and crucible is heated through electrodes by means of an impulse electric current. Vapors from enclosed chamber 1 pass through conduit to enclosed chamber 2. Chamber 2 also contains a graphite crucible that is heated by impulse electricity through electrodes. Vapors leave chamber 2 through conduit and pass into hydrogen analyzer.

In the first enclosed chamber, which contains the sample, means are provided to dissociate the water in the sample into hydrogen and oxygen. This is conventionally done by heating the sample in the presence of a dissociation catalyst such as graphite or platinum. The preferred dissociation catalyst is graphite because it can be formed into the shape of a crucible which can also be used to hold the sample, it is inexpensive, and a graphite crucible is conductive and therefore can be electrically heated by impulse heating.

While other means of heating the sample can also be used, electrical impulse heating is preferred because it heats the sample very rapidly so that any water in the sample is heated to its dissociation temperature while it is still in contact with the graphite.

We have found that a second enclosed chamber for dissociating water is necessary to achieve an accurate measurement of the hydrogen content of a sample. While the second enclosed chamber need not be the same as the first enclosed chamber, it is preferably identical for ease of construction and maintenance. The other portions of the apparatus, the source of inert carrier gas (e.g., argon, nitrogen, etc.), and the apparatus for analyzing hydrogen content are conventional and are well known in the art."

From the given text TO sequence for achieving the goal 'measure' looks as the following sequence of the verbs:

send through → heat → dissociate → pass through → heat → dissociate → pass into → measure.

SHORT INTRODUCTION IN A USC THEORY

Universal Semantic Code is a semantic theory [6]. With its help we can implement a semantic analysis of the target situation and re-define it when necessary. USC is designed as a semantic language of knowledge representation and conversion. It has own algebra using the given set of variables.

The type of the USC algebra for representation and conversion the USC formulas has been determined:

A = < **M**, *****, **-** >

Where:

M is a set of elements;

' ***** ' is a binary operation on the given set (the operation of coincidence);

' **-** ' is an unary operation on the given set the (operation of junction).

The main KR unit in USC is a complex formula with a set of variable. Semantic relations in the complex formulas are represented as:

((X*Y)*Z)*(Z*(Y*Z)) -- X by means of Y acts on Z and as a result Z keeps Y in itself (in Z).

We would like to draw attention to the fact that the left part of the formula is a cause of an action in its right part. Thus the complex formula reflects the following chain: **stimulus → reaction.**

Each complex USC formula has one and only one meaning and each conversion of the USC formula into another has one and only one corresponding conversion. The rules according to which the conversion is done are defined with USC axioms.

The USC formulas represent all possible TOs in a field of engineering. USC allows to structure the TO database, turning it into the KB and enclosing semantics in its representation.

The USC classifier contains a list of TOs and its corresponding USC formulas. In the USC classifier all TOs are analogues. There is a class name with the corresponding USC formula defining a verbal interpretation. Each action-analogue related with the corresponding class name (engineering action). Fig. 1.1 shows the USC classifier structure.

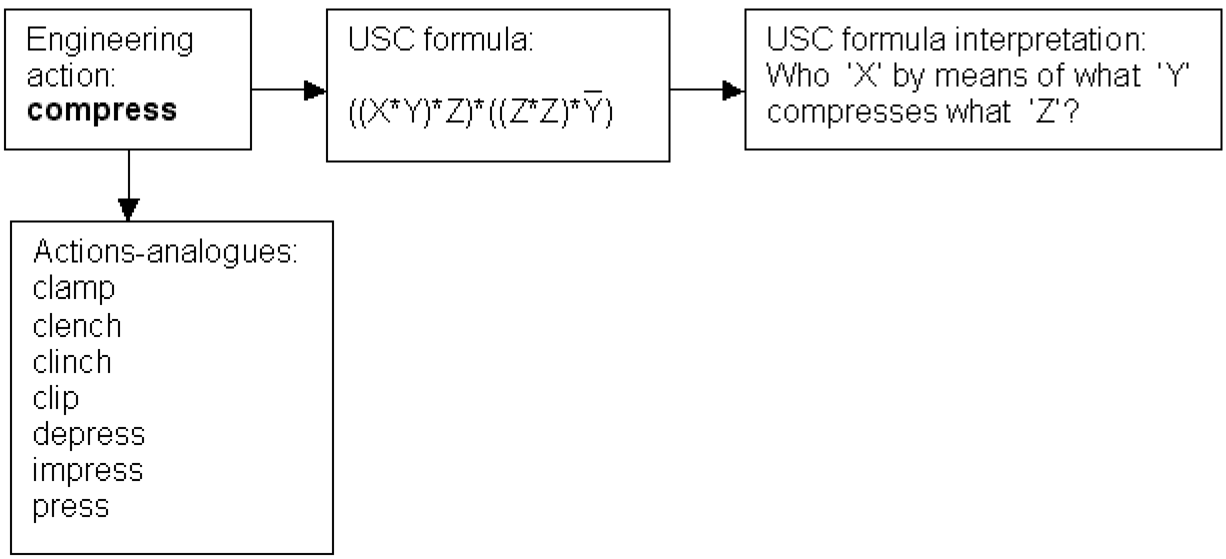


Fig.1.1 The USC classifier structure

Practically, within the limits of the given classifier the non-traditional system of TOs-analogues has been built according to the function they perform.

In the computer intellectual systems for a description of conditions of the problem it is necessary to use strictly structured KR units.

Saying that we mean the USC formulas having single sense, strictly given length, and limited set of the variables. Involuntary sequences of the USC formulas define variants of the solutions of the inventive problem.

The USC classifier is collecting technical verbs with its corresponding natural language interpretations according to the USC formulas. Since each verb expresses a relation between a subject and an object the USC classifier is representing that.

The USC formula is the knowledge representation unit.

The USC classifier defines a name of the class with a verb. Each USC formula corresponds to each class name and each USC formula has own verbal interpretation. The class name is an abstract presentation of a particular TO.

For example, the verb 'heat' is a class name and is presented by the USC formula:

heat - ((X*Y)*Z)*((Z*Z)*Y) -- 'X' by means of 'Y' heats 'Z'.

The verb 'measure out' is pointed to the verb 'measure' that is a class name and is presented by the USC formula:

measure - ((X*Y)*X)*((X*Y)*Z) -- 'X' by means of 'Y' measures 'Z'.

Within the framework of the given verb classifier the TO-analogues classifier has been built.

USC THEORY APPLICATION

The classifier was applied for analyzing of patent descriptions.

Checking the USC classifier for the patent the TO sequence is:

- the verb 'send through' is related with the class verb 'orient';
- the verb 'heat' is the class verb;
- the verb 'dissociate' is related with the class verb 'decompose';
- the verb 'pass through' is related with the class verb 'conduct';
- the verb 'pass into' is related with the class verb 'embed';
- the verb 'measure' is the class verb.

The final sequence of the abstract TOs for the patent is:

orient → heat → decompose → conduct → heat → decompose → embed → measure

Each TO is represented by the own USC formula and the verbal interpretation. Now we fill the formula variables according to the patent.

orient - ((X*Y)*Z)* ((Z*Z)*Z) -- X by means of Y orients Z

X = source;
Y = line;
Z = inert gas.

heat - ((X*Y)*Z)* ((Z*Z)*Y) -- X by means of Y heats Z

X = electrodes;
Y = impulse electric current;
Z = sample.

decompose - ((X*Y)*Z)* ((Z*Y)*Z) -- X by means of Y decomposes Z

X = graphite crucible;
Y = impulse electric current;
Z= sample.

conduct - ((X*Y)*Z)* ((Z*X)*Y) -- X by means of Y conducts Z

X = conduit 1;
Y = internal surface;
Z = vapors.

heat - ((X*Y)*Z)* ((Z*Z)*Y)) -- X by means of Y heats Z

X = electrodes;
Y = impulse electric current;
Z = vapors.

decompose - ((X*Y)*Z)* ((Z*Y)*Z) -- X by means of Y decomposes Z

X = graphite crucible;
Y = impulse electric current;
Z = vapors.

embed - ((X*Y)*Z)* ((Z*(W*Z) -- X by means of Y embeds Z into W

X = conduit 2;
Y = internal surface;
Z = vapors;
W = analyzer.

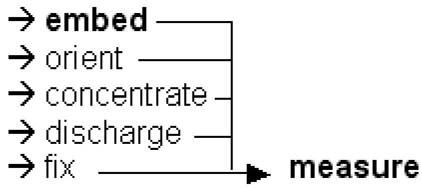
measure - ((X*Y)*X)* ((X*Y)*Z) -- X by means of Y measures Z

X = user;
Y = analyzer;
Z = hydrogen.

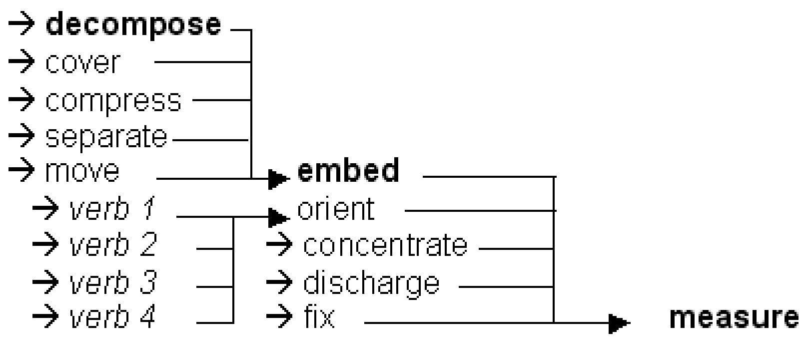
The USC formula helps to specify a subject, an instrument, and an object or objects of each TO.

CONSECUTIONS OF TECHNICAL OPERATIONS

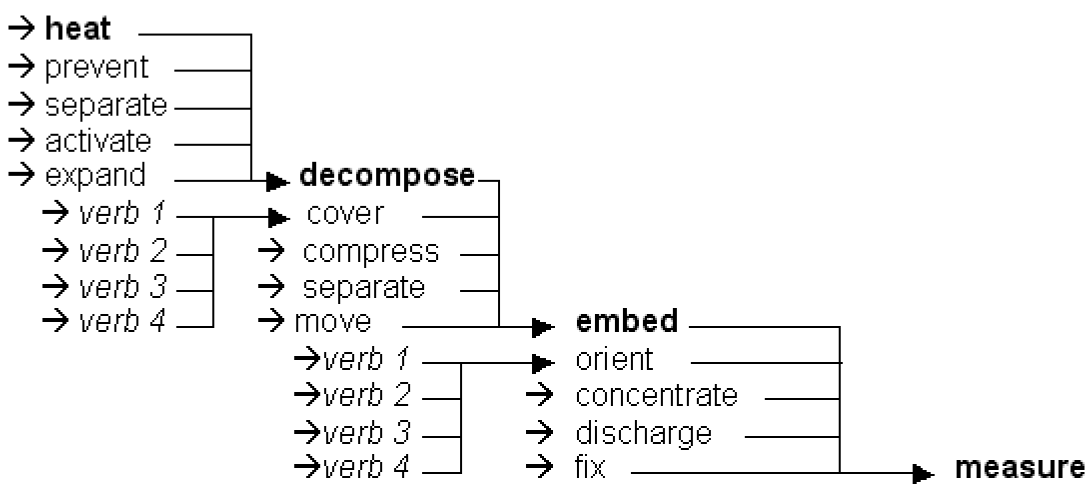
Let's consider how to reach the goal determined by the verb 'measure' with the TO sequences. For example, from the considered TO sequences we form a previous verb tree for the TO 'measure' as following:



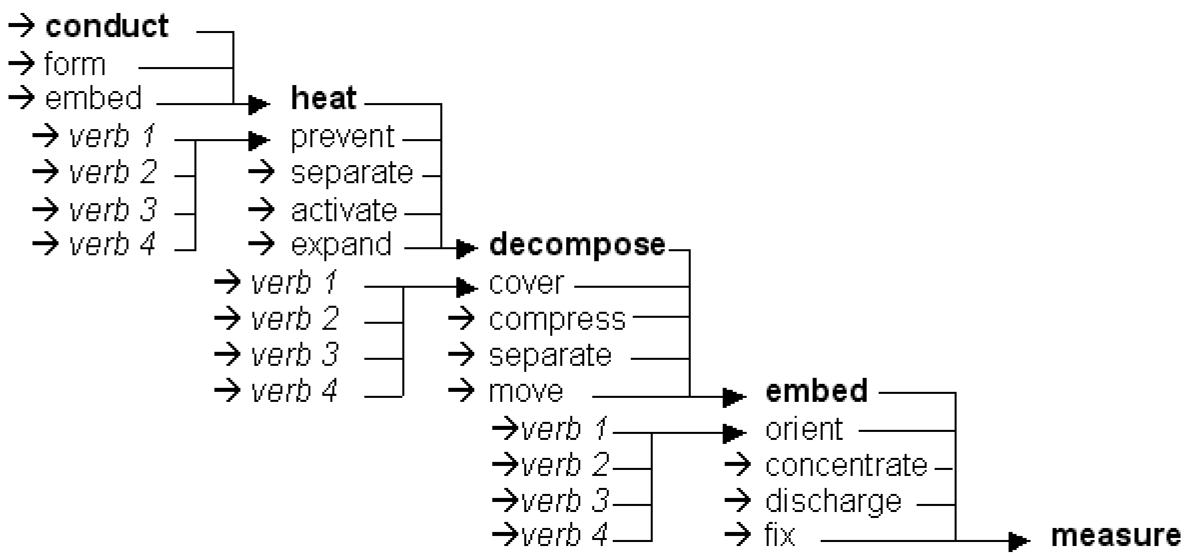
There is the previous verb 'embed' and now on the left of the verb 'embed' we can add its previous verbs. As a result, we see the verb tree:



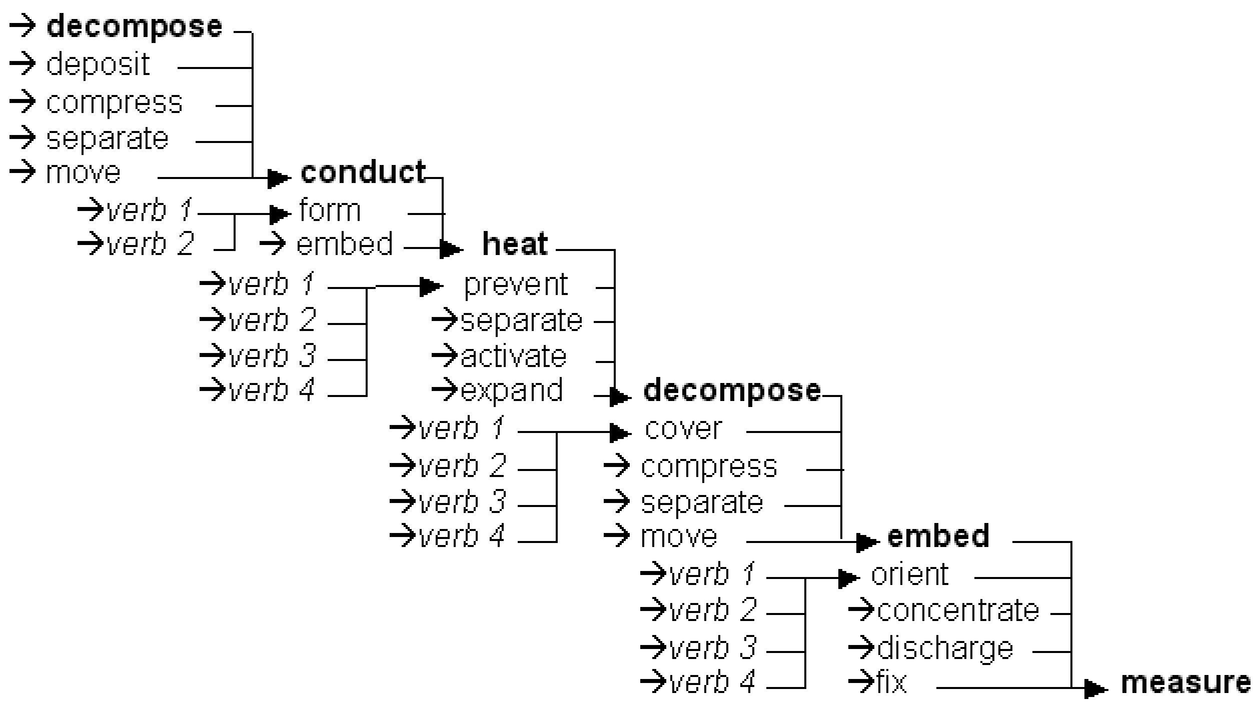
The next step is in adding new verbs to the left part of the tree at the verb 'decompose'.



A further tree with the left side to the verb 'heat'.



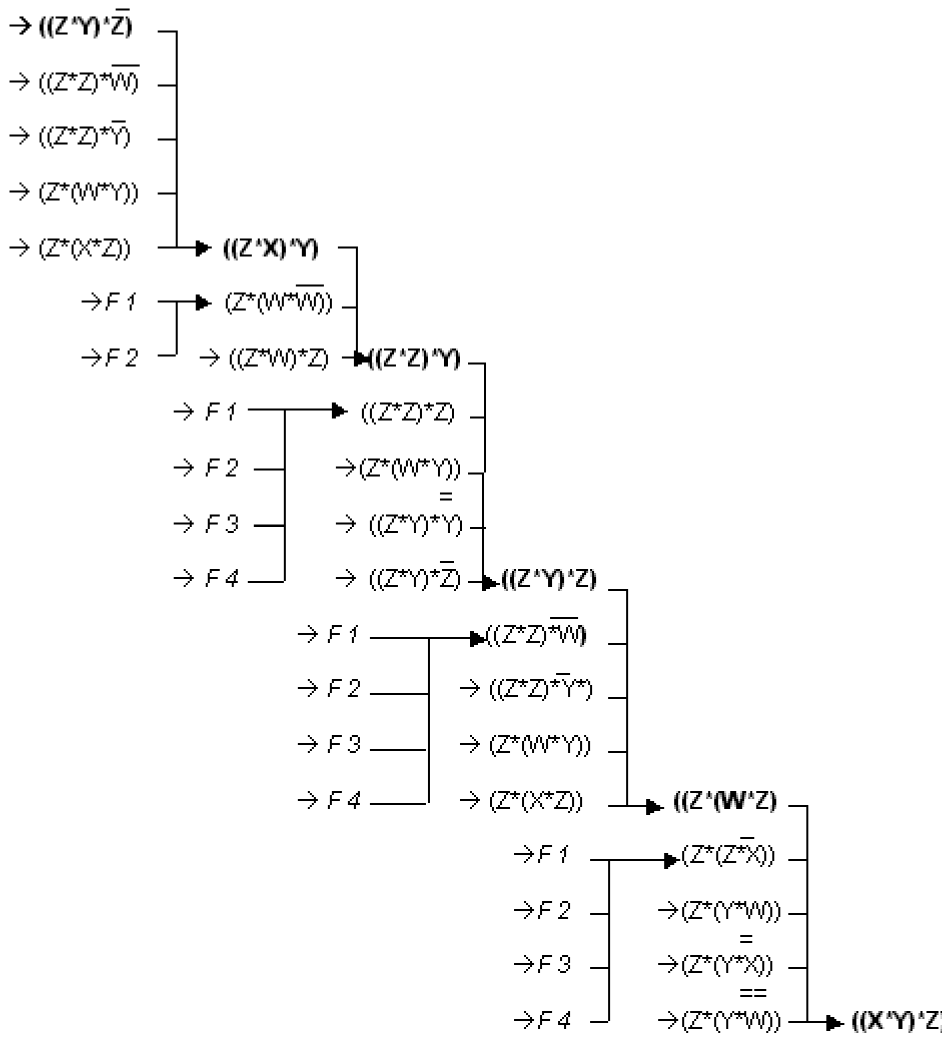
We continue to build the tree with the verb 'conduct'.



To each verb from the left part it is possible to add previous verbs.

Obviously, for reaching the TO presented by the verb 'measure' it is possible to use several ways or several verb sequences.

Now instead of the specific verbs we substitute its USC formulas. In a result, we see the following USC formula tree:



The sequences defining the order of using the USC formulas (or TOs) we name the semantic unit sequences (SUS) since each USC formula is an elementary KR semantic unit.

The USC formula tree may be considered as a graph with nodes in the form of the USC formulas.

The goal node is represented with the goal verb.

A complete sequence of the USC formulas for the considered patent is:

((Z*Z)*Z) → ((Z*Z)*Y) → ((Z*Y)*Z') → ((Z*X)*Y) → ((Z*Z)*Y) → ((Z*Y)*Z) → ((Z*(W*Z) → ((X*Y)*Z)

This shows that having rules of conversion of one USC formula into another in the manner of the USC algebra [6] it is possible to calculate the variants of the inventive problem solution. Substituting the USC formulas with its verbs it is possible to examine the solutions of the problem.

INTELLECTUAL PROBLEMS TYPES

We divide the intellectual problems on two main classes:

- Integrated-distributing problems;
- Inventive problems.

Integrated-distributing problems are problems on planning of the displacing, keeping, creating, and deleting different types of resources both physical and informational.

The inventive problems appear when an inventor develops new or improves already existing devices and technologies. As a rule, the inventive problem is easier to solve after revealing a technical, or physical, or both contradictions in the system.

The inventor begins the problem solving after formulating the initial and goal situations. I.e. a transition algorithm from initial to goal situation is a variant of solving. Ideally, both situations should be defined by single verbs.

Imagine you have the KB of the transition algorithms in the computer system and the KB is built on the basis of USC formula conversions. Comparing each of the defined situations with the KB of transition algorithms it is possible to find several alternative solutions as passes from initial to goal situations. I.e. the inventive problem is reduced to the integrated-distributing problem.

Since transition algorithms already exist we only need consecutively follow its steps for specifying the objects instead of the variables of the USC formulas.

For the inventive problem solving it is necessary to present only the goal situation. Moving from that to the previous TO the inventor decides on what of the previous steps he stops, i.e. defining this step as a preferred initial situation. Besides, in a process of the retrospective motion it is possible to solve contradictions creating the inventive problem.

A set of the semantic unit sequences presents the graph with the limited set of the elements (the verbs) strictly bound together. The graph is describing an inventive problem world.

So the number of the problems of both types is limited by the number (may be relatively huge number) of the semantic unit sequences. Appearance of new substances and materials will not enlarge this number but only will initiate using of the semantic unit sequences not used earlier.

CONCLUSION

Using the abstract verbs of the USC classifier allows reducing psychological inertia obtruded by specific terms of some domain.

Distraction from a specific description by means of abstract descriptions allows to find the inventive problem solution by analogy. Well-known inventive problem solving tools using analogies are the main solution searching tools for the identified goal [1; 5].

Filling the USC formulas with the variables specifies the subject and objects of analyzed TO and forces the inventor to think on a physical micro-level. That initiates deeper understanding of the considered situation.

References

1. Altshuller G.S. Creativity as an Exact Science. Gordon and Breach Science Publisher, 1984.

2. Ueno H., Koyama T., Matsuby M. Knowledge Representation and Using. Isidzuka. Omsya. Tokyo, 1987.

3. Martynov V.V, Boyko I.M., Gyminski A.P. Semantic Coding and Invention Problem Solving. // TRIZ journal. Obninsk, 1991, N2.1. (in Russian)

4. Martynov V.V, Boyko I.M., Gyminski A.P. Knowledge Bases Construction of Systems for Solving Intellectual Problems. // Controlling Systems and Machines. // Kiev, 1992, N5/6.

5. Altov H. And Suddenly the Inventor Appeared. Published by Technical Innovation Center. MA, USA, 1994.

6. Martynov V.V. Universal Semantic Code: USC -5. MSLU Pre-print N4. Minsk, 1995. (in Russian)

7. Boyko I.M. Semantic Search on the Stage of the Structure Designing in the CAD. PhD thesis. Institute of Engineering Cybernetics of Academy of Science of Belarus. Minsk, 1995. (in Russian)