Semantic Selection of a Newsletter Content in Accordance with Search Objectives Based on Account of User Preferences

[1]D.A. Akimov, [2]T.G. Tsunikova, [3]V.V. Karnakov and [4]A.V. Ostroukh
[1]Department of Information Systems,
[2]Department of Advanced Scientific Research, Institute of Integrated Security,
Moscow Technological University, Moscow, Russia
[3]ICSSA Moscow Technological University, Moscow, Russia
[4]Department of Automated Control Systems, Moscow Automobile and Road Construction
State Technical University, Moscow, Russia

Abstract: A method of preliminary semantic formalization based on the categories of the semantic dispositions, motives and attributes of the subjective model of awareness of events and images of the virtual world is proposed. At the same time, the semantic disposition actualizes the meaning of objects and phenomena, motive in the same interpretation serves as a source of meaning for objects and phenomena that are significant in the context of an actual unfolding activity model. As sources of attributive sense formation, not values in the generally accepted understanding but individually-specific categorical scales, serve as an instrument for identifying, classifying and evaluating the subject's significant characteristics of objects and phenomena of virtual reality. The goal of creation of models by the subject (agent) is formalized which consists in obtaining a forecast that is a new information object (message or algorithm of actions). A technological solution is proposed on the basis of the search model of the associative context in the semantic kernel.

Key words: Semantic disposition, subjective model of awareness of events, attributive sense formation, ontology, information flow, associative context

INTRODUCTION

When processing natural language processing to organize interaction with computer or robotic devices, the procedures of tokenization, normalization, stamping and lemming for the unification of word processing are used. The reverse side of unification is the simplification and emasculation of meanings in semantical constructs.

The question becomes topical: Are the semantic constructs capable of generating a source of generation of not only personal meanings but also, semantic installations of the actual activity of the actors in the network?

In the general case, the personal meaning of an object (a phenomenon, situation) can be composed of 3 components, determined by its meaning-making connections, respectively with motive, dispositions and constructs. The relative importance of these 3 components may differ in different cases and in a particular assessment situation, a decisive role may belong to either of them. Between different meaningful structures and mechanisms, conflict relations are also, possible, so-called "errors" or cognitive dissonance which is resolved in the process of compartmentalization or personal prioritization of meanings, motives and attributes.

MATERIALS AND METHODS

Method of formlization of the oretical model: According to, the research "The Psychology of Meaning: Nature, Structure and Dynamics of Reality" (Leontief, 2002) "a person is capable of preserving traces of semantic experience in the form of stable relations to objects and phenomena of reality, invariantly significant in various contexts of individual life activity and also, provoked even a single strong emotional reaction. The semantic dispositions serve as a form of semantic dispositions as a form of "fixing the subject's relationship to objects and phenomena of reality, determined by the role and place of these objects and phenomena in his life activity in a latent, inactive form." In the proposed method
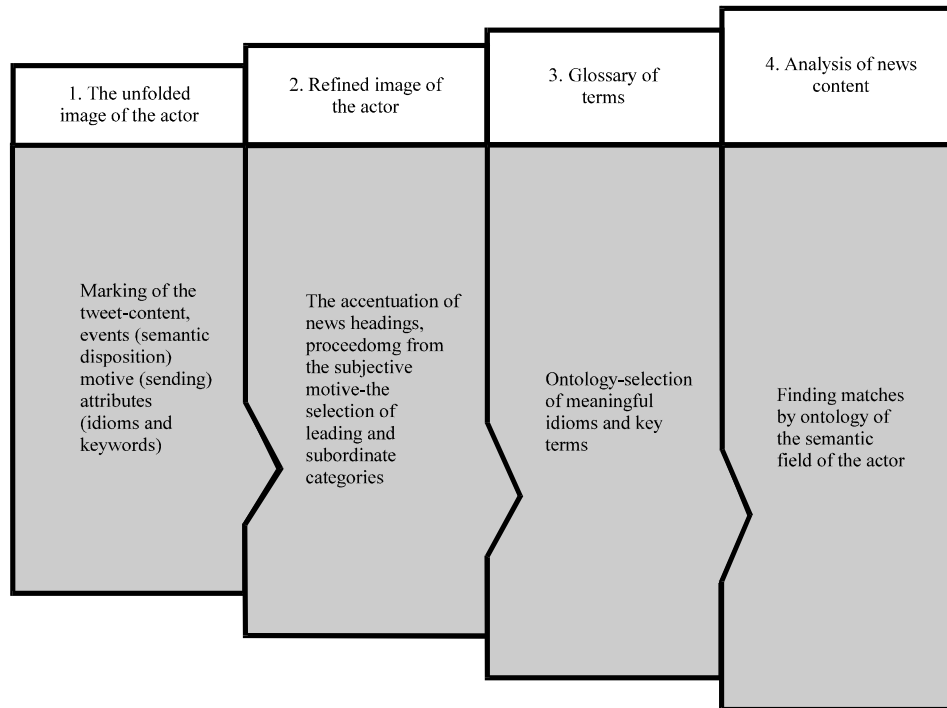

Corresponding Author: D.A. Akimov, Department of Information Systems, Institute of Integrated Security,
Moscow Technological University, Moscow, Russia

1. The unfolded image of the actor	2. Refined image of the actor	3. Glossary of terms	4. Analysis of news content
Marking of the tweet-content, events (semantic disposition) motive (sending) attributes (idioms and keywords)	The accentuation of news headings, proceedomg from the subjective motive-the selection of leading and subordinate categories	Ontology-selection of meaningful idioms and key terms	Finding matches by ontology of the semantic field of the actor

Fig. 1: Construction of the ontology of the subject's semantic field

of preliminary semantic formalization, the semantic dispositions are reflected in the form of a "short content", arising through the mechanism of actualization of this latent relationship in the relevant situation, reflecting the personal meaning of the object or phenomena associated with its pragmatic significance. Thus, the semantic disposition actualizes the meaning of objects and phenomena to which subjects of activity for some reason or another retain a stable out-of-situ attitude and impart meaning to other objects and phenomena directly related to them.

The motive in the same interpretation serves as a source of the meaning of objects and phenomena that are significant in the context of the actual unfolding activity. In the proposed technique, the motive (incentive, message) is associated with some prejudice (implied meaning) in relation to it.

So, we take as a basis the assertion that the sources of derived meanings which are given to objects and phenomena of reality that fall within the scope of the subjects are the motive and the semantic disposition.

A possible evaluation criterion and accordingly, the source of the meaning that the means acquire for the subject in the first case and the act in the second in both cases are their own significant parameters, qualities or attributes. That sets the third attributive mechanism of

sense formation-the mechanism of revealing the meaning of objects and phenomena from the point of view of personal meanings and values.

Thus, the sources of attributive sense formation are not the values in the conventional sense but individual-specific categorical scales serving as an instrument for distinguishing, classifying and evaluating the subject's significant characteristics of objects and phenomena of reality (Leontief, 2002).

Or else, the actualization of semantic constructs "from below" is connected with other psychological mechanisms: the choice of these or other constructs is "imposed" by the nature of the objects themselves. The mechanism of actualization of semantic constructs in this case is not a motive but a semantic disposition in relation to the estimated content that is actualized through the subject component and generates a categorical setting that determines the selection of semantic constructs, regardless of the motive of the actual activity. Considering the above, we define the mechanism of actualization of semantic constructs for some agent. Semantic field actor (Fig. 1) is formed in four steps the following rules formalization:

Step 1: Arbitrarily selected news events with subsequent classification and value judgments.
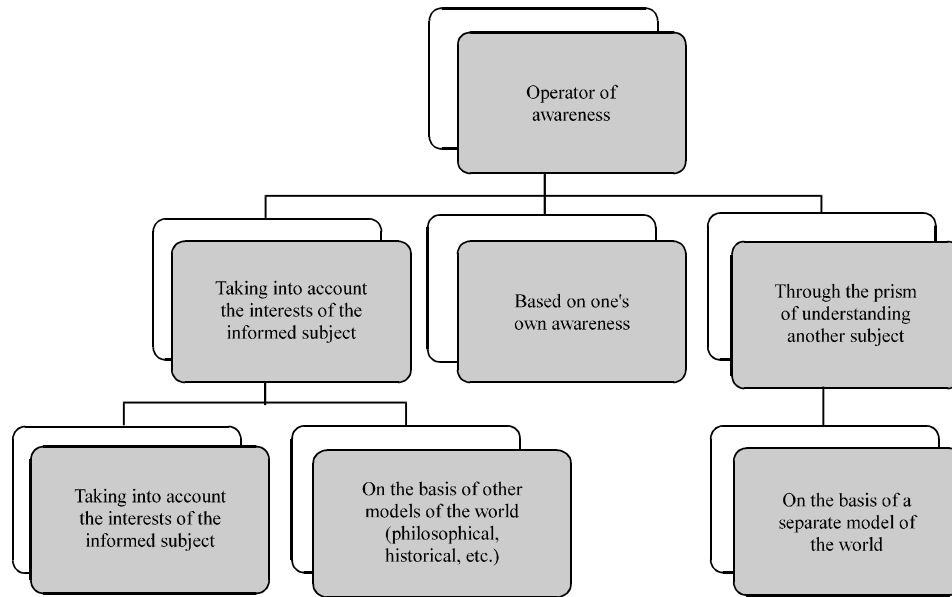
Fig. 2: Formation of a semantic event model

Rules:
- the choice of headings is made arbitrarily
- tweet content contains an indication of the domain "address/person/source"
- mask event provides as a summary
- motive, message-personal reflection of the virtual subject

Step 2: The accentuation of news headings is carried out by the user on the basis of his own preferences: the leading columns and subordinates are highlighted.

Step 3: Selection of significant idioms and key terms (semantic field of the subject) throughout the sample is carried out it comprises a glossary of terms for all subject's patterns (expert, incompetent user, experienced user).

Step 4: Taking into account the already obtained semantic field of the terms of each virtual subject (actor) and patterns of semantic preferences, an analysis of the arbitrary news content is conducted to see if all the subjects involved in the search coincide with the selected semantic field patterns: correlation of events, prioritization, accentuation according to the levels of user competence, etc.

Returning to the question posed in the introduction about the meaning of actual activities of network actors, it is necessary to answer in the affirmative, since as follows from the proposed justification of the theoretical model, "the ontological aspect as a facet of meaning, ... is not a psychological reality but rather a virtual reality" (Nosov, 1995, 1998).

Formalization of the logic of the objectives of the meaning field of the subject: Turning to the research "Mathematical models in the information confrontation", we will determine that the goal of creation by the subject (agent) of models is to obtain a forecast that is new information objects (message or algorithm of actions) (Rastorguev, 2014). Then the fact that the model is generated by the new image can be written as:

- $x_1A1 \rightarrow A2$, if the event is included in the world of A2,
- $x_1A1 \rightarrow -A2$, if the event is included in the world of A2

In this case, x, y, z is information objects or active elements in particular an algorithm; A-D is named information objects, including ordinary messages. The physical meaning of the process of awareness lies in the appearance of links between the elements of the structure and the fixing of the attributes "useful", "dangerous", etc.

The agent (acting subject/object) can be aware of the world (the formation of pseudo-images and semantic dominants, together with the attributes of conditions) on the basis of awareness of already existing models of the world (Fig. 2). We introduce the notation as suggested (Rastorguev, 2014).
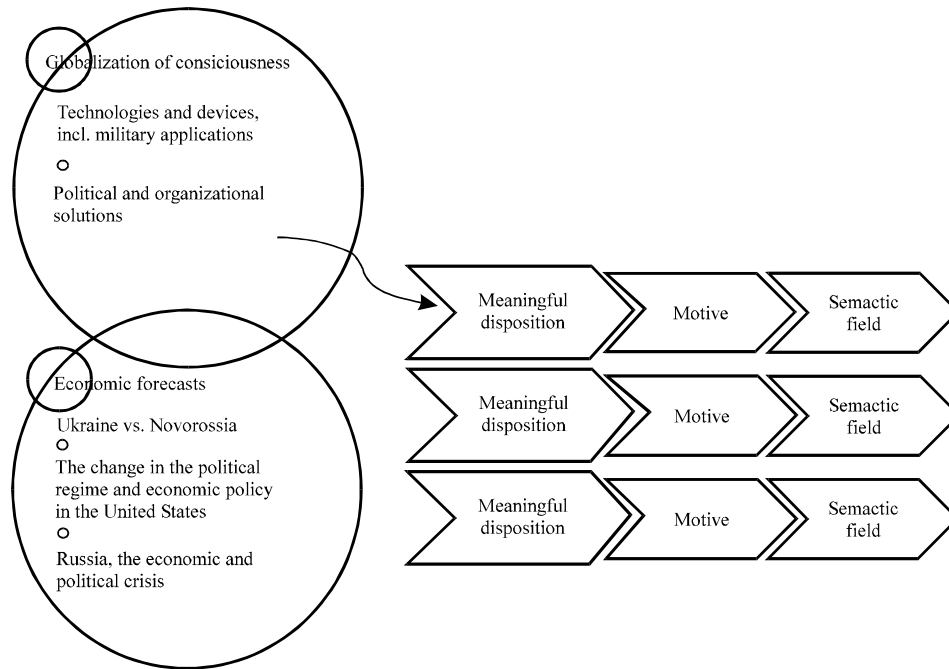
Fig. 2: Formation of the basic target ontology

If "I" is the operator of realizing the world as a non-informational agent (avatar) then during its implementation, there is no awareness of the world and a change in the picture of the world. Then reflection without awareness, then it is formalized as follows:

$$I (x+y) = x+y \qquad (1)$$

Denote w-the world transformation operator for our case, when the awareness operator includes not only the subject but also, objects or the avatar, "works" according to, the model of the subject's world.

Otherwise, the world transformation operator will include not only subjects but also, any text objects or avatars:

$$<\text{Conversion operator}>: = (* \{<\text{awareness}>\}+\{< \\ \text{inclusion/exclusion operations to/from the world}>\}) \qquad (2)$$

In this case, the transformation operator of the world allows inclusion not only the process of awareness but also an event for awareness which will be represented as a semantic tetrahedron-in the relationship of the semantic disposition (event for awareness) motive or sending (awareness) and attributes (connectors with the subject's past experience) (Fig. 1). Attributes themselves form the basic ontology (Fig. 3).

RESULTS AND DISCUSSION

Software search and selection system of the semantic field of the signed event model based on basic ontology: In automatic mode, the system analyzes the information stream of news feeds coming from the API of the information environment and identifies events based on the data received from the stream. The system annotates them with additional information (place, time, actors, etc.,) which comes in the process of analysis. On the basis of the information received, the system is able to select events according to, a certain criterion (for example, political events). Descriptions of such events are stored in the knowledge base for further analysis.

Figure 4 show the notion of "marking a tweet-content implies the allocation of an event (semantic disposition) motive (sending) attributes (idioms and keywords) as well as the accentuation of news headings with the selection of leading subordinate categories. Formation of the ontology of personified content is the selection of significant idioms and key terms on the ontology of the actor's semantic field.

Let's consider in more detail the principle of finding the associative context in a semantic kernel. The method of searching consists in comparing the semantic proximity of the incoming vector with the available vectors in the kernel dictionary. The semantic proximity of the 2 vectors
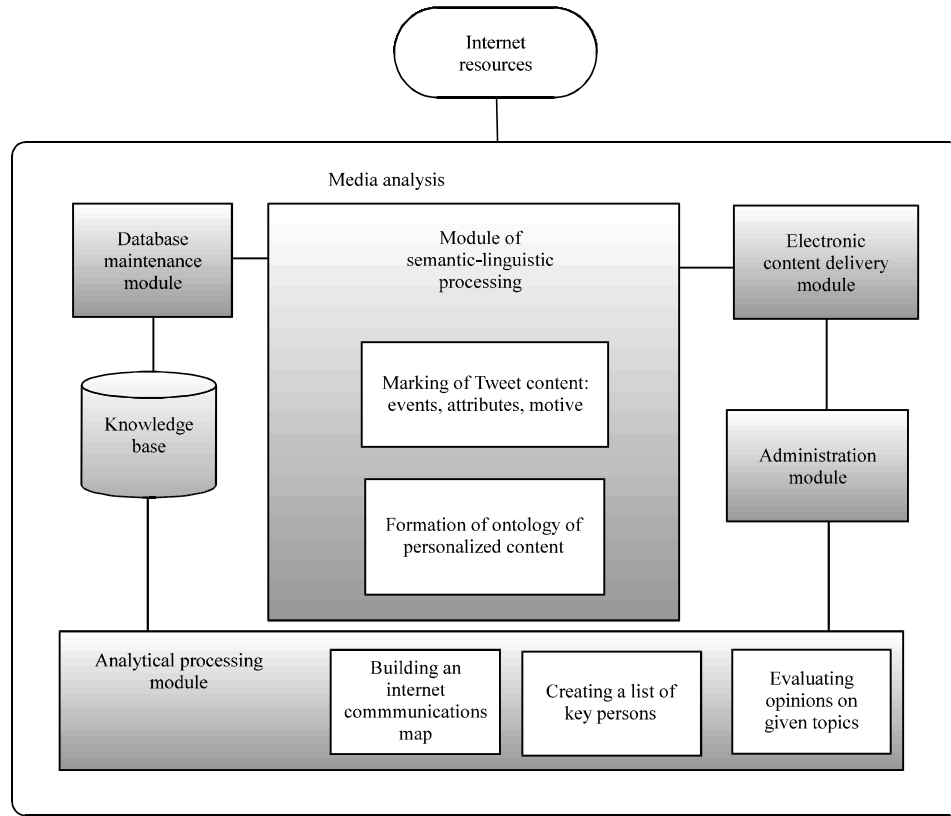
Fig. 4: Structural scheme of the collection and analysis of news feeds

A and B is calculated as the distance between the vectors, using the cosine similarity represented by the following formula (Huang *et al.*, 2012):

$$Similarity = \cos\theta = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (3)$$

The results will always be a number in the range from -1 to 1. But in this case, it is necessary to find a set of vectors with the closest distance to the vector A (the number is close to 1) so, we transform the formula to the following form:

$$Similarity = \underset{x \in V}{\arg\max}\left(\frac{\sum_{i=1}^{n} A_i x_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\sqrt{\sum_{i=1}^{n}(x_i)^2}}\right) \quad (4)$$

Где $V = \{x_1, x_2, ..., x_n\}$ is the set of vectors of the semantic kernel dictionary and x is the word vector in the dictionary. Then the set of word vectors forming the associative context will be represented by the following condition:

$$W = \{x \in V | similarity(A, x) \geq k_{min}\} \quad (5)$$

where, W is a set of association vectors and k_min is the threshold coefficient of the minimum cosine distance between vectors that is required to filter the results. Similarly, a lot of annotations consisting of n-grams will be displayed:

$$N = \{A \in D, x \in V | similarity(A,x) \geq k_{min}\} \quad (6)$$

where D is the set of vectors of incoming n-grams. The result of the search model of associative matches is the output of a lot of annotations N of the analyzed corpus or corporation, as well as the many associations W related to it. The abstract is represented by a set of n-grames, semantically closest word vectors in the vector space of the corresponding semantic kernel.

To automate the procedures associated with supporting information retrieval and preservation of information about political publications, the projected Information System (IS) is designed. To extract the network content data, the IS software is used, working with files with the scrapper.py extension and

Parser_rio.py

It downloads a news context on the scraber links in the "data" folder and saver it as a CSV file in the "Parse_data" folder with a semicolon separator

Acrappe_nov.py

It collects links to the address "http://old.novayagazeta.ru" for 2578 days (the whole archive) and saves it in the "data" folder as a CSV file with a semicolon separator;

Parser_kom_lenta.py

It uses the links collected by the scrapper in the "data" folder and downloads the content from the news feed, saving it the "parse_data" folder as a CSV file with a semicolon separator

Scrapper_kommersant_lenta.py

It collects all the links that are in the tape at the address "http://kommersant.ru/archive/news" for 5840 days (that is, the entire existing archive) and saves it in the "data" folder as a CSV file with a semicolon separator

Parser_kom_lenta.py

It uses the links collected by the scrapper in the "data" folder and downloads the content from the news feed, saving it in the "parse_data" folder as a CSV file with a semicolon separator
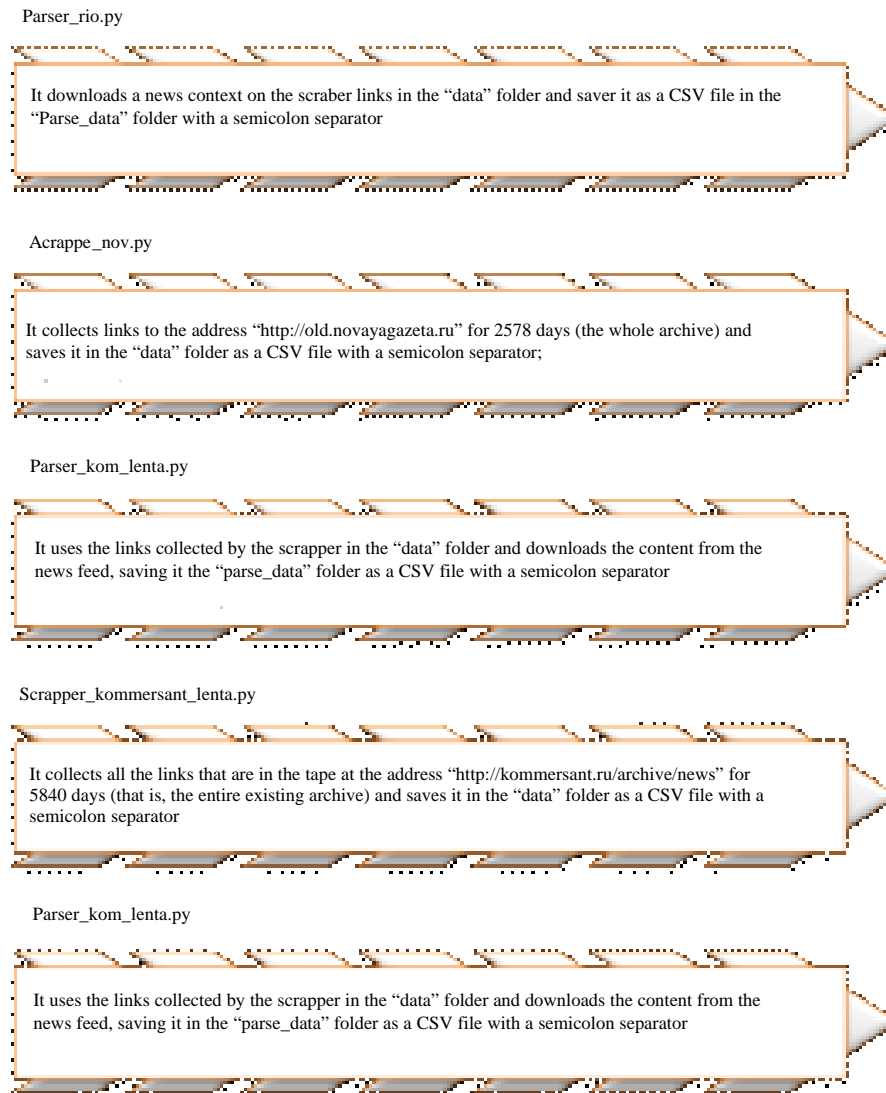
Fig. 5: Diagram of the description of the operation logic of the associative search algorithm

parse_rio.py. This software product allows the user to enter the address of the news feed site while the built-in information system will perform an automatic search and retrieval of data from the downloaded documents (Fig. 5).

Detailed block diagram of the semantic selection of news content in accordance with the search objectives based on the user preferences "The program complex for searching and highlighting the semantic field of a given event model on the basis of a basic ontology".

CONCLUSION

A method of preliminary semantic formalization based on the categories of the semantic dispositions, motives and attributes of the subjective model of awareness of events and images of the virtual world is proposed. The goal of creation of models by the subject (agent) is formalized which consists in obtaining a forecast that is a new information object (message or algorithm of actions).

It is established that the proposed solution phenomenologically coincides with the results of S.P. Rastorguev (2014) study in the part of the model of the operator of the transformation of the virtual world, since, the world transformation operator allows the inclusion not only of the awareness process but also, of the event for awareness which is a semantic tetrahedron in the relationship of the semantic disposition (an event for awareness) motive or sending (awareness) and attributes (connectors with the subject's past experience).

The software complex of search and selection of a semantic field of the given model of event on the basis of the basic ontology is developed. The method of searching for the associative context search in the semantic kernel is justified.

REFERENCES

Huang, L., D. Milne, E. Frank and I.H. Witten, 2012. Learning a concept-based document similarity measure. J. Assoc. Inf. Sci. Technol., 63: 1593-1608.

Leontief, Y., 2002. [Modern Psychology of Motivation]. Sense Publisher, Moscow, Russia, (In Russian).

Nosov, N.A., 1995. Psychology of virtual realities. Logic Method. Philosophy Sci., 9: 94-98.

Nosov, N.A., 1998. The virtual paradigm virtual reality. Proc. Lab. Virtualistics, 1: 87-94.

Rastorguev S.P., 2014. Mathematical Models in Information Confrontation (Existential Mathematics). ANO Nauka Press, Moscow, Russia, Pages: 268.