Intellectual Contribution of Specialized Thesauruses to Cognitive Technologies

Natalia P. Tuchkova

Dorodnicyn Computing Centre, FRC CSC RAS
Moscow, Russia
natalia_tuchkova@mail.ru

Abstract—The research is devoted to problems of perception of information in training, science, critical situations in the context of cognitive technologies. To consider semantics of subject domains when using cognitive technologies is one of problems of the digital environment. This problem is offered to be solved by attraction of specialized subject thesauruses. It is noted that during a digital era the role of specialized thesauruses as bases for ontology data presentation increases in information systems. On the basis of thesauruses it is easy to create ontology of a situation and to use it as model of data for cognitive agents.

Keywords—thesaurus, ontology of situation, subject domain, semantic relations, cognitive technology

I. INTRODUCTION

The feature of modern style of work consists in the speed of receiving and dissemination of data as consequences of *change of a representation form and a way of information transfer*. In passing modern reader turned into the user. At the beginning of a human civilization there were graphic and symbolical forms of representation, text also videos appeared then. Thanks to development there were cardinal changes in data carriers, speeds of transfer and perception of information. Need to acquire quickly (to perceive, read and understand) information dictates new forms of its represention and development of the appropriate cognitive technologies. The idea of *compression of information* began to predominate in these technologies, and *time becomes the main resource* which should be saved.

The problem of cognitive perception, arose as a part of information technologies of artificial intelligence. The researches connected with artificial intelligence, rather specialized and high-technical. The main problems of artificial intelligence include *creation of the software products intended for operating with formal deduction, knowledge, logic of the task solution, perception of information (recognition), training, planning, object control.*

Therefore when mechanical, and then and intellectual functions began to be transferred in production and communications to automatic machines, computers, "cognitive agents" (intelligent agents), scientists thought of process of interaction of the person with the computer (*brain-computer interface*) also among themselves (*computer -*

computer interface). Before were interested in a problem of cognitive perception in medical, psychological and social aspects.

Researches in recognition of the speech, training, planning, solution of tasks make one of key parts in technology of representation of knowledge and in cognitive technologies.

The task for developers is a creation of the information environment with the set properties and the corresponding purposes, one of which improvement of cognitive perception of information (we will not consider a case when information intentionally is distorted or disappears from the user). It belongs first of all to the systems of training (education) and scientific information, but also extends to spheres of business, medicine and technological information which accompanies engineering (electric, gas) networks, monitoring of the environment, etc. A part from listed directly influences daily existence of modern society, and delay, distortion, ignoring, incorrect understanding and perception of information can promote creation of a critical situation.

Cognitive technologies it is technologies of knowledge assimilation that, naturally, implies process, but not single action. This process begins with the *choice of a way of knowledge (data) representation and continues to assessment of assimilation result*. The way of representation and assessment of knowledge varies depending on subject domain. In some cases computer technologies when as in some fields, for example, of psychology or sociology, role-playing games and tests are actively used can be used. Nevertheless, cognitive graphics is anyway used as almost integral element of submission of information for tests and communication in the course of assimilation and assessment of knowledge.

The idea about support of subject thesauruses graphics and supports of cognitive graphics explanations and a reference material from the corresponding thesauruses began to be implemented as soon as technologies allowed. When there were first programs of visualization of communications in ontologies, at once it became clear that it is the effective working tool and it was added to the arsenal by cognitive technologies. The projects WebVOWL Protégé, OWLGrEd and many others offer technologies of visualization of subject

ontologies. But, there is a question of filling of subject domains, creation of descriptions with allocation of the main concepts and terms and communications between them which are set by logic of subject domains. The basis for such descriptions is made, as before, by dictionaries with communications - thesauruses of subject domains. In this paper it is discussed that on the basis of thesauruses, addressing technologies of languages of ontologies, it is possible to form ontologies of subject domains and the description of situations and to use them as models of data for cognitive agents.

II. FROM HISTORY

The digital community for a long time realized importance of cognitive technologies, and for data presentation addressed symbolical and graphic information again. For example, scientific applications, knowledge where it is impossible to do without symbolics. Mathematics, chemistry, physics, mechanics, nanotechnologies and others which also inherit mathematical symbolics. The modern materials science includes molecular modeling that means inheritance of chemical symbolic.

Understanding of formulas belongs to cognitive graphics. *Mechanisms* are in turn necessary *for interpretation of graphics*. Symbols and formulas in perception of the scientist, the graphics in management, signals in the equipment, signs in semiotics are various interpretations. The history of the addressing of mankind graphics is a history of communication and it contains the centuries-old period. Since *Homo sapiens* and so far (*Digital Era*) graphics is used for transfer and assimilation of information. But, at the beginning of a civilization the question of transfer of knowledge and interpretation of information had limited character. This was *not about distance learning, management or remote communication via digital means of communication*. The periods of development of visualization tools and communication of homo sapiens which include operational stages of graphics, symbols, writings, the press and computer technologies are schematically reflected in fig. 1.
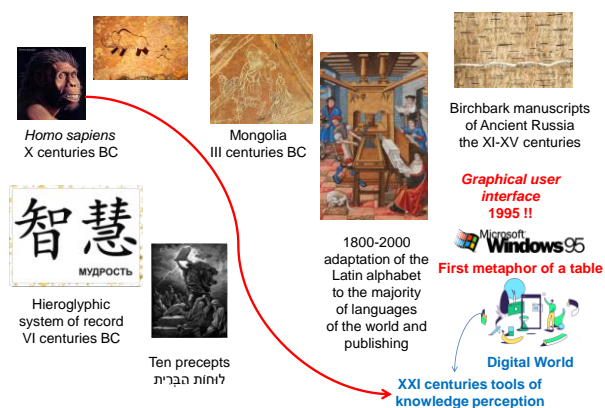

Fig. 1. Means of communication from prehistoric time to a digital era: images, pictograms, icons, symbols, writing, digital information

From this illustration (fig. 1) it is visible that visualization of information is necessary for the best perception and interaction between individuals. Interaction of an individual with devices and between devices is added to a digital era to it. The role of cognitive graphics and its interpretation increases in various spheres of human activity. For example, in work [1] the option of *identification of some mathematical abilities on the basis of the mechanism of the analysis of symbolical information is offered*.

To *consider semantics of subject domains* during creation of cognitive graphics is one of problems of the digital environment. This problem is solved by attraction of specialized subject thesauruses. Researches in this direction are conducted with the middle of last century and include a problem both a query [2], and actually representations of subject domains and expansion of a query by means of the relations between concepts [3]. In the monograph [4] the analysis of that is given as the model of information perception changed with arrival of technological progress. From model of vocational education "*training at practice*" to model "*communication and control*". On the basis of key concepts of the theory of management of information (cybernetics) perception of "*knowledge*", "*facts*", "*data*" is analyzed. It is claimed that "information", is a chain of semantic connected facts and, setting semantics of data, it is possible to receive model of a situation, the entrance exit type ("initial information" - "final information").

Thus, "management of information" and "management of behavior" become the central problems in professional perception of information, from the point of view of the author [4]. Since a release of this monograph passed more than 20 years, and in modern information systems semantic models of data are used. Work [5], and to issues of expansion of knowledge in digital representation of mathematical subject domain - work is devoted to questions of use of the specialized thesauruses containing symbolical information for identification of a conflict situation [6]. The real research is devoted to cognitive opportunities of use of thesauruses for interpretation of symbolical and graphic information on the example of adjacent subject domains and tasks of management.

III. PROBLEM OF IMPROVEMENT OF PERCEPTION OF INFORMATION

With change of speed of information transfer and receiving a reciprocal signal in the form of management the problem of improvement of perception of information increased. Traditional cognitive tools underwent changes and were complemented with new tools of artificial intelligence. One of researchers (Iqbal M., [7]) in 2015 reflected variety of interaction of modern homo sapience with various electronic devices "brain-computer interface" in the form of the map. It is interesting that the speed of development of this area

such is that now almost in 5 years this map can be added with interaction "computer - computer interface". The part of control of information flows is assigned to the person here, and the logic of management processes relies on semantics of links of objects of management.

The author [7] continued to analyze interrelations of theories of knowledge of 21 centuries in the context of the social relations, relying in many respects on works of the 80th years of the 20th century about theories of the distributed cognitive processes (Edwina Hatchinza, etc. [8]). In fig. 2 the map of theories of human-machine interaction with definition of the place of cognitive systems is shown.
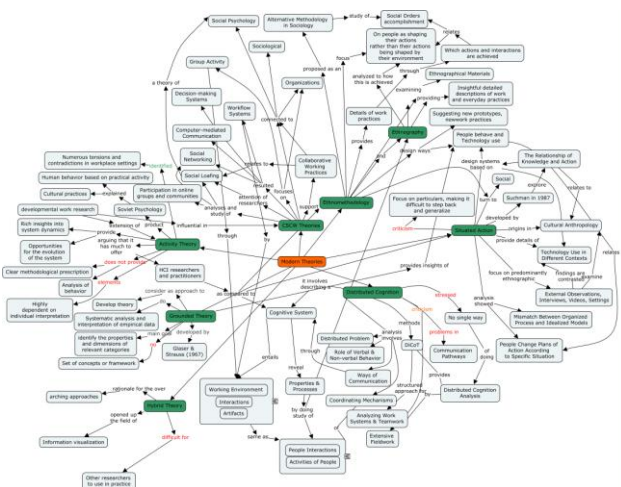

Fig. 2. Map of modern theories of human-machine interaction: problems and restrictions of theories and green - the main theories / structures are highlighted with red color

It is possible to treat such illustrations as they reflect a subjective view on a problem of dissemination of knowledge in modern information networks and on social networks, in particular differently. Nevertheless, it is one their *examples of cognitive graphics*, allowing to estimate *scales of a problem of reliability* of information which face in control systems. The system of check of reliability of information and control of assimilation of reliable information is inevitably built in a basis of improvement of perception of information (the facts, data, etc.) during a digital era. It is necessary to remember that cognitive *technologies are technologies of assimilation of knowledge* which are borrowed from social (medical, psychological) sciences and are based on a combination classical and modern (digital) methods of representation of knowledge. It is the part of information technologies allowing to consider influence of human factors in interaction of "brain-computer" and the being great importance in modeling of emergency situations. Appropriate even to mention distinction and similarity of the purposes in information transfer (and dissemination of knowledge) with transition to a digital era, fig. 3. Initial and ultimate goals of information transfer did not change, it is communication. The problems arising in this process

if it was necessary to inform and save reliable information earlier, then now, owing to ease of the process of communication changed, the problem of protection against unreliable information and unreliable sources appeared.
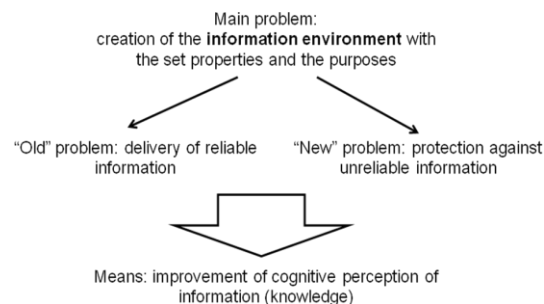

Fig. 3. Scheme of problems of communication

What was means of communication less, subjects, naturally, there are less opportunities to distort information, but also it is less than opportunities in dissemination of knowledge.

Questions of cognitive perception include:

- language problems;
- education problems;
- problems of interpretation of subject domain of knowledge;
- problems of representation of subject domain in the digital environment.

Traditional control methods of perception of information and knowledge use:

- guides and textbooks;
- cognitive graphics and video;
- role-playing games and modeling of a situation.

The artificial intelligence focuses attention to creation of intellectual means which work and react as people in the field of recognition of the speech, images, training and planning of actions.

Digital era allows to use in addition:

- specialized thesauruses for the description of subject domain (situation) and their display in subject ontologies;
- methods of virtual reality for modeling of communications and interaction of objects (cognitive agents);
- numerical modeling and forecasting;
- remote communications by the Internet.

The analysis of data and preliminary information processing makes a significant part in digital data presentation (knowledge). The purpose of such

preliminary processing is use of cognitive technologies for the maximum satisfaction of information needs of the user in the course of communications of a various type and management. Management, thus, "leans" on situation model (knowledge of subject domain) which is set by means of ontology of subject domain (ontology of adjacent subject domains). For creation of these ontologies it is necessary to reveal the main objects and their communications that at a stage of data processing is set when forming the thesaurus of subject domain. This process is schematically reflected in fig. 4.
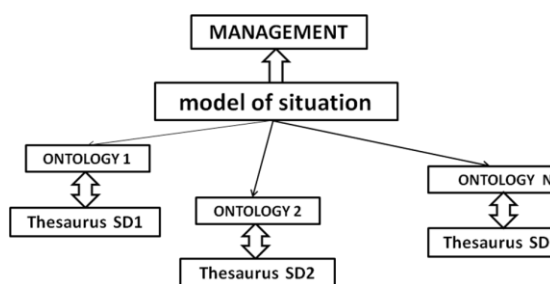

Fig. 4. The scheme of management on the basis of the description of subject domains of a situation (knowledge)

IV. ROLE OF THESAURUSES IN DETERMINATION OF RELIABILITY OF INFORMATION

Data presentation, information, knowledge, also allow to receive situations in combination with modern means of communication:

- message about an essence of a concrete event (situations, incidents);

- to correct information, using the communications set by the thesaurus relations;

- to receive assessment of reliability of information (situation);

- to create the list of possible actions and consequences.

The listed opportunities are implemented by means of information systems and create information support in adequate perception of the imparted (gained) knowledge, management of data flows and cognitive agents.

To management of communications with correction of data it is applied:

- the description of a situation with use of the dictionary, with controlled lexicon and graphic maintenance (illustrations in the form of cognitive graphics);

- options of alternative and similar situations (tasks, examples, etc.);

- options of development possible (transitional) situation;

- versions of decisions (transitions to a final situation);

- assessment of probability of options of development and transition to a final situation;

- comparison of a final most probable situation with the description in the thesaurus.

If the most probable final situation is in the description a situation - it is possible to consider initial information reliable, otherwise, check in the form of request of additional data is required, to information, etc. In the simplified look the scheme of check of reliability (verification) of information with use of the preliminary descriptions set is presented in the thesaurus of subject domain (knowledge, situations) in fig. 5.
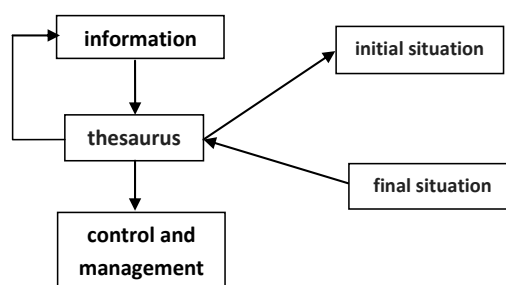

Fig. 5. The scheme of verification of information with use of controlled lexicon

To consider the same scheme realization of one-to-one correspondence:

ontology <-> situation

If the ratio at receipt of information is carried out, then it is possible to consider that reliable information, differently - No. Means that in an information system, there is a description of a situation in the form of the ontology created on the basis of communications in the thesaurus. That is, the task comes down to existence or lack of unambiguous descriptions and means of their representation in the form available to perception at the moments of management.

The method of preliminary information processing can just conform to standards of the description of thesauruses [9-10], for example, article of the thesaurus of a situation:

Descriptor : CS 1

BT: previous CS

NT: post CS

RT: related factors Cs(i), Cs(i+1)…Cs(n), HF, TF

KW: kw1, kw2, … , kwK

Here **CS** is reduction from "critical situation", the term descriptor; The **BT** - the term is "higher", wider

concept, here means the previous situation, according to **NT** - the term is "lower", the affiliated term, the subsequent situation means here; **RT** is the associative term, i.e. the similar description, a similar situation, terms, descriptions of properties (for example, transfer "technical" (technical factors, **TF**) and "human" factors (human factors, **HF**); **KW** is the list of keywords. The result of the description of a situation has to be expressed in formula **CS = F(TF,HF)** where the description of a situation comes down to function of dependence on technical and human factors. The perfect management in such description consists in assessment of probability of influence of negative factors of:

P (min F (TF,HF)=0).

Such dependences need to be calculated for all situations from the thesaurus of situations (the microthesaurus of subject domain or a situation).

Maintenance by graphics, with use of communications, the situations set in the description, will allow to construct the corresponding cognitive models of [11] events in the form of cognitive maps, event counts and schemes.

If, according to the description, **CS** set in article, at receipt of information in a system there is a transition from **BT** to **NT**, then "the system knows" what occurred, otherwise specification through **RT** is required. If there is also no compliance through associative communications, then, so that or information is incorrect, or it is a new situation which demands new management.

The description of a situation reflected in ontology according to standards of languages of ontologies, finds objects, their properties and the relations similar to articles of the thesaurus.

At correct information the control is transferred to cognitive agents. At an incorrect situation (from the point of view of available descriptions) it is necessary to analyze properties of objects of management for allocation of groups of the influencing factors, for example, from groups of associative communications and influencing factors (**TF, HF**).

V. CONCLUSION

The option of use of technology of the description of data with the help of controlled lexicon, descriptors and communications of the thesaurus in relation to a problem of management of a situation is offered. The emphasis is placed on a problem of submission of information for the purpose of the best perception by the user in the course of interaction with devices. As such means both ontologies, and cognitive descriptions and cards on the basis of thesauruses can serve. Examples of means of communication in historical development are given. Dissemination of culture, writings and emergence of digital means of representation and information transfer served as the transmission medium of knowledge. "Cultural practice is a key component of human knowledge", Edwin Hatchinz (Edwin Hutchins [12]). Throughout these words it is possible to add that it is necessary to accumulate knowledge to operate situations in the world of digital technologies. The economy of knowledge where the intelligence represents the main resource, gives advantages to economy of the developed countries. Modern information technologies, including the cognitive technologies based on real knowledge of mankind are the real capital in the form of applications in all spheres of human activity.

REFERENCES

[1] Orrantia J., Muñez, D., Matilla L. , Sanchez R., San Romualdo S. and Verschaffel L. Disentangling the Mechanisms of Symbolic Number Processing in Adults' // Mathematics and Arithmetic Achievement. Cogn Sci, 2019. 43: e12711. https://doi.org/10.1111/cogs.12711.

[2] Mooers C. Information retrieval viewed as temporal signaling // Proceedings of the International Congress of Mathematicians. 1950. Vol. 1. P.572–573.

[3] Shiri A., Revie C. Query expansion behaviour with a thesaurus-enhanced search environment: a user-cantered evaluation. // JASIS. 2006. Vol. 57. № 4. P. 462–478.

[4] Gerald E. Flueckiger Control, Information, and Technological Change 1995 Springer Netherlands 155 p. DOI 10.1007/978-94-011-0377-0.

[5] Muromskiy A.A., Tuchkova N.P. Application of thesaurus of symbolic information for the identification of the specific situations // Proceedings of the V-th International workshop "Critical infrastructures: Contingency management, Intelligent, Agent-based, Cloud computing and Cyber security" (IWCI 2018) Advances in Intelligent Systems Research. 2018. T. 158. C. 183-188. https://doi.org/10.2991/iwci-18.2018.32.

[6] Muromskiy A.A., Tuchkova N.P. Ob ontologii adresata v matematicheskoj predmetnoj oblasti // Russian Digital Libraries Journal, 2018. Vol. 21. № 6. P. 506-533 [in Russian]

[7] Iqbal M. Inter-organisational knowledge transfer in social networks, 2016 URL: https://mubashariqbalhci.wordpress.com/

[8] Hutchins E. The Distributed Cognition Perspective on Human Interaction. In Nicholas J. Enfield and Stephen C. Levinson (Eds.) Roots of Human Sociality: Culture, Cognition and Interaction. Berg Publishers: Oxford, UK. P. 375–398, 2006.

[9] Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval URL: http://www.willpowerinfo.co.uk/glossary.htm

[10] ISO 25964 URL: https://en.wikipedia.org/wiki/ISO_25964

[11] Milstead J.L. ASIS Thesaurus of Information Science and Librarianship. Second Edition. 1998. 169 p.

[12] Hutchins E. Culture and inference: A Trobriand case study. Cambridge, MA: Harvard University Press, 1980.