The Problem of Semantic Gaps in the Context of World Models and Sign-Symbolic Communication

Anfisa Chuganskaya[1,2] [0000-0001-9727-6406]

[1] [a]Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333 Moscow, Russia
[2a]MIPT, 141701 Institutsky Lane, 9, Dolgoprudny, Moscow Region, Russia
anfisa.makh@gmail.com

Abstract. The current development of artificial intelligence technologies involves many aspects of incorporating tools for analyzing and generating content related to social phenomena and communication. To develop a model that takes into ac-count the significant role of semantic gaps in the study of psychological phenomena and the formalization of social phenomena, we turned to the consideration of this problem in computer science, linguistics, psychology, and applied researches. The task of overcoming the semantic gap has a history within the field of humanities. It does not have a complete solution, which is related to the problem of formalizing knowledge. However, the theories of gaps and se-mantic processing of perceived images in psychology provide directions for building new models that include the development of tools for overcoming the semantic gap. The difficulty to explain the meaning in analyzing of social interactions is connected with the problem of the formation of the volume of the analyzed details of the picture and the need to correlate with the existing experience of the person. In machine learning such visual analyses and constructing relevant prediction of images with social actions demand including semantic level. One of the models is based on the Vector Semiotic Architecture. Another model is a Sign-Based World Model cognitive architecture which make an attempt to address the symbol grounding problem. This model is a complex architecture that takes into account both the level of building images of the objective world based on perception, and the complex images of the world based on reasoning. Sign-symbolic communication as a systemic way of exchanging information with a high degree of semantic depth and the presence of diverse semantic gaps allows us to examine the components of the formation of such semantic gaps.

Keywords: semantic gap, world model, lexical lacunae, activities, sign, artificial intelligence.

1 Introduction

The current development of artificial intelligence technologies involves many aspects of incorporating tools for analyzing and generating content related to social phenomena, communication, and human-robot interaction. One of the challenging tasks is the development of tools and models for analyzing social interaction and communication,

taking into account the psychological context. This task defines the direction of studying the semantic content of information, which includes the analysis of not only language and texts, but also images and videos. To develop a model that takes into account the significant role of semantic gaps in the study of psychological phenomena and the formalization of social phenomena, we turned to the consideration of this problem in computer science, linguistics, psychology, and applied research.

2 The problem of semantic gaps in computer science and linguistics

2.1 Semantic gaps in computer science

The problem of data formalization arose in the first half of the twentieth century during the development of the Turing machine and the formulation of Gödel's in-completeness theorem, indicating the presence of limitations in deductive operations and rule-based inferences [4]. This raised the problem of the unprovability of the truth of certain statements in formal systems. Further research in the field of modeling and recognition of complex social phenomena has concretized this fundamental problem in the topic of the semantic gap.

In information technology, semantic gaps are considered as a semantic difference in the description using different sign systems, for example, text and image. A.M. Hein describes the semantic gap as a difference in meaning between constructs formulated in different systems of representations or worldviews [17]. In the system of knowledge formalization, this problem always concerns social and psychological phenomena, because this class of phenomena has fuzzy formulations of description even in natural language and is context-dependent on many factors. In the process of formalization necessary for building algorithms for computational systems, we still simplify complex phenomena such as communication, empathy, aggression, and attitudes to a threshold level of accuracy that allows us to retain the meaning of these phenomena in most cases. A.M. Hein's work provides an example of the problem of semantic gap in engineering rectification [17]. The author notes the following difficulties: expressing user tasks and needs in natural language and formalizing an engineering solution with a truncated version of information; in multi-domain engineering, there is a difference between the functional and physical representation of data. A.M. Hein suggests solving the semantic gap by finding analogies in natural language and approximating the meaning in both data representations; by visualizing; and by increasing the overall understanding of the phenomenon by introducing documentation in a standardized language (such as SysML).

The topic of the semantic gap was addressed in the early 2000s as part of a study of media and the need to process data semantically accurately. In the work of K. Dorai and S.Venkatesh propose to consider the problem of the semantic gap as a discrepancy between the diversity, richness of meanings and interpretation by automatic systems that carry out simple, low-level descriptions [10]. The authors propose to consider increasing semantic accuracy through video analysis and a more complete understanding

of information through a mechanism for explaining and interpreting information for the audience. They called their approach the computer aesthetics of media [10].
The problem of searching for images by content became the basis for research by A.W.M. Smeulders and his colleagues in the early 2000s [27]. The authors discuss not only the issue of semantics, but also the specifics of sensory information processing. The researchers introduce parameters of cultural specificity and the significance of categorization when working with images. A.W.M. Smeulders and his colleagues note that users are looking for semantic similarities between textual content and images, which combines different modalities of information into a single system of meaning [27].

2.2 Lexical lacunae in linguistics

In linguistics, the problem of semantic gaps has been developed in the context of studying different languages, the problems of semantically adequate translation, and the direction related to lexical gaps.

A lexical lacuna is understood as the absence of a name for a particular concept in one language or the presence of a semantic difference in the scope of concepts in different languages [32]. There are several types of such lacunas:

• at the word level (the absence of a word in another language),

• at the level of meaning (an incomplete equivalent of a concept in another language, with partial overlap of meaning),

• at the level of connotation (a mismatch between the associative field and the absence of a phenomenon in another culture) [1].

Linguistic features that are related to the national worldview have also been referred to as lacunae in linguistics. Modern researchers describe lacunae as national-specific elements of culture that are reflected in language and speech. Lexical lacunae reflect the absence of concepts or structures in one language that are present in another language [29].

When comparing translations of literary works, linguists noticed that there is inevitably a loss of some semantic content due to the discrepancy between the lexical and semantic systems [18].

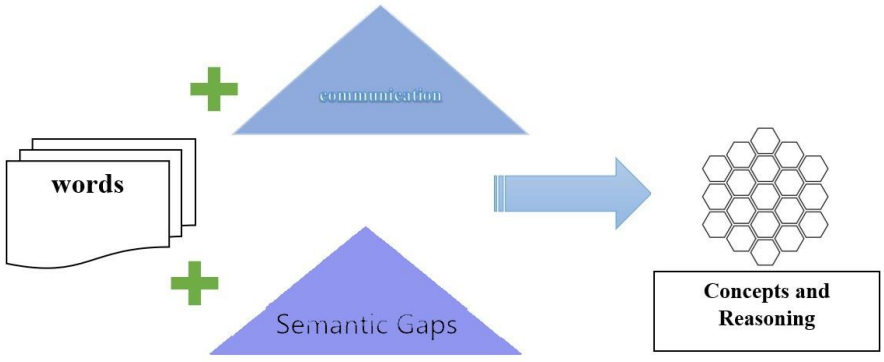

Fig. 1. The place of semantic gaps in linguistics.

Thus, the very existence of lexical lacuna is inextricably linked to the process of comparison with concepts and ideas, and therefore to the process of reasoning. Lexical objectification in relation to a concept is not mandatory and depends on communicative factors. The existence of a concept is not related to the existence of linguistic units, as concepts arise as a result of the reflection of reality by the mind. The linguistic representation of any concept (idea, gestalt, or concept) depends on the communicative needs of society and is derived from these needs [28]. The language system nominates what becomes a subject of discussion in society, but not everything that becomes a subject of thought.

3 Methods for solving problems with semantic information in psychodiagnostics and computer science in analyzing social interaction

3.1 Computer science investigations in analyzing social interaction and common-sense reasoning

One of the tasks in the development of cognition methods in artificial intelligence is connected with the problem of formalization of the area that has been called "common sense" [6]. Common sense is a set of fundamental categories that uses even a seven-year-old child [8]. Davis designates areas of specific tasks in which it will be im-possible to advance modern automated methods without solving the problem of "common sense". He refers to them natural language processing, visual interpretation and robotics. In this series, a special place is occupied by the problems of reasoning in everyday logic based on visual experience. In fact, such a perceptual experience within the framework of a child's development serves as the basis for building basic categories and serves as the foundation of the everyday picture of the world, where common sense is the leading mechanism of reasoning.

Researchers in the field of artificial intelligence also come to the formulation of such a task in the processing of visual information. Zellers, R., Bisk, Y., Farhadi, A., & Choi, Y. pose this problem as a distinction between object recognition and cognition based on visual information. For these researchers, this was called Visual common sense (Visual Commonsense Reasoning) [33], for which it is not the sensory description of objects that is significant, but the formulation of conclusions based on what they saw and understanding the meaning in the depicted material, actions and goals of people in the framework of communication.

Increasing interest to visual analysis in social media and video surveillance drew attention to the new areas of research [3, 30]. One of the most popular and relevant direction in the field of machine learning is the task of analyzing visual information and drawing conclusions on this basis [16]. Such tasks are investigated in the Visual Question Answering (VQA) [2]. It combines perceptional and linguistic levels. In VQA, an algorithm analysis an image and a dialog which contain questions to that image. As a result, it should output an answer to the given question. This task serves as the starting

point for more complicated problems concerning with social perception and linguistic analysis of scenario [5]. There are two levels of image analysis: perceptual and semantic [8]. Firstly, in psychology, the analysis of visual information to determine objects, their location and the construction of a perceived image refer to the field of perception (primarily visual) [11]. Such a task is reduced to the classification of objects and the analysis which is based on correlation with a standard (sensory standard) [31]. For example, the correlation of images of cars from real life from CCTV cameras with the reference class "car". It is possible both to identify significant features of objects of the class ("car" - wheels, car body type, doors, etc.) and set directly them during training, or using machine learning to generate image recognition by experimenting with the reference database. Within the framework of perception, psychologists identify up to 30 signs that can describe images and their mutual location [14]. The task to identify the objects within the framework of perception is now solvable and allows classifiable images in machine learning to be distinguished well enough [5]. But for the class of social actions is needed a different model of recognition and construction of learning methods.

3.2 Methods for solving problems with semantic information in psychodiagnostics

A psychological direction in which visual interpretation was used as a way to study personal processes was projective techniques [12, 13]. The method of Thematic Apperceptive Test (TAT) became a classic technique in which ambiguous pictures were given with a task for the subject to explain them or make a detailed story based on them [13, 22]. This technique was first proposed by American scientists K. Morgan and G. Murray in 1935 [22]. The TAT consists of 31 black-and-white paintings with varying degrees of uncertainty of both the objects themselves and the meaning of the situation. Each of the paintings in the interpretation has a description of the image and typical themes that manifest themselves in the stories. The task that is given to the subject is to compose a story for each of the paintings, in which a description of the situation and a specific moment would be given, what happens to the characters, what emotions they experience and have ideas, thoughts [19]. In fact, in such a methodology, an ambiguous semantic task is given, in which the subject completes the solution based on personal experience. The essential principle that is interpreted in this technique and is important for visual interpretation is the identification of the narrator with one of the actors. At the same time, the semantic content may vary from the disposition, the author's point of view. This is important when building automated systems for visual interpretation of such complex paintings, because sets the task of multi-subjectivity of positions and expansion of the semantic diversity of descriptions, depending on the name of which character, through the eyes of which actor the description is given.

In the study of the understanding of plot paintings, Russian psychologist Luria identifies three stages: the process of analyzing the plot of the painting, the selection of the main essential elements and their synthesis to identify the semantic content of the image [21]. A.R.Luria offered the subjects complex paintings in which the meaning can be understood through the synthesis of significant details and emotional content. Thus, one of the main psychodiagnostics methods of interpretation of one-act plot paintings is the

"Broken Window" [13, 26]. The task of the subject is to explain what is happening in the picture. The first analytical stage involves the study of the depicted objects (a boy, a briefcase, a man, a second boy, a broken window, a woman, a tree, etc.). The success of solving the problem will depend on the completeness of visual analysis. If you do not highlight such a detail as a snowball in the hand of the second boy, then the meaning of the picture becomes not obvious. At this stage, significant details are highlighted, such as a broken window, boys, a man and a snowball. Combining them in the right order gives a semantic solution to the visual interpretation: in the picture, a boy who broke a window hide behind a tree, while an innocent child is denounced by a man.

Fig. 2. The "Broken Window" technique (A.R.Luria) [26].

The difficulty to explain the meaning in analyzing of social interactions is connected with the problem of the formation of the volume of the analyzed details of the picture and the need to correlate with the existing experience of the person, i.e., those given semantic relations that he learned in communication with other people.

4 The World Model and the Activity approach in Psychology as a Basis for Overcoming Semantic Gaps

A qualitative leap in cognitive function, which is possible due to the use of a symbolic and semiotic system, becomes a problem point when trying to algorithmize such processes in artificial intelligence. If we turn to the problems of visual interpretation, then technical solutions based on the logical generalization of visual stimuli rest on the one-

dimensional consideration of the processes of human perception. Taking into account and logically constructing an arbitrarily large number of stimuli does not lead to a qualitative transition to the semantic level of visual interpretation.

The solution of human communication tasks with semantic gaps within the framework of robotic systems can be considered within the framework of the system approach of the sign-based world model [24]. There are three types of the world model: emotional, behavioral (scenario) world, sign-based (or semantic).

The emotional world model is based on working with meanings and emotions. The derivation of knowledge is based on the laws of emotions and feelings, which were pointed out for society by L. Levi-Brul [20]. Communication within this direction meets the tasks of maintaining emotional balance, reducing anxiety, encountering situations of stress and uncertainty. It is focused on maintaining conversation, recognizing and expressing emotions, broadcasting ideas that act as axioms and reduce human anxiety.

The tasks of the behavior, maintaining communication are significant in the scenario world model. Many human-robot interaction systems are aimed at intuitive, everyday knowledge based on experience [6]. In general, many areas of common-sense studies (VCS) and tasks within the volume investigated it. Reliance on common sense and intuitiveness of knowledge are characteristic of the behavior world model.

A sign-based world model is based on logical inference and the use of conceptual thinking, the process of rational reasoning [25]. The semantic-symbolic function of consciousness develops during this process and translates human consciousness into a special dimension of perception of the surrounding reality in a system of meanings. This step methodologically allows us to approach the construction of a rational picture of the world based on the construction of abstract connections between concepts and the use of not only a field of values, but also a holistic association within the framework of human-computer interaction tasks with meanings and images [24].

One of the views on the active formation of perception images and the worldview was proposed in Russian psychology by A.N. Leontiev, who constructed the methodological approach based on activity. The activity theory provides an understanding of the formation of semantic gaps and their structure in the context of the three components of the sign-based world model. In his theory of the world image, A. N. Leontiev specifies the causal and functional relationships between the objective world, the perceptual operations, actions of the person, and the subjective image of the world [19]. He emphasizes that the initial mental functions are external and are formed through external actions that establish significant connections between the subject and the object. The basic opposition that underlies the analysis and explanation of mental phenomena is the opposition between the mental image and the process of its construction. According to activity theory, when analyzing and explaining the processes of perception, we should move from the objective world of objects to the image of the world and the way it is generated [19, 31]. The objective world is inherently modeless. Only for the person the world appears in various modalities. All properties of the objective world revealed in interactions between objects. In relation to organisms, the properties of the objective world appear through interactions with the receptors. These interactions are the processes of direct perception. At the same time, properties that appear in interactions between objects and do not have a direct effect on the senses are indirectly

established by the subject and constitute the processes of thinking [10,19]. It is the second type of building images of the world picture through internal comparison and thinking that provides the basis for semantic gap. Tasks of simple perception are based on the objective properties of objects and the accuracy of their definition can be adjusted.

5 The Sign-Symbolic Communication

The use of images and videos allows us to fill the semantic gaps in linguistic information, which is important for solving more complex tasks. The problem of semantic gaps is discussed in humanitarian works in the context of sign-symbolic communication. Sign-symbolic communication is an exchange of sign-symbolic systems that reflect culturally conditioned values, meanings, and relationships as a worldview with unifying, identifying, resource, and protective functions. The structure of sign-symbolic communication includes five components: a sign-symbolic source, a sign-symbolic content, a sign-symbolic medium, a sign-symbolic pragmatics, and a sign-symbolic communication product. However, in modern researches, there are developments that methodologically combine different forms of the sign-symbolic model with the use of AI and human interaction.

Cultural competence becomes a specific case of tuning this semiotic system, where the robot employs probabilistic hypotheses about cultural context to refine both interpretation and message generation [7]. This architecture, uniting late fusion for perception and a semiotic model for expression, establishes a unified abstract layer of communication, agnostic to specific modalities, which is key to developing flexible and intuitive collaborative systems.

Many human-robot interaction systems are aimed to intuitive, everyday knowledge based on experience. In general, many areas of common-sense studies [33] and tasks within the volume investigated it. Davis notes that the problem of everyday reasoning and its automation is central to modern research in the field of artificial intelligence.

One of the tasks of using sign-symbolic communication in the interaction between artificial intelligence and humans is to algorithmize the semantic space (S), taking into account the complexity of semantic processes and the presence of semantic gaps in natural language. The development of semantic space modeling (S) for natural language has led to the emergence of approaches such as latent semantic analysis (LSA), latent Dirichlet allocation (LDA), Word2Vec, GloVe and transformers (BERT and others) [34]. These approaches represent the semantic space (S) as a mathematical model in the form of a system of vectors in a multidimensional space. In the researches of deep semantic structures and individual semantic assessment, this direction has been developed by the tool of semantic differentiation, introduced by Ch.Osgood. He used to evaluate the semantic field (f) in the context of the study of individual semantic categorization of various objects [23]. If we refer to the formalization of sign-symbolic communication in the context of semantic space (S) and field (f), we can present the following formula:

$$S=\Sigma((f) +(g) + (G)),$$

where the general semantic space (S) is represented by a set of semantic fields (f) and two types of semantic gaps. The first type, g, is a semantic gap within a semantic field (f) between categories that are similar in meaning, which depends on contextual factors and is largely culturally determined. The second type, G, is a semantic gap between semantic fields (f) that is caused by the difference in the category system itself.

Conclusion

The task of overcoming the semantic gap has a history within the field of humanities and cognitive sciences. It does not have a complete solution, which is related to the problem of formalizing knowledge and thoughts. However, the theories of gaps and semantic processing of perceived images in psychology provide directions for building new models that include the development of tools for overcoming the semantic gap.

The difficulty to explain the meaning in analyzing of social interactions is connected with the problem of the formation of the volume of the analyzed details of the picture and the need to correlate with the existing experience of the person, i.e., those given semantic relations that he learned in communication with other people.

In machine learning such visual analyses and constructing relevant prediction of images with social actions demand including semantic level [15, 31]. One of the models is based on the Vector Semiotic Architecture. Another model is a Sign-Based World Model cognitive architecture which make an attempt to address the symbol grounding problem [24, 25]. This model is a complex architecture that takes into account both simple solutions at the level of building images of the objective world based on perception, and complex images of the world based on reasoning. Sign-symbolic communication as a systemic way of exchanging information with a high degree of semantic depth and the presence of diverse semantic gaps allows us to examine the components of the formation of such semantic gaps. This creates the basis for modeling scenarios of interaction between an intelligent agent and a user.

References

1. Ahmad, A. O., Baluyan, S. R.: Comparative analysis of lexical interlingual lacunae in Russian and English // Philological Sciences. Issues of Theory and Practice, 3-2 (69), 60-62 (2017).
2. Akhutina, T.: L.S. Vygotsky: cultural-historical and Natural-scientific approaches to internalization. Bulletin of the Moscow University. Series 14. Psychology, 3, 41-57 (2004).
3. Antol S., A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh. VQA: visual question answering. CoRR, abs/1505.00468, (2015).
4. Berto, F.: The Gödel Paradox and Wittgenstein's Reasons. Philosophia Mathematica, 17(2), 208–219. doi:10.1093/philmat/nkp001 (2009).
5. Besold, T.R., Ku¨hnberger, K.U.: Towards integrated neural–symbolic systems for human-level AI: Two research programs helping to bridge the gaps. Biologically Inspired Cognitive Architectures 14, 97 – 110 (2015).

6. Breazeal, C.: Social interactions in HRI: the robot view // IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 34, no 2, 181–186 (2004).
7. Bruno B., R. Menicatti, C. Recchiuto, E. Lagrue, A. K. Pandey, and A. Sgorbissa: Culturally-competent human-robot verbal interaction, 06, 2018, pp. 388–395.
8. Davis, E.: Logical Formalizations of Commonsense Reasoning: A Survey. Journal of Artificial Intelligence Research 59, 651-723, (2017).
9. Dennett, D.C.: Beliefs about beliefs. Behavioral and Brain Sciences, 4, 568-570 (1978).
10. Dorai, C., Venkatesh, S.: Bridging the Semantic Gap with Computational Media Aesthetics". In IEEE MultiMedia. 10 (2): 15–17 (2003). doi:10.1109/MMUL.2003.1195157
11. Gibson, J.: The Perception of the Visual World.: Houghton Mifflin, Boston (1950).
12. Gille, R.: Le test-film; un instrument pour la mesure objective du niveau de maturité affective et de certains traits du comportement. Presses universitaires de France, Paris (1959).
13. Glozman, Zh., Potanina, A., Soboleva, A.: Neuropsychological diagnostics in preschool age. St. Petersburg, Peter Publishing House (2006).
14. Grabner, H., J. Gall, and L. Van Gool. What makes a chair a chair? In CVPR (2011).
15. Gregory, R. L.: The intelligent eye, Wiedenfeld and Nicolson London (1970).
16. Gupta, A., L. S. Davis: Beyond nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers. In ECCV (2008).
17. Hein, A. M.: Identification and Bridging of Semantic Gaps in the Context of Multi-Domain Engineering. In Conference: 2010 Forum on Philosophy, Engineering & Technology. Colorado, https://www.researchgate.net/publication/282574480_Identification_and_Bridging_of_Semantic_Gaps_in_the_Context_of_Multi-Domain_Engineering, last accessed 2026/01/01.
18. Kuznetsov, V.G.: Lexical lacunas and the limits of translatability. // Bulletin of the Moscow State Linguistic University. Humanities, 9 (781), 50-59 (2017).
19. Leontiev, A.N.: Activity. Consciousness. Personality. 2nd ed. M.: Politizdat, (1977).
20. Lévy-Bruhl L. How Natives Think. New York: Washington Square Press. (1966).
21. Luria, A.R.: Higher human cortical functions and their lesions in local brain lesions. Moscow, Moscow University Press (1962).
22. Morgan, W.: Origin and History of the Earliest Thematic Apperception Test Pictures. Journal of Personality Assessment 79:3, 422-445 (2002). DOI: 10.1207/S15327752JPA7903_03
23. Osgood, C.E.: Method and Theory in Experimental Psychology, Oxford, 1956.
24. Osipov, G.S., Chudova, N.V., Panov, A.I., Kuznetsova, Yu.M.: The Sign Picture of the World of the Subject of Behavior. Moscow: FIZMATLIT (2018).
25. Osipov, G.S.: Signs-based vs. symbolic models. In: Sidorov, G., Galicia-Haro, S.N. (eds.) Advances in Artificial Intelligence and Soft Computing. Springer International Publishing, Cham, pp. 3-11 (2015).
26. Rubinstein, S.: Experimental methods of pathopsychology and the experience of their application in the clinic. Moscow, Meditsina Publishing House (1970).
27. Smeulders, A. W. M. et al.: Content-Based Image Retrieval at the End of the Early Years. In IEEE Trans Pattern Anal Mach Intell. 22 (12): 1349–80 (2000). doi:10.1109/34.895972. S2CID 2827898.
28. Sternin, I.A., Bykova, G.V.: Concepts and Lacunas // Methods of Studying Linguistic Consciousness. Moscow: Institute of Linguistics of the Russian Academy of Sciences, 55-67 (1998).
29. Titkova, S.I.: Language lacuna in the practice of teaching Russian as a foreign language (based on the experience of teaching Russian to English-speaking students) // Russian Language Abroad, 3, 39-50 (2007).

30. Vedantam, R., Lin, X., Batra, T., Zitnick, C.L., & Parikh, D. (2015). Learning Common Sense through Visual Abstraction. 2015 IEEE International Conference on Computer Vision (ICCV), 2542-2550.
31. Velichkovsky, B.M.: Cognitive science: Fundamentals of the psychology of cognition. In 2 volumes [in Russian]. Smysl/Akademiya, Moscow (2006).
32. Vinay, J.-P., Darbelnet, J.: Stylistique comparée du français et de l'anglais. Montréal: Beauchemin (1958).
33. Zellers, R., Bisk, Y., Farhadi, A., & Choi, Y.: From Recognition to Cognition: Visual Commonsense Reasoning. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6713-6724 (2019).
34. Zhaxybayev D.O., Mizamova G.N. Natural Language Processing Algorithms for Understanding the Semantics of Text. Trudy ISP RAN/Proc. ISP RAS, vol. 34, issue 1, 2022, pp. 141-150 (in Russian). DOI: 10.15514/ISPRAS–2022–34(1)–10.