Graph Visualization of the Terminological Thesaurus by F.M. Dostoevsky's Gospel Text (using the Example of Bigram Models)

V.V. Borisova[1]

M. Akmullah Bashkir State Pedagogical University, Moscow State Linguistic University, V. I. Dahl State Museum of the History of Russian Literature (Museum Center "Moscow House of Dostoevsky"), Ufa, Russia

[1] ORCID: 0000-0002-9011-0160, vvb1604@gmail.com

Abstract

Methods and means of scientific visualization through graphs are actively used today in many areas, including the humanities. The effectiveness of their use is also evident in philological research. The article describes the graph model of the terminological thesaurus of the Gospel text by F.M. Dostoevsky, based on the clustering of digrams, taking into account their thematic belonging to different discourses. The identification and analysis of digrams as conceptual points of a terminological thesaurus is important for its network visualization, thanks to which it is clear in which conceptual conjugation key terms "work", how they are distributed among certain clusters. In one cluster there can be terms from different discourses. This is an important indicator of how the terms are conjugated, how integral in this case is the metalanguage of the studies of the Gospel text by F.M. Dostoevsky. Thus, the terminological thesaurus performs in this case a diagnostic and prognostic role.

Keywords: visualization of scientific data, graph model, digrams, terminological thesaurus, gospel text F.M. Dostoevsky.

1. Introduction

Various forms of scientific data computer visualization are widely implemented in modern interdisciplinary theoretical and applied research. This data machine analysis is necessarily coupled with the so-called "human interpretation". It is also significant that a new concept of "visualistics" has been established as a humanitarian knowledge sphere [1, 2].

Graph models are used as universal means of visualizing scientific information [3, 4]. Today they are actively used in many spheres including in the humanities. With such a network visualization, it is clearly visible in which conceptual conjugation, for example, bigram terms "work" as the most stable word combinations, how they are distributed among certain thematic clusters. The effectiveness of their use is also evident in philological research [5, 6, 7, 8, 9, 10]. So, B.V. Orekhov paid attention to the methods of solving some problems partially related to the literary terminology analysis. Among them, he highlighted the creation of graphs as an effective way of modeling of terminology [11].

In our view the use of such tools is advisable, for example, in the description of a thesaurus created on the basis of digital processing of the most complete corpus of texts. This allows to build a terminological dictionary network model in the form of a graphs systems considering the joint occurrence of basic terms. Due to this basis, it is possible to describe the thesaurus in its basic concepts and identify their overgrowth with important definitions that reveal the features of word usage.

The specificity of our research is due to the interdiscursive nature of terminology that is used today in the study of the "gospel text" by F.M. Dostoevsky (meaning the reflection of images, plots, motives of the Gospel in the texts of the writer). This is the priority subject of a

new scientific direction – ethnopoetics, which focuses on national and confessional features of Russian literature [12, 13].

The most important scientific task of ethnopoetics is to create a conceptual apparatus that consistently reflects the connections of national artistic culture and literature with the Orthodox tradition. It is necessary to "make the philological tools, applied <also to Dostoevsky. – V.B.>, more consonant with the ontological and axiological values of the writer himself" [14, p. 7]. In fact, we are talking about the Russian literature metalanguage. It is based on the thesaurus, which is structured on the value basis of the terminologically formed knowledge complex.

In its current form, the terminological dictionary of the "gospel text" includes groups of basic terms and concepts. They are increasingly used in mutual conjugation, referring to the spheres of general (theoretical), historical poetics and ethnopoetics, theology and philosophy. The accumulated experience allows us to make a full-scale explication of the thesaurus functioning in the studies of the gospel text of the author of the Brothers Karamazov. It is possible to give an up-to-date description of key concepts and terms used in Dostoevsky studies, to conduct their comparative analysis.

This is necessary for a new systematization of theoretical knowledge about Dostoevsky, a certain "revision" of his categorical apparatus for compliance with the subject of research. This refers to the adequacy of the terminological matrix of the text being studied. In this case, the gospel one. The terminological thesaurus analysis, based on its graph model, is intended to contribute to the solution of this problem.

2. Research sources and methods

As initial data, we used a corpus of scientific texts in the journal "Problems of Historical Poetics", from 1990 to 2021, (there are 333 articles mentioning the name of the writer, the number of words he uses is 1.4 million)[1]. The number of texts presented in this edition is quite representative of the gospel text of Dostoevsky and Russian classical literature at large. By means of his thesaurus, a national picture of the world opens. It is focused on the main concepts and constants of the Christian tradition.

We used the methods of automatic processing of source texts by the Python programming language to clearly show the correlation between terms in a form that is accessible for analysis and interpretation. They were cleared of irrelevant noise (links, numbers, line breaks, etc.) and automatically brought to the initial form using the console lemmatizer Mystem [15]. Service words (conjunctions, prepositions) were also removed from the texts. At the next stage frequency lists of words of the text were compiled.

The terms were selected and distributed in four discourses (different subject areas): general (theoretical) poetics; historical poetics and ethnopoetics; theology; philosophical and other discourses. To compile the dictionary, we selected the most frequent terms-unigrams (a sequence of one word, for example, "gospel") and terms-bigrams (sequences of two words, "gospel text")[2] from the entire list.

An example of the analysis of the graph model of unigrams is presented in the previous work [16]. In this article, we focus on the bigram visualization as the most stable terminological combinations fixed by the tradition of scientific word usage. When constructing graphs from unigrams, we used a limit of 50, 40, 30 and 25 joint word usage of two node terms in windows 3 and 5 (joint use of terms at a distance of up to 3 and 5 words from each other, respectively). In the case of bigrams, their minimum frequency is reduced to 5 and 3 joint mentions in window 5. We used the Gephi program to visualize the graph [17].

---

1 Archives of articles are collected and calculations are made by D. D. Buchneva (PetrSU).

2 The full list of terms is available in the repository of our project, prepared by A.V. Panasyuk (HSE): https://thesaurus-dostoevsky.github.io/slovniki She was also engaged in the creation of graph models presented below.

Gephi has a built-in Louvian modularity maximization algorithm [18], which allocates communities automatically, without a predefined number of communities, so we chose it as suitable for the task of visualizing the distribution of terms. The graph was laid by the Yifan Hu algorithm, optimal for our data amount (95 and 176 ribs in graphs with a joint occurrence constraint of 5 and 3, respectively) [19].

The coefficients of the node degree (the number of ribs coming from the node) and the degree of node mediation (the coefficient showing how important this vertex is for the integrity of the graph) were also calculated. The size of the nodes also depended on it.

The key terms ("vertices" or "nodes") appearing in graphs as central conceptual points have a conceptual meaning in the studies of Dostoevsky's "gospel text". The "ribs" connecting the "nodes" with lines clearly reflect the joint occurrence of these terms in the text. The more of them, the stronger the bond and the thicker the "ribs". It is clearly visible in which conceptual conjugation, for example, bigram terms "work" as the most stable word combinations with such network visualization, how they are distributed among certain thematic clusters. There may be concepts from different discourses in the same cluster. This is an important indicator of their interface. It shows how integral is the metalanguage of studies of F.M. Dostoevsky's gospel text in this case. That is, the terminological thesaurus performs a diagnostic and predictive role.

3. Analysis of the terminological dictionary

The analysis of the bigrams sorted by frequency made it possible to identify stable terminological phrases in the studies of Dostoevsky's gospel text, fixed by both literary and religious-philosophical traditions. We publish fifteen bigrams, the most revealing for each discourse.

General (theoretical) poetics	Historical poetics and ethnopoetics	Theology	Philosophical and other discourses
Russian literature 1470 the main character 256 Russian philology 233 artistic world 159 Dostoevsky 's poetics 119 Old Russian literature 100 Russian classics 97 literary text 92 Russian novel 80 classical literature79 the picture of the world 68 Russian poetry 68 literary genre 68	gospel text 301 Christian tradition 147 Christmas story134 Christian realism 94 quote/ reminiscence 91 motive / plot 87 Orthodox tradition 83 plot / genre 82 Russian Orthodox80 reminiscence / motive 79 Easter story 72 canonical text 71 Easter archetype 70 category of conciliarism 64	Prodigal Son 156 Jesus Christ 130 Holy Scripture 119 The Image of Christ 99 Orthodox Church 81 Old Testament 78 Mother of God 77 Holy Father 72 Great Lent 56 The Gospel Word56 The Kingdom of God 53 The Holy Spirit 53 The Gospel Parable 50 The Last Judgment48 The Word of God 47	Russian people 226 Russian man 204 religious and philosophical 96 Russian culture 332 Christian culture 74 cultural tradition 62 Orthodox tradition 83 spiritual life 68 social Christianity 55 Slavic culture 55 Russian world 51 Russian religious 44 Orthodox culture 43 Russian thought 43 Russian land 43 Christian idea 39

The final version of the dictionary included more than 200 bigrams. From a thematic point of view the analysis of the dictionary revealed several indicative features of its structure and content. Firstly, the synonymous correlation of the concepts used is obvious, indicating

the expansion of the terminological field, the addition of literary tools with categories from other discourses.

Thus, the group of key terms denoting the subject of ethnopoetics, along with its traditional formulations, includes new ones due to the axiology and methodology of the scientific direction in which Dostoevsky's gospel text is studied. Researchers actively use, for example, the terminological phrases "Russian philology", "domestic philology", "Christian culture", "Orthodox culture" alongside with the bigrams "Russian folklore", "Russian culture", "Slavic culture", "national culture", "folk culture".

The transition from the concept of "Russian literature" to the concept of "Russian philology" has a fundamental conceptual significance. It's similar to the transition from the definition of "Old Russian literature" to the terminological phrase "Old Russian philology" [20, 21]. The term "philology" is broader than the concept of "literature". According to the Gospel of John philology is the basis of life: "In the beginning was the Word...".

Statistical data of terminological word usage also indicate a predominant research attention to the classical period of Russian literature (this is revealed by the bigrams "classical literature", "Russian classics"). The connection with the Orthodox tradition is most characteristic of this research. The "foreign Russian literature " is marked by no less ethnoconfessional originality. Such bigrams as "Russian poetry", "Russian prose", "Russian novel" determine the subject of the fiction study. The definition of "Russian" is emphasized in the framework of the study of the "literary process" and the "creative path" of the writer.

The system of concepts looks similar in the terminological dictionary. The correlated categories "artistic world", "world picture", "world model", "inner world" of the writing are of prime importance. New bigrams, denoting an axiological turn in the study of Russian literature, are synonymous with them: "Russian world", "Orthodox world", "spiritual world", "God's world", "Christian world". The traditional attributes of the "image of the world" are still "artistic time" and "artistic space", the "manor topos" is also actively studied. The characteristic toponyms "Russian land", "holy place", "holy land", "Zion height", etc. are used.

The thematic group, which mainly consists of the "literary genre" and, in particular, the "Dostoevsky genre" is no less representative. Along with the concept of "Russian novel", they are accompanied by such genre formulas as "Yuletide story", "Easter story", "Christmas story", "Gospel parable", "Spiritual verse", "Epistle of the Apostle", "Lives of the Saint", "Sermon on the Mount", "Revelation of John". This also indicates the expansion of the terminological apparatus in the studies of the gospel text in Russian literature in general and in the works of Dostoevsky in particular.

This trend is also evident in the system of terms associated with the basic categories "poetic image", "national image", which is actively supplemented by the theological concepts "Christian image", "biblical image", "gospel image", "image of Christ", "Jesus Christ", "Russian Christ", "Mother of God", "Holy Mother of God", the periphrases "son of man", "image of God", "son of God", as well as the stable phrases "Holy Father", "Reverend father", "Prodigal son", "Prodigal daughter", "evil spirit", etc.

The fact that the discourse of modern literary criticism is expanding is evidenced by the "nest" of terms. Alongside with the traditional categories of general poetics "artistic text", "poetic text", "literary text", "canonical text", includes new bigrams "gospel text", "biblical text", as well as the concepts of "literary tradition", "cultural tradition", "ancient tradition" are complemented by the phrases "Christian tradition", "Orthodox tradition".

A new understanding of Russian literature in its connections with Christian tradition is also manifested in relation to plotology. Besides the traditional designations "storyline", "plot situation", "test plot", and the terminological pairs "motive / plot", "plot/ genre", plotology terminological apparatus includes bigrams "gospel plot", "the biblical plot", "The Nativity of Christ", "the resurrection of Lazarus", "the Bright Resurrection", "the end of the world", "the transfiguration of man", "salvation of the soul", etc. The thematic group of terms looks similar, so its traditional definitions of "literary quotation", "literary reminiscence" are supplemented with their new variants "gospel quote" and "bible quote".

Thus, even at the analysis and interpretation stage of the visualized data of the terminological dictionary on the material of bigrams, it is possible to draw the following conclusions regarding the current state of the metalanguage of ethnopoetics as a priority direction of Russian literary studies: there is its interdiscursive character, due to the dynamically developing conjugation of poetological, theological and philosophical categories due to genetic and typological connections between them, actualized in a new context of study Russian literature.

4. What graph models of the terminological thesaurus of Dostoevsky 's gospel text show

Let's move on to the analysis and interpretation of graph models of the terminological thesaurus of Dostoevsky's gospel text. Cases of joint occurrence of bigram terms at distances of up to 5word usage from each other (inclusive) were extracted from the article sets for their construction. Then two graphs were constructed. The terms are automatically grouped into clusters (they are highlighted in color). The size of the node in the graph depends on the value of the coefficient of the mediation degree: the more important a term is for a network, the larger its node is. To determine the key nodes in the network model, centrality was calculated according to the degree of mediation. This allows identifying the most important terms that contribute to the interconnection of clusters. The key node (term) in both graphs is the bigram "Russian literature", which has the highest betweenness centrality coefficient.

Figure 1. Graph of terms with a joint occurrence of at least 5, a distance of 5.

Visual analysis of this graph shows which clusters it is divided into. This is, firstly, the central cluster (purple), the key nodes "historical poetics", "Gospel text" are associated with the bigrams "Easter archetype", "Christian realism", "Christian tradition", "category of conciliarity", "national identity", etc. This testifies the interaction of discourses of historical poetics and ethnopoetics in this the case.

The largest number of ribs comes from the node "Russian philology", the main one in the cluster (green). Such stable phrases as "prodigal son", "prodigal daughter", "gospel parable"

appear alongside with the concept of "literary tradition". In turn, the top "gospel text" is quite strongly connected with the paired terms of historical poetics "plot/ genre", "reminiscence / motive", "motive/ plot", "quote/ reminiscence" and with the category of ethnopoetics "gospel quote". There is also a large degree of conjugation of poetical terms.

The bigram terms "philosophical thought" differ in a lesser degree of mediation in this network model. Visualization of the thesaurus also shows that a number of bigrams, in particular philosophical, cultural and theological concepts, are on the periphery of the graph, despite the clustering tendency. This is, for example, a freestanding monothematic cluster (black), which includes the bigrams "Orthodox tradition", "Orthodox church", "Orthodox culture". The cluster with the terms (red) "national culture", "national character", "Russian national", etc. functions in the same way. This indicates its locality.

Due to the discrete nature of the graph, the conjugation of bigram terms related to the discourses of theology, philosophy and ethnopoetics with the categories of general and historical poetics in modern Dostoevsky's studies is not yet complete enough. Only in the central cluster with the main node "Russian literature" there is a tendency to joint word usage of categories related to different discourses. The potential for the development of the terminological tools of Dostoevsky's gospel text is connected with him.
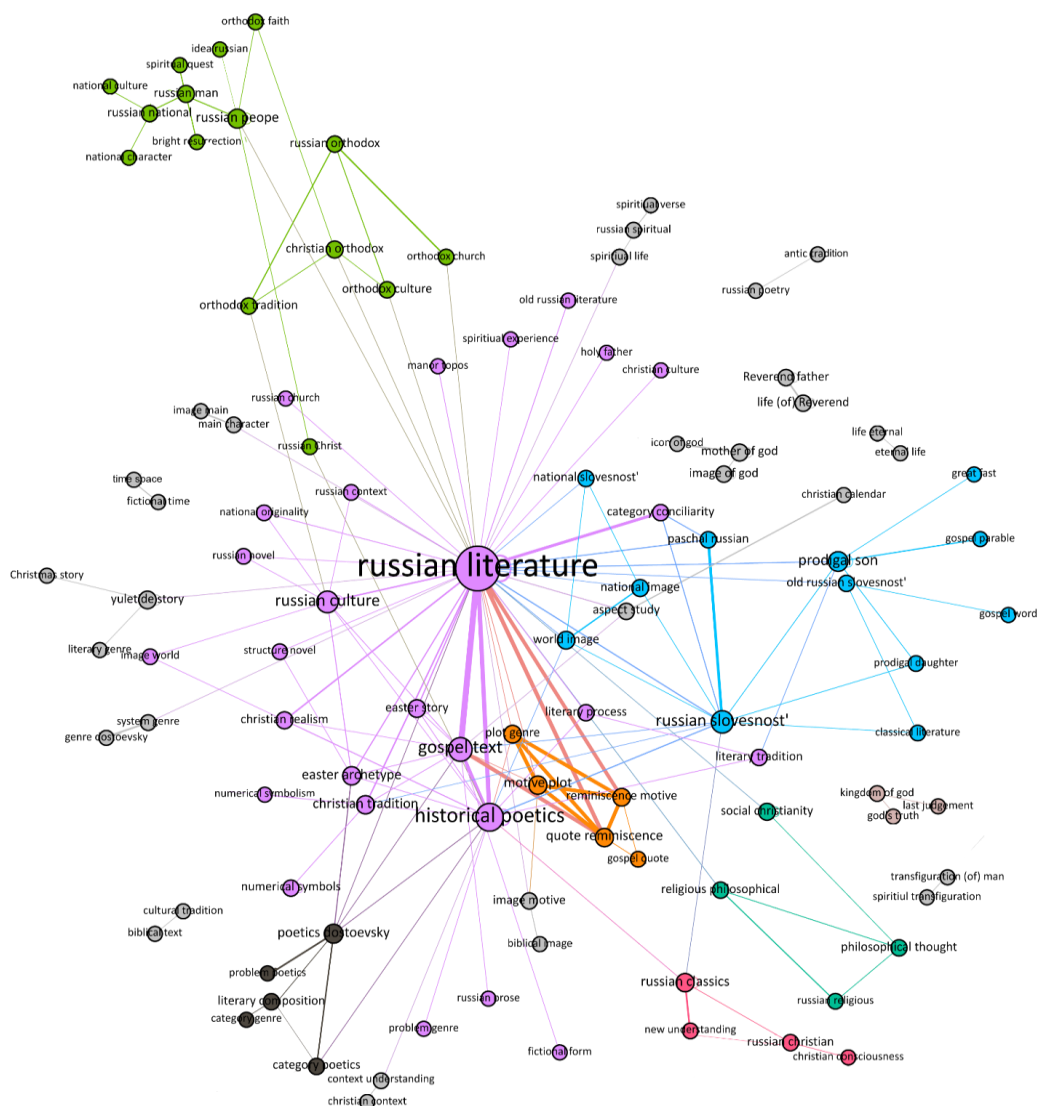

Fig. 2. A graph of terms with a joint occurrence of at least 3, a distance of 5.

In general, this graph looks less discrete, but it is distinguished by predominant mono-clusterization, unrelated components are revealed, which indicates the heterogeneity of the

thesaurus of Dostoevsky's gospel text. Although the largest number of ribs emanates from the main bigram "Russian literature" to all clusters, a number of theological concepts ("human transfiguration", "spiritual transfiguration", "eternal life", "image of God", "kingdom of God", etc.) are still presented separately on the periphery of the graph. It is no coincidence that they have a zero coefficient of mediation with other terms. Since some terms appear outside the main component of the graph, the degree of conjugation of terms, including bigrams related to different discourses, remains low in modern Dostoevsky studies.

Nevertheless, there is an obvious joint use tendency of heterogeneous concepts, of their semantic expansion. For example, when initially theological categories in the literary context acquire new meanings. In particular, the bigram "Prodigal Son" has two meanings: "prodigal son" as a character in a biblical parable and "prodigal son" as a type of literary hero, which is typical for Russian literature, so it indicates the overlap and interpenetration of discourses.

5. Conclusion

Our research allows us to conclude that graphs as a universal means of visual modeling of scientific terminology are quite effective in philological research. Thus, the analysis of the terminological thesaurus of Dostoevsky's gospel text by its graph model shows how integral in this case the metalanguage of a new direction in Russian literary criticism – ethnopoetics.

On the one hand, there is its interdiscursive character, which is dynamically developing through the coupling of pathological, theological and philosophical categories. Due to the genetic and typological connections between them, these categories are actualized in the new context of the study of Russian literature

On the other hand, due to the discrete nature of graph models, is not high enough in modern Dostoevsky studies, the degree of conjugation of bigram terms related to different discourses. However, there is also an obvious joint use tendency of dissimilar concepts, of the interpenetration of discourses.

Thus, graph visualization, the terminological thesaurus of Dostoevsky's gospel text in this case, performs an important diagnostic and prognostic function, which is necessary for the further development of the scientific apparatus in modern studies of Russian literature.

References

1. Drozdova A.V. The specifics of visual research in modern humanitarian knowledge // Yaroslavl Pedagogical Bulletin. 2015. No. 3. S. 254-259.

2. Mochalova K.A., Efremov A.V. Specifics of data visualization in the field of digital humanities // Modern Science: Theory and Practice. Collection of articles based on the materials of the All-Russian scientific-practical conference. Yoshkar-Ola, 2019, pp. 100-110.

3. Kasyanov V. N., Kasyanova E. V. Information visualization based on graph models // Scientific visualization. 2014. V. 6, N 1. C. 31-50.

4. Kasyanov V.N. Visual Graph is a system for visualizing complexly structured information of a large volume based on graph models // Proceedings of the Anniversary 25th International Scientific Conference GRAFICON'2015 (Protvino, September 22–25, 2015). 2015. S. 154-162.

5. Orekhov B.V. Modeling the terminological thesaurus of R. G. Nazirov's works on mythology and the history of folklore plots // Nazirovsky archive. 2015. No. 2 (8). P. 118–131 [Electronic resource]. URL: http://nevmenandr.net/nazirov/journal20152.php (07/10/2022)

6. Orekhov B.V. What can graph theory tell about O.M. Freidenberg? // Vestnik RGGU. Ser.: Literary criticism. Linguistics. Culturology. 2017. No. 4 (25). P. 89–98 [Electronic re-

source]. URL: https://history.rsuh.ru/jour/article/view/292 (03/10/2022). DOI: 10.28995/2686-7249-2017-4-89-98

7. Orekhov B.V. In the network of terms of M. M. Bakhtin: graph theory about dialogue, carnival and chronotope // Dialogue. Carnival. Chronotop. 2018. No. 1. P. 105–115.

8. Pilshchikov I. A., Sysoev M., Boikov V. N., Zakharov V. E. Thesaurus as a tool of poetology // Modeling and analysis of information systems. 2010. V. 17. No. 1. S. 5-24.

9. Pilshchikov I. A., Boikov V. N. Semantic model of "thesaurus in poetology" as part of an information-analytical system // Internet and modern society. Sat. scientific articles: Proceedings of the XVI All-Russian United Conference. St. Petersburg National Research University of Information Technologies, Mechanics and Optics. 2013. S. 273-279.

10. Skorinkin D. A. Extracting character networks to explore literary plot dynamics // Komp'juternaja Lingvistika i Intellektual'nye Tehnologii. - 2017. - S. 257-270.

11. Orekhov B.V. Humanitarian terminology as a network: graph theory about the patterns of scientific style // Criticism and semiotics. 2016. No. 2. P. 94–101.

12. Zakharov V.N. The idea of ethnopoetics in modern research // Problems of historical poetics. 2020. V. 18. No. 3. S. 7–19 [Electronic resource]. URL: https://poetica.pro/files/redaktor_pdf/1593805089.pdf (03/10/2022). DOI: 10.15393/j9.art.2020.8382

13. Borisova V.V. Gospel text in the works of F. M. Dostoevsky: problems and prospects of study // Problems of historical poetics. 2020. V. 18. No. 4. S. 186–208. [Electronic resource]. URL: https://poetica.pro/files/redaktor_pdf/1604078344.pdf

14. Esaulov I. A., Tarasov B. N., Sytina Yu. N. Analysis, interpretation and understanding in the study of Dostoevsky's heritage. M.: Indrik, 2021. 336 p.

15. Segalovich I. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine //MLMTA. - 2003. - S. 273-280.

16. Borisova V.V., Shaulov S.S. Terminological thesaurus of the gospel text by F. M. Dostoevsky // Unknown Dostoevsky. 2022. V. 9. No. 2. S. 117–136. [Electronic resource]. URL: https://unknown-dostoevsky.ru/files/redaktor_pdf/1657603083.pdf

17. Bastian M., Heymann S., Jacomy M. Gephi: An Open-Source Software for Exploring and Manipulating Networks // Abstract of the International AAAI Conference on Weblogs and Social Media. 2009.

18. Blondel V. D., Guillaume J., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks // Journal of Statistical Mechanics: Theory and Experiment (10), p. 1000. 2008.

19. Hu, Y. Efficient and high-quality force-directed graph drawing. // Mathematica Journal. 10.37-71. 2006.

20. Zakharov VN Do we have literature? The concepts of "literature" and "literature" in Russian criticism // Problems of historical poetics. 2016. Issue. 14. S. 7–15 [Electronic resource]. URL: URL: https://poetica.pro/files/redaktor_pdf/1482396216.pdf

21. Motorin A. V. "Literature" and "literature": the spiritual foundations of the Russian and Western European attitude to verbal creativity // Problems of historical poetics. 2008. V. 8. S. 17–36 [Electronic resource]. URL: https://poetica.pro/journal/article.php?id=2590 (03/10/2022).