

## **THE USE OF HIERARCHICAL THEMATIC MODELING METHODS FOR ASSESSING COMPETENCIES IN RECRUITING HIERARCHICAL MODELING**

**E.A. Mashina (ITMO University)**

**Scientific supervisor – Ph.D., P.V. Balakshin  
(ITMO University)**

The paper discusses issues related to the use of hierarchical thematic modeling of texts to assess the competencies of their author during his recruitment.

The purpose of the work is to identify ways to improve the objectivity of assessing the qualifications of personnel of innovative companies by creating methods for assessing their competencies based on the collections of texts generated by them.

Innovative enterprises usually spend many resources on recruiting highly qualified personnel due to the high degree of dynamism of their business in the rapidly developing fields of science and technology [1]. Since innovative research of such companies, as a rule, is multidisciplinary in nature and is conducted in areas with a yet unstated system of concepts [2], most of the competence requirements for job seekers are unique and cannot be provided only by the educational competencies of the specialist being hired. In this connection, the main attention when choosing the necessary candidate is paid to the competencies acquired during the periods of previous work [3].

Considering the fact of verbleness of knowledge and competencies of a specialist, it is proposed to evaluate his competencies acquired during the period of previous production activity using semantic analysis of a collection of texts describing his previous innovative activity using Natural Language Processing, NLP methods, which provide users with a fairly wide set of text filtering methods based, among other things, on machine learning methods. The methodological basis underlying NLP methods is the understanding that texts created in natural language assume a certain occurrence of words in the text.

At the same time, the task of comparing the innovative competencies of a researcher hired in an innovative company with the requirements of the employer can be successfully solved based on the analysis of the correspondence between the collection of texts generated by him and the collection of documents describing his future work [4].

In several cases (for recruiting specialists of mass specialties, for the most part falling under the "young/novice researcher" characteristic), it will be sufficient to determine the degree of compliance with the semantic proximity of texts based on a comparison of vector representations of texts generated by a neural network model trained on a collection of scientific publications of a candidate applying for a vacancy and a collection of texts describing job requirements [5]. At the same time, the proximity of the cosine measure of semantic proximity to the unit will indicate that the candidate's knowledge meets the requirements of the vacancy.

At the same time, it should be borne in mind that the use of such a procedure for comparing the semantic content of collections of texts has a significantly limited application since the user still understands the reasons for highlighting certain topics, as well as their internal structure [6].

The methods of thematic modeling of texts are devoid of such a disadvantage, which makes it possible to ensure the interpretability of the result acceptable for managerial decision-making [7], based on the assumption that each occurrence of a specific term in scientific work is associated with a certain topic, and the appearance of a specific term in a specific document belonging to a specific topic depends only on the topic to which the document relates (and does not depend on the document), which can be described by a single probability distribution [8]. At the same time, the developed hierarchical thematic models allow us to present the structure of the author's collection under study in the form of a topic tree, at the upper levels of which larger topics are located, and more specialized topics are in the leaves [9].

The use of such an approach makes it possible to ensure the interpretability of the result obtained at a sufficiently prominent level in the form of a multidisciplinary tree of employee competencies, which can be used in intelligent decision support systems for recruiting employees.

### **References:**

1. Korkina T.A., Zotova E.N. Zarubezhnyj i otechestvennyj opyt podbora personala// Obshhestvo, jekonomika, upravlenie. 2021. Tom 6, № 4. 58–63. DOI 10.47475/2618-9852-2021-16408.
2. Mashina E.A. Application of statistical methods to solve the problem of enriching ontologies of developing subject areas // Al'manah nauchnyh rabot molodyh uchenyh Universiteta ITMO - 2022, Vol. 1, pp. 301-305
3. Mashina E.A. Preliminary Systematization of Corporate Knowledge Objects for The Use of Prescriptive Analytics Methods When Creating an Innovative Product by Small and Medium-sized Companies // Proceedings of FRUCT'32 - 2022, Vol. 2, pp. 366-373
4. Balakshin PV, Mashina EA. Formalization of implicit knowledge based on educational competencies and background knowledge [In Russian]. *Ontology of designing*. 2022; 12(4): 481–494. DOI:10.18287/2223-9537-2022-12-4-481-494.
5. Howard J, Ruder S. Universal Language Model Fine-tuning for Text Classification // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers), Melbourne, Australia, 2018, P. 328–339
6. Shmatov G.P., Fomina E.E. Neironnye seti i geneticheskii algoritm [Neural networks and the genetic algorithm]. Tver, Tver State Technical University, 2019, 200 p.
7. Blei, David M. (April 2012). Introduction to Probabilistic Topic Models. *Comm. ACM* 55 (4): 77–84
8. Rychagov S. A. Ispol'zovanie latentno-semanticheskogo analiza dlja avtomaticheskoy klassifikacii tekstov // *Mezhdunarodnyj zhurnal informacionnyh tehnologij i jenergojefektivnosti*. 2017. № 2. С. 28–33.
9. Vorontsov K. V., Potapenko A. A. Tutorial on Probabilistic Topic Modeling: Additive Regularization for Stochastic Matrix Factorization // AIST'2014, Analysis of Images, Social networks and Texts. — Springer International Publishing Switzerland, 2014. *Communications in Computer and Information Science (CCIS)*. Vol. 436. pp. 29–46