

**Approaches to enriching collections of texts used to determine the competence of a job applicant**

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

The paper deals with issues related to the enrichment of collections of texts used to determine the competence of a job applicant when analyzing collections of scientific texts of a job applicant used for an objective assessment of a specialist's qualifications during his recruitment. The purpose of the work is to expand the volume of analyzed information used for an objective assessment of the qualifications of personnel of innovative companies.

To date, the sphere of recruiting is one of the most promising areas for the introduction of artificial intelligence in the field of operational management of a modern innovative enterprise [1].

One of the most promising areas for the introduction of such tools in the field of recruiting is the automation of the preliminary selection of candidates for vacancies from an extremely considerable number of candidates who have submitted documents.

This is explained by the fact that due to the high dynamism of the development of innovative companies and sufficiently high wages for their employees, HR departments of such companies must process hundreds of sets of documents of job seekers. However, due to the innovativeness of the ongoing developments, an employee of the HR department cannot have a set of knowledge that allows for a correct assessment of the knowledge and qualifications of the required candidate. In this connection, such work is delegated to the research departments of the company in the form of professional interviews, taking their specialists away from their main activities [2].

Since innovative enterprises, for objective reasons, are in a state of constant recruitment of increasingly new specialists, it is necessary to create automated systems for preliminary assessment of the compliance of candidates for vacancies and the requirements of their future jobs based on the intellectual analysis of information about the job applicant.

In most of the systems used, the texts of candidates' resumes are used as the basis for such intellectual analysis [3], from which information about their education and qualification level is highlighted. However, as numerous tests of such solutions show, such an approach, which works well at classical enterprises, does not apply to innovative enterprises operating in expanding fields of science and technology [4]. This is explained by the fact that for such enterprises it is necessary to consider personal competencies in highly specialized areas of developing areas, which are usually not considered by certification documents.

In this regard, it seems necessary to use collections of documents describing his previous innovative activity as information sources about the level of competence of a specialist, processing them using Natural Language Processing, NLP methods. NLP provides users with a wide set of text filtering methods, including those based on machine learning methods using vector representation of texts. The methodological basis underlying NLP methods is the understanding that texts created in natural language assume a certain occurrence of words in the text [5].

At the same time, it should be borne in mind that obtaining an acceptable result in accuracy can be achieved only when using significant collections of documents generated by a candidate for a vacancy for analysis. And as the real practice of recruiting shows, the standard set of documents provided by the applicant does not have the necessary volume to conduct work on the analysis of the author's texts by NLP methods. In this regard, it is necessary to conduct procedures for

information enrichment of the applicant's package of documents with materials that allow expanding the base for analysis.

This study examines several practical methods for determining the objects-carriers of parameters containing information about the competencies of specialists and the competence requirements of workplaces of enterprises, the choice of methods for separating information from the objects-carriers necessary for comparison, describes the procedures for bringing the objects-carriers of primary data to a form that allows an adequate identification procedure, as well as possible procedures internal and external data enrichment, allowing to significantly expand the list of sources, being analyzed by NLP methods [6].

All the materials listed in the work are based on a series of expert studies of arrays of documentary information used for preliminary assessment of the competencies of qualified specialists of research centers in the shipbuilding industry.

The materials presented in the paper can be used to create automated decision support systems for recruiting personnel for innovative enterprises.

### **References:**

1. Chupina I.P., Zarubina E.V., Simakova N.N. Recruiting technologies in the modern personnel management system// Moscow Economic Journal, No. 9, 2021. pp. 444-453.
2. Korkina T.A., Zotova E.N. Foreign and domestic experience of personnel selection// Society, economy, management. 2021. Volume 6, No. 4. 58-63. DOI 10.47475/2618-9852-2021-16408.
3. Prytkov R.M. Formation of an effective recruitment system based on a systematic approach//Intelligence. Innovation. Investment. No. 9,2017, pp. 49-53.
4. Kalachanov D.V., Efimova N.S., Novikov A.N., Maksimov V.S. Methods of automation of personnel management processes at aircraft manufacturing enterprises in the digital economy// Innovation and investment. No. 11, 2020, pp. 292-296.
5. Giménez M., Palanca J., Botti V. Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis // Neurocomputing. 2020. Vol. 78. P. 315-323.
6. 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.