DOI: 10.17323/2587-814X.2023.2.71.84

An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus*

Ivan E. Nikolaev
E-mail: ivan_nikolaev@csu.ru

Chelyabinsk State University
Address: 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia

Abstract

Online recruitment systems have accumulated a huge amount of data on the real labor market in recent years. Of particular interest to the study are the data on the real requirements of the labor market contained in the texts of online vacancies, as well as the process of extracting and structuring them for further analysis and use. The stage of compiling an up-to-date list of requirements for a position profile in the recruitment process is very time-consuming and requires a large amount of effort from an HR specialist related to monitoring changes in entire industries and professions, as well as analyzing relevance of existing requirements on the market. In this article, the author proposes a conceptual model of a recommendation system that allows one to reduce the burden on an HR specialist at the stage of forming an up-to-date list of requirements for a position profile in the recruitment process. The model is based on a combination of the following components: a graph model of labor market requirements based on the ESCO taxonomy adapted for the Russian language; and an intelligent method of forming

* The article is published with the support of the HSE University Partnership Programme

recommendations for compiling an up-to-date list of requirements in the recruitment process based on neural network models of the language on the architecture of transformers, ESCO skills taxonomy and corpus online vacancies of the Russian labor market. The article also provides a conceptual algorithm for the work of the recommendation system and possible options for recommendations on updating the list of requirements of the position profile in the recruitment process based on an analysis of the needs of the real labor market.

Keywords: labor market analysis, labor market requirements, human resources, job profile, data mining, natural language processing, neural network language models

Citation: Nikolaev I.E. (2023) An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus. *Business Informatics*, vol. 17, no. 2, pp. 71–84. DOI: 10.17323/2587-814X.2023.2.71.84

Introduction

Currently, the vast majority of companies cover a significant part of their staffing needs by posting online job advertisements in online recruitment systems. In such systems, a huge amount of structured and semi-structured data about vacancies and applicants (resumes) is generated and accumulated daily. For example, there are two Russian-language online recruitment systems hh.ru and superjob: the first one has about 57 million resumes and more than 40 million vacancies for the period from 2010 to 2020; the second contains more than 12 million vacancies for the period from 2010 to 2020.

In this regard, the problem of processing and extracting information from online job data becomes particularly relevant, since its solution will allow modeling and understanding complex phenomena in the labor market (see, for example, [1–6]).

Although online recruitment systems have de facto become the main source for co-founders and recruitment managers, they still show shortcomings in search, relevance and accuracy, since job offers are presented in natural language and often in several syntactically and lexically different, but semantically similar forms. This leads to the fact that in the search process search queries are subject to natural language ambiguity and do not compare well with job descriptions in online vacancies. In particular, queries that are overly defined or inconsistent often do not return matches, while relevant job offers could still be found if the problem of consistency or specificity of search queries was solved. If there are not enough exact matches, it is often necessary to accept the worst alternatives or compromise with the initial requirements.

Another problem is the lack of tools that allow a person to use the data extracted from online vacancies in their professional activities. For example, when developing a position profile in the process of creating a vacancy for a position in a company, an HR specialist or the head of the corresponding department needs to spend significant efforts in order to select several dozen, or even hundreds, of similar vacancies, analyze them, extract from them a list of published requirements and responsibilities, conduct their analysis and ranking, and compare them with the job responsibilities of the position for which a new vacancy is being developed.

Modern recommendation systems in general, and in the field of the labor market in particular, largely depend on large amounts of manual data processing and expert knowledge, which makes them expensive, difficult to update and error-prone.

This article proposes a conceptual model of a recommendation system based on the following components:

- graph model of labor market requirements based on the ESCO taxonomy adapted for the Russian language;
- an intelligent method of forming recommendations for compiling a list of requirements in the recruitment process based on neural network language models using the ESCO skills taxonomy and the corpus of online vacancies of the Russian labor market;
- a model and conceptual algorithm of the automated recommendation system for the formation of recommendations;
- possible options for recommendations on updating the list of requirements of the job profile based on an analysis of the needs of the labor market.

1. Analysis of the results of previous works

In recent years, there has been an increasing interest in the use of artificial intelligence (AI) methods for analyzing data on the labor market – "labor market intelligence" (LMI). LMI means the development and use of AI methods, algorithms and structures for analyzing labor market data that help with policy planning and decision-making [7–9].

For example, in [10, 11], research is aimed at creating recommendation systems that determine the correspondence between the applicant's resume and the vacancy for specific competencies at the level of determining the position. Other works are aimed at determining the demand for certain skills [12], which can help students determine their educational trajectory or direction of retraining and increase their level of competitiveness.

Due to the rapid development of computational linguistics and tools for analyzing texts in natural language, some scientists are trying to analyze changes in the labor market based on the texts of online vacancies at the level of individual competencies [13–16]. This approach has many advantages, as it allows us to identify changes at the level of specific professions and specialties, as well as the requirements of employers. For example, it allows one to monitor online vacancies in different regions and countries in real time, predicting the demand for individual skills, competencies and technologies within specific professions or industries, as well as quickly comparing similar labor markets in different countries and regions.

The project of the European Center for the Development of Vocational Education (Cedefop) deserves special attention, since its goal is to collect and classify online vacancies for the entire EU using machine learning [17–19]. In addition, within the framework of this project, research is being conducted to identify trends in the labor market and predict the demand for individual skills. For example, in [20], the authors, using the methods of intellectual text analysis, analyze the literature in the category "fourth technological revolution" and compare the results with the new version of the ESCO skills classification to determine to what extent the new version of the ESCO skills classification, created by experts manually, reflects the trends occurring in the real market labor.

2. Methods and materials

2.1. Overview of the European Taxonomy of ESCO Skills

The ability to extract valuable knowledge from large amounts of data, such as online recruitment systems, strongly depends on the availability of up-to-date knowledge bases, taxonomies and thesauri. Such resources are necessary for the effective application of machine learning methods and for solving most NLP (natural language processing) and NLU (natural language understanding) tasks.

Currently, a large number of labor market analysis projects are based on the European ESCO Skills Classification. ESCO (European Skills, Competencies and Professions) is a multilingual classification of European skills, competencies, qualifications and professions. It defines and classifies skills, competencies, qualifications and professions corresponding to the EU labor market, education and vocational training, in 25 European languages. The system provides professional profiles showing the relationship between professions, skills, competencies and qualifications. ESCO was developed in an open IT format, is available for free use by everyone and is available through an online portal.

ESCO is structured on the basis of three interrelated components, representing a searchable database in 28 languages. These main elements are: a) Professional Profiles (professions), b) Skills/competencies/knowledge and c) Qualifications, as shown in *Fig. 1* of the ESCO data model. The first component – professional profiles (or professions) contains the name, description of the profession and shows whether skills and competencies and knowledge are necessary or optional, and which qualifications are relevant to each profession. The second component contains information about knowledge, skills and competence, as well as some group concepts. ESCO vl contains about 13 500 concepts (and if you include alternative names, then almost 100k formulations) organized into a hierarchy and is also structured through communication with professions. The third component, the qualifications component, allows States and assigning authorities to provide data on qualifications that are collected in ESCO. Qualifications are structured using the European Qualifications Framework (EQF) and ISCED Fields of Education and Training 2013.

Currently, separate groups of scientists are proposing approaches and algorithms for automatically expanding the taxonomy of ESCO skills based on open online job data [21, 22].

For example, for the profession "computer equipment engineer", ESCO code 2152.1.1, 22 alternative names of the profession are defined in the taxonomy; an incomplete list of alternative names is given in *Table 1.* Also, a number of entities skill/competence and knowledge that are necessary for this profession are defined for this profession, each of which also includes a list of alternative names in natural language. For example, 47 basic and 25 additional skills/competencies, 16 basic and 20 additional knowledge entries are defined for this profession; examples of names for the entities of skills/competencies and knowledge, as well as their alternative names are presented in *Table 2.*

The most important advantage of this classification is that it uses the wording of the names of the profession, the names of skills / competencies and knowledge, as well as their alternative names – in natural

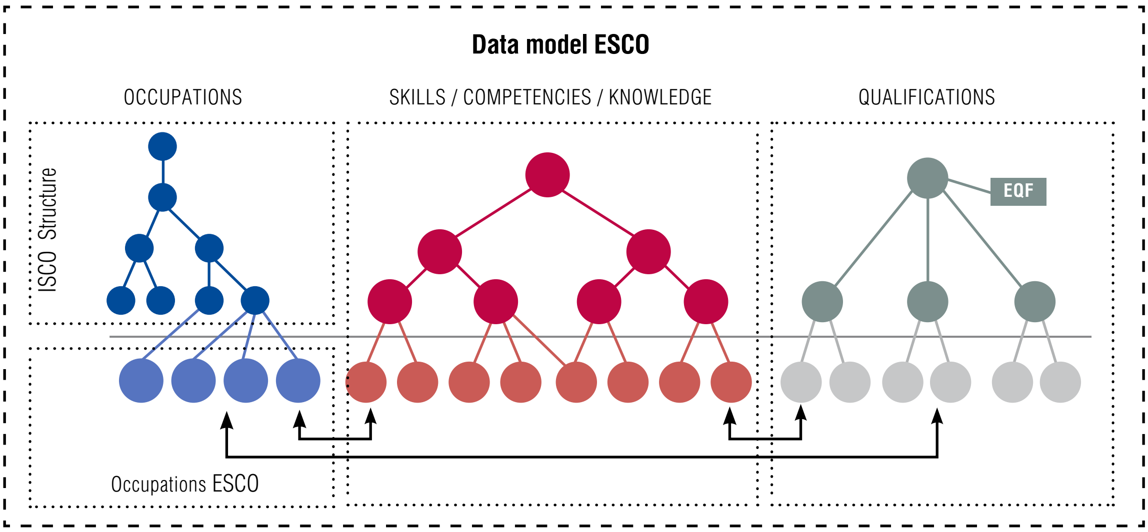

Fig. 1. Data Model ESCO.
Source: https://ec.europa.eu/esco/portal/escopedia/ESCO_data _model

Table 1.

Examples of alternative names for the profession "computer equipment engineer"

Name
computer equipment specialist computer
computer engineer
engineer PC hardware engineer
IT equipment specialist

language, which greatly simplifies and expands the possibilities of its application for analyzing the texts of online vacancies of the real labor market, which are also presented in natural language, modern machine learning methods.

2. The model of the recommendation system for the formation of current requirements of the profile of the position

2.1. Generalized graph model of labor market requirements based on ESCO classification and data from online vacancies

Let's imagine a labor market model as a directed graph. Let's take the ESCO taxonomy as a basis.

Definition of a graph model of the labor market based on the ESCO taxonomy. The graph model is represented as a tuple of three elements $E = (O, R, S)$, where $O = \{o_1, ..., o_n\}$ – set of occupations, $S = \{s_1, ..., s_m\}$ – set of multiple entities of skills/competencies and knowledge, and $R : O \cdot S \to B$ – the relation that connects occupation o with skill s, namely $r(o, s) = 1$ if skill s is associated with occupation o in ESCO, and 0 otherwise.

It is worth noting that one profession can be associated with several entities of skills/competencies and knowledge, and one entity of skills/competencies or knowledge can relate to several professions (*Fig. 2*).

Definition of an online vacancy. Online vacancy j is represented as a tuple $V = (i, c, p, t)$, where $i \in N$ is a unique identifier, $c \in C$ is a unique identifier of the industry, $p \in P$ is a unique identifier of the profession of the vacancy, $t \in T$ is the text of the requirements from the vacancy.

2.2. Calculating the importance of skills for a profession

Since the same skill in the ESCO structure can be associated with several professions, a tool is needed to assess the importance of a skill for a particular profession.

The importance of skills for each profession can be assessed using the RCA tool, originally used in the context of research in the USA [23], where the authors used the O*NET skill classification (the American equivalent of ESCO) to take into account the importance of each skill for each profession. The importance (frequency) of skills for professions $o_i \in O$ and skills $s_l \in S$ is determined by the formula (1), where I denotes the indicator function. The rca function is calculated by the formula (2), where sf is the frequency of the skill s_l for the profession o_i.

To get a more understandable measure, we calculate the normalized rca by formula (3), normalizing the *rca* with respect to the maximum value obtained for the profession in question, so that the most sought-after skill for each profession has a normalized *rca* equal to 1.

$$sf(o_i, s_l) = \frac{\sum_{k=1}^{m} I(o_k = o_i) \cdot I(s_i = s_l)}{\sum_{k=1}^{m} I(o_k = o_i)}, \tag{1}$$

$$rca(oi, sl) = \frac{sf(o_i, s_l) / \sum_{j=1}^{p} sf(o_{ij}, s)}{\sum_{k=1}^{m} sf(o_k, s_l) / \sum_{k=1}^{m} \sum_{j=1}^{p} sf(o_k, s_j)}, \tag{2}$$

$$rca_N(o_i, s_l) = \frac{rca(oi, sl)}{max_j rca(oi, sl)}. \tag{3}$$

Table 2.

Examples of the name for the entities skill/competence and knowledge for the profession "computer equipment engineer"

Priority name	Alternative names	Type
assemble hardware components	assembly of computer equipment, installation of equipment, assembly of computer components, assembly of computer components	skill / competence
installing the software	computer software installation, software download, computer software download, computer software, installation, software installation, software download, …	skill / competence
create technical plans	create plans for technical details, create industrial plans, create technical drawings	skill / competence
principles of electricity	electric current, voltage, electricity physics, electricity science, electricity theory, resistance, voltage	knowledge
hardware components	hardware components of the system, types of hardware components, hardware components, components for hardware systems, parts for hardware systems, components of hardware systems, hardware parts of the system, typology of hardware components	knowledge

2.3. Formation of recommendations based on semantic comparison of the initial list of requirements from the position profile with the graph model of the labor market

The idea of the method of forming recommendations for updating the list of requirements when compiling a position profile comes from the semantic comparison of individual entities from the initial list of requirements and entities from the graph model of the labor market.

In the proposed method, the following stages can be distinguished:

1. Creating a graph model of the labor market based on the classification of ESCO skills.

2. Expansion of the graph model due to information from the texts of online vacancies of the real labor market.

3. Comparison of the initial list of requirements with the entities of the graph model of the labor market.

4. Ranking of matching results based on the RCA metric.

5. Formation of recommendations for inclusion in the initial list of requirements.

Steps 3–5 can be repeated several times, which will allow you to form a more accurate and up-to-date list of requirements with each new iteration.

The recommendation system model is shown in *Fig. 3.*

3. Conceptual algorithm of the system's functioning

A conceptual algorithm is an abstract description of the process of solving a problem or performing a certain action without specifying detailed instructions or

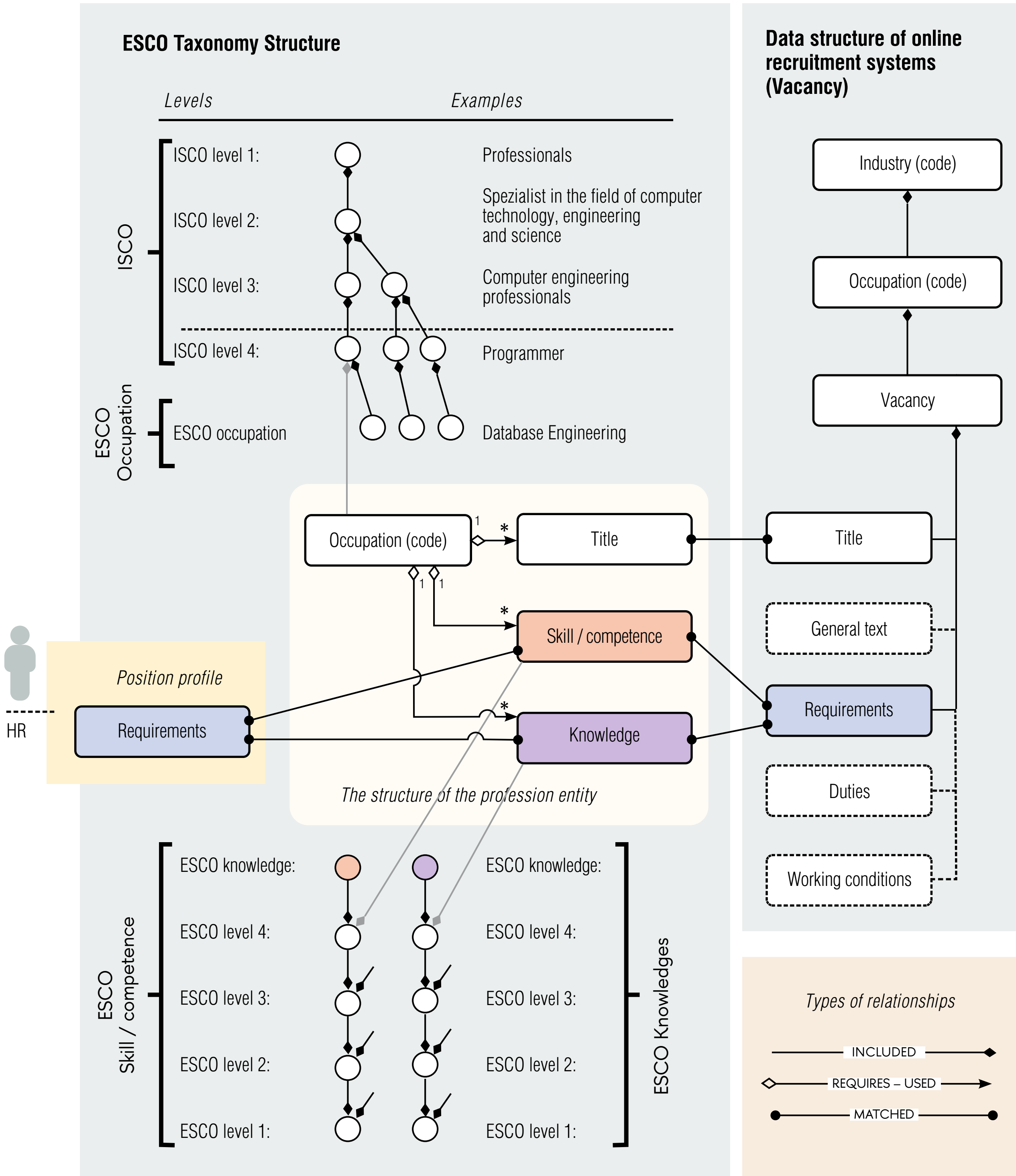

Fig. 2. The main entities and relationships in the generalized graph model of labor market requirements based on the ESCO classification.

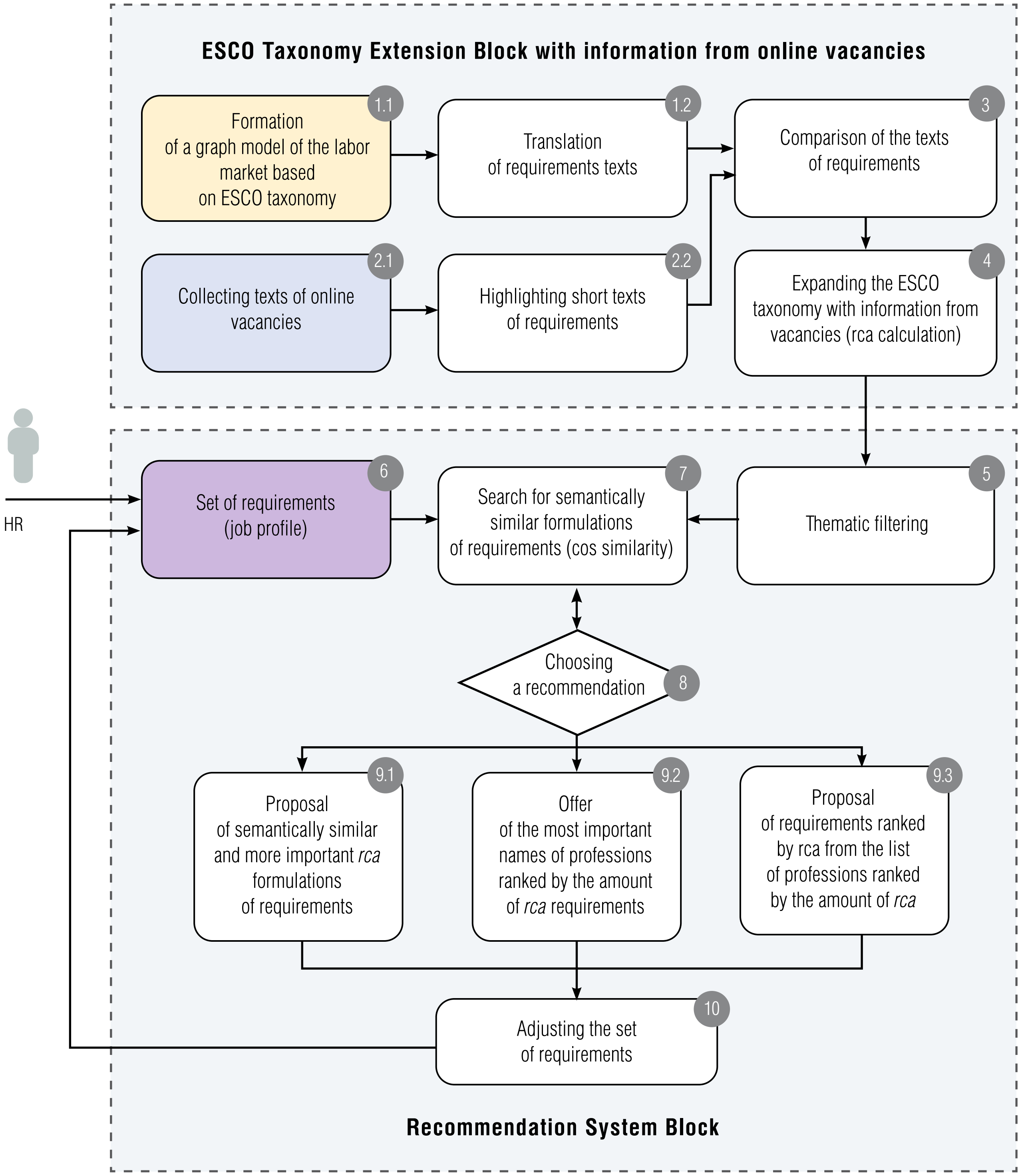

Fig. 3. Model of the recommendation system for updating the list of requirements when compiling a position profile.

programming language. This concept is used in computer science, mathematics and other scientific fields where the process of solving a problem is important, and not a specific code or programming language. In scientific publications, this term can be used to discuss general approaches to solving problems, without reference to specific technologies or implementations.

The result of our work was the development of a conceptual algorithm for the operation of the recommendation system:

1.1. Formation of a graph model of the labor market based on the European taxonomy of ESCO skills (see description of the graph model).

1.2. Adaptation of the graph model of labor market skills for the Russian language. With the help of automatic translation services, all formulations of the European taxonomy of skills are translated, while the structure of relations between the entities of the taxonomy is preserved.

2.1. Collecting online vacancies from Russian online recruitment systems. Many online recruitment systems support an open API for obtaining job data (for example api.hh.ru, api.superjob.ru and others).

2.2. Selection of short texts of requirements from texts of vacancies. There are several possible options for selecting short texts of requirements from the texts of vacancies:

- search for direct matches of entity formulations from the graph model of the labor market with texts from online vacancies;
- developing rules and finding matches based on rules (for example, using the yargy parser from the natasha library for python (https://natasha.github.io/);
- and the third option, preparing a training dataset and training a model for the NER task. For example, using neural network models to analyze natural language texts from the deeppavlov library (https://deeppavlov.ai /) developed by an innovative AI company – iPavlov, a spin-off of MIPT. Continuation of the successful project "Neurointellect iPavlov", implemented within the framework of the NTI, with the industrial support of Sberbank.

3. Comparison of the texts of requirements from online vacancies with the entities of the graph model of labor market requirements. At this step, it is supposed to use a simple comparison and search for matches between normalized (reduced to normal form) texts of the requirements extracted from the texts of online vacancies and the entities of the graph model.

One of the possible improvements of this step can be suggested using the Russian version of the ruwordnet thesaurus [24], which contains relations of hyponyms, hyperonyms, as well as a dictionary of synonyms for the Russian language. Using this information will allow you to expand the list of variations of the texts of requirements when comparing.

4. For all formulations of the graph model, the *rca* parameter is considered, which actually indicates the importance of a skill for a particular profession.

It is worth noting that the expansion of the ESCO taxonomy may occur not just with statistical data, but may also represent a more complex process, for example, the search for new formulations of skills/competencies or knowledge and their integration into an existing graph model [21, 22].

5. Thematic filtering. At this step, using thematic modeling tools, the user can be offered terms and concepts automatically grouped by topic. The user can select a list of words or concepts that must necessarily be contained or absent in the texts of the final output. The possibility of using such a thematic filter was demonstrated by the author in the article [25].

6. The user of the system (for example, an HR specialist) forms an initial set of requirements and this is submitted to the system input.

7. Comparison of the texts of requirements from the user's request and entities from the graphical model of the labor market. In this case, it is assumed that we use a more intelligent matching process which can be divided into two stages: with the help of modern neural network models built on the architecture of transformers (RuBERT, Robert and others), vector representations are obtained for the texts of user requirements and for the entities of the graph model. Then, using a cone proximity measure, vector rep-

resentations are matched in pairs. Further, all texts of requirements that lie beyond a certain permissible distance (determined experimentally) are cut off, and which are the most semantically close to the text of the original requirement. This is how semantically similar texts of requirements from the graph model are determined for all texts from the user's original list of requirements.

The algorithm for choosing the most effective neural network model that would allow generating the best (from the point of view of compactness) vector representations for semantically similar texts of requirements was considered by the author in the article [26].

8. The user chooses which type of recommendation he would like to form for the initial list of requirements.

9.1. Semantically similar texts from the graph model selected in step 7 are ranked by the *rca* parameter. The arranged formulations are demonstrated to the user, indicating the *rca* parameter and profession. In fact, at this stage, the user gets the opportunity to select the most important and semantically similar requirements, and he decides to add the proposed requirements to his initial list.

9.2. Semantically similar texts from the graph model selected in step 6 are ranked by the *rca* parameter. For the totality of the requirements of the initial list, the formulas with the highest *rca* are selected. For all requirements and professions, the amount is considered. The sessions are ranked by the amount of *rca* for the initial list of requirements (*Fig. 4*). From the graph model, *n* profession names are selected based on the largest amount of *rca* for the initial list of user requirements and are shown to the user.

9.3. Just like in 9.2, the *rca* sum is calculated. *N* profiles are selected from the graph model based on the largest amount of *rca* for the initial list of user requirements (*Fig. 4*). From each selected profession, the *M* most important *rca* texts of skills/competencies or knowledge are selected and demonstrated to the user.

10. After studying the proposed recommendations, the system user can choose the options that will be added to the initial list of requirements.

Steps 5 through 10 can be repeated over again, which will iteratively refine and improve the initial list of requirements set by the user.

Examples of possible options for modifying the set of requirements of the position profile based on the recommendations formed:

- Propose a new alternative formulation for the existing requirement based on a higher *rca*.
- Propose to include a new requirement in the position profile based on a high *rca*, taking into account the relationship with existing requirements.
- Recommend to exclude a requirement from the position profile based on a low *rca*, taking into account the relationship with existing requirements.
- Break down the requirements into categories by profession
- Suggest the title of the position based on the list of requirements of the position profile.
- Ranking of requirements according to their demand based on the *rca* indicator.

The list of recommendations can be adjusted towards expansion by adding new functional blocks to the system.

The computational complexity of the entire system is estimated as low. The most time-consuming stages, such as extracting the entities of knowledge, skills and competencies from the texts of vacancies, as well as their vector representation, occur once and can occur in the background. Thematic filtering and *rca* counting are relatively simple computational operations. The most difficult, from the point of view of computational complexity, is the operation of ranking a large number of requirements relative to each other based on a cosine measure of proximity. To perform a highly productive search for similar vectors of requirements, there are plans to use the FAISS library (Facebook AI Similarity Search) [26, 27]. This library provides a set of algorithms for indexing large sets of vectors and quickly searching for nearest neighbors in these sets. The library was developed by Facebook AI Research and is distributed under the terms of the Apache 2.0 license.

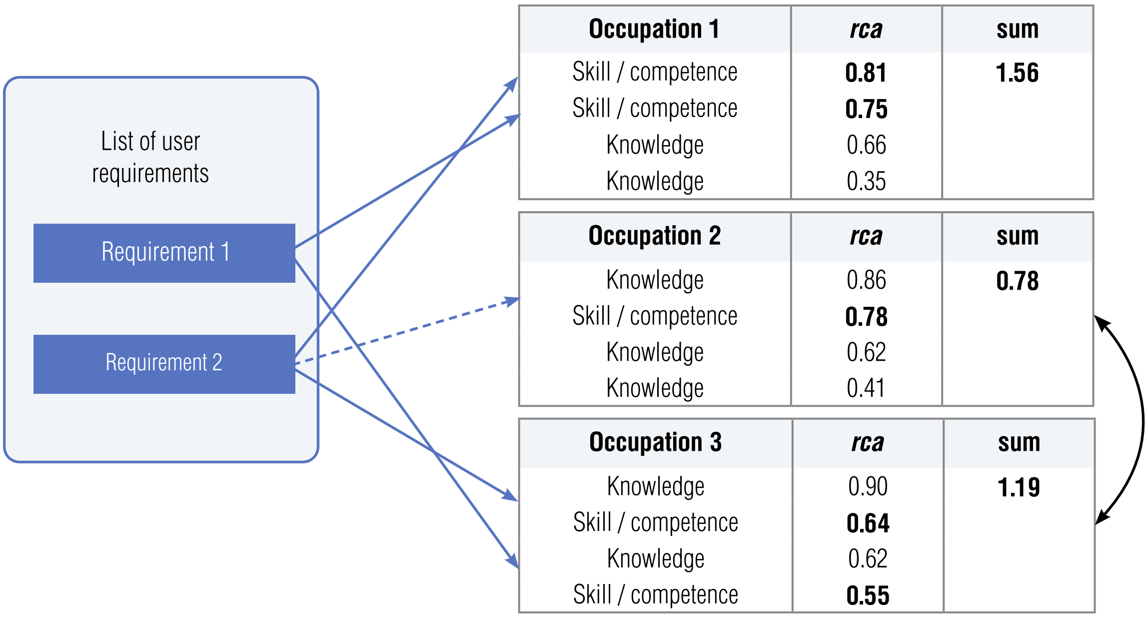

Fig. 4. Ranking of the profession by the amount of *rca* for the initial list of requirements.

Conclusion

The development of a recommendation system for the formation of an up-to-date list of job profile requirements based on an analysis of the real requirements of the labor market is an important step in the development of HR technologies and will allow us to: significantly reduce the labor costs of HR specialists; more accurately and systematically determine the requirements for candidates for various positions; identify potential opportunities for retraining employees; better understand how changes in the requirements of the labor market affect companies and their personnel; form more flexible and adaptive strategies for attracting and managing personnel.

The proposed conceptual model of the recommendation system includes:

- Graph model of labor market requirements based on ESCO taxonomy adapted for the Russian language;
- An intelligent method of forming recommendations for compiling a list of requirements in the recruitment process based on neural network language models using the ESCO skills taxonomy and the corpus of online vacancies of the Russian labor market. Within the framework of the method, we propose to use neural network models of the language built on the architecture of transformers (models of the BERT family) to assess the semantic proximity of the entities of the initial list of requirements with the graph model of the labor market;
- Model and conceptual algorithm of the automated recommendation system for the formation of recommendations.

This article also provides possible options for recommendations on updating the list of requirements of the position profile based on an analysis of the needs of the real labor market.

To improve this model, the author additionally plans to: develop a method for extracting individual short texts of knowledge and skills from the texts of the requirements of online vacancies of the real labor market; develop a system for automatically expanding the graph model with texts of knowledge and skills; integrate an extended graph model of labor market requirements through the international system of classification of occupations (ISCO) with all-Russian classifiers (OKZ, the all-Russian classifier of occupations, OKVED, the all-Russian classifier of economic activities), and professional standards of the Russian Federation. In addition, a separate and important task is to develop a method for assessing the quality and investigating the effectiveness of using a recommendation system based on the proposed model in various sectors of the economy and in various labor markets. ■

References

1. Boselli R., Cesarini M., Mercorio F. (2017) Using machine learning for labour market intelligence. *Machine Learning and Knowledge Discovery in Databases: European Conference*, pp. 330–342. https://doi.org/10.1007/978-3-319-71273-4_27

2. Colombo E., Mercorio F., Mezzanzanica M. (2019) AI meets labor market: Exploring the link between automation and skills. *Information Economics and Policy*, vol. 47, pp. 27–37. https://doi.org/10.1016/j.infoecopol.2019.05.003

3. Giabelli A., Malandri L., Mercorio F. (2022) GraphLMI: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, vol. 81, pp. 3061–3090. https://doi.org/10.1007/s11042-020-09115-x

4. Mezzanzanica M., Boselli R., Cesarini M., Mercorio F. (2015) A model-based approach for developing data cleansing solutions. *Journal of Data and Information Quality*, vol. 5, no. 4, pp. 1–28. https://doi.org/10.1145/2641575

5. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. Proceedings of the *AAAI conference on artificial intelligence*, vol. 32, no. 1. https://doi.org/10.1609/aaai.v32i1.11847

6. Zhang D., Liu J., Zhu H., Liu Y., Wang L., Wang P., Xiong H. (2019) Job2Vec: Job title benchmarking with collective multi-view representation learning. Proceedings of the *28th ACM International Conference on Information and Knowledge Management*, pp. 2763–2771. https://doi.org/10.1145/3357384.3357825

7. *The importance of LMI* (2015) UK Commission for Employment and Skills. Available at: https://www.gov.uk/government/publications/the-importance-of-labour-market-intelligence (accessed 30.01.2023).

8. Mezzanzanica M., Mercorio F. (2019) Big data enables labor market intelligence. *Encyclopedia of Big Data Technologies*, pp. 226–236. https://doi.org/10.1007/978-3-319-63962-8_276-1

9. Eggertsson T. (1990) *Economic behavior and institutions: Principles of neoinstitutional economics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609404

10. Qin C., Zhu H., Xu T., Zhu C., Jiang L., Chen E., Xiong H. (2018) Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. Proceedings of the *41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*, pp. 25–34. https://doi.org/10.1145/3209978.3210025

11. Zhu C., Zhu H., Xiong H., Ma C., Xie F., Ding P., Li P. (2018) Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Transactions on Management Information Systems*, vol. 9, no. 3, pp. 1–17. https://doi.org/10.1145/3234465

12. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. Proceedings of the *AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. https://doi.org/10.1609/aaai.v32i1.11847

13. Vinel M., Ryazanov I., Botov D., Nikolaev I. (2019) Experimental comparison of unsupervised approaches in the task of separating specializations within professions in job vacancies. *Artificial Intelligence and Natural Language. AINL 2019. Communications in Computer and Information Science* (eds. D. Ustalov, A. Filchenkov, L. Pivovarova), vol. 1119. Springer, Cham. https://doi.org/10.1007/978-3-030-34518-1_7

14. Nikolaev I., Ryazanov I., Botov D. (2020) The comparison of distributive semantics models applied to the task of short job requirements clustering for the Russian labor market. Proceedings of the *8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)*. Atlantis Press, pp. 295–301. https://doi.org/10.2991/aisr.k.201029.056

15. Colombo E., Mercorio F., Mezzanzanica M. (2018) Applying machine learning tools on web vacancies for labour market and skill analysis. *Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence. Technology Policy Institute Conference on the Economics and Policy Implications of Artificial Intelligence*. Available at: https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo_paper.pdf (accessed 30 May 2023).

16. O'Kane L., Narasimhan R., Nania J., Taska B. (2020) *Digitalization in the German labor market: Analyzing demand for digital skills in job vacancies*. Gütersloh: Bertelsmann Stiftung.

17. *Real-time labour market information on skill requirements: Setting up the EU system for online vacancy analysis (2016)* Cedefop. Available at: https://www.cedefop.europa.eu/it/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-setting-eu (accessed 30 January 2023).

18. Boselli R., Cesarini M., Mercorio F., Mezzanzanica M. (2018) Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, vol. 86, pp. 319–328. https://doi.org/10.1016/j.future.2018.03.035

19. Boselli R., Cesarini M., Marrara S., Mercorio F., Mezzanzanica M., Pasi G., Viviani M. (2018) WoLMIS: A labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, vol. 51, pp. 477–502. https://doi.org/10.1007/s10844-017-0488-x

20. Chiarello F., Fantoni G., Hogarth T., Giordano V., Baltina L., Spada I. (2021) Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change*, vol. 173, article 121177. https://doi.org/10.1016/j.techfore.2021.121177

21. Giabelli A., Malandri L., Mercorio F., Mezzanzanica M., Seveso A. (2020) NEO: A tool for taxonomy enrichment with new emerging occupations. *The Semantic Web – ISWC 2020. Lecture Notes in Computer Science*, vol. 12507, pp. 568–584. https://doi.org/10.1007/978-3-030-62466-8_35

22. Malandri L., Mercorio F., Mezzanzanica M., Nobani N. (2020) Meet: A method for embeddings evaluation for taxonomic data. Proceedings of the *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 31–38. Sorrento, Italy. https://doi.org/10.1109/ICDMW51313.2020.00014

23. Alabdulkareem A., Frank M.R., Sun L., AlShebli B., Hidalgo C., Rahwan I. (2018) Unpacking the polarization of workplace skills. *Science Advances*, vol. 4, no. 7, eaao6030. https://doi.org/10.1126/sciadv.aao6030

24. *The Russian language Thesaurus RuWordNet*. Available at: https://ruwordnet.ru/ru (accessed 30 May 2023).

25. Nikolaev I., Botov D., Dmitrin Y., Klenin J., Melnikov A. (2019) Use of topic modelling for improvement of quality in the task of semantic search of educational courses. Proceedings of the *21st International Workshop on Computer Science and Information Technologies (CSIT 2019)*, pp. 104–111. Atlantis Press. https://doi.org/10.2991/csit-19.2019.18

26. Nikolaev I., Melnikov A. (2022) Comparison of neural network models based on transformer architecture in the context of the task of evaluating the compactness of vector representations of semantically similar texts of the requirements of the European ESCO skills classification. *Bulletin of the South Ural State University. Series: Computer Technology, Control, Radio Electronics*, vol. 22, no. 3, pp. 19–29.

27. *Faiss. A library for efficient similarity search and clustering of dense vectors*. Available at: https://faiss.ai/ (accessed 30 May 2023).

28. Johnson J., Douze M., Jégou H. (2019) Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547.

About the author

Ivan E. Nikolaev

Senior Lecturer, Department of Information Technology, Institute of Information Technologies, Chelyabinsk State University, 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia;

E-mail: ivan_nikolaev@csu.ru

ORCID: 0000-0002-9686-2435