

Сравнение языковых моделей в задаче извлечения навыков из вакансий и резюме

Л. А. Комарова^{1*}, В. И. Соловьев¹, А. М. Колосов²

¹ ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации», г. Москва, Российская Федерация

Адрес: 125167, Российская Федерация, г. Москва, Ленинградский проспект, д. 49/2

*229388@edu.fa.ru

² ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», г. Москва, Российская Федерация

Адрес: 119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1

Аннотация

Способность больших языковых моделей (LLM) «понимать» большие объемы текстовых данных дает возможность к улучшению качества подбора кандидатов на открытые в компании вакансии. Цель данной работы заключается в сравнении способностей языковых моделей и численного метода, использующего векторные представления, в задаче извлечения навыков из текстов вакансий и резюме. Особое внимание уделяется использованию методов ранжирования навыков с помощью LLM и численного метода, использующего косинусное расстояние между векторными представлениями навыков. Исследование состоит из трех экспериментов: первый эксперимент направлен на извлечение фраз, которые являются навыками, из описаний опыта работы из текста резюме; второй включает определение соответствия навыков из текста резюме эталонному набору навыков из требований к вакансии; третий эксперимент направлен на оценку степени схожести между двумя наборами навыков. Итогом исследования является выбор наилучшей модели и способа извлечения навыков из текста резюме, а также сравнения двух наборов навыков между собой. Проведенные эксперименты показали, что языковые модели превосходят численные методы в точности и гибкости определения навыков из текстов. Использование LLM для ранжирования навыков с последующей оценкой с помощью косинусного расстояния продемонстрировало низкую эффективность и точность в сопоставлении навыков между вакансиями и резюме. Однако, численный метод с использованием векторных представлений навыков показал более высокий результат по качеству ранжирования и стабильность с возрастанием количества примеров для сопоставления. Результаты данного исследования имеют практическое значение для разработки более точных и эффективных систем подбора персонала. Внедрение языковых моделей в процессы HR может существенно повысить качество и скорость обработки больших объемов данных, что приведет к более точному и быстрому подбору квалифицированных специалистов.

Ключевые слова: большие языковые модели, извлечение навыков, сопоставление навыков, семантический анализ, автоматизация в рекрутинге

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Комарова Л. А., Соловьев В. И., Колосов А. М. Сравнение языковых моделей в задаче извлечения навыков из вакансий и резюме // Современные информационные технологии и ИТ-образование. 2024. Т. 20, № 1. С. 157-163. <https://doi.org/10.25559/SITITO.020.202401.157-163>

© Комарова Л. А., Соловьев В. И., Колосов А. М., 2024



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Comparison of Language Models in Skills Extraction From Vacancies and Resumes

L. A. Komarova^a, V. I. Soloviev^a, A. M. Kolosov^b

^a Financial University under the Government of the Russian Federation, Moscow, Russian Federation
Address: 49/2 Leningradsky Prospekt, Moscow 125167, Russian Federation
*229388@edu.fa.ru

^b Lomonosov Moscow State University, Moscow, Russian Federation
Address: 1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation

Abstract

The ability of large language models (LLMs) to «understand» large volumes of text data allows for consistent quality selection of candidates for company openings. The purpose of the work is to consider the capabilities of language models (LLM) and the specified method using vector representations in the tasks of extracting skills from texts of vacancies and resumes. Particular attention is paid to the use of skill ranking methods using LLM and the following method using the cosine distance between vector representations of skills. The study consisted of three experiments: the first experiment aimed to extract skill phrases from described work experiences from a text resume; the second involves assigning skills from the resume text to a reference set of functions from the job requirements; The third experiment aims to evaluate the best performance between the two skill sets. The result of the study is the selection of the best model and the derivation of functions from the summary text, as well as a comparison of the two sets of functions with each other. Experiments have shown that language models are superior to numerical methods in terms of accuracy and flexibility in determining the capabilities of a text. Using LLM to rank features using cosine distance has shown poor performance and accuracy in measuring features between job openings and resumes. However, the numerical method using vector representation methods showed better results in quality ranking and stability with increasing number of parliamentary examples. The results of this study have practical implications for the development of more accurate and efficient personnel selection systems. The introduction of language models into human resource management processes can improve the quality and speed of processing large volumes of data, which will lead to a more accurate and faster selection of qualified specialists

Keywords: LLM, Skill Extraction, Job Matching, Semantic Analysis, Recruitment Automation

Conflict of interests: The authors declare no conflict of interest.

For citation: Komarova L.A., Soloviev V.I., Kolosov A.M. Comparison of Language Models in Skills Extraction From Vacancies and Resumes. *Modern Information Technologies and IT-Education*. 2024;20(1):157-163. <https://doi.org/10.25559/SITITO.020.202401.157-163>



Введение

В современной экономической ситуации принятие быстрых и точных решений при подборе персонала приобретает решающее значение для эффективной работы бизнеса. Компании стремятся привлечь кандидатов с наиболее подходящими навыками для решения задач, которые требует открытая позиция.

Задача оценки соответствия кандидата требованиям из вакансии является фундаментальной проблемой, с которой сталкиваются рекрутеры в процессе найма. Исторически эта задача выполнялась вручную: рекрутеры просматривали резюме и сравнивали заявленный опыт и навыки кандидата с должностными требованиями. Очевидно, этот процесс требует много времени, подвержен человеческим ошибкам и предвзятости. Необходимость автоматизации и повышения точности сопоставления опыта кандидата с требованиями вакансии становится все более очевидной по мере расширения рынков труда и диверсификации набора навыков.

В последние годы были достигнуты значительные успехи в области обработки естественного языка (NLP), что позволило быстро продвинуться в области интеллектуального анализа рынка труда. Развитие больших языковых моделей (LLM) представляет собой набор инструментов для повышения эффективности и результативности процессов анализа резюме и отбора талантливых кандидатов [1, 2]. Внедрение больших языковых моделей (LLM) в деятельность рекрутера направлено на оптимизацию оценки резюме за счет автоматизации извлечения информации из текста резюме, тем самым обеспечивая более объективное и эффективное сопоставление навыков кандидатов с требованиями вакансии.

Целью данной статьи является изучение возможностей LLM в автоматизации задачи скрининга кандидатов, которая традиционно трудоемка и подвержена человеческим ошибкам

Обзор подходов к извлечению навыков из неструктурированного текста

Рынок труда находится в состоянии постоянного развития — часто из-за изменений в технологиях, появления новых видов трудовой деятельности или изменений требований к существующим должностям [3].

Для следования трендам, без потери качества моделей по отбору кандидатов, целесообразно использовать информацию, полученную из текстов о требованиях к вакансии и текстов резюме (набору навыков), что может помочь, в сопоставлении кандидатов на актуальные вакансии [4].

Задача автоматического извлечения навыков (skill extraction) заключается в извлечении компетенций, необходимых для любой профессии, из неструктурированного текста.

Ранние работы по извлечению набора навыков из текста резюме или вакансий представляли собой последовательность из задачи разметки и последующий классификации с использованием меток-классов [5-10]. Подход часто основывался на методах машинного обучения, таких как Random Forest, SVM, Linear Regression. Эти методы требовали значительных усилий для ручной разметки данных (аннотирования) – дорогостоящего и

трудоемкого процесса, требующего большого количества примеров для разметки и последующего обучения модели.

Также ранее был распространен подход с использованием кластеризации.

С развитием глубокого обучения и обработкой естественного языка (NLP), подходы к извлечению навыков существенно изменились. Одним из ключевых изменений стало использование предобученных языковых моделей, таких как BERT (Bidirectional Encoder Representations from Transformers) и его вариаций¹ [11-14]. Данные модели позволяют учитывать контекст слова. Для задачи извлечения навыков модель дообучают на специализированных данных вакансий и резюме, чтобы лучше понимать контекст и выделять релевантные навыки. Существует также комбинированный подход для улучшения результатов. Например, использование BERT для начального извлечения навыков и последующее применение специализированных классификаторов для уточнения и категоризации навыков. Так, в исследованиях авторы используют BERT для извлечения навыков из текстов вакансий и резюме и дополнительная модель для уточнения навыков, что показало превосходство над традиционными методами машинного обучения [15-18].

В последнее время интерес вызывают вопросы выделения и классификации навыков с помощью больших языковых моделей (генеративных моделей), что подтверждается большим количеством публикаций, основанных на достижениях в области технологии обработки естественного языка (NLP) [19, 20]. Например, с помощью больших языковых моделей (LLM) можно решать задачи по извлечению навыков с использованием синтетических обучающих данных, требующих больших ресурсов [21]. Также, появляются исследования, касающиеся извлечения навыков без учителя [22-25]. Эти модели не только понимают контекст, но и могут генерировать текст. В задаче извлечения навыков они могут использоваться для создания аннотаций и предложений, что может помочь в более точной классификации навыков.

Описание данных

Были собраны наборы вакансий и резюме посредством API портала hh.ru. Всего удалось собрать 31000 резюме и 18000 вакансий по рабочим специальностям.

Тексты собранных наборов вакансий и резюме были разделены на фразы при помощи LLM модели от openAI gpt-4 и библиотеки Langchain для создания единой структуры и возможности проведения ряда экспериментов.

Также из json-файлов вакансий были извлечены стандартные навыки, которые составили корпус из 30000 навыков.

Извлечение описания опыта работы из текста резюме

Постановка задачи

Дан набор фраз, на которые разделен текст резюме. Необходимо определить является ли каждая фраза навыком или нет. Проверка качества решения определяется с помощью экспертной разметки и расчета меры F1.

¹ Chernova M. Occupational skills extraction with FinBERT : Master's Thesis. Helsinki, Finland : Metropolia University of Applied Sciences, 2020. 49 p.



Подход к решению, основанный на запросе к языковой модели

Текст описания опыта работы направляется в языковую модель с запросом определить является ли фраза из текста резюме профессиональным навыком.

Для проверки эффективности извлечения навыков, в работе используется показатель микро-F1 (1), гармоническое среднее значение точности и полноты, которое широко используется при оценке задач определения именованных сущностей (NER - name entities recognition).

$$2 \frac{Precision_{micro} \cdot Recall_{micro}}{Precision_{micro} + Recall_{micro}} \quad (1)$$

Для реализации эксперимента выбраны три модели: gpt-3.5-turbo-1106, gigachat, saiga_mistral_7b_merged, в которые был подан следующий запрос (Рис. 1)

```
template_string = """Ты опытный hr специалист

Для каждой фразы в списке   ```{context}``` выведи фразу и выведи + если она является навыком и выведи -, если она не является навыком
Обязательно выводи фразу и знак, больше ничего не генерируй.
В mark должно быть значение + или -, мы не можем оставлять другой знак или пустую строку

Example Output Format:
[{"phrase": " ", "mark": "+"}, {"phrase": " ", "mark": "-"}]

"""
```

Р и с. 1. Промпт для подачи в LLM для определения навыка

F i g. 1. Prompt for submission to the LLM to determine the skill

Источник: здесь и далее в статье все таблицы и рисунки составлены авторами.

Source: Hereinafter in this article all tables and figures were made by the authors.

Результаты

Получены результаты для F1-микро (Таблица 1):

Т а б л и ц а 1. Результат эксперимента по определению навыков с использованием LLM

T a b l e 1. Result of an experiment to determine skills using LLM

Модель	F1-micro
gpt-3.5-turbo-1106	0,85
Gigachat	0,87
saiga_mistral_7b_merged	0,78

Выводы

На основании эксперимента делается вывод об эффективности модели gigachat при определении профессиональных навыков, написанных человеком в произвольной форме, из текста резюме, с помощью запроса к языковой модели.

Извлечение дополнительных навыков из текста вакансий**Постановка задачи**

Даны: набор фраз, на которые разделен текст резюме (аналогичен эксперименту 1), и база стандартных навыков. Необходимо на основании базы навыков определить является ли каждая фраза профессиональным навыком или нет. Проверка качества решения определяется с помощью экспертной разметки и расчета меры F1.

F1 позволяет измерить точность и полноту извлеченных данных, гарантируя, что навыки правильно определены и актуальны.

Определение навыков с помощью LLM

В эксперименте сравнивается эффективность трех различных языковых моделей в задаче автоматического распознавания профессиональных навыков и опыта работы содержащихся в тексте резюме.

Эксперимент поможет определить языковую модель, которая наиболее эффективно справляется с поставленной задачей, что важно для улучшения инструментов автоматической обработки резюме.

Подход к решению, основанный на мере близости фразы из резюме и стандартного набора навыков

Фраза из описания опыта работы направляется в языковую модель для получения векторного представления текста (эмбединг). Далее полученный эмбединг сравнивается с каждым эмбедингом из базы стандартных навыков по косинусному расстоянию. Для определения является ли фраза навыком задается порог для косинусного расстояния и, если расстояние больше, чем заданный порог, фраза определяется, как не навык. Для проверки эффективности извлечения навыков, в работе используется показатель F1, аналогично первому эксперименту.

Измерение косинусного расстояния между фразой и навыком из набора навыков

В эксперименте сравнивается эффективность трех различных языковых моделей в задаче автоматического распознавания профессиональных навыков и опыта работы из текста резюме. Этот эксперимент поможет сравнить эффективность использования численного метода при определении навыков с результатом языковой модели из первого эксперимента.

Получены результаты для F1-микро (Таблица 2):

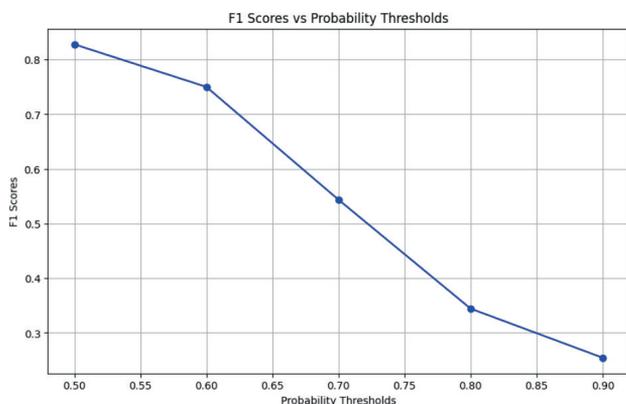
Т а б л и ц а 2. Результаты эксперимента по измерению косинусного расстояния между фразой и навыком из набора навыков

T a b l e 2. Results of an experiment to measure the cosine distance between a phrase and a skill from a skill set

Модель	F1-micro
gpt-3.5-turbo-1106	0,34
Gigachat	0,84
saiga_mistral_7b_merged	0,84



По результатам видно, что модель gpt-3.5-turbo-1106 показала худший результат. В связи с этим решено провести дополнительный эксперимент для определения порога (Рис. 2), при котором качество модели может вырасти.



Р и с. 2. Вариация порога отсечения косинусного расстояния для определения навыка моделью gpt-3.5-turbo-1106

Fig. 2. Threshold variation of cosine cut for defining skill with gpt-3.5-turbo-1106 model

На основании графика можно сделать вывод о способности качественно определять навыки только при пороге ниже 0,6, что говорит о чувствительности модели и не адаптированности к русскому языку.

Вывод

На основании эксперимента делается вывод об одинаковой эффективности моделей gigachat и saiga_mistral_7b_merged при определении профессиональных навыков, написанных человеком в произвольной форме, из текста резюме, с помощью измерения косинусного расстояния между фразой и навыком из набора навыков.

Сопоставление навыков из резюме и требований из вакансий

Постановка задачи

Даны список навыков из текстов резюме и список навыков из текста вакансии. Необходимо определить, насколько эффективно могут быть сопоставлены навыки, указанные в резюме, с требованиями, предъявляемыми в вакансии.

Подход к решению, основанный на определении порядков сходства вакансии и резюме

Подход состоит в сравнении способность языковой модели, выбранной в результате экспериментов 1 и 2 упорядочить набор пар по степени соответствия навыков внутри пар со степенью соответствия внутри каждой пары и упорядочивания по косинусному расстоянию.

Для реализации эксперимента выбрана модель, в которую подается запрос (Рис. 3).

```
template = ""Упорядочи навыки по возрастанию сходства с вакансией.
rank = 1 - наибольшее сходство

Вакансия: {job_description}

Резюме: {cv}

Верни список порядков

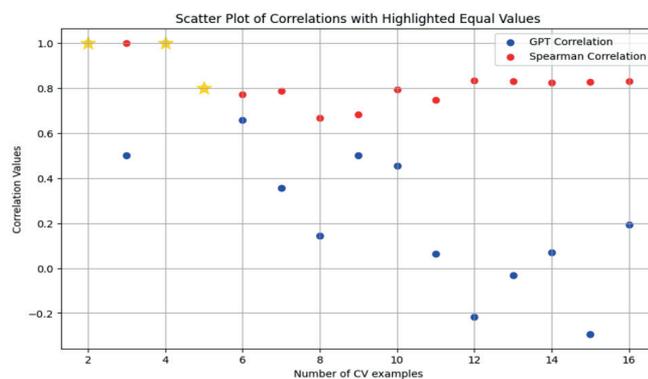
Answer:
""
```

Р и с. 3. Промпт для подачи в LLM для ранжирования резюме относительно требований вакансии

Fig. 3. Prompt for submission to the LLM to rank resumes in relation to vacancy

Определение порядков с помощью LLM и косинусного расстояния

В результате проведенных экспериментов получен график (Рис. 4) значений корреляции Спирмена для подхода с большей языковой моделью и численного метода.



Р и с. 4. Промпт для подачи в LLM для ранжирования резюме относительно требований вакансии

Fig. 4. Prompt for submission to the LLM to rank resumes in relation to vacancy

На графике звездочкой отмечены эквивалентные значения корреляций в обоих подходах.

Вывод

На основании проведенных экспериментов можно сделать вывод о большей эффективности и стабильности использования численного метода при ранжировании резюме относительно требований из вакансии.

Заключение

Проведенные эксперименты показали, что языковые модели превосходят численные методы в точности и гибкости определения навыков из текстов. Использование LLM для ранжирования навыков с последующей оценкой с помощью косинусного расстояния продемонстрировало низкую эффективность и точность в сопоставлении навыков между вакансиями и ре-



зюме. Однако, численный метод с использованием векторных представлений навыков показал более высокий результат по качеству ранжирования и стабильность с возрастанием количества примеров для сопоставления.

Результаты данного исследования имеют практическое значе-

ние для разработки более точных и эффективных систем подбора персонала. Внедрение языковых моделей в процессы HR может существенно повысить качество и скорость обработки больших объемов данных, что приведет к более точному и быстрому подбору квалифицированных специалистов.

References

- [1] Dakhel A.M., Nikanjam A., Majdinasab V., Khomh F., Desmarais M.C., Effective test generation using pre-trained Large Language Models and mutation testing. *Information and Software Technology*. 2024;171:107468. <https://doi.org/10.1016/j.infsof.2024.107468>
- [2] Kalyan K.S. A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*. 2024;6:100048. <https://doi.org/10.1016/j.nlp.2023.100048>
- [3] Deming D.J., Kahn L.B., Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*. 2018;36(S1):S337-S369. <https://doi.org/10.1086/694106>
- [4] Komarova L.A., Zolkin A.L., Kornetov A.N., Pestin V.A. Research Methods and Mechanisms of Decision-Making in HR Management (Literature Review). *Scientific and Technical Volga region Bulletin*. 2023;(5):136-141. (In Russ., abstract in Eng.) EDN: ITNRZB
- [5] Tamburri D.A., Van Den Heuvel W.-J., Garriga M., Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). Las Vegas, NV, USA: IEEE Press; 2020. p. 391-394. <https://doi.org/10.1109/IRI49571.2020.00063>
- [6] Zhang M., Jensen K.N., Plank B., Kompetenzer: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning. *arXiv:2205.01381*. 2022. <https://doi.org/10.48550/arXiv.2205.01381>
- [7] Zhang M., Jensen K.N., Sonniks S., Plank B., SkillSpan: Hard and soft skill extraction from English job postings. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics. Seattle, United States: IEEE Press; 2022. p. 4962-4984. <https://doi.org/10.48550/arXiv.2204.12811>
- [8] Jiechieu K.F.F., Tsopze N., Skills prediction based on multi-label resume classification using CNN with model predictions explanation. *Neural Computing and Applications*. 2021;33(10):5069-5087. <https://doi.org/10.1007/s00521-020-05302-x>
- [9] Fareri S., Melluso N., Chiarello F., Fantoni G., SkillNER: Mining and mapping soft skills from any text. *Expert Systems with Applications*. 2021;184:115544. <https://doi.org/10.1016/j.eswa.2021.115544>
- [10] Gaur B., et al. Semi-supervised deep learning based named entity recognition model to parse education section of resumes. *Neural Computing and Applications*. 2021;33:5705-5718. <https://doi.org/10.1007/s00521-020-05351-2>
- [11] Paaß G., Giesselbach S. Pre-trained Language Models. In: Foundation Models for Natural Language Processing. *Artificial Intelligence: Foundations, Theory, and Algorithms*. Cham: Springer; 2023. p. 19-78. https://doi.org/10.1007/978-3-031-23190-2_2
- [12] Komarova L.A., Cheremuhin A.D. Increasing the efficiency of recruitment based on deep neural networks. *Journal of Applied Informatics*. 2024;2(110):10-22. (In Russ., abstract in Eng.) <https://doi.org/10.37791/2687-0649-2024-19-2-10-22>
- [13] Wings I., Nanda R., Adebayo K.J., A Context-Aware Approach for Extracting Hard and Soft Skills. *Procedia Computer Science*. 2021;193:163-172. <https://doi.org/10.1016/j.procs.2021.10.016>
- [14] Botov D., Klenin J., Melnikov A., Dmitrin Y., Nikolaev I., Vinel M. Mining Labor Market Requirements Using Distributional Semantic Models and Deep Learning. In: Abramowicz W., Corchuelo R. (eds.) Business Information Systems. BIS 2019. *Lecture Notes in Business Information Processing*. Vol. 354. Cham: Springer; 2019. p. 177-190. https://doi.org/10.1007/978-3-030-20482-2_15
- [15] Tian X., Pavur R., Han H., Zhang L. A machine learning-based human resources recruitment system for business process management: using LSA, BERT and SVM. *Business Process Management Journal*. 2023;29(1):202-222. <https://doi.org/10.1108/BPMJ-08-2022-0389>
- [16] Emary E. A proposed Emergent Skill Extraction Methodology from Unstructured Text. In: Proceedings of the Federated Africa and Middle East Conference on Software Engineering (FAMECSE '22). New York, NY, USA: Association for Computing Machinery; 2022. p. 26-30. <https://doi.org/10.1145/3531056.3531071>
- [17] Nikolaev I.E. Knowledge and skills extraction from the job requirements texts. *Ontology of Designing*. 2023;13(2):282-293. (In Russ., abstract in Eng.) <https://doi.org/10.18287/2223-9537-2023-13-2-282-293>
- [18] Nikolaev I.E. 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*. 2023;17(2):71-84. <https://doi.org/10.17323/2587-814X.2023.2.71.84>
- [19] Nguyen K.C., Zhang M., Montariol S., Bosselut A. Rethinking Skill Extraction in the Job Market Domain using Large Language Model. *arXiv:2402.03832*. 2024. <https://doi.org/10.48550/arXiv.2402.03832>
- [20] Bhola A., Halder K., Prasad A., Kan M.-Y. Retrieving skills from job descriptions: A language model based extreme multi-label classification framework. In: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics. Barcelona, Spain (Online): IEEE Press; 2020. p. 5832-5842. <https://doi.org/10.18653/v1/2020.coling-main.513>
- [21] Haq M.U.U., Frazzetto P., Sperduti A., Da San Martino G. Improving Soft Skill Extraction via Data Augmentation and Embedding Manipulation. In: Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing (SAC '24). New York, NY, USA: Association for Computing Machinery; 2024. p. 987-996. <https://doi.org/10.1145/3605098.3636010>



- [22] Clavié B., Soulié G. Large Language Models as Batteries-Included Zero-Shot ESCO Skills Matchers. arXiv:2307.03539. 2023. <https://doi.org/10.48550/arXiv.2307.03539>
- [23] Decorte J.-J., Verlinden S., Hautte J.V., Deleu J., Develder C., Demeester T. Extreme Multi-Label Skill Extraction Training using Large Language Models. In: The International workshop on AI for Human Resources and Public Employment Services (AI4HR&PES) as part of ECML-PKDD. 2023. p. 1-12. Available at: https://ai4hrpes.github.io/ecmlpkdd2023/papers/ai4hrpes2023_paper_173.pdf (accessed 12.01.2024).
- [24] Nguyen K., Zhang M., Montariol S., Bosselut A., Rethinking Skill Extraction in the Job Market Domain using Large Language Models. In: Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024). St. Julian's, Malta: Association for Computational Linguistics; 2024. p. 27-42. Available at: <https://aclanthology.org/2024.nlp4hr-1.3/> (accessed 12.01.2024).
- [25] Fang C., Qin C., Zhang Q., Yao K., Zhang J., Zhu H., Zhuang F., Xiong H. RecruitPro: A Pretrained Language Model with Skill-Aware Prompt Learning for Intelligent Recruitment. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23). New York, NY, USA: Association for Computing Machinery; 2023. p. 3991-4002. <https://doi.org/10.1145/3580305.3599894>

*Поступила 12.01.2024; одобрена после рецензирования 14.02.2024; принята к публикации 18.03.2024.
Submitted 12.01.2024; approved after reviewing 14.02.2024; accepted for publication 18.03.2024.*

Об авторах:

Комарова Любовь Александровна, аспирант, ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации» (125167, Российская Федерация, г. Москва, Ленинградский проспект, д. 49/2), **ORCID: <https://orcid.org/0000-0001-5277-8234>**, 229388@edu.fa.ru

Соловьев Владимир Игоревич, руководитель Департамента анализа данных, принятия решений и финансовых технологий, ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации» (125167, Российская Федерация, г. Москва, Ленинградский проспект, д. 49/2), доктор экономических наук, профессор, **ORCID: <https://orcid.org/0000-0003-0338-1227>**, vsoloviev@fa.ru

Колосов Алексей Михайлович, младший научный сотрудник кафедры математической теории интеллектуальных систем механико-математического факультета, ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова» (119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1), **ORCID: <https://orcid.org/0000-0002-9474-9666>**, akolosov@cs.msu.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Lyubov A. Komarova, Postgraduate Student, Financial University under the Government of the Russian Federation (49/2 Leningradsky Prospekt, Moscow 125167, Russian Federation), **ORCID: <https://orcid.org/0000-0001-5277-8234>**, 229388@edu.fa.ru

Vladimir I. Soloviev, Head of the Department of Data Analysis, Decision-Making and Financial Technology, Financial University under the Government of the Russian Federation (49/2 Leningradsky Prospekt, Moscow 125167, Russian Federation), Dr. Sci. (Econ.), Professor, **ORCID: <https://orcid.org/0000-0003-0338-1227>**, vsoloviev@fa.ru

Alexey M. Kolosov, Junior Researcher of the Chair of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University (1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation), **ORCID: <https://orcid.org/0000-0002-9474-9666>**, akolosov@cs.msu.ru

All authors have read and approved the final manuscript.

