

РАЗРАБОТКА ИНСТРУМЕНТА ДЛЯ КЛАСТЕРИЗАЦИИ ВАКАНСИЙ НАЧАЛЬНОГО УРОВНЯ ПО СКРЫТЫМ ПАТТЕРНАМ ТРЕБОВАНИЙ

Данилевский А.А., Дмитриева Е.А. (ИТМО)

Научный руководитель – ассистент ФПИИ Чернышева А.В. (ИТМО)

Введение

Качество данных на рынке труда является актуальной проблемой: наличие «призрачных» вакансий, завышенных требований и противоречивых условий снижает эффективность поиска работы для начинающих специалистов [1, 2]. Существующие инструменты агрегации вакансий [3–5] фокусируются на фильтрации по формальным параметрам (опыт, зарплата, локация), но не предоставляют возможности интерпретации скрытых паттернов в описаниях, таких как эмоциональная манипуляция, несоответствие требований уровню позиции или скрытые барьеры для новичков.

Цель данной работы – разработать инструмент для кластеризации вакансий начального уровня в сфере IT, позволяющий выявлять скрытые требования и «красные флаги», а также адаптировать подбор вакансий под профиль соискателя с использованием возможностей больших языковых моделей.

Основная часть

Анализ работ с тематикой кластеризации вакансий [6–9] показал, что соответствующие методы в основном использовались для тематической кластеризации в контексте различных профессиональных направлений, что не приносит принципиально новой информации для соискателя, открытые источники [1, 2], содержащие общую информацию о скрытых паттернах, не предоставляют возможности проанализировать большое количество вакансий, что, в частности, не позволяет удостовериться в корректности предоставляемой информации. Поэтому был разработан собственный пайплайн кластеризации, включающий в процессе создания перечисленные упрощения.

Изучение методов кластеризации, векторизации [10, 11] и понимание данных позволили сформулировать несколько гипотез по выявлению скрытых паттернов, указывающих на сложности в рабочих процессах. Для их проверки и достижения цели работы были решены следующие задачи:

- Формирование набора признаков для кластеризации, значимых для пользователя,
- Проведение экспериментов на основании гипотез о детекции негативных скрытых паттернов в описаниях вакансий разными инструментами,
- Обоснование наиболее релевантного метода кластеризации,
- Интерпретация полученных кластеров.

Отдельно был разработан web-интерфейс, агрегирующий результаты предыдущих этапов в удобном и логичном для пользователя формате. Данный этап включил в себя:

- Выбор инструментов, проектирование системы,
- Определение функционала, доступного пользователю,
- Проектирование UI,
- Итоговую реализацию.

Выводы

Разработанный инструмент демонстрирует возможность автоматического выявления скрытых паттернов в описаниях вакансий. Решение позволяет подобрать вакансию начинающему соискателю по гибкому набору параметров. Более того,

благодаря агрегации статистических показателей в визуальном интерфейсе, соискатель может ознакомиться с ситуацией на фрагменте рынка труда в целом. Подход при создании позволяет провести масштабирование на большие объемы данных и адаптацию под различные домены.

Список использованных источников

1. Неоднозначные требования в вакансиях [Электронный ресурс] / Unisender. – URL: <https://www.unisender.com/ru/blog/neodnoznachnye-trebovaniya-v-vakansiyah/#anchor-2> (дата обращения: 10.02.2026).
2. Призрачные вакансии: как их распознать [Электронный ресурс] / Habr. – URL: https://habr.com/ru/companies/beeline_cloud/articles/952526/ (дата обращения: 10.02.2026).
3. Ghost Jobs in Online Labor Markets: Prevalence and Detection [Электронный ресурс] // arXiv. – 2023. – URL: <https://arxiv.org/pdf/2304.02019> (дата обращения: 11.02.2026).
4. 47% россиян готовы поменять работу в 2025 году [Электронный ресурс] / Работа.ру. – URL: <https://press.rabota.ru/47-rossiyan-gotovy-pomenyat-rabotu-v-2025-godu> (дата обращения: 13.02.2026).
5. ANR-Lab представили программу для мультимодального сетевого анализа вакансий [Электронный ресурс] // ВШЭ. – URL: <https://economics.hse.ru/ecjourn/news/1103172640.html> (дата обращения: 14.02.2026).
6. Кластеризация вакансий по их описанию с использованием методов машинного обучения и методов анализа текста [Электронный ресурс] // ЛЭТИ. – URL: <https://scm.etu.ru/assets/files/2023/sbornik/201-204.pdf> (дата обращения: 11.02.2026).
7. Проектирование цифрового сервиса извлечения из текстов вакансий структурированной информации о требованиях к соискателю с использованием технологий обработки естественного языка. [Электронный ресурс] // УРФУ. – URL: https://elar.urfu.ru/bitstream/10995/140536/1/m_th_s.v.savoskina_2024.pdf (дата обращения: 11.02.2026).
8. Терников А. А. Алгоритм кластеризации на основе навыков для онлайн-объявлений о вакансиях // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. — 2022. — Т. 22 — вып. 2. — С. 250-265. (дата обращения: 11.02.26).
9. Гарафиев, И. З., Гарафиева Г. И. Кластеризация вакансий инженеров: методы k-средних и модель гауссовской смеси // Управленческий учет. — 2024. — № 9. — С. 234-240. (Дата обращения 11.02.2026).
10. scikit-learn. [Электронный ресурс] // scikit-learn. – URL: <https://scikit-learn.ru/stable> (дата обращения: 11.02.2026).
11. Sentence transformers. [Электронный ресурс] // Sentence transformers. – URL: <https://www.sbert.net> (дата обращения: 11.02.2026).

Автор _____ Данилевский А.А.
Научный руководитель _____ Чернышева А.В.