

**УДК 004.8**

**ПОДГОТОВКА ДАННЫХ И ВЫЯВЛЕНИЕ СКРЫТЫХ ПАТТЕРНОВ В  
ВАКАНСИЯХ НАЧАЛЬНОГО УРОВНЯ В ИТ-СФЕРЕ**

**Куделенская Э.С., Соклаков Е.А., Суханова М.В. (ИТМО)**

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

**Введение**

Начинающие специалисты на рынке ИТ – одна из уязвимых категорий соискателей. Обилие «призрачных» вакансий, завышенных требований и противоречивых условий делает поиск работы для разработчиков без опыта неэффективным и сложным [1, 2]. Существующие инструменты агрегации вакансий [3–5] содержат фильтрацию в основном по формальным параметрам (опыт, зарплата, локация), но не позволяют увидеть скрытые паттерны в описаниях, такие как эмоциональная манипуляция, несоответствие требований уровню позиции или скрытые барьеры для новичков.

Цель данного проекта – создание пайплайна обработки вакансий начального уровня в сфере ИТ для выявления скрытых требований и «красных флагов». В работе использовался ряд нестандартных признаков вакансий и механизмы больших языковых моделей (LLM, Large Language Model).

**Основная часть**

Был проведен анализ современных инструментов для обработки данных о вакансиях, в ходе которого выяснилось, что большинство решений не поддерживает семантическую интерпретацию текстов описаний и не учитывает контекстные маркеры, важные для начинающих разработчиков. Чтобы устранить эти недостатки, был спроектирован пайплайн обработки данных, который включает:

- Сбор данных. Парсинг вакансий осуществлялся через API HeadHunter с фильтрацией по ключевым словам, а именно позициям начального уровня и ИТ-специальностям. Для каждой вакансии извлекались не только структурированные метаданные из полей JSON (уровень опыта, тип занятости, зарплатная вилка), но и неструктурированные текстовые описания.

- Предобработку текстов. Текстовые поля прошли очистку от HTML-разметки, были приведены к нижнему регистру и нормализованы. С использованием частотного анализа и определения важных для ИТ технических терминов удалось очистить вакансии от стоп-слов, не потеряв основной смысл текстов.

- Feature engineering. Благодаря лингвистическому анализу вакансий и промпт-инжинирингу (на основе LLM-модели DeepSeek-V3) были разработаны словари маркеров для поиска скрытых паттернов, которые не найти в чистом виде на HeadHunter (упоминания в текстах вакансий наставничества, портфолио, софт скиллов и т.д.).

- Поиск «красных флагов». С применением анализа рынка и текстов вакансий был составлен список из 13стораживающих сигналов для начинающего соискателя (неаккредитованность работодателей, несоответствие требуемого опыта заявленному, эмоциональная манипуляция в описаниях, занятость по ГПХ вместо трудового договора и т.д.)

- Статистический анализ. Изучен рынок вакансий с точки зрения количества «красных флагов», их корреляции с зарплатой, с требуемыми навыками и т.д.

**Выводы**

Разработанный пайплайн позволяет автоматически выявлять скрытые паттерны и противоречия в описаниях вакансий еще на этапе подготовки данных. Полученные маркеры и правила определения «красных флагов» позволяют ранжировать предложения по уровню рискованности для начинающих специалистов. Результаты этапа feature engineering и статистического анализа служат основой для последующей

кластеризации и персонализации подбора вакансий, а также позволяют лучше понять специфику рынка IT, актуальную на данный момент. Более того, решение может быть масштабировано на большие объемы данных и адаптировано под различные предметные области, в которых также важен глубинный анализ противоречий.

#### **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/](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).
- Автор \_\_\_\_\_ Куделенская Э.С.  
Научный руководитель \_\_\_\_\_ Чернышева А.В.