

УДК 504.064.36

doi 10.54708/22259309\_2025\_334100

## АНАЛИЗ РЕГИОНАЛЬНЫХ ДИСБАЛАНСОВ НА РЫНКЕ ТРУДА РОССИИ С ИСПОЛЬЗОВАНИЕМ КЛАСТЕРНОГО АНАЛИЗА И ГИС-ТЕХНОЛОГИЙ

С. К. ГАВРИЛЕНКО<sup>1</sup>, Ш. М. МИНАСОВ<sup>2</sup>

<sup>1</sup>job.gavrilenko@ya.ru, <sup>2</sup>minasov@ufanet.ru

ФГБОУ ВО «Уфимский университет науки и технологий» (УУНИТ)

**Аннотация.** Рассматривается проблема территориального дисбаланса между предложением и спросом на рынке труда России. Разработан комплексный программный продукт, включающий модуль сбора данных с сайта «Работа в России», процедуры их нормализации, алгоритмы кластеризации (*TF-IDF + LSA + K-means* с динамическим определением числа кластеров), а также веб-приложение с интерактивной картографической визуализацией. Приложение построено по принципу трехзвенной клиент-серверной архитектуры.

**Ключевые слова:** рынок труда; региональные дисбалансы; кластерный анализ; *LSA*; ГИС; визуализация данных; веб-приложение; трудовая мобильность.

### ВВЕДЕНИЕ

Проблема баланса между спросом и предложением рабочей силы на рынке труда является одной из ключевых социальных и экономических проблем современной России. Наше государство столкнулось с рядом серьезных проблем, вызванных изменениями демографической ситуации, быстрым темпом технологического роста в мире, отраслевыми дисбалансами, внутренней миграцией населения и все еще актуальным для России оттоком высококвалифицированных специалистов за рубеж и притоком рабочей силы из стран постсоветского пространства, квалификация которых оставляет желать лучшего.

Рынок труда в России сегодня характеризуется ощутимым разрывом между квалификацией соискателей рабочих мест и потребностью реального сектора экономики. При этом наблюдаются как нехватка квалифицированных специалистов, так и высокая безработица среди менее квалифицированных граждан.

При этом еще одной проблемой является внутренняя миграция в России, которая создает проблемы с распределением трудовых ресурсов в регионах. Высокая разница в уровне жизни и заработной плате поощряет внутреннюю миграцию и усугубляет социально-экономическую обстановку. Низкая привлекательность удаленных от Москвы и Санкт-Петербурга регионов способствует оттоку трудоспособного населения, усугубляя проблему дефицита кадров даже в успешных отраслях экономики.

Таким образом, эффективное решение вопросов сбалансированного развития рынка труда требует комплексного подхода, включающего меры по повышению квалификации персонала, развитию инфраструктуры, стимулированию инноваций и привлечению инвестиций в регионы.

Актуальность исследования определяется дисбалансом между числом вакансий и соискателей на рынке труда регионов России.

## ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Целью настоящей работы является повышение точности анализа состояния рынка труда путем разработки инструментария, позволяющего наглядно визуализировать общую картину в разрезе регионов России. Гипотезой исследования является идея о том, что визуализация картины спроса и предложения на рынке труда позволит точнее прогнозировать тенденции и разрабатывать меры реагирования для обеспечения региональной экономики кадрами требуемой квалификации.

Для достижения поставленной цели поставлены следующие задачи:

- предложить технологию сбора и обработки данных об имеющихся на рынке труда вакансиях и предложении рабочей силы;
- разработать математическую модель консолидации сырых данных;
- разработать алгоритм решения задачи группировки исходных данных;
- разработать программное обеспечение для наглядной визуализации данных.

## АРХИТЕКТУРА

Разрабатываемая система построена по мультиагентной технологии, где каждый из агентов строится по двухзвенной (клиент-сервер базы данных) или классической трехзвенной архитектуре (клиент – сервер приложения – сервер базы данных). Агенты представляют собой независимые модули сбора, хранения, анализа и визуализации данных. Сервер приложения выполнен на базе веб-сервера *Apache*, передающего интерпретатору *PHP* запросы на обработку сценариев по запросу пользователей веб-приложения, где клиентской частью выступает браузер. Кроме демонстрации сгенерированных интерпретатором *PHP* веб-страниц, сервер возвращает клиенту результат обработки с интерпретируемым кодом на *JavaScript*. Реализованные на *PHP* с использованием *MVC*: контроллеры обрабатывают *HTTP*-запросы. Центральным элементом модели выступает сервер базы данных *MariaDB* (*MySQL*). Данные, извлекаемые из базы данных, формируются в структуру в формате *JSON*. Выбор *MariaDB* обусловлен ее производительностью и масштабируемостью. Клиентская часть – *SPA* на *JavaScript* (*ES6+*), где *ApiClient* обеспечивает асинхронный обмен данными и обновление при изменении фильтров. Структура агентов представлена на рис. 1.

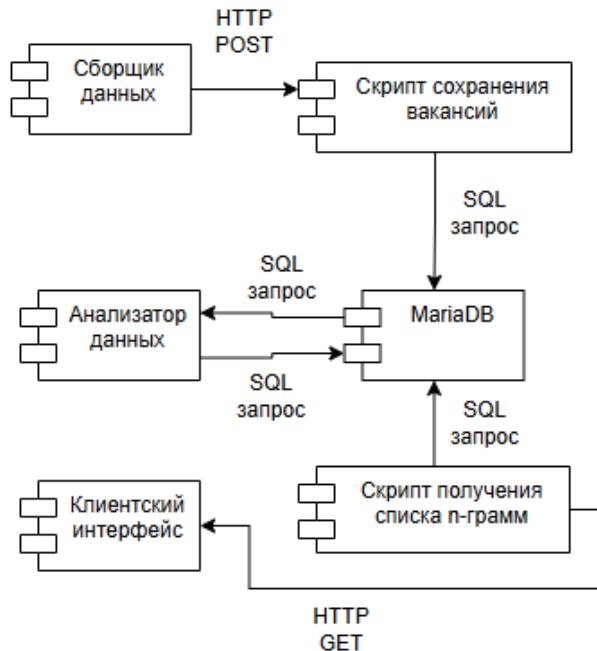


Рис. 1 Диаграмма компонентов

Объектно-ориентированная модель включает сущности: «Вакансия», «Регион», «Кластер» и «Пользовательский запрос». Принцип открытости/закрытости реализован через иерархию: базовый «Кластер» обрабатывает сходство и статистику, подклассы – «Кластер n-грамм» и «Кластер тематических векторов» – извлекают специфические признаки. Диаграмма классов веб-формы представлена на рис. 2.

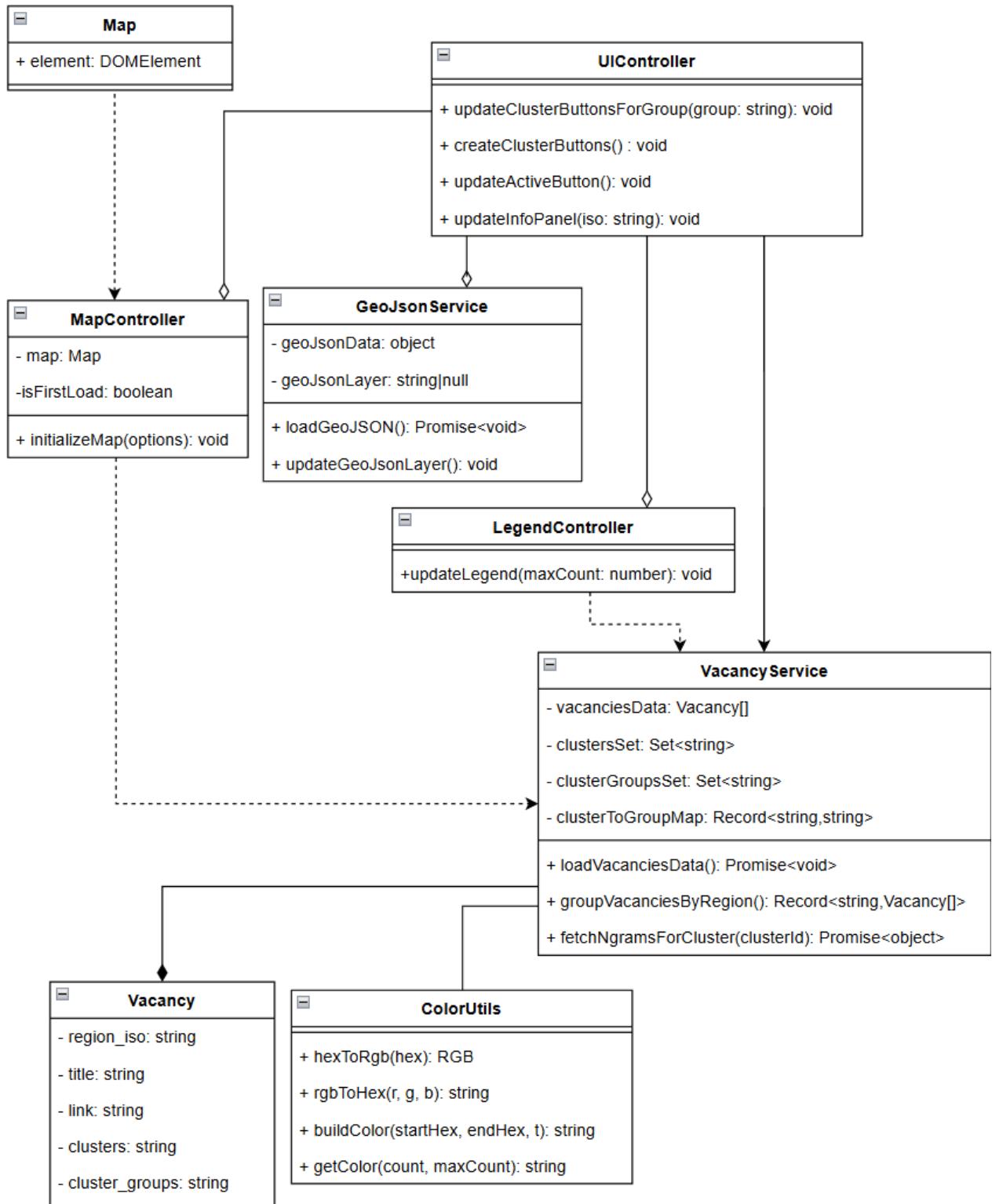


Рис. 2 Диаграмма классов веб-формы

## ИНФОРМАЦИОННАЯ МОДЕЛЬ БАЗЫ ДАННЫХ

Информационная модель, представленная *ER*-диаграммой на рис. 3, отражает сущности и отношения между ними, а оптимизированные составные индексы по полям «регион», «дата обновления» и «уровень зарплаты» позволяют быстро формировать тепловые карты и временные ряды.

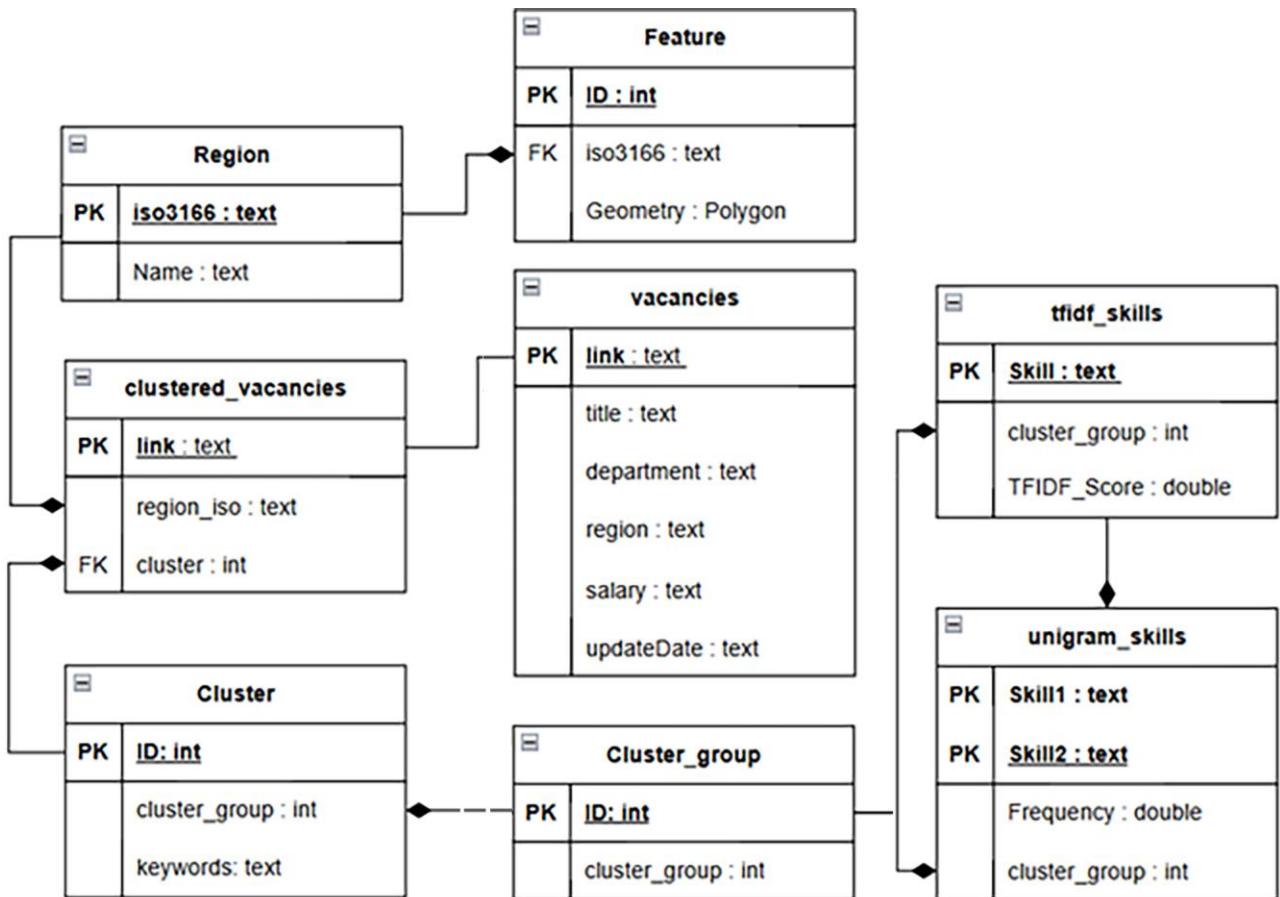


Рис. 3 *ER*-модель базы данных проекта

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Математическая модель анализа региональных дисбалансов рынка труда базируется на формализации текстов вакансий в виде векторного пространства и последующей кластеризации с учётом пространственного распределения. Каждая вакансия  $v_i$  представляется как набор ключевых компетенций, извлекаемых из текстов с применением *TF-IDF* и снижения размерности через латентно-семантический анализ (*LSA*).

Для построения модели сначала формируется матрица "документ–термин"  $D \in R^{n \times m}$ , где  $n$  – число вакансий, а  $m$  – число уникальных терминов после токенизации и очистки. Значения ячеек  $d_{ij}$  определяются по формуле:

$$tf - idf(t_j, v_i) = \frac{f_{ij}}{\sum_k f_{ik}} \log \left( \frac{N}{n_j} \right),$$

где  $f_{ij}$  – частота термина  $t_j$  в вакансии  $v_i$ ,  $N$  – общее число вакансий,  $n_j$  – число вакансий, содержащих термин  $t_j$  [3].

Для устранения синтаксического шума и повышения обобщающей способности применяется сингулярное разложение:  $D \approx U_k \Sigma_k V_k$ , где сохраняются только  $k$  наиболее

значимых компонент. Это даёт компактное представление вакансий в  $k$ -мерном семантическом пространстве, где расстояния между точками отражают тематическую близость. Полученные векторы подаются на вход алгоритму  $K$ -средних, который минимизирует внутрикластерное отклонение:

$$\arg \min_c \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

где  $\mu_i$  – центр кластера  $C_i$ ,  $x$  – вектор вакансии. Число кластеров  $k$  подбирается динамически по максимальному значению силуэтного коэффициента на интервале  $k \in [3, 10]$ .

Каждая вакансия также содержит координату региона (в виде ISO-кода), что позволяет агрегировать результаты кластеризации по регионам. На основе долей кластеров по субъектам строятся карты региональных специализаций. Это создаёт модель распределения компетенций по территории страны, выявляя структурные несоответствия между спросом и предложением.

## РЕАЛИЗАЦИЯ

Рабочий процесс начинается с парсинга HTML-страниц портала «Работа всем» [1] с помощью асинхронной IIFE-функции *JavaScript*, имитирующей нажатие «Загрузить еще» до исчерпания вакансий [2]. Из DOM-дерева извлекаются данные: название, регион, зарплата, дата, URL. Далее данные нормализуются и валидируются.

Кластеризация проводится по навыкам и компетенциям. Тексты описаний токенизируются, лемматизируются, после чего строится *TF-IDF*-матрица [3]. Метод *LSA* снижает размерность, устраняя шум. Оптимальное число кластеров определяется по индексу силуэта, затем применяется *K-means*. Результаты связываются с *GeoJSON*-регионами: рассчитываются доли вакансий по кластерам и используется градиентная окраска [4].

*SPA*-клиент реализует всплывающие подсказки и топ-3 n-граммами по региону. n-граммы запрашиваются через *get\_top\_ngrams.php* с *SQL*-лимитом в 3 записи. Обмен с сервером построен на *REST* и *Fetch API* с *async/await*, обеспечивает надежность и читаемость кода.

Экранные формы разработаны с учётом базовых принципов человека-машиинного взаимодействия и архитектуры *MVC*, что обеспечивает интуитивную доступность и консистентность элементов управления.

Перечень компонентов экранных форм включает: Заголовок и навигационную панель, Поле глобального поиска; Интерактивную карту; Панель фильтрации и кластеризации; Панель списка вакансий по региону; Легенду карты (рис. 4).

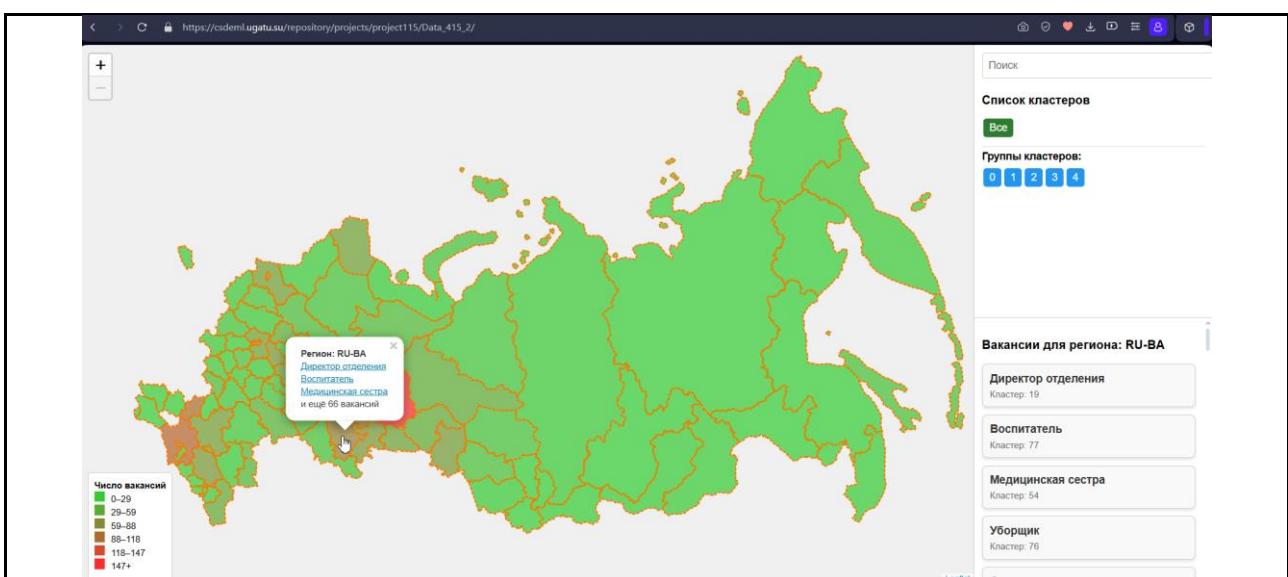


Рис. 4 Скриншот интерфейса веб-страницы

## ЗАКЛЮЧЕНИЕ

Проведенная работа подтвердила исходную гипотезу: интеграция тематической кластеризации вакансий с их географической привязкой позволяет выявить скрытые региональные диспропорции и значительно повысить информативность анализа рынка труда РФ. В результате были созданы адаптивные алгоритмы кластеризации (*K-means + LSA + динамический подбор числа кластеров*), обеспечены надежность и непротиворечивость данных благодаря унифицированной базе вакансий с точными региональными метками, а также разработан отказоустойчивый веб-сервис на основе *PHP/MariaDB* и *SPA*-приложения с собственной картографической системой. Механизмы виртуализации слоев и кэширования сократили время отклика интерфейса, а историческая версионизация данных позволила анализировать динамику изменений. Итоговая платформа предоставляет региональным властям и кадровым агентствам средства для оперативного выявления «узких мест» рынка труда и разработки целевых программ по стимуляции трудовой мобильности и адаптации образовательных курсов в перспективных субъектах федерации.

## СПИСОК ЛИТЕРАТУРЫ

1. Бызов А. А. Интеллектуальный анализ текстов в социальных науках / А. А. Бызов // Социология: методология, методы, математическое моделирование. 2019. № 49. С. 131–160. EDN GCIIVL.
2. Дудина Н. Н. Геоинформационная система как средство визуализации данных о безработице в Республике Беларусь / Н. Н. Дудина, Ю. В. Фролова // Научные записки молодых исследователей. 2019. Т. 7. № 6. С. 66–74. EDN FSHWFD.
3. Макаревич Т. И. Интеллектуальный анализ текстовой информации в специализированных областях в системе электронного правительства / Т. И. Макаревич // Цифровая трансформация. 2019. № 2. С. 46–52. EDN JPWQQI.
4. Долгодворова Е. В. Кластерный анализ: базовые концепции и алгоритмы / Е. В. Долгодворова // Вопросы науки и образования. 2018. № 7(19). С. 73–76. EDN XPICXB.
5. Сидорчук Р. Р. Геоинформационные системы в маркетинге предпринимательских структур / Р. Р. Сидорчук // Маркетинг MBA. Маркетинговое управление предприятием. 2019. Т. 10. № 1. С. 97–114. EDN SNBJGU.

## ОБ АВТОРАХ

**ГАВРИЛЕНКО Сергей Константинович**, студ. каф. ГИС, УУНиТ.

**МИНАСОВ Шамиль Маратович**, доцент, к. техн. наук., доцент кафедры ГИС, УУНиТ.

## METADATA

**Title:** Analysis of Regional Labor Market Imbalances in Russia Using Cluster Analysis and GIS Technologies

**Author:** S.K. Gavrilenko<sup>1</sup>, Sh.M Minasov<sup>2</sup>

**Affiliation:**

<sup>1,2</sup> Ufa University of Science and Technology (UUST), Russia.

**Email:** <sup>1</sup>job.gavrilenko@ya.ru, <sup>2</sup>minasov@ufanet.ru

**Language:** Russian.

**Source:** Molodezhnyj Vestnik UGATU (scientific journal of Ufa University of Science and Technology), no. 3 (34), pp. 100-104, 2025. ISSN 2225-9309 (Print).

**Abstract:** This paper addresses the problem of territorial imbalance between supply and demand in the Russian labor market. We have developed an integrated software tool comprising a data-collection module for the “Rabota v Rossii” portal, data normalization procedures, clustering algorithms (TF-IDF, LSA, K-means with dynamic determination of cluster count), and a web application featuring interactive cartographic visualization. The system architecture follows a client-server model implemented in PHP, MariaDB and JavaScript (SPA).

**Key words:** labor market; regional imbalances; cluster analysis; LSA; GIS; data visualization; web application; labor mobility.

**About authors:**

**GAVRILNKO Segey Konstantinovich**, student, Dept. of GIS (UUST).

**MINASOV Shamil Maratovich**, associate professor, candidate of technical sciences, associate professor of the Dept. of GIS.