УДК 004-032-26
https://www.doi.org/10.47813/rosnio.5.2026.19.2012
EDN DIIRNK

Разработка веб-приложения для совместного проживания

О.А. Митина*, **А.В. Шрейдер**

МИРЭА – Российский технологический университет, проспект Вернадского, 78, Москва, 119454, Россия

*E-mail: alogmi@yandex.ru

Аннотация. В статье рассматривается проблема поиска соседей для совместной аренды жилья в условиях высокой миграции населения в крупные города и отсутствия специализированных инструментов подбора. Актуальность исследования обусловлена ростом спроса на совместную аренду и низкой эффективностью существующих способов поиска через социальные сети, где объявления представлены в неструктурированном виде и требуют значительных временных затрат на анализ. Научная новизна работы заключается в разработке веб-приложения, объединяющего механизмы структурированного хранения объявлений, интеллектуальной фильтрации и ранжирования кандидатов на основе оценки сходства пользовательских профилей. Предложен подход к построению рекомендательной системы, включающий предварительную фильтрацию несовместимых пользователей и последующее вычисление комплексной функции сходства с использованием категориальных, множественных и числовых признаков. Особое внимание уделено процессу агрегации данных из социальных сетей, применению LLM-моделей для семантического анализа текстовых описаний постов и автоматического извлечения характеристик пользователей без необходимости формирования обучающей выборки. Оценка качества извлечения категориальных признаков рассмотрена через призму многоклассовой классификации.

Ключевые слова: веб-приложение, совместное проживание, агрегация данных, семантический анализ, многоклассовая классификация, ранжирование.

Developing a web application for co-living

O.A. Mitina*, **A.V. Shreider**

MIREA – Russian Technological University, Vernadsky Avenue, 78, Moscow, 119454, Russia

*E-mail: alogmi@yandex.ru

Abstract. The article examines the problem of finding neighbors for joint rental housing in conditions of high population migration to large cities and the lack of specialized selection tools. The relevance of the study is due to the growing demand for shared rentals and the low efficiency of existing search methods through social networks, where ads are presented in an unstructured form and require significant time for analysis. The scientific novelty of the work lies in the development of a specialized web application that combines mechanisms for structured ad storage, intelligent filtering and candidate ranking based on an assessment of the similarity of user profiles. An approach to building a recommendation system is proposed, including preliminary filtering of incompatible users and the subsequent calculation of a complex similarity function using categorical, multiple and numerical features. Special attention is paid to the process of aggregating data from social networks, the use of LLM models for semantic analysis of text descriptions of posts and automatic extraction of user characteristics without the need to form a training sample. The assessment of the quality of categorical feature extraction is considered through the prism of a multiclass classification.

Keywords: web application, cohabitation, data aggregation, semantic analysis, multiclass classification, ranking.

1. Введение

Поиск соседа – проблема, с которой сталкиваются студенты, молодые специалисты и другие группы населения. Их объединяет желание сэкономить на аренде. Наиболее остро это проблема стоит в городах-миллионниках. По данным Росстата, за 2023 год в одну только Москву преехало сыше 290 тысяч мигрантов в пределах России, однако финансовые возможность арендовывать жилье в одиночку есть далеко не у всех [1].

Популярные базы данных (БД) недвижимости (Циан, Авито) не предоставляют такой опции. В подавляющем большенстве случаев поиск ведётся в тематических сообществах в социальных сетях, в частности, в ВКонтакте суммарная аудитория подобных групп свыше 45 тысяч пользователей [2]. Важнейший недостаток подхода – большое количество сообществ и отсутсвие классифицирующей информации о потенциальном соседе, из-за чего пользователю приходится тратить много времени на просмотр нерелевантных объявлений.

Решением проблемы является разработка специализированного веб-приложения, которое позволит хранить данные объявлений в структурированном виде – с выделением ключевых признаков, как числовых (бюджет, возраст), так и категориальных (пол, наличие или отсутсвие квартиры, вредные привычки и т. д.) для последующей фильтрации, сортировки по запросу пользователя, а также агрегировать объявления из открытых социальных сетей. Помимо этого, необходимо реализовать возможность ранжировать выборку по уровню сходства с объявлением для персонализированных рекомендаций.

Результатом исследования является разработанное и готовое к введению в эксплутацию MVP (Minimum Viable Product) – «минимально жизнеспособный продукт» для тестирования функционала реальными пользователями.

2. Постановка задачи (Цель исследования)

Цель исследования – проанализировать наиболее востребованные характеристики, на которые обращают внимания пользователи при отборе кандидатов; спроектировать и разработать базу данных для хранения объявлений, рассматривая эти характеристики как таблицы-справочники для центральной таблицы-фактов (объявлений), а также реализовать бизнес-логику, включая извлечение признаков из текстового описания постов.

В случае введения в эксплуатацию, разработанное решение позволит существенно сократить время поиска наиболее подходящего соседа и, как следствие, финансовые потери, связанные с временным проживанием в одиночку.

3. Методы и материалы исследования

В 2025 году сервисы «Литрес» и «Авито» совместно проводили опрос об аренде жилья, в том числе о факторах, влияющий на выбор соседа [3]. Среди них можно выделить следующие: финансовые ожидания, образ жизни, вредные привыки, психологическая совместимость, географические предпочтения. На рисунке 1 продемонстрирована модель базы данных, где полезные для отбора признаки выделены в отдельные таблицы-справочники.

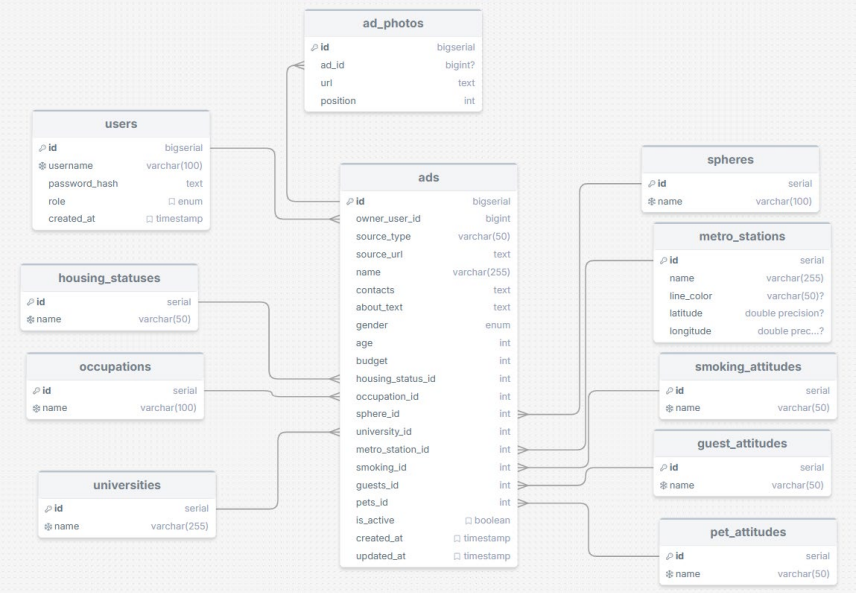

Рисунок 1. Модель базы данных.

Для веб-приложения используется клиент-серверная архитектура. Парсинг и семантический анализ текста вынесены в отдельный микро-сервис аграгации. Серверное приложение написано на языке Java с помощью фреймворка Spring Boot, клиентская часть на Vue.js, а сервис аграгации с помощью Python.

Серверная часть реализует стандартную CRUD-логику средствами экосистемы Spring с учётом особенностей хранения данных. Детали шаблонного кода не рассматриваются, как не представляющие научный интерес, фокус исследования сделан на модуле рекомендации и сервиса агрегации данных.

Поскольку в контексте данной разработки отсутствуют исторический данные, о взаимодействиях, используем статические методы рекомендации. Модуль рекомендаций работает в два этапа: на первом из выборки исключаются однозначно несовместимые пользователи (например, исключаются курящие пользователи для тех, кто резко против курения). Формируется подмножество A_{valid} из множества A:

$$A_{valid} = \{ a \in A \mid \bigwedge_{k=1}^{K} P_k(u, a) \},$$

где P_k – логический предикат, проверяющий соответствие k-го критического атрибута, A – совокупность всех активных объявлений. Фильтрация описывается системой бинарных предикатов.

Второй этап – ранжирования полученной выборки A_{valid} на основании комплексной оценки сходства $S(u,a)$ между пользовательским объявлением и объявлением из списка [4]. Функция представляет собой взвешенную сумму частных критериев (функций подобия каждого признака):

$$S(u,a) = \sum_{i=1}^{n} w_i \cdot s_i(u,a),$$

где n – общее количество частных критериев подобия;

w_i – весовой коэффициент i-го критерия;

$s_i(u,a)$ — нормированное значение i-го частного критерия подобия.

Весовые коэффициенты w_i отражают приоритетность признаков для предметной области (например, бюджет и локация получают более высокие веса, чем университет или конкретные хобби). Значения $S(u,a)$ лежат в интервале [0,1], что обеспечивает интерпретируемость и устойчивость к выбросам. Ранжирование выполняется по убыванию значения функции $S(u,a)$ на множестве A_{valid}.

Для параметров с фиксированным количеством значений категориальных признаков критерием подобия выступает бинарная функция совпадения. Для множественных признаков используется коэффициент Жаккара, показывающий отношение между количеством совпавших признаками и суммарным количеством признаков пользовательского и целевого объявления [5]. Для числовых признаков (бюджет, возраст) рассчитывается нормированное линейное сходство в упрощенном виде, как отношение диапазона рассматриваемой пары объявлений к максимальному диапазону по всей выборке [6]. Каждый частный критерий подобия нормирован, где 1 означает полное совпадение.

Оценку качества рекомендаций можно дать только после введения MVP в эксплуатацию для получения обратной связи в виде данных о пользовательских взаимодействиях для корректировки весовых коэффициентов.

Модуль семантического анализа направлен на изучение и интерпретацию значений слов, фраз и предложений в их контексте. Это позволяет машинам не просто распознавать текст, но и понимать его смысл. В контексте данной работы под семантическим анализом подразумевается извлечение из текстовых описаний постов в социальных сетях различных полезных признаков, таких как пол, сфера дейтельности, отношение к курению и т. д. [7].

Значения признаков рассматриваются по каждой категории как классы, где null также допустим (в исходном посте может отсутсвовать информация в признаке). В такой парадигме извлечение допустимого значения является задачей многоклассовой классификации.

Существует множество методов семантического анализа, основанных на моделях машинного обучения: Bow, TF-IDF, Word2Vec, FastText, BERT и другие [8]. Однако для извлечения сразу девяти категориальных признаков и двух числовых применение представленных моделей требует тонкой настройки и комбинации различных практик. Перспективным направлением в рамках подобной задачи является использование LLM (Large Language Model)-модели [9]. Данный подход подходит в контексте разрабатываемого решения, поскольку отсутсвует выбучающая выборка достаточных объемов.

Гипотеза заключается в том, что LLM-модель, сможет показать сопоставимый результат по точности При этом разработчику даже не обязательно знать детали реализации методов извлечения. LLM инкапсулирует всю логику и возращает готовый результат, поддающийся оценке.

4. Полученные результаты

В качестве метрик оценки извлечения использовалась метрика макро-F1 по каждому категориальному признаку, где каждое значение признака рассматривалось как отдельный класс с соответсвующей F1-мерой. Итоговая оценка определялась как среднее арифметическое по всем девяти атрибутам. Для числовых признаков применяется метрика Exact Match Ratio (EMR), отражающая долю полностью совпавших значений, поскольку отклонения диапазанов в рамках задачи не важны [10]. Дополнительно рассчитывался EMR для объявления в целом с учётом всех признаков.

Результаты оценки на подвыборке из 50 размеченных вручную объявлениях результаты приведены в таблице 1.

Таблица 1. Результаты классификации.

Метрика	Фактическое значение
Средняя Макро-F1 категориальных признаков	0,8989
Средняя EMR числовых признаков	0,9700
EMR полного объявления	0,6200

LLM продемонстрировала высокое качество извлечения, несмотря на отсутствие дополнительного обучения. Применение данного подхода особенно эффективно при работе с небольшими датасетами, поскольку скорость извлечения ниже традиционных подходов.

Наглядно процесс парсинга объявлений и извлечения признаков представлен на рисунках 2-3.

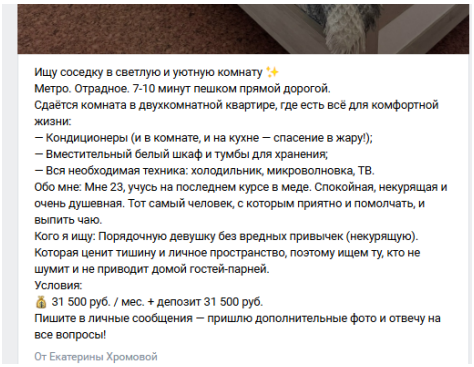

Рисунок 2. Исходный пост из социальной сети VK.

Исходный текст поста содержит как полезную информацию, так и эмоциональные оценки, затрудняющие автоматическое извлечение признаков.

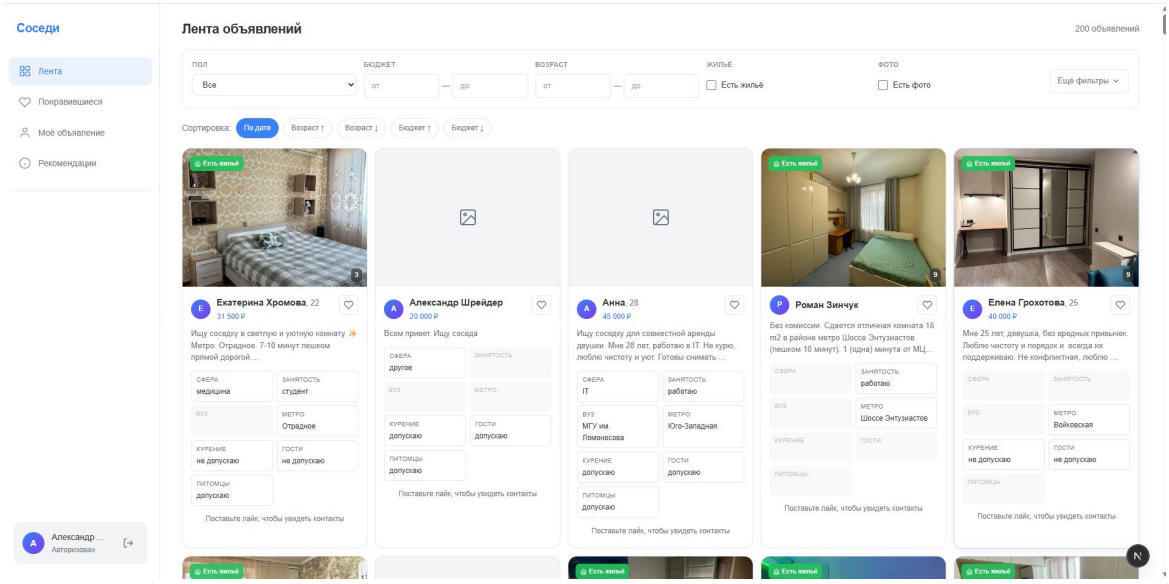

Рисунок 3. Главная страница веб-приложения.

На главной странице веб-интерфейс отображает агрегированные объявления из социальных сетей с успешно извлеченными категориальными и числовыми признаками, а также оригинальные объявления, созданные в веб-приложении. Например, первое слева объявление – результат агрегации приведённого ранее поста из ВКонтакте.

Верхняя панель предоставляет пользователю доступ к функции фильтрации и сортировки по ключевым параметрам, включая бюджет, возраст, наличие жилья, отношение к курению и другие, позволяя исключить нерелевантные объявления еще на этапе поиска.

Дополнительно реализован модуль рекомендаций, выполняющий ранжирование на основании рассчитанного уровня сходства между пользователями. Навигационная панель обеспечивает быстрый доступ к основным разделам веб-приложения, включая страницу рекомендаций, список понравившихся объявлений и форму создания собственных публикаций.

5. Выводы

Разработано веб-приложение для поиска соседей по аренде жилья, реализующее хранение структурированных объявлений, фильтрацию, сортировку и механизм персонализированных рекомендаций. Достигнутые значения метрик подтверждают применимость использование LLM-модели для задачи извлечения признаков в случае отсутствия обучающей выборки.

Предложенные методы и архитектурные решения могут использоваться в других предметных областях, связанных с поиском и подбором пользователей на основе текстовых описаний и критериев совместимости

Список литературы

1. Численность и миграция населения Российской Федерации // Росстат. – URL: https://www.rosstat.gov.ru/compendium/document/13283 (дата обращения: 20.05.2026).
2. Поиск подписчиков групп ВКонтакте // BARKOV. – URL: https://vk.barkov.net/members.aspx (дата обращения: 20.05.2026).
3. Большинство студентов (64%) в 2025 году выбирали обособленные форматы проживания // ВЕДОМОСТИ. – URL: https://www.vedomosti.ru/press_releases/2025/12/12/bolshinstvo-studentov-64-v-2025-godu-vibirali-obosoblennie-formati-prozhivaniya-erid-2VfnxvVfoa1 (дата обращения: 20.04.2026).
4. Грубый, С.В. Оптимизация механической обработки / С.В. Грубый. – 1-е изд. – СПб. Издательство Лань, 2020. – 140 с.
5. Баженов, А.Н. Обобщение коэффициента Жаккара для анализа данных с интервальной неопределённостью / А.Н. Баженов, А.Ю. Тельнова // Измерительная техника. – 2022. – № 12. – С. 15-22. – DOI: 10.32446/0368-1025it.2022-12-15-22. – EDN: LHKUAN.
6. Гайдамакин, Н.А. Мера сходства последовательностей одинаковой размерности / Н.А. Гайдамакин // Математические структуры и моделирование. – 2016. – № 4 (40). – С.

5-16. – EDN: XBSJMV. – URL: https://cyberleninka.ru/article/n/mera-shodstva-posledovatelnostey-odinakovoy-razmernosti (дата обращения: 22.05.2026).

7. Recent Advances in Computational Methods in Science and Technology / S. Kaur, A. Kaur, M. Kumar // CRC Press. – 2026. – Vol. 48. – P. 330-336.

8. Кушнер, К.В. Семантический анализ / К.В. Кушнер, О.И. Захарова // Научный Лидер. – 2024. – № 36 (186). – URL: https://scilead.ru/article/7061-semanticheskij-analiz (дата обращения: 22.04.2026)

9. Шабля, В.О. Методы семантического анализа на основе моделей машинного обучения с использованием искусственных нейронных сетей / В.О. Шабля, С.А. Коноваленко, Е.О. Орлов, Н.А. Галямин // Наука и реальность. – 2025. – № 1(21). – С. 113-122. – EDN: VWEZNG. – URL: https://cyberleninka.ru/article/n/metody-semanticheskogo-analiza-na-osnove-modeley-mashinnogo-obucheniya-s-ispolzovaniem-iskusstvennyh-neyronnyh-setey (дата обращения: 22.04.2026).

10. MatchXML: An Efficient Text-Label Matching Framework for Extreme Multi-Label Text Classification / H. Ye, R. Sunderraman, Sh. Ji // IEEE Access. – 2024. – Vol. 36. – P. 4781-4793. – DOI: 10.48550/arXiv.2308.13139.