

ОРИГИНАЛЬНАЯ СТАТЬЯ



DOI: 10.26794/2408-9303-2026-13-4-97-110

УДК 338.27(045)

JEL C31, C55, C81

Сбор и обработка пространственных данных для решения задач размещения и планирования

Н.В. Гринева^а, А.Д. Топыркин^б^а Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация;^б Российская академия народного хозяйства и государственной службы при Президенте РФ, Москва, Российская Федерация

АННОТАЦИЯ

Актуальность статьи состоит в необходимости использования качественных пространственных данных для решения задачи размещения объектов торговли и инфраструктуры в условиях ограниченной доступности открытых геоаналитических источников. **Целью** исследования является разработка и систематизация подходов к сбору, обработке и агрегации пространственных данных из различных источников для последующего построения экономико-математических моделей в рамках решения поставленной задачи. В статье рассмотрены основные типы пространственных данных: сведения о домохозяйствах (ГИС ЖКХ), стоимости недвижимости (Авито недвижимость), зданиях и дорожной сети (OpenStreetMap), организациях и конкуренции (Яндекс Карты), а также о пешеходном и автомобильном трафике и социально-демографической структуре населения (Яндекс Геоаналитика). Для каждого источника описан процесс получения данных, необходимые инструменты и методы обработки, обеспечивающие пространственную привязку и возможность расчета факторов в заданных радиусах. **Практическая значимость.** Предложенный в исследовании подход позволяет формировать богатое пространство объясняющих переменных для моделирования экономических показателей локаций. Выявленные факторы будут использованы в дальнейших работах для сравнительного анализа результатов моделирования.

Ключевые слова: пространственные данные; проблема размещения; домохозяйства; стоимость недвижимости; организации; дорожная сеть; автомобильный трафик; пешеходный трафик

Для цитирования: Гринева Н.В., Топыркин А.Д. Сбор и обработка пространственных данных для решения задач размещения и планирования. *Учет. Анализ. Аудит = Accounting. Analysis. Auditing.* 2026;13(4):97-110. DOI: 10.26794/2408-9303-2026-13-4-97-110

ORIGINAL PAPER

Collection and Processing of Spatial Data for Solving Location and Planning Problems

N.V. Grineva^а, A.D. Topyrkin^б^а Financial University under the Government of the Russian Federation, Moscow, Russian Federation;^б Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation

ABSTRACT

The relevance of the article lies in the need to use high-quality spatial data for solving location problems of retail and infrastructure facilities under conditions of limited access to open geospatial sources. **The main objective** of the research is to develop and systematise approaches to collecting, processing, and aggregating spatial data from various sources for subsequent construction of economic-mathematical models within the framework of the posed location problem. The article examines the main types of spatial data via household information (GIS Housing and Utilities), real estate prices (Avito Real Estate), data on organisations and competition (Yandex Maps), building and road network data (OpenStreetMap), as well as pedestrian and car traffic data and socio-demographic structure of the population (Yandex Geoanalytics). For each source, the data collection process is described, along with the necessary tools and processing methods that ensure georeferencing and the ability to calculate features within given radii. **Practical implications.** The approach proposed in the study allows for the formation of a rich set of explanatory variables for modeling the economic indicators of locations. The obtained factors will be used in future work for comparative analysis of modeling results.

Keywords: spatial data; location problem; households; real estate prices; organizations; road network; car traffic; pedestrian traffic

For citation: Grineva N.V., Topyrkin A.D. Collection and processing of spatial data for solving location and planning problems. *Uchet. Analiz. Audit = Accounting. Analysis. Auditing.* 2026;13(4):97-110. DOI: 10.26794/2408-9303-2026-13-4-97-110

© Гринева Н.В., Топыркин А.Д., 2026

ВВЕДЕНИЕ

Современные исследования в области задач пространственного размещения все чаще опираются на количественный анализ данных, привязанных к местоположению. Принятие решений о размещении торговых объектов, точек оказания услуг, объектов социальной инфраструктуры, жилых комплексов требует учета множества факторов: от структуры локального спроса и уровня конкуренции до доступности транспортной сети и характеристик пешеходного трафика. В последние десятилетия распространение географических информационных систем (ГИС), открытых картографических платформ и публичных API создало принципиально новые возможности для получения и обработки пространственных данных. Однако на практике реализация таких проектов сталкивается с рядом методических и технических ограничений.

Во-первых, отсутствует универсальное определение и классификация пространственных данных, а также единый перечень их источников. Сведения могут поступать из государственных реестров (кадастровой и налоговой информации, реестра жилищного фонда), открытых картографических платформ (OpenStreetMap), коммерческих платформ (агрегаторы объявлений о продаже недвижимости, сервисы геоаналитики), а также формироваться на основе обращений к различным API и инструментам. Каждый источник имеет собственную структуру данных, степень их детализации, формат пространственной привязки и т.д. Во-вторых, даже при наличии сырых данных, возникает проблема их обработки и агрегации до уровня, пригодного для решения конкретной задачи размещения. Например, точечные сведения об организациях, адреса домохозяйств или гексагональные сетки трафика необходимо преобразовывать в набор факторов, вычисляемых в заданных радиусах или изохронах. В-третьих, значительная часть ценной информации (социально-демографический профиль населения, интенсивность автомобильного и пешеходного трафика, транзакционная активность) не находится в открытом доступе и может быть получена только через коммерческие геоаналитические сервисы или их публичные демоверсии с ограниченным покрытием.

ПРОСТРАНСТВЕННЫЕ ДАННЫЕ

Определений и трактовок для понятия «пространственные данные» много. Каждый источник выделяет какие-то свои особенности, которые должны быть

отмечены в определении, поэтому единого мнения по этому вопросу не существует. Зачастую под этим термином понимаются и другие, схожие с ним. Для примера рассмотрим, как трактуют это понятие различные государственные документы и авторы научных работ.

Министерство промышленности и энергетики Российской Федерации в лице Федерального агентства по техническому регулированию и метрологии (Росстандарт) Приказом от 29.12.2005 № 423-ст¹ утвердило национальный стандарт ГОСТ Р 52438–2005² «Географические информационные системы. Термины и определения», действующий с 1 июля 2006 г. В нем устанавливается следующее: «пространственные данные (Нрк. геоинформационные данные, геопространственные данные, географические данные, геоданные) — данные о пространственных объектах и их наборах».

Технический комитет по стандартам ISO/TC 211, созданный в рамках Международной организации по стандартизации (ISO) и занимающийся вопросами цифровой географической информации в стандарте ISO 19109:2025³ «Geographic information — General feature model and rules for application schema», действующем с 8 июля 2025 г., утвердил следующее определение: «географические данные — данные с неявной или явной ссылкой на местоположение относительно Земли».

Теперь рассмотрим примеры того, как различные исследователи использовали пространственные данные для решения задачи размещения в области экономики, торговли, общества, недвижимости и строительства.

С. Дибб и Л. Симкин из Уорикского университета в своей статье описали историю теоретических подходов, которые применялись при решении указанной задачи для размещения магазина или другого объекта в рамках сферы торговли. В роли пространственных данных в их исследовании выступили: количество конкурентов, расстояние до них, параметры каннибализации (количество собственных магазинов и расстояние до них) [1].

М.Х. Сатман и М. Алтунбей из Стамбульского университета занимались решением проблемы размещения и открытия новых книжных магази-

¹ URL: https://www.consultant.ru/document/cons_doc_LAW_258522

² URL: <https://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=EXP&n=406294#P06SEQV4ngiT3b2C>

³ URL: <https://www.iso.org/standard/84700.html>

нов в рамках сети NT Book Stores в Турции. Авторы использовали географическую информацию о количестве различных организаций и объектов в радиусе 500 метров от книжных магазинов. Источником данных был Google Places API [2].

М. Сюй и его коллеги из университета Цинхуа решали задачу по определению потенциального места для открытия нового магазина и гостиницы в Пекине. В процессе моделирования авторы использовали данные запросов из Baidu Maps, точки интереса (POI), сведения о подключении к WiFi [3].

Д. Карамшук и его коллеги из Школы передовых исследований ИМТ в Лукке и Кембриджского университета описывают выбор оптимальных мест для розничных магазинов с использованием данных из сервиса Foursquare и методов машинного обучения. Сами данные представляют собой информацию по чекинам пользователей в различных местах в центре Нью-Йорка, собранную с помощью API Twitter [4].

Н. Ройг-Тьерно и его коллеги из политехнического университета Валенсии занимались разработкой методологии выбора местоположения для розничных магазинов с использованием ГИС на примере ArcGis и метода аналитической иерархии (АНР). В работе использовались: положения и расстояния до 100 магазинов конкурентов, кадастровые данные для расчета количества квартир на городской квартал, муниципальные сведения о переписи населения для оценки количества жителей на городской блок [5].

Э. В.Л. Ченг, Х. Ли, Л. Юй из технологического университета Квинсленда применяли геоинформационные системы (GIS) для выбора оптимального местоположения торговых центров. В качестве пространственных данных в работе используется информация о границах районов, улиц; дорожной сети; расположении торговых центров; среднемесячных доходах домохозяйств по районам; среднемесячных расходах в точках концентрации населения; количестве магазинов; конкуренции; посещаемости других ТЦ; транспортной доступности. Источниками части данных выступили различные государственные департаменты Гонконга [6].

Ч. Сюэ, Ю. Цзюй, Ш. Ли, Ц. Чжоу и Ц. Лю из университета Чанъань (Китай) на примере города Сиань демонстрируют использование пространственных данных при решении задачи планирования в сфере недвижимости. При этом были задействованы данные о недвижимости, ценах на жилье (сайт Lianjia), а также о дорожной сети (API карт Gaud) в главном городском районе города [7].

К. Адольф Мозер и В. Скотт из школы экономики и политических наук Лондона собрали данные по 157 городам Англии и Уэльса, население которых, по данным переписи 1951 г., превышало 50 тыс. чел. Для каждого города учитывалось 60 переменных, полученных на основе данных переписи населения и статистики его естественного движения, а также — информации из Министерства торговли и секции социальных обследований Центрального информационного управления [8].

Д. Коутс, Н. Доэрти, А. Френч и М. Киркуп из бизнес-школы университета Лафборо сравнивали точность нейронных сетей и методов множественной регрессии. Среди пространственных факторов, которые использовались авторами: население, демографические характеристики рыночной зоны магазина, конкуренция, доступность объекта, качество расположения по отношению к пешеходным потокам [9].

Я. Цюй и Ц. Чжан из Мичиганского университета анализировали торговые зоны (Trade Area Analysis, ТАА) и потенциальный охват клиентов для точек четырех торговых сетей (Whole Foods (супермаркет, Нью-Йорк), IKEA (мебельный магазин, Мичиган), Starbucks (кофейня, Нью-Йорк), Macy's (универсам, Сан-Франциско)), учитывая данные о местоположении, генерируемые пользователями мобильных устройств (таких как чекины Foursquare), собранные с помощью API Twitter [10].

Таким образом, сегодня не существует единого определения для понятия «пространственные данные», так же как нет единого понимания, какие источники и форматы могут быть отнесены к этой группе. Однако большинство источников сходятся в своих трактовках к тому, что к пространственным данным может быть отнесена только та информация, которая имеет прямую привязку к местоположению — как на нашей планете, так и в абстрактной системе координат. А это автоматически позволяет называть пространственными большинство данных, собранных с поверхности Земли и имеющих какую-то физическую принадлежность к месту. В этом случае становится понятно, почему невозможно указать все источники такого рода данных — ведь их бесчисленное множество, и даже из информации, не имеющей реальной привязки к местности, можно создать таковую, ведь большинство экономических, финансовых, социально-демографических и других показателей так или иначе имеют какую-то косвенную принадлежность к пространству, в рамках которого они были собраны.

Далее подробно остановимся на способах получения пространственных данных и их последующем использовании в рамках исследования.

СБОР И ОБРАБОТКА ДАННЫХ О ДОМОХОЗЯЙСТВАХ

Наиболее полную информацию (с точки зрения как количества, так и описания) касательно объектов жилья в Российской Федерации можно получить из открытого источника — Государственной информационной системы жилищно-коммунального хозяйства (ГИС ЖКХ)⁴. В разделе «Реестр объектов жилищного фонда» содержится база объектов по всем субъектам РФ, в которой находится свыше 200 млн записей в разрезе квартир/частных домов, содержащих сведения о кадастровом номере, адресе, уникальном идентификаторе здания, общей/жилой площади, типе дома, управляющей организации, типе помещения, дате сноса и т.д. (всего порядка 20 полей).

Однако для одного и того же дома в этой базе может быть сразу несколько записей. Например, в случае многоквартирного здания количество строк будет соответствовать числу квартир и других нежилых помещений, находящихся в нем, отмеченных одинаковым уникальным идентификатором здания (около 20 млн). Таким образом, имея только этот набор данных, уже можно получить сведения о количестве квартир и общей/жилой площади здания или частного дома и использовать их при реализации дальнейших этапов исследования. Но ГИС ЖКХ обладает расширенной информацией: ее нельзя получить в формате готового датасета — для этого необходимо воспользоваться собранными ранее уникальными идентификаторами зданий и внутренним API сайта. Это, в свою очередь, позволяет дополнительно получить данные по, как минимум, 50 полям, среди которых: год ввода в эксплуатацию, год постройки, год реконструкции, стадия жизненного цикла, тип проекта здания, количество этажей наземных/подземных, количество подъездов, лифтов, жилых/нежилых помещений, общая площадь жилых/нежилых/общих помещений, количество балконов/лоджий, износ здания. Также имеется и более специфическая информация, касающаяся конструктивных особенностей здания; систем отопления, водоснабжения, водоотведения, газоснабжения; фасада дома, лифтов, приборов

учета, земельного участка. Каждый пункт может быть детализирован до минимального уровня (лифта, квартиры, помещения и т.д.): таким образом исходные данные будут достаточно сильно обогащены. Безусловно, какой-то части детальной информации у большинства объектов ЖКХ не будет, так как цифровизация отрасли началась не так давно, и всей проектной документации по старым домам в электронном виде не существует. Подробная схема сбора данных об объектах ЖКХ представлена на *рис. 1*.

При всем объеме информации, у ГИС ЖКХ все же имеется один большой недостаток: как отмечалось ранее, пространственные данные должны иметь параметры, гарантирующие строгую привязку информации к конкретному положению (обычно в этой роли выступают координаты объектов), но у данных из реестра подобной привязки нет. Есть только два поля, способные стать косвенным источником: можно либо попытаться превратить адрес в набор координат, либо сопоставить кадастровый номер с реальными границами здания. Теоретически второй подход проще, так как для него нужна лишь кадастровая карта, но здесь возникает целый ряд проблем. Во-первых, не у всех объектов в базе имеется кадастровый номер, а, во-вторых, в открытом доступе нет инструмента, который бы позволил по имеющимся кадастровым номерам в промышленных масштабах получать границы домов. Безусловно, существует Национальная система пространственных данных, где по кадастровому номеру можно найти границы конкретного здания, но эту операцию придется проделывать для каждого объекта, и нет никакой гарантии, что она будет соответствовать действительности. Поэтому первый подход более реальный, так как Яндекс Геокодер⁵ с помощью API Яндекс Карт может преобразовать практически любой адрес в набор координат. Так, например, при запросе «Москва, Театральная площадь, 1» на выходе мы получим *latitude* (широта) — 55.760188 и *longitude* (долгота) — 37.618571 градусов. И если кадастровый номер имеется далеко не у каждого здания, то адрес есть у всех. На *рис. 2* представлена подробная схема работы с API Геокодер.

Здесь, конечно, тоже есть свои недостатки, главным из которых можно считать вариативность написания адреса и — как следствие — далеко не

⁴ URL: <https://dom.gosuslugi.ru/>

⁵ URL: <https://yandex.ru/maps-api/docs/geocoder-api/index.html>

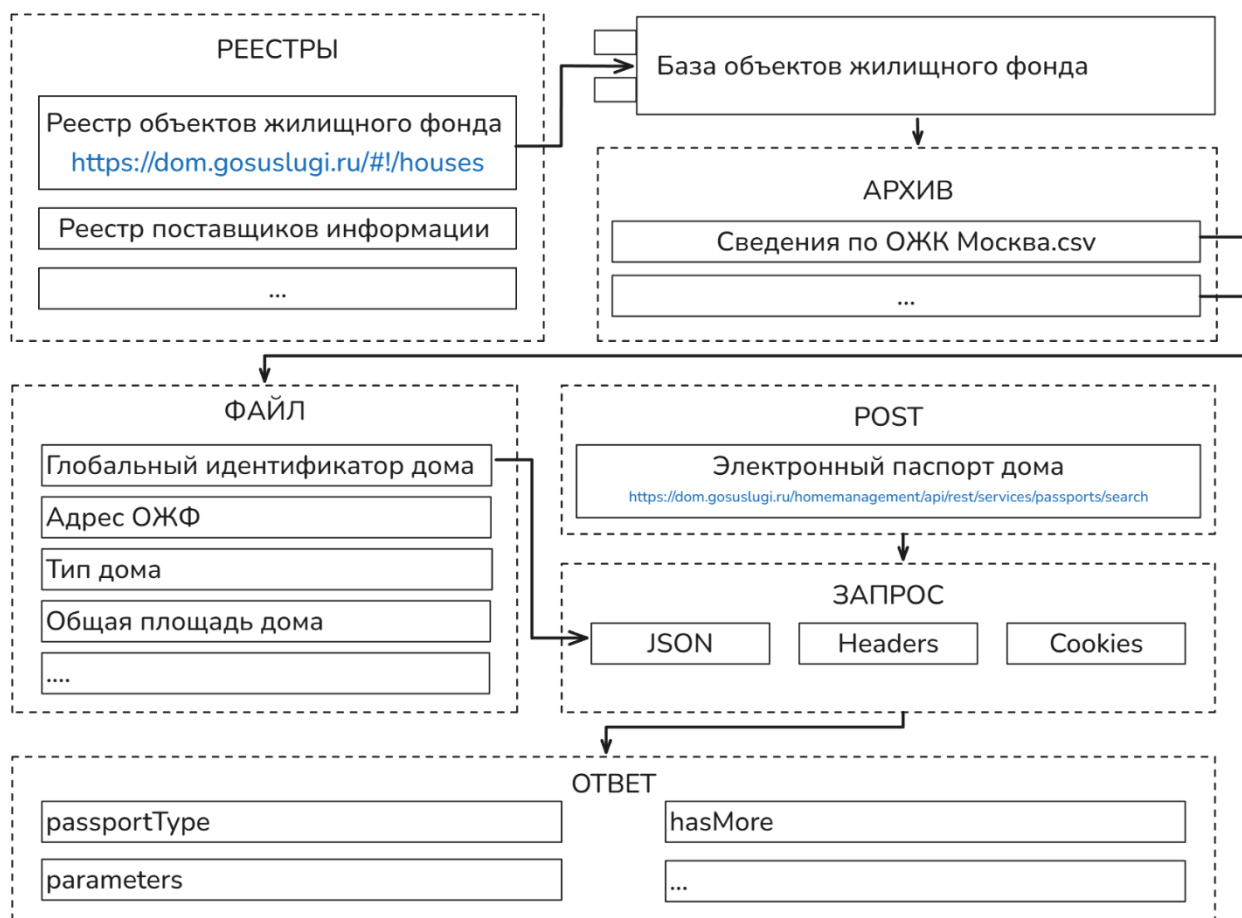


Рис. 1 / Fig 1. Схема взаимодействия с государственной информационной системой жилищно-коммунального хозяйства (ГИС ЖКХ) / Scheme of Interaction with the State Information System of Housing and Communal Utilities (GIS HCS)

Источник / Source: разработано авторами / developed by the authors.

100%-ю точность геокодирования. Но при работе с таким объемом данных в любом случае придется идти на компромиссы.

СБОР И ОБРАБОТКА ДАННЫХ О СТОИМОСТИ НЕДВИЖИМОСТИ

Сбор данных о стоимости недвижимости осуществляется за счет парсинга об объявлениях о продаже квартир на определенных сайтах-агрегаторах, например, Авито. Через открытый доступ или посредством API имеющиеся объявления посмотреть невозможно, а каждая компания в этой сфере постоянно ведет борьбу с попытками неконтролируемого скачивания данных. Поэтому в рамках поставленной задачи это обычно реализуется за счет изменения HTML-разметки, добавления CAPTCHA, динамической подгрузки информации или же усложнения процесса доступа к картам

объявлений. И если удастся создать какой-то универсальный инструмент для сбора данных, то без должного контроля и своевременного обновления уже через несколько месяцев он может утратить актуальность.

В случае с Авито есть два подхода. Следуя первому, можно открыть раздел с объявлениями о продаже квартир; получить ссылку на страницу, заголовки и cookie файлы, необходимые для ее загрузки; после чего в цикле собрать информацию об объявлениях и ссылки на них по всем страницам и пройти по каждой из них, обогатив выборку недостающей информацией. Но у этого метода есть несколько недостатков. Страницы с объявлениями обычно отдаются под конкретные регионы, весь список которых изначально не виден (только территориально близкие к исходному). Причем количество объявлений достаточно велико, а на одну

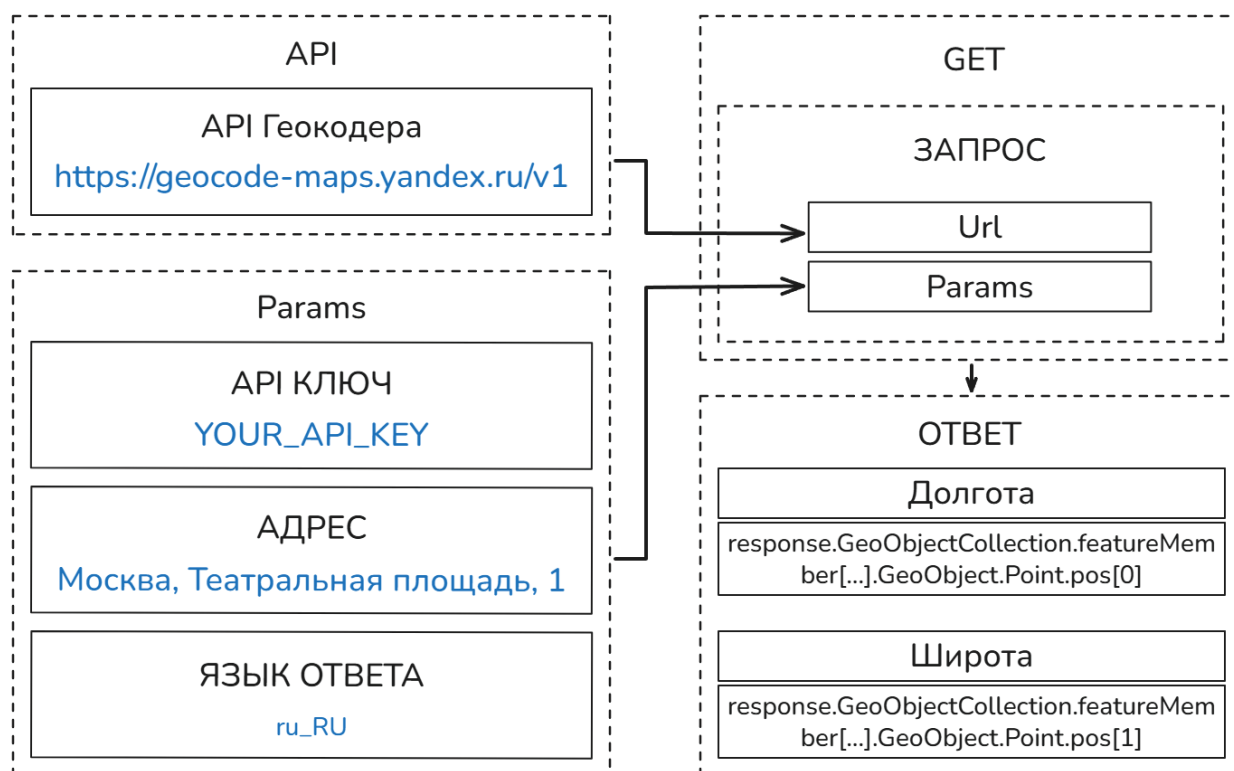


Рис. 2 / Fig 2: **Схема взаимодействия с API Геокодер Яндекс Карты /**
Scheme of Interaction with the Yandex Maps Geocoder API

Источник / Source: разработано авторами / developed by the authors.

страницу выдачи приходится не более 40. Отсюда возникает вторая проблема: даже если в запросе указать все регионы (а такая возможность есть), то количество объявлений будет свыше 1 млн, но число страниц, которые можно просмотреть, — на уровне 100. Таким образом, получить больше 4000 объявлений просто невозможно. Второй путь является более сложной версией первого, и его основное отличие — процесс получения ссылок на объявления. На сайте есть возможность просмотра объявлений прямо на карте: запрос на получение данных в таком случае отправляется прямо на внутренний API сайта. В качестве главных параметров при этом выступают координаты верхней левой и нижней правой границы произвольного четырехугольника. Таким образом можно собрать все объявления из выделенной области. Подобный подход гарантирует большую интерпретируемость, но с его помощью можно получить очень ограниченную информацию об объявлении: координаты и уникальный идентификатор. Обычно для итоговой ссылки последнего достаточно, но в нашем случае требуется еще и название объявления, записанное латиницей с заменой определенных символов

и знаков пунктуации, чего во внутреннем API не предусмотрено. Следовательно, дополнительно надо получить название объявления в нужном формате при том, что известен лишь один уникальный идентификатор. На рис. 3 представлена схема взаимодействия с сайтом «Авито Недвижимость».

Однако на сайте есть строка поиска с html-разметкой. И если в качестве запроса указать уникальный идентификатор объявления, то появится страница с нужной карточкой, из которой уже можно получить итоговую ссылку на объявление. По ней находится необходимая информация: стоимость квадратного метра, площадь, этажность, тип ремонта и т.д., а также характеристики дома (год постройки, количество этажей, тип, наличие парковки и т.д.). Таким образом, данный вариант, хоть и сложнее, но при этом позволяет получить более полную информацию по объявлениям.

СБОР И ОБРАБОТКА ДАННЫХ ОБ ОРГАНИЗАЦИЯХ И КОНКУРЕНЦИИ

Данные о бизнесе также представляют собой точечную информацию, которая собирается аналогично — с помощью парсинга и запросов в Яндекс.API.

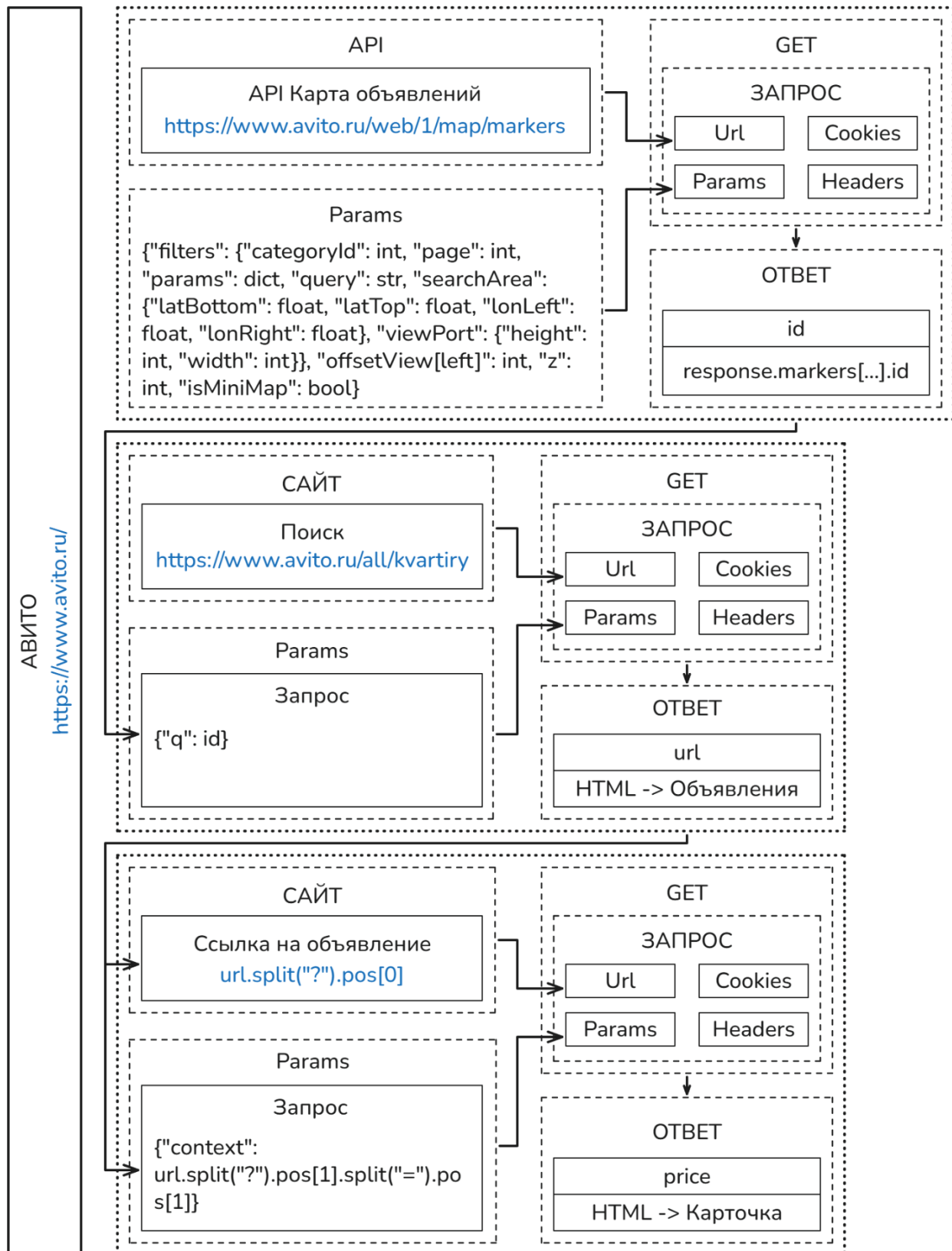


Рис. 3 / Fig 3. Схема взаимодействия с сайтом Авито недвижимость /
Scheme of Interaction with the Avito Real Estate Website

Источник / Source: разработано авторами / developed by the authors.

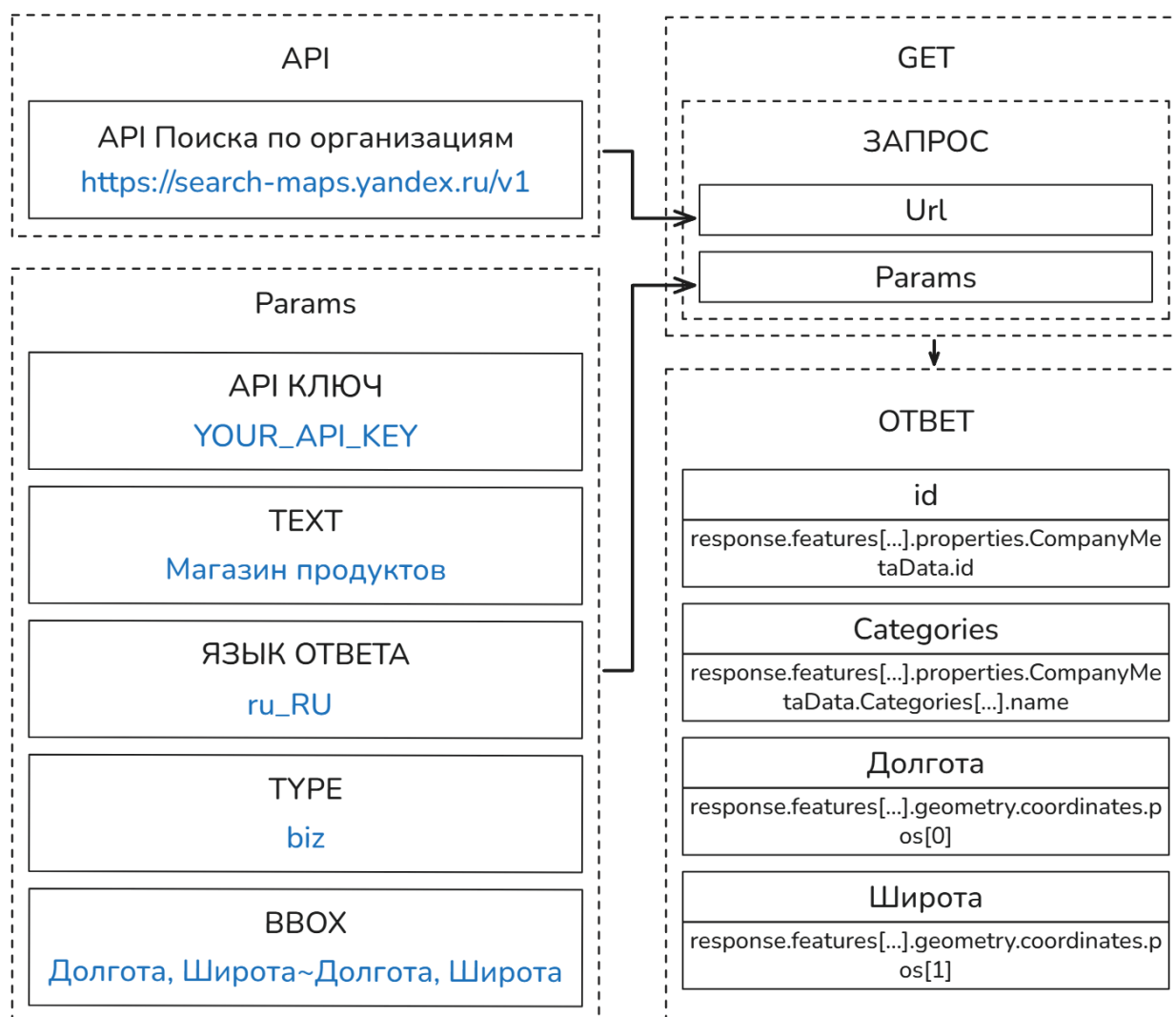


Рис. 4 / Fig 4. Схема взаимодействия с API Поиск по организациям Яндекс Карты /
Scheme of Interaction with the Yandex Maps Geosearch API

Источник / Source: разработано авторами / developed by the authors.

Так как API поиска по организациям⁶ не предоставляет сведения об организации (отзывы, рейтинги), то процесс их получения усложняется. Для начала необходимы списки организаций, а в рамках Яндекс.АРІ есть возможность узнать, присутствует ли какая-либо из них в заданной области. Причем не обязательно указывать конкретное название точки: можно ограничиться классом (ресторан, торговый центр, магазин продуктов и т.д.). Структура подобного запроса представлена на рис. 4. После его отправки в ответ должен прийти JSON, содержащий в себе все организации с указанным классом или названием в определенной области.

⁶ URL: <https://yandex.ru/maps-api/docs/geosearch-api/index.html>

Для каждой точки будет указан целый спектр информации, в которой в основном интересны три элемента: класс, к которому принадлежит организация (продуктовый магазин, кафе, ТЦ и т.д.), координаты (долгота и широта), а самое важное — ID (у каждой организации на карте есть свой определенный идентификационный номер, причем разный даже в рамках точек, принадлежащих одной компании). Он и будет основным инструментом для сбора сведений об отзывах, среднем рейтинге, а также количестве организаций. Таким образом, для каждой заданной области можно получить информацию о точках, принадлежащих к определенным классам, и сформировать большую базу данных организаций, в которой условно есть 4 столбца (ID, класс, широта и долгота). Затем с помощью

библиотек Selenium, Playwright, Cypress, Puppeteer и т.п. необходимо получить информацию об оценках и рейтингах с Яндекс Карт⁷. Здесь основную роль сыграют ID, полученные ранее. После того, как парсинг будет выполнен, исходная база данных расширится на 2 столбца (количество оценок и рейтинг).

СБОР И ОБРАБОТКА ДАННЫХ О ЗДАНИЯХ И ДОРОЖНОЙ СЕТИ

Основным источником информации о зданиях и дорожной сети является платформа OpenStreetMap (OSM)⁸. Это открытый сервис, содержащий обширную базу пространственных данных для всего мира, которую поддерживают отдельные энтузиасты и добровольцы. Главным плюсом OSM в сравнении с другими картографическими сервиса-

ми является открытый доступ к данным на условиях лицензионного соглашения Open Database License, что позволяет их использовать для неограниченного круга исследовательских и коммерческих задач.

Однако получить, например, информацию обо всех дорогах или зданиях какой-либо страны напрямую на сайте зачастую невозможно — как из-за объемов информации, так и отсутствия соответствующих разделов. Но это не является проблемой, поскольку сайт зачастую указывает сторонние сервисы, где можно скачать необходимые данные. Одним из таких является Geofabrik⁹. Компания Geofabrik GmbH предоставляет бесплатный сервер для всех континентов, большинства стран и многих административных единиц. При этом все файлы ежедневно обновляются, что обеспечивает актуальность данных для оперативных исследований (рис. 5).

⁷ URL: <https://yandex.ru/maps>

⁸ URL: <https://www.openstreetmap.org/>

⁹ URL: <https://download.geofabrik.de/>

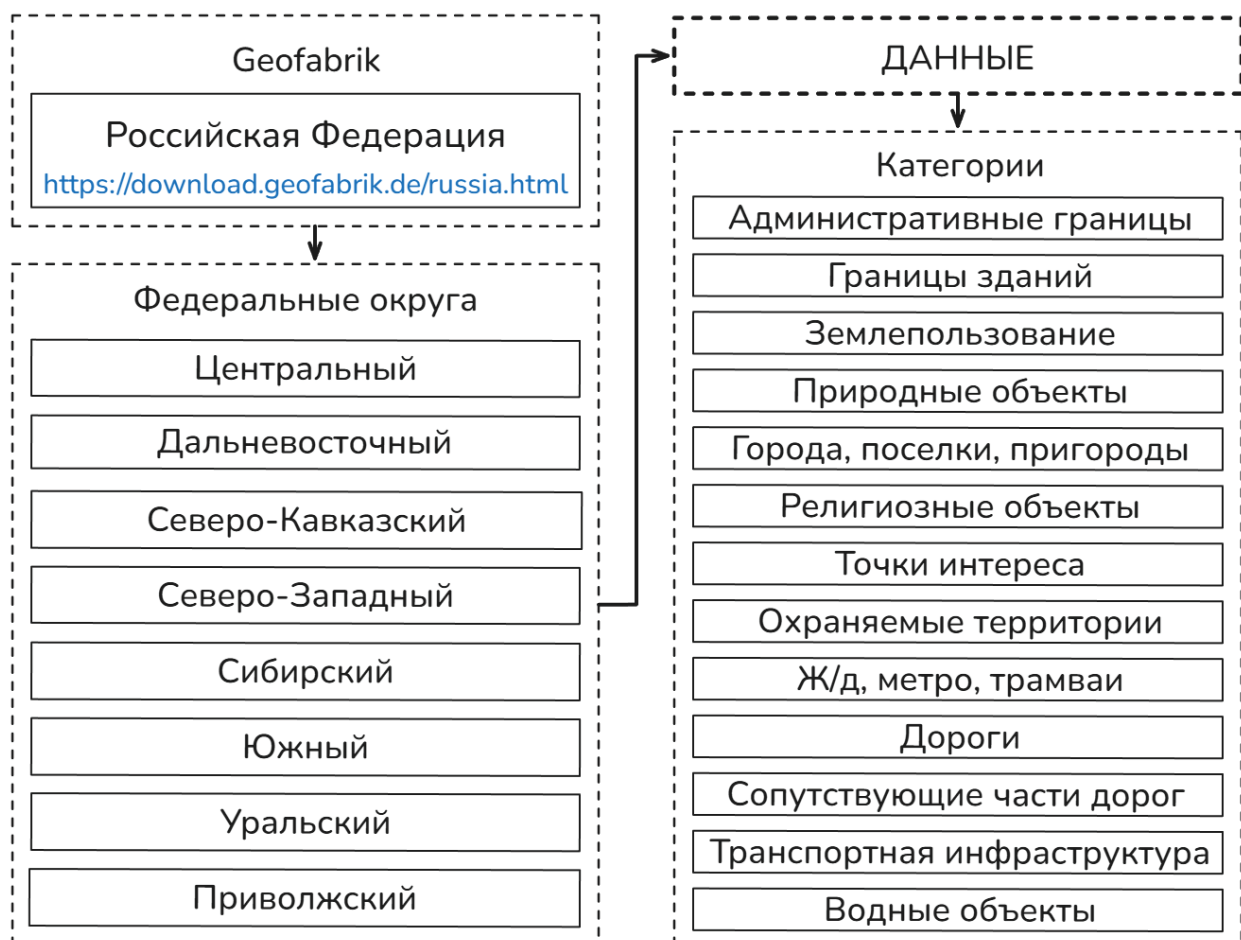


Рис. 5 / Fig 5. Схема взаимодействия с данными OpenStreetMap / Scheme of Interaction with OpenStreetMap data

Источник / Source: разработано авторами / developed by the authors.

Для Российской Федерации доступны дампы как стран, так и отдельных федеральных округов. Внутри архива могут содержаться файлы с информацией о геометрии здания, его названии, типе, годе постройки, серии, количестве квартир и т.д. Касательно дорожной сети будет указана категория дороги, ее наименование, геометрия, скоростное ограничение и т.д. Помимо этого, имеются данные о водных и лесных объектах, точках интереса, достопримечательностях, железных дорогах и т.д.

СБОР И ОБРАБОТКА ДАННЫХ О ТРАФИКЕ И СОЦИАЛЬНО- ДЕМОГРАФИЧЕСКОЙ СТРУКТУРЕ НАСЕЛЕНИЯ

Для эффективного решения задачи размещения в офлайн бизнесе недостаточно понимать, сколько людей проживает в городе, какое количество автомобилей на них приходится и каким средним доходом они обладают. Важно понимать, какой объем трафика передвигается в непосредственной близости от места, где может быть размещена потенциальная организация, а также то, какими социально-демографическими характеристиками описывается конкретная целевая аудитория. В открытом доступе информации о пешеходном или автомобильном трафике в таком низком разрешении нет. Государственные же источники статистики, если и обладают похожими данными, то только на уровне города, области или страны. При такой детализации их полезность для решения задачи размещения торговых объектов в рамках города не очень высока, так как любая потенциальная локация внутри одного и того же города будет описываться одним набором значений пространственных факторов. Сведений о социально-демографической структуре трафика (пол, возраст, доход) также не существует. Но при игнорировании такого пласта информации высок риск пропустить значимые переменные и, как следствие, получить модели с низкой предсказательной способностью.

Большинство сервисов, обладающих такими данными (сотовые операторы, социальные сети, карты, агрегаторы информации мобильных приложений, банки), предоставляют их только на коммерческой основе. Однако некоторые из них имеют открытые платформы.

Одним из подобных сервисов является «Геоаналитика»¹⁰, представленная в Яндекс Картах. В об-

ласть покрытия пространственными данными входят: Москва, Московская область, Санкт-Петербург, Ленинградская область, Волгоград, Воронеж, Екатеринбург, Казань, Краснодар, Красноярск, Нижний Новгород, Новосибирск, Омск, Пермь, Ростов-на-Дону, Самара, Уфа, Челябинск.

Сбор информации по пешеходному и автомобильному трафику осуществляется в 2 этапа. На первом реализуется генерация гексагональной сетки для заданной области с получением уникальных идентификаторов гексагонов, входящих в нее. Подробнее это показано на *рис. 6*.

На втором этапе полученные ID гексагонов декодируются в целочисленный формат. А затем последовательно осуществляется отправка запросов к внутреннему API с целью получения информации по всем имеющимся слоям данных (пешеходный и автомобильный трафик; численность населения с указанием долей по полу, возрасту и доходу, а также поисковые запросы и организации по заданной категории бизнеса). Главным недостатком такого подхода является то, что задаваемая гексагональная сетка на первом этапе включает в себя все шестиугольники, входящие в какую-то область, однако в действительности данные будут далеко не по всем гексагонам, так как некоторые из них включают леса, поля, озера и т.д. Именно поэтому требуется дополнительная проверка полученной информации.

РАСЧЕТ ПРОСТРАНСТВЕННЫХ ФАКТОРОВ

Опираясь на описанные ранее подходы, авторы реализовали отдельные программные решения на языке программирования Python для сбора и обработки информации о следующих пространственных группах данных: «организации», «социальная демография», «жилой фонд», «автомобильный трафик» и «пешеходный трафик». Каждая из групп при этом представлена как минимум одним источником данных по причине схожести некоторых факторов. Например, к группе «жилой фонд» могут быть отнесены данные сразу из трех источников: ГИС ЖКХ, Яндекс Геоаналитика и Авито. Теперь каждую группу рассмотрим более детально.

Пространственная группа факторов «организации» представлена одним источником данных, в рамках которого существуют два отдельных сервиса: API Яндекс Карт, а также сама веб-версия Яндекс Карт. Сначала производится сбор карточек организаций по категориям с помощью API Яндекс Карт

¹⁰ URL: <https://yandex.ru/geoanalytics>

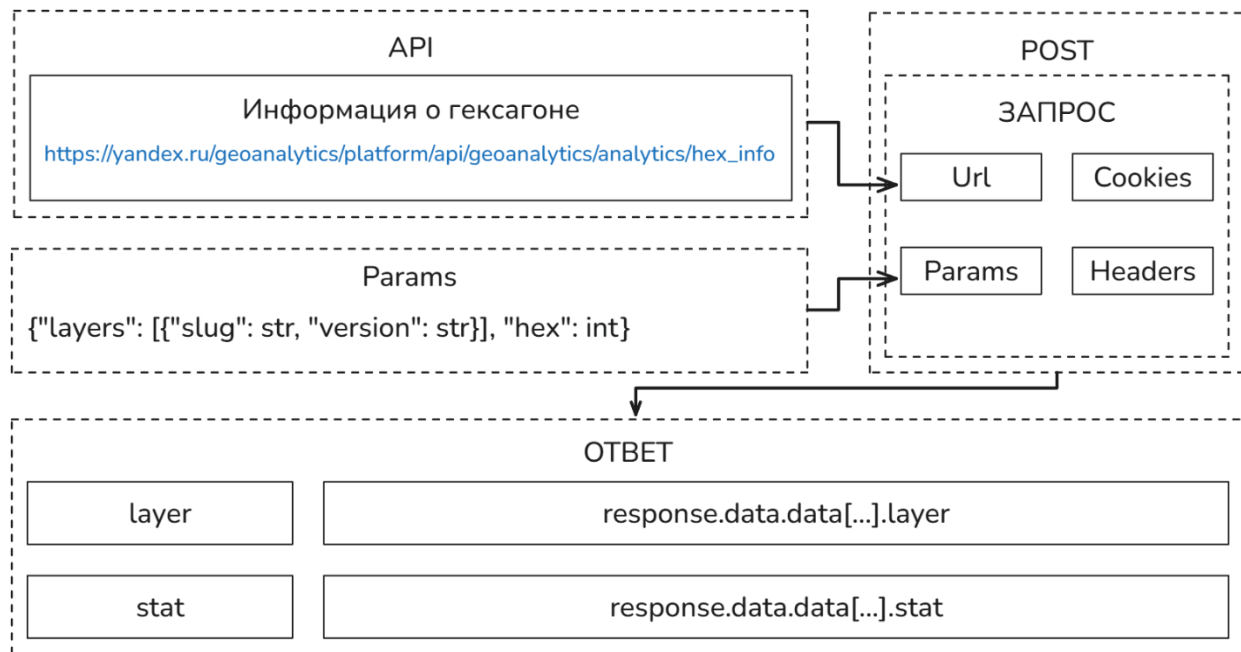


Рис. 6 / Fig 6: **Схема взаимодействия с API Яндекс Геоаналитика / Scheme of Interaction with Yandex Geoanalytics API**

Источник / Source: разработано авторами / developed by the authors.

(точнее — с помощью поиска по организациям). Затем эта информация дополняется количеством оценок и показателями рейтинга непосредственно с самих Яндекс Карт с использованием ID-карточек прошлого этапа. Данные по организациям собирались для 22-х основных категорий объектов: АЗС; автосервис, автотехцентр; аптека; банк; банкомат; быстрое питание; гостиница; жилой комплекс; кафе; кофейня; магазин автозапчастей и автотоваров; магазин одежды; магазин продуктов; магазин электроники; медцентр, клиника; офис организации; платежный терминал; ресторан; салон красоты; стоматологическая клиника; супермаркет; торговый центр. Когда база была готова, для каждой произвольной точки было подсчитано количество организаций в определенной категории и заданном радиусе от исходной позиции, их средний рейтинг или суммарное количество оценок.

Пространственная группа факторов «социальная демография» также представлена одним источником данных — платформой Яндекс Геоаналитика. Сбор информации осуществляется в два этапа. Сначала реализуется генерация гексагональной сетки для заданной области, после чего по полученным уникальным идентификаторам шестиугольников осуществляется отправка запросов к внутреннему API сервиса с указанием необходимых слоев данных. Информация о социально-демографической струк-

туре населения представлена в следующих разрезах: пол (доля населения женского/мужского пола); возраст (доля населения в возрасте от 0 до 17 лет/возрасте от 18 до 24 лет/возрасте от 25 до 34 лет/возрасте от 35 до 44 лет/возрасте от 45 до 54 лет/возрасте от 55 до 99 лет), доход (доля населения с низким доходом/ доходом ниже среднего/доходом выше среднего/высоким доходом/премиум-доходом).

Пространственная группа факторов «жилой фонд», как уже упоминалось ранее, формируется на основе трех различных источников: ГИС ЖКХ; Авито и Яндекс Геоаналитика.

Пространственная группа факторов «автомобильный трафик» формируется на основе двух источников. Сами данные о трафике и его количестве в разрезе уникальный/неуникальный/сигналы собираются с использованием платформы Яндекс Геоаналитика, а информация о протяженности дорог, их категориях и времени проезда по ним — на основе OpenStreetMap. Если человек прибывает в какой-то области несколько раз за определенный интервал времени, то такой трафик считается уникальным и принимается за «1», а если его визиты нерегулярны (неуникальный трафик), то считается их число. Количество сигналов связывается со временем, которое пользователь провел в текущей области. Вся информация о трафике и его количестве привязана к гексагональной сетке, а данные

о дорожной сети представлены в виде линий и их множеств, которые отражают реальные транспортные магистрали.

Пространственная группа факторов «пешеходный трафик» представлена платформой Яндекс Геоаналитика. Сами данные по аналогии с автомобильным трафиком включают в себя информацию об уникальном/неуникальном трафике и сигналах. Пешеходный трафик, как и автомобильный, является агрегированной информацией и представлен в виде гексагонов.

Длина основных радиусов для расчета пространственных факторов составляет 75, 150, 250, 350 и 500 м. Уменьшение шага или увеличение размеров приводит к нелинейному росту затрат времени на расчет, последующий анализ и моделирование. Именно поэтому используются пять радиусов, а не изохроны.

В *таблице* представлены основные группы и категории пространственных факторов, рассчи-

танные в рамках исследования. По каждой паре дополнительно приводится детализация в разрезе атрибутов, агрегаций и радиусов, которые использовались при формировании пространства объясняющих переменных. В колонке «количество» содержится информация о числе факторов, полученных для такого набора условий.

ВЫВОДЫ

В работе описан подход к сбору и обработке пространственных данных по следующим категориям: домохозяйства, стоимость недвижимости, организации и конкуренция, здания и дорожная сеть, трафик и социально-демографическая структура населения.

Всего выделено 1045 пространственных признаков. Все они будут использованы в дальнейших работах в целях построения моделей для оценки экономических показателей локаций на примере российской сферы торговли.

Таблица / Table

Структура пространственных факторов в исследовании /
The Structure of Spatial Factors in the Study

Группа факторов / Group of factors	Категория факторов / Category of factors	Атрибут факторов / Attribute of factors	Агрегация факторов / Aggregation of factors	Радиус, м / Radius, m	Количество / Quantity
Организации	1) АЗС; 2) автосервис, автотехцентр; 3) аптека; 4) банк; 5) банкомат; 6) быстрое питание; 7) гостиница; 8) жилой комплекс; 9) кафе; 10) кофейня; 11) магазин автозапчастей и автотоваров; 12) магазин одежды; 13) магазин продуктов; 14) магазин электроники; 15) медцентр, клиника; 16) офис организации; 17) платежный терминал; 18) ресторан; 19) салон красоты; 20) стоматологическая клиника; 21) супермаркет; 22) торговый центр	1) количество; 2) количество оценок; 3) рейтинг	1) количество; 2) максимум; 3) среднее; 4) сумма	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	770

Окончание таблицы / Table (continued)

Группа факторов / Group of factors	Категория факторов / Category of factors	Атрибут факторов / Attribute of factors	Агрегация факторов / Aggregation of factors	Радиус, м / Radius, m	Количество / Quantity
Социальная демография	1) население в возрасте от 0 до 17 лет; 2) население в возрасте от 18 до 24 лет; 3) население в возрасте от 25 до 34 лет; 4) население в возрасте от 35 до 44 лет; 5) население в возрасте от 45 до 54 лет; 6) население в возрасте от 55 до 99 лет; 7) население женского пола; 8) население мужского пола; 9) население с низким доходом; 10) население с доходом ниже среднего; 11) население с доходом выше среднего; 12) население с высоким доходом; 13) население с премиум доходом	Доля	1) максимум; 2) среднее	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	130
Автомобильный трафик	1) повторяемый; 2) сигналы; 3) уникальный	Количество	1) максимум; 2) среднее; 3) сумма	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	45
Автомобильный трафик	1) время проезда; 2) категория дорог; 3) протяженность дорог	Количество	1) среднее; 2) сумма	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	30
Автомобильный трафик	1) количество дорог	Количество	Количество	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	5
Жилой фонд	1) домохозяйства; 2) население	Количество	1) максимум; 2) минимум; 3) среднее; 4) сумма	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	40
Жилой фонд	1) дома	Количество	Количество	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	5
Жилой фонд	1) стоимость квадратного метра	Количество	Среднее	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	5
Пешеходный трафик	1) повторяемый; 2) сигналы; 3) уникальный	Количество	Количество	1) 75; 2) 150; 3) 250; 4) 350; 5) 500	15

Источник / Source: разработано авторами / developed by the authors.

REFERENCES

1. Dibb S., Simkin L. Retail store performance model the impact of own-branch “friendly fire”. *International Transactions in Operational Research*. 1994;1(4):479-487. DOI: 10.1111/j.1475-3995.1994.00324.x
2. Satman M.H., Altunbey M. Selecting Location of Retail Stores Using Artificial Neural Networks and Google Places API. *International Journal of Statistics and Probability*. 2014;3(1):67-77. DOI: 10.5539/ijsp.v3n1p67
3. Xu M., Wang T., Wu Z., Zhou J., Li J., Wu H. Demand driven store site selection via multiple spatial-temporal data. *SIGSPATIAL '16: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2016;(40):1-10. DOI: 10.1145/2996913.2996996
4. Karamshuk D., Noulas A., Scellato S., Nicosia V., Mascolo C. Geo-spotting: Mining online location-based services for optimal retail store placement. *KDD '13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013:793-801. DOI: 10.1145/2487575.2487616
5. Roig-Tierno N., Baviera-Puig A., Buitrago-Vera J., Mas-Verdu F. The retail site location process using GIS and the analytical hierarchy process. *Applied Geography*. 2013;(40):191-198. DOI: 10.1016/j.apgeog.2013.03.005
6. Cheng E., Li H., Yu L. A GIS approach to shopping mall location selection. *Building and Environment*. 2007;42(2):884-892. DOI: 10.1016/j.buildenv.2005.10.010
7. Xue C., Ju Y., Li S., Zhou Q., Liu Q. Research on accurate house price analysis by using GIS technology and transport accessibility: A case study of Xi'an, China. *Symmetry*. 2020;12(8):1329-1350. DOI: 10.3390/sym12081329
8. Moser C.A., Scott W. British towns: A statistical study of their social and economic differences. Edinburgh and London: Oliver & Boyd; 1961. 169 p.
9. Coates D., Doherty N., French A., Kirkup M. Neural networks for store performance forecasting: An empirical comparison with regression techniques. *The International Review of Retail, Distribution and Consumer Research*. 1995;5(4):415-432. DOI: 10.1080/09593969500000024
10. Qu Y., Zhang J. Trade area analysis using user generated mobile location data. *WWW '13: Proceedings of the 22nd international conference on World Wide Web*. 2013:1053-1064. DOI: 10.1145/2488388.2488480

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Наталья Владимировна Гринева — кандидат экономических наук, доцент, доцент кафедры информационных технологий, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Natalia V. Grineva — Cand. Sci. (Econ.), Assoc. Prof., Assoc. Prof. of the Department of Information Technology, Financial University under the Government of the Russian Federation, Moscow, Russian Federation

<https://orcid.org/0000-0001-7647-5967>

Автор для корреспонденции / Corresponding author:

ngrineva@fa.ru

Алексей Дмитриевич Топыркин — аспирант, аналитик-разработчик компании Яндекс Карты, Российская академия народного хозяйства и государственной службы при Президенте РФ, Москва, Российская Федерация

Alexey D. Topyrkin — Post-graduate Student, Analyst-Developer at Yandex Maps, Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation

<https://orcid.org/0009-0005-3862-2651>

topyrkinalexei@yandex.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 05.05.2026; после рецензирования 28.05.2026; принята к публикации 15.06.2026.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was received on 05.05.2026; revised on 28.05.2026 and accepted for publication on 15.06.2026.

The authors read and approved the final version of the manuscript.