

УДК 004.6, 004.9

ИНФОРМАЦИОННАЯ СИСТЕМА ИССЛЕДОВАНИЯ ГОРОДСКОГО ПРОСТРАНСТВА ЧЕРЕЗ ПУБЛИЧНЫЕ ДАННЫЕ НА ПРИМЕРЕ Г. ИРКУТСКА

Косогоров Д.Е. (федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный университет»)

Научный руководитель – Пестова Ю.В.

(федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный университет»)

Введение.

Одним из способов изучения городского пространства является исследование городских точек притяжения. Точки притяжения – это локации, объединяющие вокруг себя людей. В работе рассматривается подход исследования городского пространства на основе анализа данных информационных веб-ресурсов и цифрового следа, оставленного жителями города. Цифровой след (или цифровой отпечаток, digital footprint) — это уникальный набор действий в Интернете или на цифровых устройствах. Активный цифровой след появляется, когда пользователь осознанно публикует информацию на веб-сайтах и в социальных медиа. Понятие цифрового следа в настоящее время активно используется при решении задач планирования и развития городов, например, при оценке качества жизни городского населения, исследования эмоционального восприятия городской среды, реализации концепции «умного города» [1]. Целевым городом для изучения выбран город Иркутск. В качестве информационных веб-ресурсов выбраны Google Maps, Яндекс.Карты и 2ГИС. Работа выполняется при поддержке проекта Российского научного фонда №23-28-00844 «Мониторинг сферы регионального туризма на основе анализа данных из открытых источников».

Основная часть.

Целью данной работы является разработка информационной системы анализа и визуализации точек притяжения городского пространства на основе данных из открытых источников (публичные данные).

Для реализации цели необходимо решить задачи:

- определить модель данных о городском пространстве, необходимых для оценки его качества,

- сформировать наборы данных о городском пространстве посредством методов сбора данных (парсинг, API) из веб-ресурсов,

- провести анализ полученных данных с применением методов: обработки данных, количественного и качественного анализов, текстового анализа для оценки привлекательных объектов городского пространства посредством ранжирования [2],

- разработать панели визуализации данных для предоставления и анализа информации пользователям без специализированных навыков программирования.

Модель данных содержит следующие понятия: локации городского пространства – идентификатор локации, название, категории локации, источник данных, URL-адрес, физический адрес, координаты, количественные показатели (фото, отзывы, рейтинг); посещаемость локаций – идентификатор локации, день, время, загрузка в процентах; отзывы о локациях – идентификаторы отзыва, локации и пользователя, текст отзыва, оценка, дата публикации; фотографии из отзывов – идентификаторы фотографии и отзыва, URL – адрес фотографии.

Анализируемые источники открытых данных (Google Maps, Яндекс.Карты и 2ГИС) являются динамическими: все данные подгружаются в веб-ресурс в процессе работы пользователя с браузером. Поэтому для автоматизации процесса сбора данных была использована библиотека Selenium (Python) для создания агента, который имитирует действия пользователя на странице и извлекает информацию с таких сайтов из их DOM-структуры.

На основе информационной модели и количественного и текстового методов были сформированы показатели, характеризующие привлекательность мест. В качестве целевой функции используется сумма нормированных показателей: количества отзывов, фотографий, рейтинга места и результат классификации сентимент-анализа отзывов. Сформированы рекомендации посредством ранжирования списка мест по оптимальности их посещения на основе отношения показателей привлекательности и загруженности.

Панели визуализации отражают результаты анализа. Они реализованы с помощью BI платформы Tableau Public. Панели представляют собой объединение нескольких диаграмм, которые являются интерактивными. Это позволяет управлять ими через взаимодействие со всеми элементами панели для фильтрации, сортировки и выбора параметров [3]. Детальные и целевые характеристики отражены на картах различными методами пространственных данных: тепловая карта для отображения плотности, сетчатая карта, полигональная карта по районам по агрегированным характеристикам, точечная карта пропорциональных объектов для оценки мест по категориям.

Информационная система состоит из следующих модулей: модуль получения данных для каждого источника данных (Google Maps, Яндекс.Карты и 2ГИС), каждый из которых состоит из модулей: сбора списка локаций, метаданных локаций (описание и количественные метрики, посещаемость, отзывы и фотографии); объединения данных и модуля предварительной обработки данных; модуль анализа данных: текстовых, пространственных и количественных, а также модуля панелей визуализации результатов анализа.

Выводы.

Разрабатываемая информационная система может предоставить более полные и достоверные данные о состоянии городского пространства, что позволит повысить уровень ориентированности жителей в жизни города и эффективность городского управления. Собранные данные могут использоваться в дальнейших исследованиях городского пространства г. Иркутска.

Список использованных источников:

1. Ненько А.Е., Курилова М.А., Подкорытова М.И. Анализ эмоционального восприятия территории и развитие «Умного города» [Электронный ресурс] – Режим доступа: URL: <https://cyberleninka.ru/article/n/analiz-emotsionalnogo-vospriyatiya-territoriy-i-razvitiie-umnogo-goroda/viewer> (Дата последнего обращения: 19.02.2023)
2. Waala Medhat, Ahmed Hassan, Hoda Korashy Sentiment analysis algorithms and applications: A survey [Электронный ресурс] – Режим доступа: URL: <https://www.sciencedirect.com/science/article/pii/S2090447914000550> (Дата последнего обращения: 19.02.2023)
3. Пестова Ю.В. Анализ данных из открытых источников для мониторинга объектов Байкальского региона туристской инфраструктуры // Коммуникационные технологии: социально-экономические и информационные аспекты : Материалы всерос. молодежной науч.-практ. конф, (г. Иркутск, 06–20 апреля 2022 г.). Иркутск: ООО ЦентрНаучСервис, 2022. С. 52-56.