

## Концепция сервиса мониторинга сферы туризма территории на основе анализа веб-ресурсов

О. А. НИКОЛАЙЧУК\*, Ю. В. ПЕСТОВА, А. И. ПАВЛОВ, Д. Е. КОСОГОРОВ  
Институт динамики систем и теории управления им. В.М. Матросова СО РАН, 664033,  
Иркутск, Россия

\*Контактный автор: Николайчук Ольга Анатольевна, e-mail: nikoly@icc.ru

Поступила 10 октября 2023 г., принята в печать 17 октября 2023 г.

Развитие устойчивого туризма территории требует осуществления мониторинга всех разнородных аспектов сферы туризма и создание соответствующей информационной технологии. Целью работы является описание концепции сервиса мониторинга туризма в Иркутской области, определение методов и средств реализации данного сервиса, а также представление отдельных результатов ее прототипирования. Создание сервиса включает разработку и программную реализацию методов сбора, обработки и визуализации данных, а также обеспечение использования функциональных возможностей основных процессов обработки информации существующих облачных платформ. Используются технологии API и web-scraping, методы статистики и искусственного интеллекта, технологии BI-систем для визуализации. В настоящее время реализован прототип функции “формирование туристского профиля территории” на примере Байкальской природной территории.

*Ключевые слова:* мониторинг туризма, туристический профиль природной территории, онтология, сбор данных из открытых источников, анализ отзывов, веб-сервис, облачная платформа.

*Цитирование:* Николайчук О.А., Пестова Ю.В., Павлов А.И., Косогоров Д.Е. Концепция сервиса мониторинга сферы туризма территории на основе анализа веб-ресурсов. Вычислительные технологии. 2023; 28(6):118–134. DOI:10.25743/ICT.2023.28.6.011.

### Введение

Туризм — одна из наиболее быстро растущих отраслей мировой экономики, важный источник валютных поступлений и рабочих мест, тесно связанный с социальным, экономическим и экологическим благополучием страны.

Согласно определению Всемирной туристской организации, под устойчивым туризмом понимают “туризм, в полной мере обеспечивающий учет его нынешних и будущих экономических, социальных и экологических последствий при удовлетворении потребностей туристов, индустрии туризма, окружающей среды и принимающих общин”.

Всемирная туристская организация ООН (UNWTO, United nations world tourism organization) провозглашает (<https://www.unwto.org/sustainable-development>), что устойчивый туризм должен: оптимально использовать ресурсы окружающей среды, уважать социокультурную самобытность принимающих сообществ, обеспечивать жизнеспособную, долгосрочную экономическую деятельность, предоставляя социально-экономические выгоды всем заинтересованным сторонам. Устойчивое развитие туризма

требует информированного участия всех соответствующих заинтересованных сторон, а также сильного политического руководства для обеспечения широкого участия и достижения консенсуса. Достижение устойчивого туризма — это непрерывный процесс, требующий постоянного мониторинга воздействия и принятия необходимых превентивных и/или корректирующих мер, когда это необходимо.

В рамках UNWTO создана международная сеть обсерваторий устойчивого туризма, которая отслеживает экономическое, экологическое и социальное воздействие туризма на уровне дестинации. С момента создания в 2004 г. к данной сети присоединились в общей сложности 36 обсерваторий в Китае, Греции, Бразилии, Индонезии, Хорватии, США, Гватемале, Италии, Аргентине, Австралии, Португалии, Испании, Канаде, Колумбии и Мексике (<https://www.unwto.org/insto/observatories>). Среди основных индикаторов мониторинга выделены: сезонность туризма, данные о занятости в сфере туризма, управление потреблением энергии и воды, управление утилизацией отходов, влияние на климат и др.

Существуют региональные системы мониторинга (обсерватории), например *shapetourism*, которые в виде интерактивных карт предоставляют инструмент интерпретации динамики туризма на основе четырех измерений: репутации, привлекательности, конкурентоспособности и устойчивости — и охватывают 52 страны средиземноморского региона [1].

Необходимо отметить еще системы мониторинга отдельных достопримечательностей или территорий, например, в Австралии [2], Португалии [3], Китае [4], Франции [5], России [6]. Причем данные могут быть представлены не в виде интерактивных карт, а в отдельных файлах с графиками и таблицами за конкретный месяц по некоторому индикатору, так что для пользователя подобный способ получения информации является малоэффективным.

Таким образом, в мире определена цель развития устойчивого туризма, отмечено, что реализация этой цели возможна только при осуществлении постоянного мониторинга данной сферы на основе всесторонних данных, что в свою очередь возможно только с привлечением современных информационных технологий.

Отметим еще один фактор — значительное увеличение объема данных в сфере туризма, которые создаются в среде Интернет туристическими агентствами, туристическими объектами (гостиницы, рестораны, мероприятия и т. п.) и потребителями туристических услуг. Имеющиеся данные характеризуются массовостью, разнородностью, иногда слабой структурированностью, неполнотой и противоречивостью. Взрывное увеличение количества данных, доступных через Интернет, создает потребность в разработке методов и программных инструментов для интеграции автономных и разнородных источников данных, поиска релевантных данных, их извлечения и интерпретации [5, 7].

Цели настоящей работы — описание концепции сервиса мониторинга туризма Иркутской области как части Байкальской природной территории, определение методов и средств технологии реализации данного сервиса и представление отдельных результатов его прототипирования. Для их достижения необходимо решить следующие задачи:

- описать концепцию сервиса мониторинга сферы туризма;
- сформировать и изучить структуру содержания источников открытых данных в сфере туризма;
- на основе выявленной структуры источников данных построить онтологию объектов туризма;

- определить метод сбора данных и разработать алгоритм для сбора структурируемых наборов данных;
- обработать полученные данные, произвести унификацию, идентификацию объектов из различных источников и агрегацию их мер и измерений;
- разработать визуализации результатов описательной статистики полученных наборов данных.

## 1. Туризм в Иркутской области

Иркутская область является одним из крупнейших регионов России. Область расположена в Восточной Сибири и занимает юго-восточную часть Среднесибирского плоскогорья, плато и кряжи которого имеют высоты от 500 до 1000 м. Крайняя южная точка области располагается на  $51^\circ$  с. ш., северная оконечность почти достигает 65-й параллели. С севера на юг область протянулась почти на 1450 км, с запада на восток — на 1318 км. Юго-восточная граница Иркутской области проходит по оз. Байкал ([https://en.wikipedia.org/wiki/Irkutsk\\_Oblast](https://en.wikipedia.org/wiki/Irkutsk_Oblast)), которое в 1996 г. было внесено в Список объектов Всемирного наследия ЮНЕСКО ([https://en.wikipedia.org/wiki/Lake\\_Baikal](https://en.wikipedia.org/wiki/Lake_Baikal)).

Безусловно оз. Байкал является объектом притяжения туристов. В Национальном туристическом рейтинге-2021 Иркутская область заняла 15-е место и вошла в золотую двадцатку регионов, наиболее привлекательных для отечественных и иностранных туристов (<http://www.irk.ru/news/20220113/rating>).

В течение последних 20 лет численность туристов в коллективных средствах размещения неуклонно росла и в 2019 г. превысила 1 млн человек, а их доходы в 2021 г. составили 6 877 621.3 тыс. руб. (рис. 1) (<https://rosstat.gov.ru/statistics/turizm>). Этот факт свидетельствует, с одной стороны, об активном развитии сферы туризма в Иркутской области, а с другой — о повышении различных экологических рисков на территории уникального природного объекта. Мониторинг и управление данными рисками является насущной задачей.

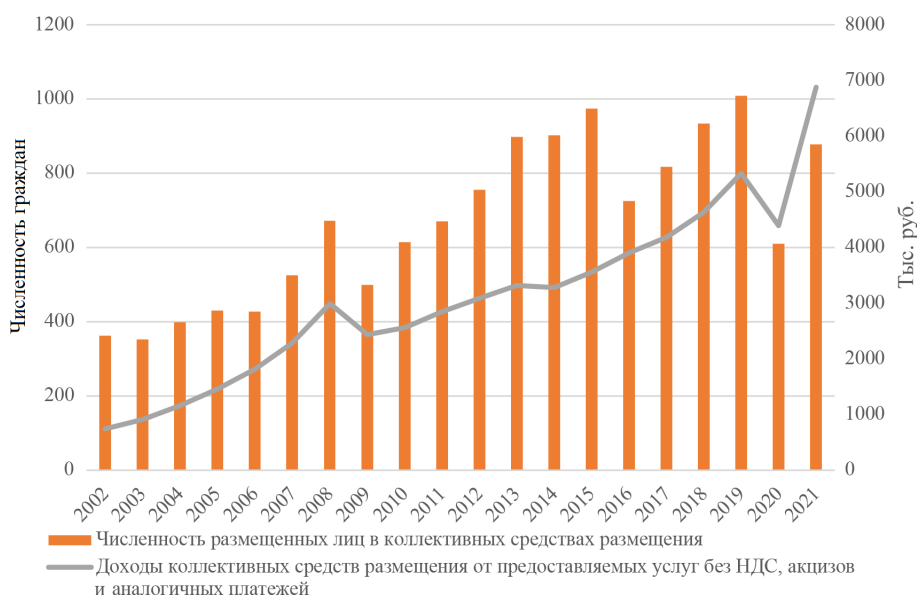


Рис. 1. Статистические данные в сфере туризма по Иркутской области (2021 г.)

Fig. 1. Tourism statistics in the Irkutsk Region (2021)

## 2. Концепция сервиса мониторинга сферы туризма

Для решения задачи мониторинга сферы туризма на заданной территории на основе открытых данных предлагается разработать сервис, обеспечивающий выполнение следующих функций:

1. Оперативное получение информации о социально-эколого-экономических показателях регионального туризма (объем показателей шире набора показателей официальной статистики и может собираться оперативно в режиме онлайн в отличие от возможностей сбора данных в результате административного управления):

- формирование туристического профиля территории (региона, района, местности):
  - виды туризма (рекреационный, событийный экологический, этнографический, активный, деловой, сельский, детский водный и круизный, социальный, самодельный, лечебно-здоровительный, экстремальный, паломнический, культурно-познавательный, гастрономический и др.);
  - “бренды” территории, достопримечательности (музеи, религиозные объекты, объекты природно-заповедного фонда, охотничье-рыболовные объекты, объекты сельского, промышленного, делового, военно-патриотического туризма, горнолыжные объекты);
  - площадки для наблюдения за природой;
  - туристические события;
  - туристические маршруты;
  - экологические туристские тропы;
  - экскурсии;
  - туристическо-информационные центры;
  - объекты размещения;
  - транспортные услуги, транспортная доступность;
  - объекты общественного питания;
  - климат;
- оперативный мониторинг (в том числе неорганизованного туризма):
  - видов туристических услуг (продуктов);
  - туристических потоков (их направления и плотность);
  - туристических маршрутов;
  - отзывов для выявления тональности и проблем в сфере туризма (экологических, транспортных, качества оказываемых услуг, их безопасности);
  - рекреационной (антропогенной) нагрузки территории, определение экологического риска;
  - населения, занятого в туризме (количество рабочих мест, образование);
  - рейтинга региона (района, местности);
  - влияния экстремальных событий (пандемия, лесные пожары — наличие дыма, гарей);
- мониторинг:
  - мест неорганизованного отдыха на основе распознавания космических снимков (размещение в палатках);
  - неучтенных в официальном реестре туристических баз (мест размещения туристов) на основе распознавания космических снимков;
  - информационной популярности достопримечательностей;

- районирование (ранжирование) территорий:
  - специализация видов туризма;
  - плотность туристических потоков;
  - рекреационная нагрузка.
- 2. Поддержка принятия решений для малого и среднего предпринимательства:
  - определение территории для бизнеса;
  - выбор, обоснование и формирование туристических продуктов для бизнеса.
- 3. Поддержка принятия решений для клиентов туристических услуг — подбор туристических продуктов.

Существуют различные способы реализации сервиса и организации к нему общего доступа, начиная от служб, предоставляющих различные услуги по аренде или размещению собственного оборудования в дата-центрах и заканчивая облачными платформами, предлагающими доступ к вычислительным ресурсам или сервисам по схеме Infrastructure as a Service (IaaS) или Platform as a Service (PaaS). При создании программной системы небольшой группой разработчиков (исследователей) наиболее предпочтителен подход, основанный на использовании сервисов облачных платформ. Данный выбор обусловлен такими факторами, как: отсутствие необходимости установки, настройки и сопровождения используемых при функционировании целевой программной системы средств и библиотек; отсутствие необходимости закупки оборудования и создания собственной инфраструктуры; снижение риска проекта, так как процесс исследования предполагает высокую вероятность существенных изменений требований и функциональности целевой программной системы; возможность автоматического масштабирования при увеличении нагрузки при использовании программной системы.

Наиболее популярными облачными платформами в настоящее время являются Amazon Web Services (<https://aws.amazon.com>), Google Cloud Platform (<https://cloud.google.com>), однако с точки зрения удобства использования, в частности поддержки русского языка, обеспечения требований ФЗ-152 РФ по защите персональных данных, наиболее целесообразно применение облачных платформ, учитывающих специфику Российской Федерации, например VK Cloud (<https://mcs.mail.ru/cloud-platform>) или Яндекс.Облако (<https://cloud.yandex.ru>). Последняя платформа наиболее предпочтительна исходя из ее функциональности. В процессе разработки сервиса предлагается проблемно-ориентированные функции оформить в виде независимых вычислительных блоков — элементов бессерверной (serverless) инфраструктуры, интеграция которых, вместе с решением стандартных задач по обеспечению хранения данных и отображения результатов их обработки, будет осуществлена с помощью средств облачной платформы. Данный подход хорошо укладывается в популярную концепцию микросервисной архитектуры, когда каждая функция программной системы оформляется в виде отдельного небольшого сервиса.

В рамках предложенной концепции сервиса реализуется функция “формирование туристского профиля территории” на примере Байкальской природной территории.

### 3. Источники информации

В настоящее время официальные статистические данные о состоянии сферы туризма Иркутской области не отражают полной информации о средствах размещения, объектах питания, оказываемых услугах и их качестве и т. д. [8, 9]. Среди формируемых статистических показателей отметим: численность размещенных лиц, численность

ночевок, доходы коллективных средств размещения и др. (<https://rosstat.gov.ru/statistics/turizm>). Один из главных недостатков этих данных — отсутствие их точной территориальной привязки для обеспечения возможности оценки рекреационной нагрузки рассматриваемой территории.

На сайте агентства по туризму Иркутской области (<https://irkobl.ru/sites/tour/index.php?type=special>) находится следующая информация: туристские паспорта муниципальных образований (8 из 32) и городских округов (7 из 10), туристические маршруты Иркутской области, актуальная информация для организаций туристической сферы и др. Информация представлена в слабоструктурированном и разрозненном текстовом виде в различных форматах. Необходимо отметить, что данные, доступные для дальнейшей автоматизированной обработки, отсутствуют.

Много информации о Байкальской природной территории находится на веб-ресурсах различной направленности:

- описание флоры и фауны Байкальской природной территории (<https://nature.baikal.ru/objs.shtml?obj=waterfall>);
- коллективные места размещений (Островок.ру, 101 Отель, Мир Турбаз);
- объекты общественного питания (сайт Туристического агентства Иркутской области);
- достопримечательности (сайт Туристического агентства Иркутской области, [https://www.russiadiscovery.ru/news/dostoprimechatelnosti\\_baikala](https://www.russiadiscovery.ru/news/dostoprimechatelnosti_baikala), социальные сети ВКонтакте и ОК);
- туристические маршруты (сайт Туристического агентства Иркутской области);
- мониторинг Байкальской природной территории (<https://baikalake.ru>).

В настоящей работе предлагается провести анализ отдельных указанных источников данных, выявить информационную модель данных, осуществить их сбор, интеграцию и визуализацию по различным критериям.

## 4. Онтология

Анализ источников информации позволил сформировать онтологию туристической сферы для дальнейшей унификации понятий, используемых в разрабатываемых информационных моделях. Для ее создания использовались онтологии сферы туризма, представленные в различных работах, например онтология TITERIA для описания туризма в Теруэле, которая включает основные понятия: party, heritage, history, cultural center, nature, gastronomy [10]; онтология для исследования туризма в Тайланде включает понятия: cultural and traditional event, event content, cultural activity, attraction, way of life, location, data time; онтология OntoTouTra (for the above, we propose OntoTouTra, an ontology that uses formal specifications to represent knowledge of tourist traceability systems), где используются понятия: accommodation, city, provider, tourist, experience, cultural, adventure, nature, attraction и др. [11]; онтология для формирования рекомендаций для туристов, где основное понятие tourist attraction [12].

Необходимо также отметить, что ЮНЕСКО разработал словарь терминов (<https://vocabularies.unesco.org/browser/thesaurus/ru>). В онтологии туристической сферы представлены основные классы: размещение, транспорт, достопримечательности, мероприятия, услуги, рестораны и культурное наследие (its ontology has the main classes: accommodation, transportation, attractions, activities, services, restaurants, and cultural

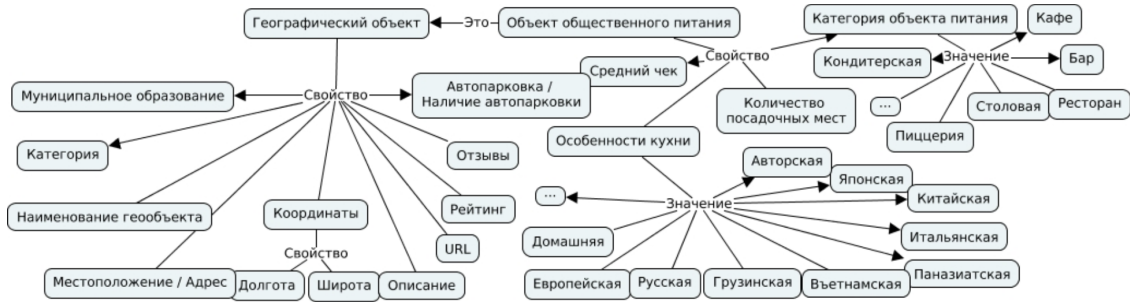


Рис. 2. Фрагмент онтологии объекта общественного питания

Fig. 2. Fragment of the ontology of the public catering object



Рис. 3. Онтология коллективных средств размещения

Fig. 3. Ontology of public accommodation facilities

heritage). Перечисленные онтологии обобщены и дополнены отношениями и свойствами понятий.

В рамках настоящей работы предлагаются онтологии объектов общественного питания (рис. 2), коллективных средств размещения (рис. 3), предоставляемых ими услуг и маршрутов (экскурсий).

## 5. Методы сбора и обработки данных

Процесс анализа данных состоит из следующих основных этапов (cross-industry process for data mining): определение проблемы, извлечение данных (сбор, описание, изучение, проверка качества), подготовка данных (очистка, трансформация, интеграция, форматирование), исследование и визуализация данных, создание модели прогнозирования (выбор алгоритма, обучение модели, оценка качества), проверка модели и ее тестирование, интерпретация результатов.

Для реализации функции сервиса формирования туристического профиля территории необходимо решить задачи сбора данных, их статистической обработки и визуализации.

### 5.1. Методы сбора данных

Получение данных веб-ресурсов в основном выполняется с использованием таких методов, как:

- онлайн опросы;
- запросы к базам данных;

- API (application programming interface) — это интерфейс обмена данными между приложениями;
- web-scraping (скрапинг).

Получение данных на основе онлайн опросов требует много времени и не дает возможность обеспечить полноту информации, наличие готовых баз данных позволяет получить уже структурированные данные, однако найти подходящие базы в свободном доступе довольно сложно. Метод API прост в реализации, но сервисы редко предоставляют возможность получения данных по API бесплатно и без ограничений, также метод не позволяет провести анализ отзывов и их оценку. Метод web-scraping предназначен для извлечения данных из структуры разметки динамических веб-источников и основан на автоматическом переходе страниц веб-источника и извлечении необходимой информации из DOM-структуры, имитируя поведение пользователя. Для решения поставленных задач этот метод наиболее перспективен.

Процесс web-scraping состоит из основных этапов [13, 14]:

1. Извлечение разметки страницы веб-источника, что осуществляется с использованием HTTP-запросов к ресурсу и сохранением полученных данных в переменную или файлы кода HTML веб-страницы.
2. Извлечение информации из структуры кода HTML осуществляется на основе поиска по заданному пути к элементам разметки кода: названию тегов, атрибутов и их значений.
3. Сохранение данных в структурируемом виде и дальнейшая их обработка.
4. Опционально: повтор действий.

В рамках настоящей работы реализацию сбора данных можно описать как сбор ссылок на необходимые объекты, а затем их обход с целью извлечения детальной информации и ее записи в таблицу об объектах. Также можно отметить особенности алгоритма реализации метода: его уникальность для каждого источника данных и необходимость с течением времени внесения изменений, так как разметка сайта может меняться его разработчиками.

## 5.2. Методы очистки и интеграции данных

На основе предлагаемых онтологий собранные данные из различных источников подвергаются обработке, в том числе опционально геокодингу (если неизвестны координаты объекта), и преобразованиям с целью получения списка уникальных объектов (рис. 4).

Ручная проверка набора данных необходима, поскольку в разных источниках встречаются записи об одном и том же объекте с ошибками в написании значений рассматриваемых свойств, например встречаются разночтения в названии объекта, номере дома, координатах, могут быть одинаковые названия объектов, но разные населенные пункты и т. д. По идентифицированным значениям находится максимальная дистанция между координатами записей, которую определяем как пороговое значение. Далее осуществляется поиск ближайшей записи (близость определяется на основе пороговых значений) по среднему значению нормированных показателей дистанций: расстояние Левенштейна используется при поиске названий объекта, населенного пункта, адреса, а дистанция между координатами — по формуле Хаверсина.

После идентификации объектов выполняются расчеты средних значений стоимости, пользовательских оценок (рейтинги), суммирование количества отзывов, определение категорий услуг и их популярности.



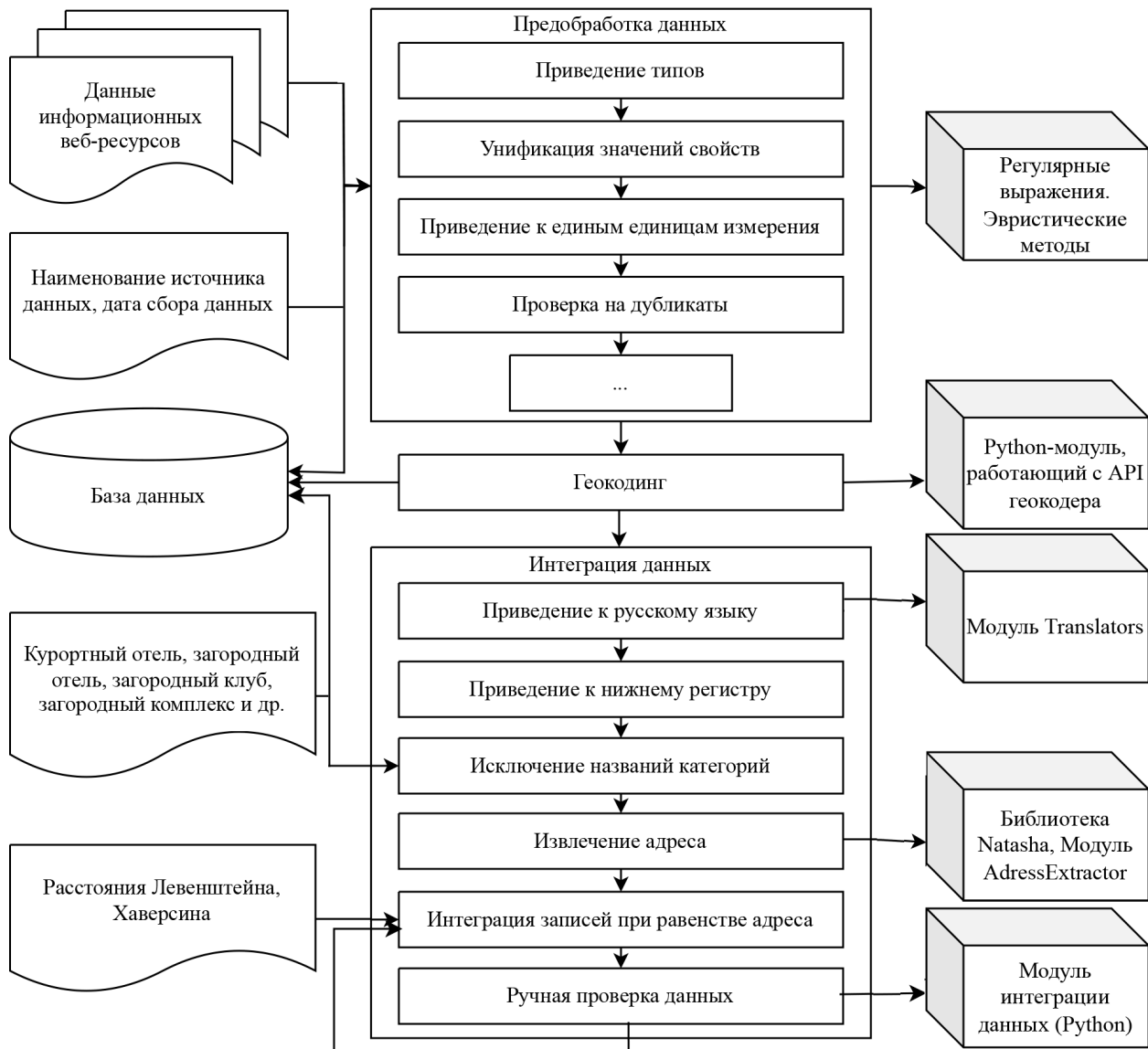


Рис. 4. Этапы алгоритма интеграции данных из нескольких источников  
 Fig. 4. Steps of the algorithm for integrating data from multiple sources

### 5.3. Методы анализа текстов (отзывов)

Полученные данные в виде коротких текстов (отзывов) из социальных сетей подвергаются следующим этапам обработки:

1. Подготовка данных к анализу предполагает исключение иностранных текстов или их перевод, например с использованием модуля Translator (<https://github.com/UlionTse/translators>).
2. Токенизация текста — разбиение текста на отдельные слова, исключая все остальные элементы (знаки пунктуации, смайлы и другие символы). Выполняется на основе регулярных выражений и специализированных методов.
3. Выделение именных сущностей (named entity recognition — NER), в контексте решаемой задачи определение локаций популярных мест отдыха с помощью библиотеки Stanza (<https://stanfordnlp.github.io/stanza>) [15].

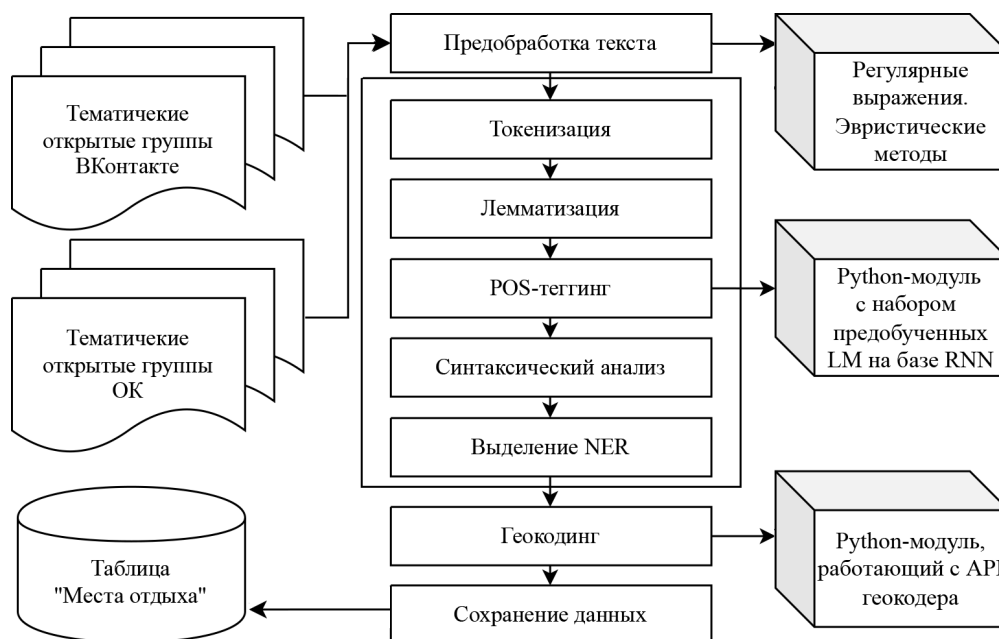


Рис. 5. Этапы анализа текста

Fig. 5. Text analysis stages

4. Лемматизация текста — приведение слов в начальную форму или стемминг — выделение основы слова. Лемматизация осуществляется с помощью библиотеки `Rmorphology2`, которая также позволяет провести морфологический анализ (задача POS-теггинга, *part-of-speech tagging*) — определение части речи. В результате из дальнейшего анализа исключаются предлоги.
5. Удаление стоп-слов (высокочастотные слова, которые не несут существенной информации).
6. Формирование словаря решаемой задачи, что позволяет провести семантический анализ текста для определения: сущностей, действий, описаний с привязкой к локациям. Это база знаний о флоре и фауне, объектах инфраструктуры и других характеристиках территории.
7. Сентимент-анализ текста — определение тональности на основе классификации характеристик на классы “негатив” и “позитив”, что позволит дополнительно выделить проблемы и предпочтения по локациям (территориям). Для этого используется метод преобразования токенов в числовые векторы (эмбединг), а далее осуществляется задача классификации с помощью нейросети.

Каждый этап анализа текста (токенизация, лемматизация, синтаксический анализ, выделение именованных сущностей) осуществляется с помощью нейронных сетей (рис. 5).

#### 5.4. Методы визуализации данных

Основной задачей визуализации является обеспечение поддержки пользователя в процессе восприятия, понимания и осмысления информации и формирования новых знаний, а также обеспечение минимизации усилий по выполнению когнитивных задач в сравнении с текстовым представлением данных. Для визуализации и обработки данных все большую актуальность приобретают технологии облачных вычислений — VI-

системы/платформы (business intelligent system), которые обеспечивают: возможность анализировать и отображать различные данные и информацию, связанные с рассматриваемой задачей. Среди существующих платформ выделим Tableau, Power BI, Yandex DataLens и др. Эти платформы обладают следующей основной функциональностью: визуализация данных, автоматический поиск зависимостей и скрытых взаимосвязей, создание интерактивных отчетов, глубокая аналитика, интеграция с другими инструментами (Microsoft Excel, Google Sheets), машинное обучение. Системы позволяют эффективно визуализировать данные с помощью информационных панелей (так называемых дашбордов, с англ. dashboard — “приборная панель”), состоящих из нескольких графиков и диаграмм с возможностью интерактивного взаимодействия с пользователем.

В случае, когда сервис создается на многофункциональной платформе, такой как Яндекс.Облако, целесообразно использовать средство визуализации этой платформы — Yandex DataLens.

## 6. Реализация сервиса мониторинга

Предлагаемую технологию сбора, обработки и представления социально-эколого-экономических данных об оперативном состоянии региональной сферы туризма планируется реализовать с помощью средств облачной платформы Яндекс.Облако на основе подхода микросервисной архитектуры. В качестве средства реализации вычислительных блоков предлагается использовать облачные функции (cloud function, <https://cloud.yandex.com/en/docs/functions/service-comparison>), для организации каналов обмена данными — очереди сообщений (message queue, <https://cloud.yandex.com/en/docs/message-queue/concepts/queue>), для общей интеграции сервис — API Gateway (<https://cloud.yandex.com/en/docs/api-gateway/concepts>), для хранения данных — PostgreSQL.

Вид архитектуры программной системы, реализующей предлагаемую технологию, отображен на рис. 6. Пользователи программной системы подразделяются на две основные группы, первая — это разработчики, которые после авторизации имеют полный доступ к функциональности, вторая — обычные пользователи, которым не требует-

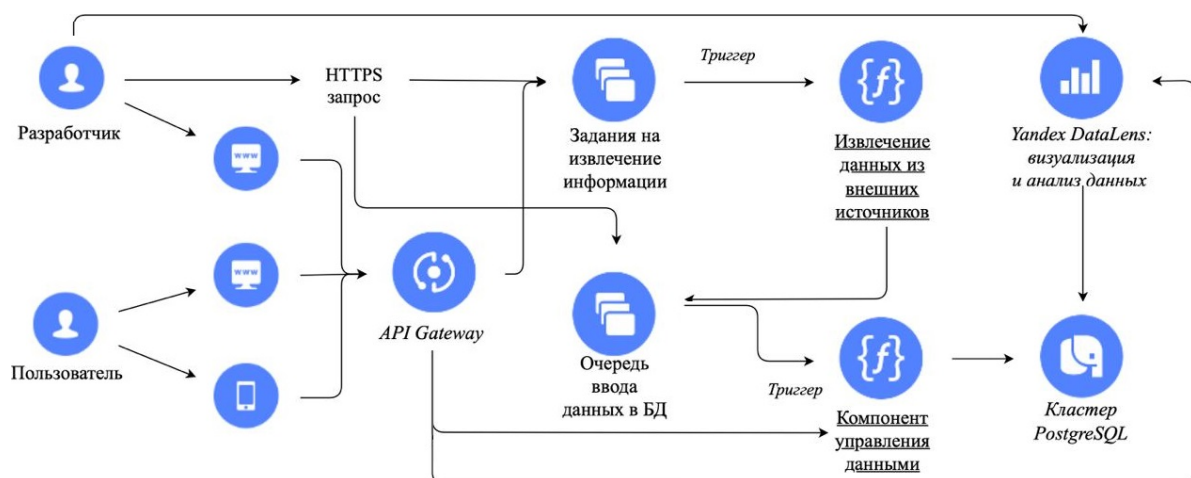


Рис. 6. Архитектура программной системы

Fig. 6. Software System Architecture

ся авторизация и доступны только информационные ресурсы, содержащие результаты анализа собранных данных. Эта классификация не является полной, но достаточна для текущего этапа процесса разработки.

Процесс сбора информации организован на основе двух очередей сообщений, где первая очередь содержит информацию об источниках, которые требуется обработать в виде “одно сообщение — один источник”, например запрос к API социальной сети ВКонтакте и его параметры (рис. 6, Задания на извлечение информации), а вторая содержит результаты обработки, которые требуется записать в базу данных PostgreSQL в виде “сообщение — одна строка целевой таблицы” (рис. 6, Очередь ввода данных в БД). Для обработки сообщений в очереди используется встроенный механизм триггеров, который обеспечивает передачу сообщений в облачную функцию-обработчик в виде массива с фиксированной максимальной длиной. Для извлечения информации из внешних источников используются функции `get_vk()` и `get_ok()`, чтобы сформировать данные постов сообществ и комментариев к ним из социальных сетей ВК и Одноклассники соответственно. Для записи данных в базу данных используется вариант компонента `data-controller` из платформы автоматизации создания интеллектуальных систем как сервиса (AI PaaS), использующий снимок состояния структуры базы данных для минимизации требуемых прав доступа (не нужен доступ к метаданным) и повышения быстродействия. В штатном режиме управление процессом добавления сообщений в очередь “Задания на извлечение информации” реализовано в рамках спецификации API Gateway, однако для нужд тестирования и отладки у разработчиков существует возможность доступа к очередям в режиме “только запись” (выделено пунктиром на рис. 6) на основе встроенного HTTPS-интерфейса очереди.

Для визуализации результатов анализа собранных данных, в том числе с применением возможностей ГИС, предложено использовать специализированное средство платформы Яндекс.Облако — сервис Yandex DataLens. Этот сервис предоставляет возможность построения различных графиков, диаграмм, слоев для отображения геоинформации в рамках специализированного пользовательского интерфейса, обеспечивающего аналитику поддержку на всем процессе формирования нужной панели визуализации — от формирования запроса к базе данных до выбора способа визуализации результатов. Возможность параметризации созданных средств отображения результатов обработки данных позволяет использовать их в качестве способа реализации функции анализа данных в разрабатываемой программной системе.

Интеграция отдельных блоков в программную систему выполняется на основе сервиса API Gateway путем создания спецификации в формате OpenAPI 3.0, которая позволяет в декларативной форме представить функции целевой программной системы в виде набора URL-адресов и указать, каким образом будет осуществляться обработка поступающих на эти адреса запросов. В данной работе использованы следующие варианты: интеграция с Object Storage, которая используется для представления статического содержимого программной системы (HTML-код, Javascript- и CSS-файлы), вызовы облачных функций, которые обеспечивают взаимодействие с базой данных и работу с методами сбора и обработки информации (подготовка данных, геокодирование), интеграция с очередями сообщений, которая позволяет управлять процессом сбора информации из внешних источников. Кроме того, спецификация API Gateway обеспечивает решение хоть и рутинной, но обязательной и достаточно трудоемкой задачи организации разграничения доступа к программной системе, в данной работе используется механизм авторизации на основе oauth Яндекса [16].

## 7. Результаты сбора и анализа данных

К настоящему времени собрана и обработана информация о туристических объектах: коллективных средствах размещения — 1163 (685) (количество объектов, о которых собрана информация, в скобках — количество идентифицированных объектов, включая геолокацию), услугах — 61 (16), объектах общественного питания — 809, достопримечательностях — 395 (352), туристических маршрутах и экскурсиях — 94.

Реализованы информационные панели (дашборды):

- 1) о коллективных средствах размещения и их услугах с возможностью фильтрации по рейтингу (от “плохо” до “великолепно”), населенным пунктам/районам и категории услуг (рис. 7):
  - отображение количества средств размещений и средней стоимости по населенным пунктам/районам в виде комбинированного графика;
  - отображение коллективных средств размещения в виде карты пропорциональных объектов по номерному фонду (количеству номеров);
  - отображение количества средств размещений по категории/подкатегории услуг в виде столбчатой диаграммы и тепловой карты;
- 2) об объектах общественного питания с возможностью фильтрации по районам и типу кухни (рис. 8):
  - отображение в виде карты пропорциональных объектов общественного питания по среднему чеку и типу кухни;
  - отображение агрегированных количественных показателей по населенным пунктам/районам для описания территорий по заданной мере (дорогие — дешевые, по количеству объектов общественного питания) в виде комбинированного графика;
  - отображение тепловой карты по количеству объектов общественного питания в разрезе районов и типа кухни;

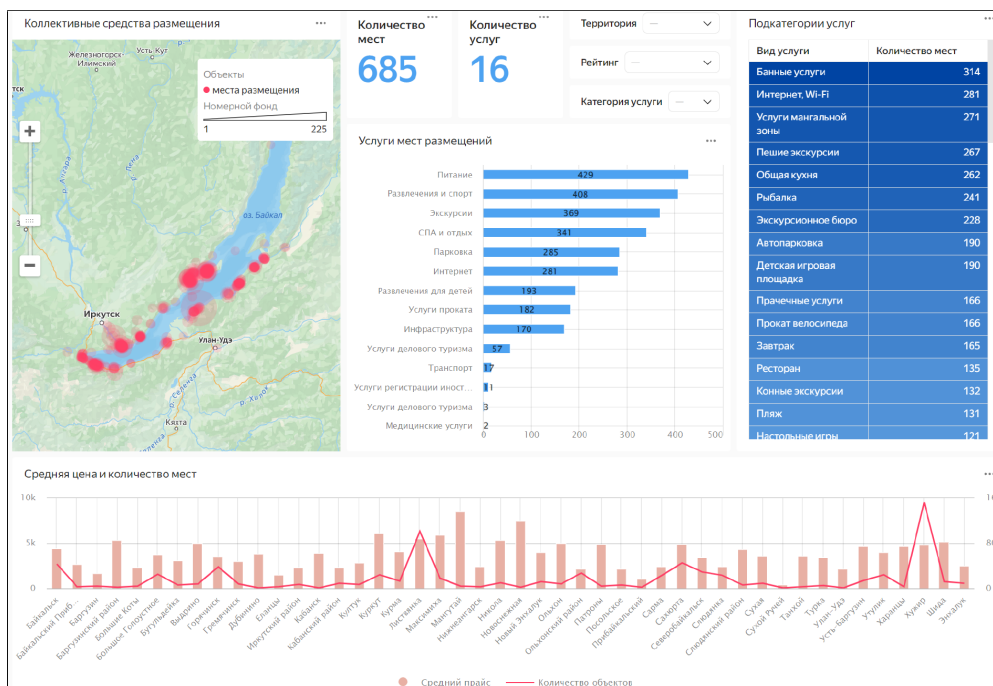


Рис. 7. Фрагмент информационной панели “Коллективные средства размещения”  
 Fig. 7. Fragment of public accommodation dashboard

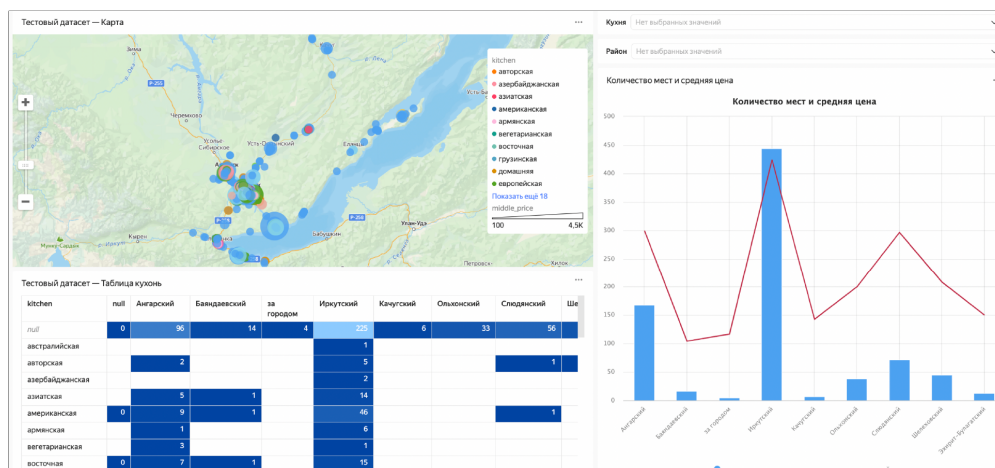


Рис. 8. Фрагмент информационной панели “Объекты питания”  
 Fig. 8. Fragment of catering establishments dashboard

- 3) о достопримечательностях и популярных местах отдыха:
  - отображение на карте и обеспечение возможности получения описательной информации о достопримечательностях;
  - отображение на карте плотности точек достопримечательностей;
  - отображение на карте популярных мест отдыха;
- 4) о туристических маршрутах и экскурсиях с возможностью фильтрации по названию маршрута:
  - отображение на карте населенных пунктов, где организованы туристические маршруты или экскурсии;
  - отображение на карте схем маршрутов и экскурсий.

Так, среди 339 мест с рейтингом “великолепно” более распространены услуги категории “развлечения и спорт”, а при детализации категорий становится известно, что чаще всего средства размещения предлагают “пешие экскурсии”. На такой сценарий формирования перечня услуг может ориентироваться бизнес, так как он охарактеризован высокой пользовательской оценкой.

Наибольшее количество объектов размещения сосредоточено в Ольхонском районе, в поселке Хужир. Наибольший средний чек у средств размещения в населенных пунктах — Мангутае и Новоснежной, что находится в южной части Байкала. В них также наблюдается низкое число объектов, что можно интерпретировать как наличие низкой конкуренции и низкую нагрузку рекреационной зоны.

## Заключение

Существующие программные сервисы решают задачи мониторинга сферы туризма территории на основе официальной статистической информации, полученной государственными органами управления, или данных, полученных от туристов в результате специальных договоренностей. Необходимо отметить, что органы территориального управления не владеют полной и актуальной информацией о туризме. Данные из открытых источников, в основном из социальных сетей, используются в рамках отдельных исследований туристических городов или отдельных достопримечательностей.

В настоящем исследовании начата работа по созданию информационной технологии, позволяющей на основе данных из открытых источников, информационных веб-

ресурсов сформировать туристический профиль территории, который в дальнейшем станет основой для исследований по созданию методов и средств поддержки принятия решений для туристов, бизнеса и органов территориального управления.

В работе мы поставили задачу разработки средств информационных технологий для мониторинга территориального туризма, описали концепцию сервиса, обосновали методы анализа и визуализации данных, построили онтологическую модель, определили наполнение дашбордов, реализовали прототип сервиса и продемонстрировали результаты сбора, обработки и представления данных, полученных в ходе исследования.

В рамках предлагаемой концепции сервиса разработан прототип функции формирования туристического профиля территории на примере района оз. Байкал.

Необходимо отметить важную особенность, что разработанные методы, реализованные в рамках облачной платформы, позволят собирать и обрабатывать данные о территории с заданной периодичностью, позволяя формировать различные тренды, характеризующие динамику сферы туризма.

В дальнейшем усилия авторов будут направлены на анализ данных социальных сетей для получения детальной информации о туристических маршрутах: популярности, качестве оказания услуг, социальных и экологических проблемах. Результаты исследования позволят оценить рекреационную нагрузку территории и принять решения для дальнейшего развития туризма.

**Благодарности.** Работа выполнена при поддержке РНФ (проект № 23-28-00844, <https://rscf.ru/project/23-28-00844/>).

## Список литературы

- [1] SHAPETOURISM OBSERVATORY. Available at: <https://www.quantitas.it/data/shapetourism/build/index.php> (accessed 12.01.2023).
- [2] Destination NSW. Available at: <https://www.destinationnsw.com.au/about-us> (accessed 15.01.2023).
- [3] **Lobao F., Aparicio M., Neto M.** Smart tourism — city tourism radar: a tourism monitoring tool at the city of Lisbon. Proc. Conf. the Portuguese Association of Information Systems (CAPSI'2019). Portugal; 2019: 1–21.
- [4] **Li H., Hu M., Li G.** Forecasting tourism demand with multisource big data. *Annals of Tourism Research*. 2020; (83):102912. DOI:10.1016/j.annals.2020.102912.
- [5] **Soualah-Alila F., Coustaty M., Rempulski N., Doucet A.** DataTourism: designing an architecture to process tourism data. Proc. Conf. Information and Communication Technologies in Tourism. Switzerland; 2016: 751–763. DOI:10.1007/978-3-319-28231-2\_54.
- [6] **Рубцова Н.В.** Формирование системы мониторинга эффективности сферы туристско-рекреационных услуг региона. *Мир экономики и управления*. 2019; 19(3):101–110. DOI:10.25205/2542-0429-2019-19-3-101-110.
- [7] **Li J., Xu L., Tang L., Wang S., Li L.** Big data in tourism research: a literature review. *Tourism Management*. 2018; (68):301–323. DOI:10.1016/j.tourman.2018.03.0090261-5177.
- [8] **Котельников Д.А.** Формирование системы показателей для ведения мониторинга устойчивого развития туристских территорий. *Перспективы развития науки в современном мире: сборник статей по материалам Всероссийского научно-исследовательского конкурса научных инноваций*. 2020: 41–50.
- [9] **Лебедева Ю.А.** Организация мониторинга качества туристских услуг на муниципальном уровне. *Чебоксары: Среда*; 2020: 132. DOI:10.31483/a-202.

- [10] Garrido P., Barrachina J., Martinez F.J., Seron F.J. Smart tourist information points by combining agents, semantics and AI techniques. *Computer Science and Information Systems*. 2017; 14(1):1–23. DOI:10.2298/CSIS150410029G.
- [11] Mendoza-Moreno J.F., Santamaria-Granados L., Fraga V.A., Ramirez-Gonzalez G. OntoTouTra: tourist traceability ontology based on big data analytics. *Applied Sciences*. 2021; 11(22):11061. DOI:10.3390/app112211061.
- [12] Pai M.-Y., Wang D.-C., Hsu T.-H., Lin G.-Y., Chen C.-C. On ontology-based tourist knowledge representation and recommendation. *Applied Sciences*. 2019; 9(23):5097. DOI:10.3390/app9235097.
- [13] Mitchell R. *Web scraping with Python: collecting data from the modern*. O'Reilly Media; 2018: 256.
- [14] Москаленко А.А., Лапоница О.Р., Сухомлин В.А. Система управления доступом к ресурсам веб приложений на основе анализа поведения пользователя. *International Journal of Open Information Technologies*. 2020; 8(9):30–35.
- [15] Qi P., Zhang Y., Zhang Y.J., Bolton C.D. Manning Stanza: a Python natural language processing toolit for many human languages. Available at: <https://arxiv.org/pdf/2003.07082.pdf> (accessed 11.07.2023).
- [16] Pavlov A.I., Stolbov A.B., Lempert A.A. Towards extensibility features of knowledgebased systems development platform. *CEUR Workshop Proceedings: 4th Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS'2021)*. Irkutsk; 2021: 2984.

## A concept of the tourism industry monitoring service based on the analysis of web resources

O. A. NIKOLAICHUK\*, YU. V. PESTOVA, A. I. PAVLOV, D. E. KOSOGOROV

Matrosov Institute of System Dynamics and Control Theory SB RAS, 664033, Irkutsk, Russia

\*Corresponding author: Olga A. Nikolaichuk, e-mail: [nikoly@icc.ru](mailto:nikoly@icc.ru)

Received October 10, 2023, accepted October 17, 2023.

### Abstract

Progress in sustainable tourism requires monitoring of all diverse aspects of the tourism sector for any territory. Expansion of information technologies provides tools for collection, analysis, processing and visualization of relevant information. This study addresses a concept of the tourism monitoring service in the Irkutsk Region within the Lake Baikal area. We define methods and software for implementing this service and present individual results of its prototyping. The design of such service implies the development and software implementation of methods for collecting, processing and visualizing data. Additionally, it should be able to exploit the functionality of the existing cloud platforms for the main information processes. The data collection methods are based on Application Programming Interface and web scraping technologies. Processing methods include statistical methods and artificial intelligence methods. Visualization methods are based on techniques from Business Intelligent Systems. This paper describes the concept of the addressed service and justifies methods for data analysis and visualization, builds an ontological model, defines the content of dashboards, implements a prototype of the service and demonstrates the results of



data collection, processing, and presentation obtained from the research. To date, information has been collected and processed for the following tourist attractions: 685 accommodation facilities, 16 services, 809 catering establishments, 352 attractions, and 94 tourist routes and excursions. The proposed technology (methods and software) of the tourism monitoring service has been proved to be effective.

*Keywords:* tourism monitoring, tourist profile of the territory, ontology, collection of data from open sources, analysis of reviews, web service, cloud platform.

*Citation:* Nikolaichuk O.A., Pestova Yu.V., Pavlov A.I., Kosogorov D.E. A concept of the tourism industry monitoring service based on the analysis of web resources. 2023; 28(6):118–134. DOI:10.25743/ICT.2023.28.6.011. (In Russ.)

**Acknowledgements.** The research was supported by Russian Science Foundation (project No. 23-28-00844, <https://rscf.ru/project/23-28-00844/>).

### References

1. SHAPETOURISM OBSERVATORY. Available at: <https://www.quantitas.it/data/shapetourism/build/index.php> (accessed 12.01.2023).
2. Destination NSW. Available at: <https://www.destinationnsw.com.au> (accessed 15.01.2023).
3. **Lobao F., Aparicio M., Neto M.** Smart tourism — city tourism radar: a tourism monitoring tool at the city of Lisbon. Proc. Conf. the Portuguese Association of Information Systems (CAPSI'2019). Portugal; 2019: 1–21.
4. **Li H., Hu M., Li G.** Forecasting tourism demand with multisource big data. Annals of Tourism Research. 2020; (83):102912. DOI:10.1016/j.annals.2020.102912.
5. **Soualah-Alila F., Coustaty M., Rempulski N., Doucet A.** DataTourism: designing an architecture to process tourism data. Proc. Conf. Information and Communication Technologies in Tourism. Switzerland; 2016: 751–763. DOI:10.1007/978-3-319-28231-2\_54.
6. **Rubtsova N.V.** Formation of the system for monitoring the efficiency of regional tourist-recreational services. World of Economics and Management. 2019; 19(3):101–110. DOI:10.25205/2542-0429-2019-19-3-101-110. (In Russ.)
7. **Li J., Xu L., Tang L., Wang S., Li L.** Big data in tourism research: a literature review. Tourism Management. 2018; (68):301–323. DOI:10.1016/j.tourman.2018.03.0090261-5177.
8. **Kotelnikov D.A.** Formation of a system of indicators for monitoring the sustainable development of tourist areas. Proceedings of the Competition of Scientific Innovations: Prospects for the Development of Science in the Modern World: Collection of Articles Based on the Materials of the All-Russian Research Competition. 2020: 41–50. (In Russ.)
9. **Lebedeva Yu.A.** Organizatsiya monitoringa kachestva turistskikh uslug na munitsipal'nom urovne [Organization of monitoring of the quality of tourist services at the municipal level]. Cheboksary: Sreda; 2020: 132. DOI:10.31483/a-202. (In Russ.)
10. **Garrido P., Barrachina J., Martinez F.J., Seron F.J.** Smart tourist information points by combining agents, semantics and AI techniques. Computer Science and Information Systems. 2017; 14(1):1–23. DOI:10.2298/CSIS150410029G.
11. **Mendoza-Moreno J.F., Santamaria-Granados L., Fraga V.A., Ramirez-Gonzalez G.** OntoTouTra: tourist traceability ontology based on big data analytics. Applied Sciences. 2021; 11(22):11061. DOI:10.3390/app112211061.
12. **Pai M.-Y., Wang D.-C., Hsu T.-H., Lin G.-Y., Chen C.-C.** On ontology-based tourist knowledge representation and recommendation. Applied Sciences. 2019; 9(23):5097. DOI:10.3390/app9235097.
13. **Mitchell R.** Web scraping with Python: collecting data from the modern. O'Reilly Media; 2018: 256.
14. **Moskalenko A.A., Laponina O.R., Sukhomlin V.A.** System for managing access to web application resources based on user behaviour analysis. International Journal of Open Information Technologies. 2020; 8(9):30–35. (In Russ)
15. **Qi P., Zhang Y., Zhang Y.J., Bolton C.D.** Manning Stanza: a Python natural language processing toolkit for many human languages. Available at: <https://arxiv.org/pdf/2003.07082.pdf> (accessed 11.07.2023).
16. **Pavlov A.I., Stolbov A.B., Lempert A.A.** Towards extensibility features of knowledge-based systems development platform. CEUR Workshop Proceedings: 4th Scientific-Practical Workshop Information Technologies: Algorithms, Models, Systems (ITAMS'2021). Irkutsk; 2021: 2984.