

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет информационных технологий и управления
Кафедра интеллектуальных информационных технологий

К защите допустить:

Заведующий кафедрой ИИТ

_____ Д. В. Шункевич

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к курсовому проекту
по дисциплине «Проектирование баз знаний»
на тему:

**РЕШАТЕЛЬ ЗАДАЧ ЭЛЕКТРОННОГО ГИДА ПО
РЕСПУБЛИКЕ БЕЛАРУСЬ**

БГУИР КП5 1-40 03 01 02 042 ПЗ

Студент гр. 221703
Руководитель

Д. Н. Вечорко
С. А. Никифоров

Минск 2024

СОДЕРЖАНИЕ

Перечень условных обозначений	3
Введение	4
1 Анализ подходов к разработке	6
1.1 Концептуальный анализ системы	6
1.1.1 Архитектура базы знаний системы	7
1.2 Подходы к разработке	9
1.2.1 Технология OSTIS	9
1.2.2 Альтернативные подходы	10
1.3 Анализ аналогов	11
1.3.1 Общеознакомительная информация об аналогах	11
1.3.2 Перечисление достоинств и недостатков аналогов	13
1.3.3 Вывод по анализу аналогов	14
1.4 Вывод	15
2 Проектирование электронного гида РБ	17
2.1 Постановка задачи	17
2.2 Описание пользователя Электронного гида РБ	18
2.3 Проектирование парсера	19
2.4 Вывод	20
3 Разработка электронного гида	21
3.1 Используемые средства разработки	21
3.1.1 Git и GitHub	21
3.1.2 Visual Studio Code	21
3.1.3 WINK-vscode-extention	21
3.1.4 PyCharm IDE	21
3.1.5 Инструменты для парсера	22
3.2 Демонстрация работы системы	23
3.3 Вывод	24
Заключение	26
Список использованных источников	27

ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ

В курсовой работе используются следующие условные обозначения:

- БЗ — База знаний;
- РЗ — Решатель задач;
- ПрО — Предметная область;
- РБ — Республика Беларусь;
- OSTIS — Open Semantic Technology for Intelligent Systems, Открытые Семантические Технологии для Интеллектуальных Систем;

ВВЕДЕНИЕ

В современном информационном обществе создание и развитие интеллектуальных информационных систем (ИИС) является одним из ключевых направлений технологического прогресса. Использование таких систем имеет огромный потенциал для улучшения эффективности работы в различных областях, включая туризм, экономику, культуру и образование. В рамках данного курсового проекта рассматривается создание базы знаний ИИС для электронного гида Республики Беларусь, что является актуальной задачей в контексте стремления к интеграции современных информационных технологий в повседневную жизнь общества.

Основная проблема заключается в неэффективном поиске и доступе к информации о Республике Беларусь из-за разрозненности и разнообразия источников данных. Пользователи, пытаясь найти необходимую информацию о мероприятиях, достопримечательностях, предприятиях и других объектах в Республике Беларусь, сталкиваются с трудностями из-за необходимости обращаться к разным источникам, часто несовместимым друг с другом. Недостаток централизованной базы знаний и отсутствие единого стандарта для представления информации приводят к дублированию данных, устареванию информации и недостаточной достоверности содержания.

Решение данной проблемы позволит повысить эффективность работы пользователей электронного гида Республики Беларусь, улучшить качество принимаемых ими решений и сделать процесс поиска информации более быстрым и удобным.

Целью данного проекта является создание интегрированной информационной системы, которая позволит эффективно собирать, хранить, обрабатывать и предоставлять пользователю актуальную информацию о Республике Беларусь через электронный гид, тем самым решая вышеупомянутую проблему.

Для успешной реализации поставленной цели необходимо определить и детализировать ряд задач, которые будут направлены на достижение этой цели:

- Анализ потребностей пользователей и особенностей информационной среды Республики Беларусь.
- Проектирование структуры и функциональности ИИС электронного гида Республики Беларусь.
- Реализация программного обеспечения ИИС с учетом специфических требований и характеристик.
- Тестирование и отладка созданной системы для обеспечения ее надежной и эффективной работы.

Вышепредставленные основные задачи, будут разбиты на логические этапы, которые позволят систематизировать процесс разработки и

обеспечить эффективное выполнение поставленной цели. Каждая задача направлена на выполнение конкретных шагов, начиная с анализа потребностей пользователей и заканчивая тестированием и отладкой созданной системы.

1 АНАЛИЗ ПОДХОДОВ К РАЗРАБОТКЕ

В данном разделе пояснительной записки проводится концептуальный анализ системы, в рамках которого будут рассмотрены основные составляющие приложения, которые помогут достигнуть поставленной цели проекта. Также проведём комплексный анализ существующих подходов к разработке и аналогичных систем что позволит создать полноценное представление о текущем состоянии и возможностях для нашего приложения. Разрабатываемое интерактивное веб-приложение для исследования достопримечательностей Беларуси направлено на решение актуальных задач, связанных с доступом к информации о туристических объектах и персонализированными рекомендациями для пользователей.

В условиях информационной перегруженности и растущих потребностей современного туриста важно понимать, как различные системы справляются с предоставлением структурированной и актуальной информации. Анализ аналогичных приложений поможет выявить их сильные и слабые стороны, а также определить лучшие практики, которые можно интегрировать в наше решение. Кроме того, мы рассмотрим подходы к разработке, чтобы обеспечить успешную реализацию и функциональность приложения.

1.1 Концептуальный анализ системы

Систему «Электронный гид по Республике Беларусь» разобьём на несколько составляющих, которые обеспечат основную функциональность системы и будут решать различные задачи.

Для реализации цели, то есть разработки интерактивного электронного гида, можно выделить следующие подзадачи, выполняемые различными подсистемами:

- 1 Хранение знаний о достопримечательностях в системе.
- 2 Сбор информации о достопримечательностях.
- 3 Авторизация и отслеживание действий пользователя.
- 4 Анализ и рекомендации достопримечательностей.

В рамках курсового проекта предыдущего семестра была создана обширная база знаний для хранения информации о достопримечательностях. Эта база знаний обладает четко заданной структурой и иерархией предметных областей, что позволяет организовать данные в строгом и систематизированном виде. Благодаря такой структуре обеспечивается целостность данных, их однозначность и удобство поиска, что делает базу знаний надежным хранилищем информации о достопримечательностях.

Рассмотрим структуру базы знаний и декомпозицию её предметных областей.

1.1.1 Архитектура базы знаний системы

При проектировании базы знаний электронного гида по Республике Беларусь раздел предметной области электронного гида был декомпозирован как показано на изображении 1.1

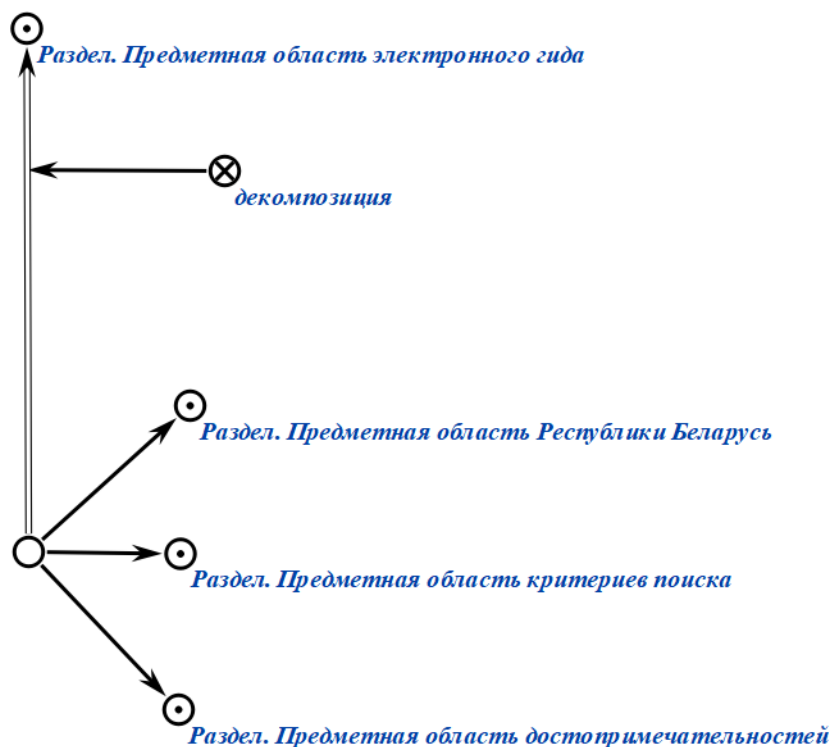


Рисунок 1.1 – Декомпозиция раздела предметной области достопримечательностей

С целью выделения классов достопримечательностей с общими свойствами была выполнена декомпозиция раздела предметной области достопримечательностей на более мелкие разделы предметных областей как показано на изображении 1.2

С целью указания географической принадлежности каждой достопримечательности раздел предметной области Республики Беларусь также подвергся декомпозиции по административно-территориальному признаку на разделы предметных областей районов Республики Беларусь как показано на изображении 1.3

Для более быстрого и удобного расширения базы знаний требуется подсистема автоматизированного сбора данных. В её собственные подзадачи включен сбор и обновление сведений о достопримечательностях в базе знаний. В рамках данного курсового проекта работа будет сосредоточена на выполнении задачи сбора информации и представления информации в стандартных структурах данных для возможности дальнейшей обработки

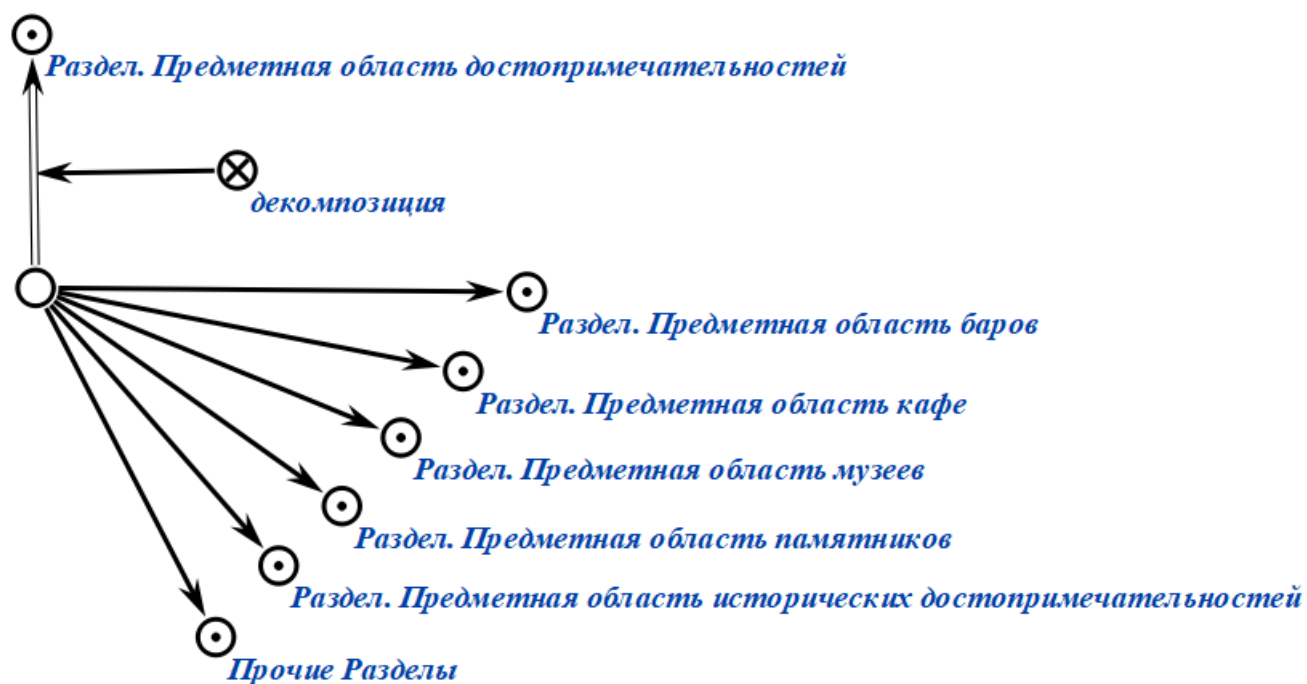


Рисунок 1.2 – Декомпозиция раздела предметной области достопримечательностей

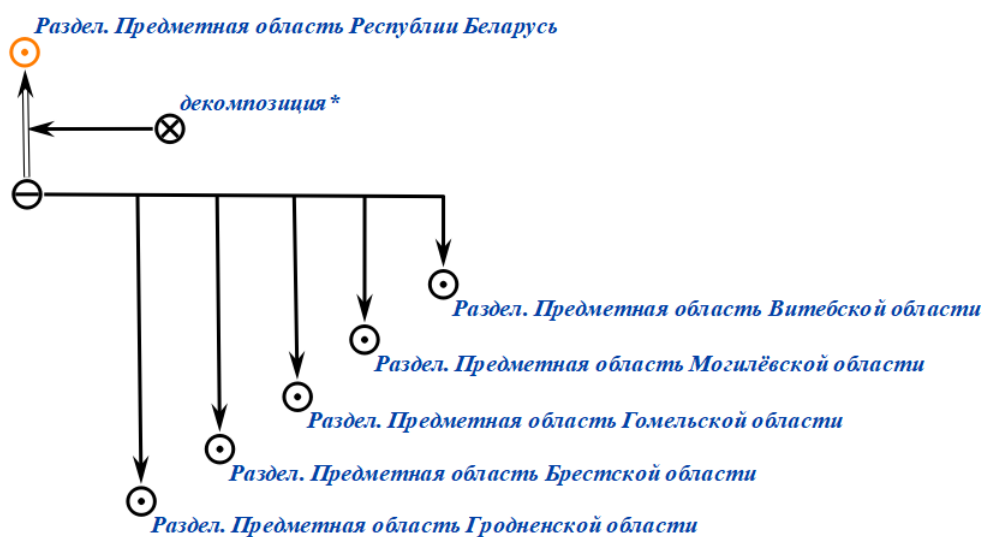


Рисунок 1.3 – Декомпозиция раздела предметной области Республики Беларусь

и загрузки в базу знаний, для обновления сведений. Реализация остальных подсистем электронного гида, которые будут рассмотрены далее, возложена на коллег по проекту.

Система регистрации и авторизации будет обеспечивать персонализированный доступ к информации, позволяя пользователям сохранять и отслеживать избранные места, получать уведомления об обновлениях и изменениях.

Подсистема анализа и рекомендаций предназначена для анализа интересов и предпочтений пользователей, а также использования их активности для формирования персонализированных рекомендаций.

Далее рассмотрим альтернативные подходы к разработке, проанализируем как аналогичные системы, схожие по назначению, решают поставленные задачи, в рамках заданной предметной области.

1.2 Подходы к разработке

1.2.1 Технология OSTIS

Для разработки интерактивного веб-приложения электронного гида по Республике Беларусь используется технология OSTIS, которая основана на SC-code (SCs, SCg, SCp) для хранения и обработки знаний.

Технология OSTIS представляет собой мощный инструмент для проектирования системы электронного гида по Республике Беларусь по нескольким причинам:

- **Универсальность и гибкость:** OSTIS позволяет создавать универсальные и адаптивные структуры данных, что особенно важно для приложения, где информация о достопримечательностях, маршрутах и туристических объектах может быть разнообразной и требовать гибкого представления.

- **Семантическое представление:** SC-code обеспечивает семантическое представление информации, что упрощает интеграцию и обработку различных типов данных, таких как текст, изображения, аудио и видео. Это особенно полезно для приложения, где информация о туристических объектах и маршрутах может включать текстовые описания, фотографии и видеоматериалы.

- **Масштабируемость:** OSTIS позволяет легко масштабировать систему, добавляя новые данные и расширяя существующие структуры. Это важно для приложения, где информация о достопримечательностях и туристических объектах постоянно обновляется и расширяется.

- **Интеграция с другими системами:** OSTIS может быть легко интегрирована с другими системами и платформами, что позволяет создавать комплексные решения для приложения, включая интеграцию с

социальными сетями, платформами бронирования и другими сервисами.

- **Открытость и совместимость:** OSTIS является открытой технологией, что дает возможность разработчикам и организациям свободно использовать и модифицировать её для своих нужд. Это может быть полезно для приложения, где требования могут изменяться в зависимости от времени и потребностей пользователей.

SC-код, в свою очередь, обладает рядом особенностей:

- **Формальная семантика:** это позволяет системе делать выводы и создавать новые знания на основе логических правил.

- **Поддержка семантического поиска:** что позволяет получать более точные результаты, опираясь на содержание, поскольку знания семантически связаны между собой.

- **Интеграция с другими языками программирования:** что расширяет возможности системы.

- **Гибкость и масштабируемость:** для моделирования сложных знаний и структур с описанием объектов, отношений между ними и их атрибутов, а также созданием иерархий и композиций.

1.2.2 Альтернативные подходы

Для построения приложения электронного гида по Республике Беларусь могут использоваться различные технологии, каждая из которых имеет свои достоинства и недостатки. Вот несколько примеров:

Базы данных (DBMS)

Достоинства:

- **Структурированность:** системы управления базами данных позволяют хранить и организовывать информацию в структурированном виде.

- **Безопасность:** могут использовать механизмы шифрования и аутентификации для защиты данных.

- **Производительность:** оптимизированные запросы и индексы могут значительно ускорить доступ к информации.

Недостатки:

- **Сложность в обслуживании:** требуется наличие специалистов для управления и обслуживания баз данных.

- **Ограниченность в масштабировании:** возможны проблемы с производительностью при большом объеме данных.

Системы управления контентом (CMS)

Достоинства:

- **Удобство использования:** большинство CMS предлагают интуитивно понятные интерфейсы для управления контентом.

- **Гибкость:** позволяют легко добавлять, изменять и удалять контент без необходимости изменения кода.

Недостатки:

- **Безопасность:** некоторые CMS могут быть уязвимы для атак, если не обновляются и не настраиваются должным образом.
- **Ограниченность функциональности:** не все CMS могут поддерживать сложные запросы или интеграцию с внешними системами.

Вики-платформы (Wiki)

Достоинства:

- **Открытость:** вики-платформы позволяют любому пользователю редактировать и добавлять информацию.
- **Коллаборативность:** позволяют сообществам совместно создавать и обновлять контент.

Недостатки:

- **Качество информации:** вики-платформы могут содержать неточную или непроверенную информацию.
- **Проблемы с достоверностью:** не все вики-платформы имеют строгие процедуры проверки информации.

1.3 Анализ аналогов

1.3.1 Общеознакомительная информация об аналогах

При попытке найти системы схожие по принципу с нашей, мы обнаружили несколько потенциальных аналогов, которые могут решить наши задачи и соответствовать нашим требованиям. Вот перечень некоторых из них:

- **Vetliva**[1]. Это популярный белорусский туристический портал, который предоставляет информацию о гостиницах, отелях, развлечениях, экскурсиях и других туристических услугах по всей Беларуси. Развитие Vetliva с уклоном в расширение информации и функционала может помочь решить проблему разрозненности информации. Ниже представлен перечень некоторых возможностей данного портала:

- *Информация о размещении:* Vetliva предоставляет обширную базу данных о гостиницах, отелях, апартаментах и других местах размещения по всей стране. Это полезно для туристов, которым нужно найти подходящее место для проживания во время поездки.

- *Описание достопримечательностей:* Портал содержит информацию о различных достопримечательностях, культурных объектах, природных заповедниках и других интересных местах в Беларуси. Это позволяет туристам лучше планировать свои маршруты и изучать местную культуру и историю.

- *Бронирование экскурсий и активных мероприятий:* Vetliva предлагает возможность бронирования экскурсий, туров и других активных

мероприятий по всей стране. Это помогает туристам организовать свое время и получить более насыщенный опыт путешествия.

- *Отзывы и рейтинги*: Пользователи могут оставлять отзывы о различных местах и услугах, что помогает другим туристам делать более информированный выбор.

- *Информация о событиях*: Vetliva предоставляет информацию о различных событиях, фестивалях, концертах и мероприятиях, проходящих по всей Беларуси.

- *By Guide*[2]. "By Guide это интернет-платформа, предназначенная для туристов, желающих узнать больше о различных аспектах жизни и деятельности в Беларуси. Вот несколько возможностей, которые он предоставляет, и их соответствие нашим потребностям и целям:

- *Информация о достопримечательностях*: By Guide предоставляет информацию о различных туристических объектах, культурных и исторических достопримечательностях, что может быть полезным для желающих познакомиться с культурным наследием и историей Беларуси.

- *События и развлечения*: Платформа предоставляет информацию о культурных событиях, выставках, концертах и других развлекательных мероприятиях, что может помочь туристам планировать свое время в Беларуси.

- *Туристические маршруты*: "By Guide" может предложить различные туристические маршруты и предложения, что удобно для путешественников, желающих исследовать различные регионы страны.

- *Информация о размещении и услугах*: Платформа может содержать информацию о гостиницах, ресторанах, транспортных услугах и других сервисах, необходимых для комфортного пребывания в Беларуси.

- *Рейтинги и отзывы*: "By Guide" может также содержать рейтинги и отзывы пользователей о различных заведениях и услугах, что поможет туристам делать более информированный выбор.

- *Go to Belarus*[3] - это путеводитель по различным городам и странам, включая Беларусь. Вот некоторые возможности, которые предоставляет данный сайт, и их соответствие вашим потребностям и целям:

- *Путеводитель по городам*: Сайт предоставляет информацию о различных городах Беларуси, их истории, культуре, архитектуре и достопримечательностях. Это может быть полезно для туристов, планирующих посетить различные города страны и узнать больше о их уникальных особенностях.

- *Статьи и рекомендации*. Go to Belarus содержит статьи, рекомендации и советы от путешественников, которые уже посетили Беларусь. Это может помочь вам найти интересные места для посещения, узнать о местных традициях и обычаях, а также получить практические советы для комфортного путешествия.

– *Фотографии и видеоматериалы*: Сайт содержит большое количество фотографий и видеоматериалов о различных местах в Беларуси. Это позволяет вам получить визуальное представление о местах, которые вы планируете посетить, и выбрать наиболее интересные для себя.

– *Интерактивные карты*: Go to Belarus предоставляет интерактивные карты с отмеченными на них местами и достопримечательностями. Это помогает вам лучше ориентироваться в городах и регионах Беларуси и планировать маршруты своих поездок.

– *Отзывы и комментарии*: Посетители сайта могут оставлять отзывы и комментарии о посещенных ими местах, что помогает другим туристам делать более информированный выбор.

1.3.2 Перечисление достоинств и недостатков аналогов

В данном разделе следуют отметить индивидуальные достоинства и недостатки каждой из платформ, чтобы выделить общие проблемы, нуждающиеся в решении при разработке нашей системы.

	Достоинства	Недостатки
Vetliva	Широкий спектр туристических услуг; Онлайн-бронирование размещения; Отзывы и рейтинги для оценки качества услуг; Интерактивные карты.	Ограниченность выбора в некоторых регионах; Недостаточная актуальность информации; Недостаток международной ориентации.
By Guide	Подробная информация о туристических объектах; Визуальные материалы для визуализации мест; Практические советы и рекомендации; Возможность взаимодействия с пользователями.	Ограниченность выбора в некоторых регионах; Недостаточная актуальность информации.
Go to Belarus	Обширная информация о туристических объектах Беларуси; Визуальные материалы для визуализации мест; Практические советы для путешественников; Интерактивные карты.	Некоторые регионы могут быть недостаточно представлены; Недостаточная актуальность информации в некоторых случаях.

Исходя из данных таблицы можно выделить следующие общие достоинства и недостатки вышеупомянутых аналогов:

Достоинства:

- *Информационное покрытие*: Все три платформы предоставляют обширную информацию о туристических объектах, местах размещения, событиях и других достопримечательностях в Беларуси.

- *Визуальные материалы*: На платформах часто представлены фотографии и видеоматериалы, что позволяет пользователям визуально оценить интересующие их места.

- *Практические советы*: Возможность получить практические советы и рекомендации помогает путешественникам сделать осознанный выбор при планировании поездки.

- *Интерактивные карты*: Наличие интерактивных карт помогает путешественникам лучше ориентироваться в местности и планировать свой маршрут.

- *Отзывы и рейтинги*: Возможность оставлять отзывы и читать рейтинги помогает путешественникам делать более информированные решения при выборе места для посещения или размещения.

Недостатки:

- *Ограниченность информации*: Некоторые регионы или места могут быть менее подробно описаны или не представлены на платформе, что может затруднить получение полной картины о Беларуси.

- *Актуальность информации*: Информация на платформах может быть не всегда актуальной или полной, что создает риск недостоверной информации для пользователей.

- *Недостаток международной ориентации*: Платформы могут быть менее ориентированы на иностранных туристов, что может затруднить их использование для международных посетителей.

1.3.3 Вывод по анализу аналогов

Вышеупомянутые туристические платформы - Go to Belarus, By Guide и Vetliva - представляют собой ценные ресурсы для получения информации о туристических возможностях в Беларуси. Однако, при анализе их особенностей становится ясно, что они не полностью соответствуют всем требованиям, предъявляемым к электронному гиду по Беларуси.

Несмотря на обширную информацию и удобство использования, все три платформы имеют ограничения в покрытии регионов и актуальности представленной информации. Некоторые регионы могут быть недостаточно представлены на платформах, что создает риск упущения важной информации для путешественников. Также отмечается недостаток международной ориентации у этих платформ, что может затруднить их использование для иностранных посетителей.

Более того, отсутствие полного взаимодействия пользователей и необ-

ходимости дополнительной проверки актуальности информации делает эти платформы менее удобными для использования в качестве базы знаний для электронного гида.

Таким образом, при разработке электронного гида по РБ необходимо учитывать недостатки вышеперечисленных туристических платформ. Ограничения в покрытии регионов и актуальности представленной информации создают риск упущения важных сведений для путешественников. Недостаток международной ориентации затрудняет использование платформ иностранными посетителями. Также отсутствие полноценного взаимодействия пользователей и необходимости дополнительной проверки актуальности информации снижает удобство использования этих платформ. Для успешного создания электронного гида необходимо расширить региональный охват, обеспечить регулярное обновление информации, ориентироваться на международную аудиторию, повысить интерактивность и пользовательский опыт, а также обеспечить качественную поддержку пользователей. Следовательно, решение данных недостатков является требованием для создания системы.

1.4 Вывод

При анализе подходов к разработке было установлено, что использование технологии OSTIS обеспечит универсальность, семантическое представление и гибкость системы. Это позволит интегрировать различные типы данных, масштабировать базу знаний и осуществлять совместимость с другими платформами. Альтернативные подходы, такие как базы данных, системы управления контентом и вики-платформы, показали свои достоинства и недостатки, однако ни один из них не обеспечивает необходимую семантическую гибкость, что подчеркивает важность выбора OSTIS в качестве основной технологии.

Анализ аналогичных систем, таких как Vetliva, BY Guide и Go to Belarus, выявил недостатки, связанные с узкой специализацией, недостаточной информацией о белорусских достопримечательностях и отсутствием продвинутых систем рекомендаций. Несмотря на их сильные стороны, такие как мультимедийный контент и удобные интерфейсы, они не полностью соответствуют потребностям пользователей, интересующихся туристическими объектами в Республике Беларусь.

Таким образом, на основе проведенного анализа были сформулированы следующие общие требования к системе:

- 1 Гибкость и универсальность для адаптации к разнообразным типам данных и пользовательским запросам.

- 2 Семантическое представление информации для улучшения поиска и обработки данных.

- 3 Масштабируемость базы знаний с возможностью интеграции новых

данных и структур.

4 Разработка системы рекомендаций, основанной на предпочтениях пользователей и контексте запросов.

5 Обеспечение доступа к актуальной и глубокой информации о достопримечательностях всех регионов Республики Беларусь.

6 Удобный и интуитивно понятный интерфейс для пользователей с различным уровнем подготовки.

Так как моя курсовая работа ориентирована на генерацию scs-файлов для БД на основе полученных JSON-файлов, то выявим следующие требования к подсистеме сбора данных:

1 Подсистема должна быть универсальной и легко адаптируемой, чтобы эффективно работать с различными структурами входных данных и обеспечивать возможность добавления поддержки новых форматов без значительных изменений в её архитектуре.

2 Важным требованием является точность обработки данных, что включает корректное извлечение ключевых полей, таких как название, тип заведения и другие важные атрибуты. Это обеспечит соответствие между содержанием JSON файлов и итоговыми SCS файлами.

3 Сгенерированные SCS файлы должны быть в удобном и стандартизированном формате, что позволит их легко интегрировать в другие части системы, а также обеспечивать удобный поиск и обработку информации.

2 ПРОЕКТИРОВАНИЕ ЭЛЕКТРОННОГО ГИДА РБ

2.1 Постановка задачи

В рамках разработки системы электронного гида по Республике Беларусь необходимо решить ряд ключевых задач, связанных с обработкой и представлением знаний о достопримечательностях. Основной целью проекта является создание многоагентной системы, которая будет эффективно работать с онтологически структурированным графом знаний. Для достижения этой цели будут применены концепции графов знаний, онтологического подхода к структурированию данных, событийно-ориентированной архитектуры и многоагентного подхода к обработке информации.

Для реализации многоагентной системы электронного гида предполагается использовать разнообразные методы решения задач, включая машинное обучение (ML). Агенты должны обладать функциональностью генерации новых знаний, проведения вычислений и реализации полезного функционала для системы. Это позволит системе не только собирать и обрабатывать информацию, но и активно участвовать в анализе и интерпретации данных, обеспечивая более глубокое понимание контекста достопримечательностей и улучшая пользовательский опыт.

Использование методов машинного обучения в системе рекомендаций позволит автоматически адаптировать предоставляемую информацию к предпочтениям пользователей. Система сможет анализировать данные о действиях и предпочтениях пользователей, а затем генерировать персонализированные рекомендации на основе этих данных. Это повысит вероятность того, что пользователи найдут интересные для себя достопримечательности и маршруты, что, в свою очередь, улучшит их опыт взаимодействия с приложением.

Как описывалось в предыдущих разделах, одной из ключевых задач проекта является разработка инструмента для автоматизированного сбора информации о достопримечательностях. Задача данного курсового проекта заключается в создании веб-скрапера, который будет собирать данные из открытого источника и преобразовывать их в стандартный формат данных.

Для успешной реализации веб-скрапера, необходимо выполнить несколько ключевых задач, каждая из которых обеспечит корректную работу и эффективность системы. Эти задачи являются основой для построения функционального и надежного инструмента для автоматизированного сбора информации о достопримечательностях. В следующих пунктах рассмотрены основные этапы, которые предстоит выполнить для реализации веб-скрапера:

1 Определение источника данных На этом этапе необходимо

выбрать и проанализировать доступные открытые источники данных, из которых будет собираться информация о достопримечательностях. Это может быть как один источник, так и несколько. Источник должен быть доступен для скрапинга (например, веб-страницы, API), а также содержать актуальную и структурированную информацию. Важно также учитывать ограничения, установленные для работы с этими данными, такие как частота запросов, ограничения на объем данных и юридические аспекты.

2 Определение формата выходного представления данных

После того как будет выбран источник, необходимо решить, в каком формате будет представлена собранная информация. Для эффективного хранения и последующей обработки данных предпочтительным вариантом является использование структурированного формата. Формат должен включать необходимые поля, такие как название достопримечательности, адрес, ссылка на соответствующую достопримечательность в источнике, описание и другие элементы, которые будут полезны для базы знаний. Также важно определить правила для обработки отсутствующих данных, например, что делать, если какой-то параметр не был найден на странице.

3 Определение инструментария Для реализации веб-скрапера нужно выбрать подходящие инструменты и библиотеки. Это включает в себя выбор языка программирования, а также фреймворков и библиотек, которые обеспечат удобную работу с HTML-страницами, в том числе которые используют динамический контент с помощью средств JavaScript.

4 Определение архитектуры системы (основных классов)

Для разработки веб-скрапера нужно спроектировать архитектуру системы, которая будет модульной и масштабируемой.

Таким образом, в проекте будет реализован многоуровневый подход к обработке знаний, который позволит создать гибкую и масштабируемую систему, способную адаптироваться к изменениям и новым требованиям пользователей.

2.2 Описание пользователя Электронного гида РБ

Электронный гид РБ может быть полезен для различных типов пользователей. Ниже перечислены способы использования системы для различного рода пользователей:

Туристы могут использовать электронный гида РБ для поиска актуальной информации о туристических объектах, достопримечательностях, музеях, ресторанах и других местах интереса в Беларуси. Это помогает туристам планировать свои поездки и маршруты.

Специалисты, работающие в сфере туризма и гостеприимства, могут использовать Электронный гид для получения актуальной информации о достопримечательностях, отелях, ресторанах и других объектах туристиче-

ской инфраструктуры.

Представители бизнеса, особенно малого и среднего, могут использовать Электронный гид для размещения информации о своих товарах и услугах, привлечения клиентов и партнеров

2.3 Проектирование парсера

Перечислим основные классы, описывающие функционал системы. Связь между классами наглядно отображена на диаграмме классов на рисунке 2.1.



Рисунок 2.1 – Диаграмма классов

Класс **FileHandler** предназначен для работы с файлами в рамках парсера, включая операции по чтению входных данных из файлов и записи результатов в выходные файлы. Этот класс инкапсулирует все взаимодействия с файловой системой, что упрощает модульность и поддержку кода

Входящим параметром является путь к директории, содержащей файлы, которые необходимо обработать, а на выходе получаем данные, считанные из файлов в виде строк или объектов Python.

Класс **ScsGenerator** отвечает за преобразование структурированных данных в строки формата SCS, которые используются для представления знаний в базах данных на основе концепции семантических сетей. Этот класс инкапсулирует логику генерации корректных SCS-представлений из входных данных.

Входящим параметром являются структуры данных Python, содержащие информацию о месте или объекте, а на выходе получаем готовый scs-файл.

Класс **Parser** предназначен для обработки входных данных из файлов формата JSON и их преобразования в файлы формата SCS. Он служит связующим звеном между чтением исходных данных и их конвертацией в структурированный формат, понятный системе.

Класс **Main** представляет собой точку входа для запуска всего процесса парсинга JSON-файлов и генерации SCS-файлов с помощью класса Parser. Этот класс управляет выполнением всех необходимых операций по обработке данных и обеспечивает взаимодействие между различными компонентами модуля.

Алгоритм работы парсера следующий:

1. Инициализация парсера.
2. Проверка выходной директории.
3. Чтение файлов из входной директории.
4. Обработка каждого элемента.
5. Генерация scs-файла.
6. Повторение процесса пока не обработаются все JSON-файлы.
7. Завершение работы.

2.4 Вывод

В рамках проектирования системы электронного гида по Республике Беларусь была проделана значительная работа по разработке архитектуры и функционала, который выполнит задачи сбора, обработки и представления данных о достопримечательностях. Основное внимание было уделено проектированию парсера, который будет обрабатывать данные из заданных файлов и структурировать их в необходимом формате, а также формированию системы, способной адаптироваться к изменениям структуры входных данных и новым требованиям пользователей.

В ходе проектирования были выполнены ключевые задачи проектирования:

Поставлена задача требующая решения. Описаны основные пользователи системы. Был спроектирован парсер позволяющий формировать scs-файлы на основе полученных JSON-файлов.

3 РАЗРАБОТКА ЭЛЕКТРОННОГО ГИДА

В рамках данного раздела будут рассмотрены выбранные средства разработки для реализации всей системы и модуля веб-скрапинга информации.

3.1 Используемые средства разработки

3.1.1 Git и GitHub

Git[4] — это бесплатная распределенная система контроля версий с открытым исходным кодом, предназначенная для быстрой и эффективной обработки любых проектов, от небольших до очень крупных. Также она помогает отслеживать изменения, вносимые в файлы со временем, облегчает командную работу, позволяет создавать ветвления и объединять код, а также предоставляет историю изменений для легкого отката при необходимости.

GitHub[5] же является сервисом для совместной разработки и хостинга проектов, использующий Git. С помощью GitHub над кодом проекта может работать неограниченное количество программистов из любых точек мира.

3.1.2 Visual Studio Code

Visual Studio Code (VS Code)[6] — текстовый редактор, разработанный Microsoft для Windows, Linux и macOS. Включает в себя отладчик, инструменты для работы с Git, подсветку синтаксиса, IntelliSense и средства для рефакторинга. Имеет широкие возможности для кастомизации: пользовательские темы, сочетания клавиш и файлы конфигурации. Распространяется бесплатно, разрабатывается как программное обеспечение с открытым исходным кодом, но готовые сборки распространяются под проприетарной лицензией.

3.1.3 WINK-vscode-extention

WINK-vscode-extension[7] - расширение для текстового редактора Visual Studio Code, реализующее поддержку синтаксиса языка SCs. Включает проверку и подсветку синтаксиса, авто-дополнение, загрузку scs-файла прямо в базу знаний работающего sc-сервера и др.

3.1.4 PyCharm IDE

PyCharm IDE[8] представляет собой мощную интегрированную среду разработки (IDE), специально предназначенную для разработки на

Python. PyCharm обеспечивает разработчиков множеством инструментов для удобного написания, тестирования и отладки кода, включая поддержку работы с большими проектами, управление зависимостями и интеграцию с системами контроля версий.

3.1.5 Инструменты для парсера

В процессе реализации парсера использовались следующие основные инструменты:

1. OS[9]

- **Описание:** Библиотека `os` в Python является стандартной библиотекой, предназначенной для работы с операционной системой. Она предоставляет функциональность для взаимодействия с файловой системой, выполнения системных команд, работы с процессами и многими другими задачами, связанными с операционной системой.

- **Преимущества:**

- Библиотека `os` обеспечивает совместимость между операционными системами, скрывая различия в путях и системных вызовах. Это позволяет разработчикам писать код, который работает как на Windows, так и на Unix-подобных системах (Linux, macOS) без необходимости учитывать специфичные различия между ними.

- Библиотека `os` предоставляет простой и удобный интерфейс для работы с файлами и директориями.

- С помощью функций, таких как `os.path.join()` и `os.path.abspath()`, легко управлять путями, получая абсолютные пути и автоматически адаптируя их под нужды разных операционных систем.

2. Json[10] (для обработки JSON)

- **Описание:** `Json` — это библиотека, которая считывает JSON из файла и десериализует его в объект Python (например, список или словарь).

- **Преимущества:**

- Простота использования для сериализации и десериализации данных.

- Быстрая работа библиотеки.

- Позволяет настраивать поведение через дополнительные параметры.

3. Язык программирования Python[11].

- **Преимущества использования Java для веб-скраппинга:** Java является основным языком программирования для создания этого веб-скраппера. Он предоставляет мощные библиотеки для работы с сетью и веб-страницами, а также обеспечивает безопасность и стабильность работы, что является важным при автоматизации процессов сбора данных.

- **Большое количество библиотек:** Python предоставляет множе-

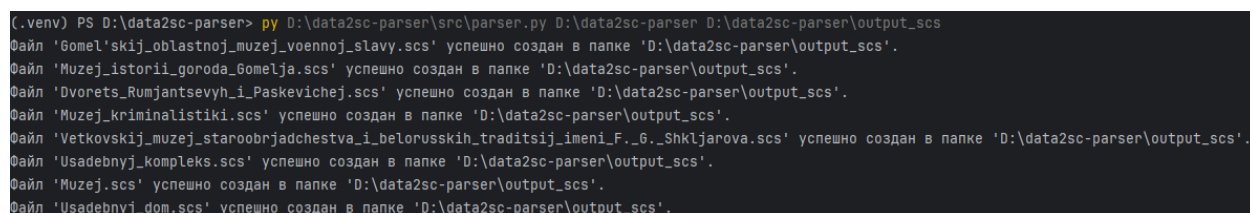
ство готовых библиотек и фреймворков для парсинга данных, включая json для обработки JSON-файлов и re для работы с регулярными выражениями.

– **Гибкость и масштабируемость:** С помощью Python можно легко адаптировать и расширять функциональность парсера, добавляя новые модули и функции без существенных изменений в архитектуре.

– **Кроссплатформенность:** Python работает на различных операционных системах (Windows, macOS, Linux), что делает его универсальным для разработки и развертывания парсеров.

3.2 Демонстрация работы системы

Демонстрация работоспособности программы будет проведена с помощью консоли. Для того чтобы запустить парсер, нужно запустить файл и передать ему аргументы - входная директория содержащая JSON-файлы и выходная директория, в которую будут помещены сгенерированные scs-файлы. По окончании исполнения программы, в консоли печатаются логи, содержащие отладочную информацию о том, какие файлы были созданы. Рисунок 3.1



```
(.venv) PS D:\data2sc-parser> py D:\data2sc-parser\src\parser.py D:\data2sc-parser D:\data2sc-parser\output_scs
Файл 'Gomel'skiy-oblastnoj-muzej-voennoj-slavy.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Muzej-istorii-goroda-Gomelja.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Dvorets-Rumjantsevvyh-i-Paskevichej.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Muzej-kriminalistiki.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Vetkovskij-muzej-starooobrazchestva-i-belorussskih-traditsij-imeni-F._G._Shkljarova.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Usadebnyj-kompleks.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Muzej.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
Файл 'Usadebnyj-dom.scs' успешно создан в папке 'D:\data2sc-parser\output_scs'.
```

Рисунок 3.1 – Логирование программы

На выходе получили множество scs-файлов для каждой достопримечательности. Рисунок 3.2

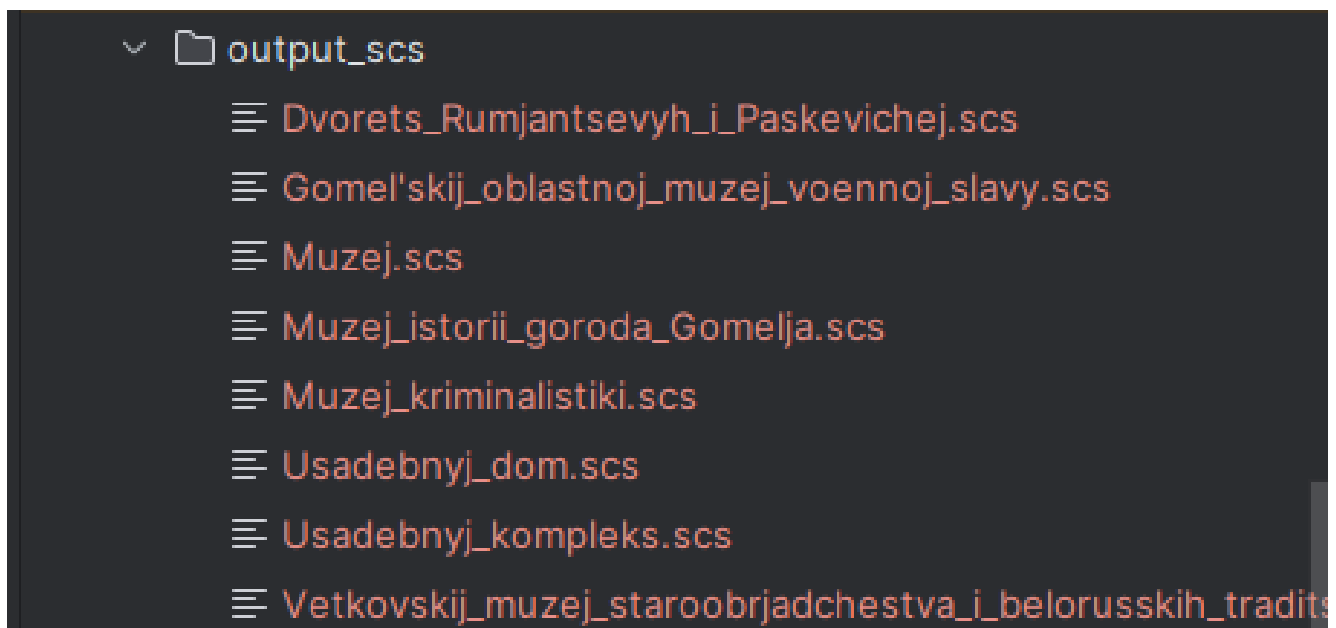


Рисунок 3.2 – Файлы

Scs-Файл имеет следующий вид(рисунок 3.3)

```
Dvorets_Rumjantsevyh_i_Paskevichej

=> nrel_main_idtf:
    [Дворец Румянцевых и Паскевичей]
    (* <- lang_ru;; *);
    [Дворец Румянцевых и Паскевичей] (* <- lang_en;; *);

=> nrel_reference_link:
    [https://yandex.by/maps/org/palats_rumiantsevykh_i_paskevychiv/44974196830/] (* <- url;; *);

=> nrel_address:
    [Гомель, площадь Ленина, 4] (* <- lang_ru;; *);

<= nrel_inclusion:
    concept_palace;;
```

Рисунок 3.3 – Содержимое SCS-файла

3.3 Вывод

В результате разработки электронного гида для Республики Беларусь были использованы современные и эффективные средства разработки, которые обеспечили успешную реализацию системы и модуля парсинга. В процессе разработки использовались инструменты для автоматизации работы с данными, такие как Json и OS, что позволило эффективно собирать и обрабатывать информацию.

Демонстрация работы системы парсера показала её успешную реализацию: парсер корректно обработал данные о достопримечательностях из

предоставленных JSON-файлов, преобразовав их в формат SCS и сохранив их в указанной директории.

Таким образом, разработанная система полностью соответствует поставленной задаче автоматизированной обработки данных о туристических объектах, эффективно извлекая, анализируя и преобразуя информацию о достопримечательностях для дальнейшего использования в базе данных, что делает её идеальной для применения в контексте сбора и обработки информации о туристических местах в Республике Беларусь.

ЗАКЛЮЧЕНИЕ

В рамках данного курсового проекта была проделана значительная работа по разработке модуля для интерактивного веб-приложения «Электронный гид по Республике Беларусь». Основной целью проекта стало создание инструмента для автоматического сбора, структурирования и представления информации о туристических достопримечательностях, что позволит пользователям легко находить и изучать интересные места по всей стране.

На начальном этапе был проведен детальный анализ существующих решений и подходов к разработке, а также рассмотрены аналоги, позволяющие оценить их функциональные возможности, достоинства и недостатки. На основе проведенного анализа были сформулированы основные требования к системе. Одним из ключевых компонентов стал парсер, который должен обеспечивать пополнение БД с помощью полученных JSON-файлов.

Проектирование системы включало определение архитектуры парсера, выбор технологий и методов, которые оптимально соответствуют поставленным задачам, а также описание форматов входных и выходных данных.

В результате была разработана система парсинга, которая эффективно извлекает и обрабатывает данные из JSON-файлов, фильтрует и структурирует информацию, а затем сохраняет её в формате SCS. Этот подход значительно упрощает дальнейшую обработку данных и их интеграцию с другими компонентами системы. Созданная система продемонстрировала свою надежность и стабильность, обеспечивая корректную работу даже при изменениях в структуре исходных данных и требований к форматированию.

Таким образом, данный проект закладывает основу для создания полноценного электронного гида, способного автоматизировано обновлять и пополнять базу данных туристических объектов, предоставляя пользователям актуальную и структурированную информацию о достопримечательностях Республики Беларусь.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] By Guide Маршруты по Беларуси [Электронный ресурс]. — Режим доступа. <https://www.by-guide.by/>.
- [2] Go to Belarus. — Mode of access: <https://34travel.me/gotobelarus>. — Date of access: 05.03.2024.
- [3] Vetliva. — Mode of access: <https://vetliva.by/>. — Date of access: 05.03.2024.
- [4] Git official cite [Электронный ресурс]. — Режим доступа. <https://git-scm.com/>.
- [5] GitHub [Электронный ресурс]. — Режим доступа. <https://github.com/>.
- [6] Visual Studio Code official cite [Электронный ресурс]. — Режим доступа. <https://code.visualstudio.com/>.
- [7] WINK-vscode-extention. — Режим доступа. <https://marketplace.visualstudio.com/items?itemName=artrayme.wink-vscode-extension&ssr=false#overview>.
- [8] PyCharm. — Режим доступа. <https://www.jetbrains.com/pycharm/>.
- [9] OS. — Режим доступа. <https://docs.python.org/3/library/os.html>.
- [10] Json. — Режим доступа. <https://docs.python.org/3/library/json.html>.
- [11] Python. — Режим доступа. <https://docs.python.org/3/index.html>.
- [12] OSTIS Applications. — Mode of access: <https://git.codeproxy.net/ostis-apps> . — Date of access: 28.04.2024.
- [13] OSTIS-example-app. — Mode of access: <https://github.com/ostis-apps/ostis-example-app> . — Date of access: 28.04.2024.
- [14] SCS language. — Mode of access: <https://ostis-dev.github.io/sc-machine/other/scs/> . — Date of access: 28.04.2024.
- [15] Tripadvisor [Электронный ресурс]. — Режим доступа. <https://www.tripadvisor.com/Tourism-g294448-Minsk-Vacations.html>.
- [16] What is a database management system? — Режим доступа. <https://marketplace.visualstudio.com/items?itemName=artrayme.wink-vscode-extension&ssr=false#overview>.

[17] SC-machine [Электронный ресурс]. — Режим доступа. <https://github.com/ostis-ai/sc-machine.git>.

[18] OSTIS [Электронный ресурс]. — Режим доступа. <https://ims.ostis.net>.