

HOTEL AI

Ассистент для создания описания отеля для сайта
travel.mts.ru

Состав проектной команды



Бубнов Кирилл

Тимлид

Python разработчик,
учится в Яндекс Лицее.
Участник летней
программы Сириуса



Громова Юлия

Разработчик, Аналитик
Python разработчик,
специалист анализа
данных.
Финалист олимпиады НТО



Бузмаков Климентий

Разработчик

Python разработчик,
физик-математик.
учится в Оц Орион



Дроботов Владислав

Разработчик

Node.js, Python
разработчик.
Специалист разработки
моделей

Проблема

Отельеру при публикации своего отеля в системе необходимо заполнить его описание. От того, насколько оно будет удачно составлено, зависит привлекательность предложения. При описании важно учитывать такие характеристики как: расположение отеля, сервисы и услуги, предоставляемые посетителям, или, например, уровень находящихся вблизи заведений. А главная цель хорошего описания помочь пользователю, читая, сразу выделять главное именно для него.

Текущее описание на mtc travel

Четырехзвездочный отель входит в состав группы компаний CITYTEL Group. Расположен в историческом центре северной столицы, на берегу реки Невы. Из большинства номеров и общественных зон открывается уникальный панорамный вид на акваторию реки, самые красивые петербургские мосты и главные достопримечательности города: Крейсер Аврора, Петропавловская крепость, Эрмитаж, знаменитые мосты. Номерной фонд состоит из 554 современных номеров, оформленных и оснащенных в соответствии с европейскими стандартами гостеприимства. Все номера удобны как для отдыха, так и для работы. В каждом из них есть ванная комната, телефон, телевизор, холодильник. На крыше отеля расположен видовой ресторан премиального уровня La Vue, предлагающий гостям современную гастрономическую кухню, оригинальную винную карту и завораживающий вид на Неву и город. Также в отеле находится панорамный ресторан «Беринг» с европейской кухней. Выпить кофе и отдохнуть можно в лобби-баре Meeting Point. В отеле расположен современный деловой центр Конгресс-холл, располагающий всем необходимым для проведения конференций, выставок, презентаций любого масштаба. Гости могут посетить SPA-центр Oceanclub Fitness & Spa, бассейн и воспользоваться тренажерным залом. На территории отеля работает салон красоты. На парковке перед отелем установлена быстрая зарядная станция для электромобилей ZEVS мощностью 50 кВт.

Анализ ситуации

Сервис

Определённая
структура

Наличие лишней
информации

Читабельность

<https://travel.mts.ru>

Отсутствует

Присутствует

Плохая

<https://ostrovok.ru>

Присутствует

Присутствует

Хорошая

<https://travel.yandex.ru>

Присутствует

Отсутствует

Нормальная

Анализ ситуации

Сервис	Определённая структура	Наличие лишней информации	Читабельность
https://www.onetwotrip.com	Присутствует	Присутствует	Нормальная
https://www.booking.com	Присутствует	Присутствует	Нормальная
https://ru.skyscanner.com	Присутствует	Отсутствует	Нормальная

Цель

Разработать ассистента-помощника, который будет на основе собранных
данных составлять описание отеля для отельера.

Задачи

- 
- Провести анализ существующих решений
 - Составить прототип описания
 - Собрать данные и провести анализ для LLM
 - Реализовать работу Mistral 7B LoRA и на её основе собрать данные для обучения
 - Реализовать обучение нейросети Bert

Задачи

Провести
тестирование и
оценить
эффективность
обучения

Запустить
обученную
нейросеть для
получение уже на
основе данных с
сайта
<https://travel.mts.ru/>
и сторонних
источников готового
описания

Собрать отзывы
у отельеров

Исправить
ошибки,
переобучить
модель

Предлагаемое решение

Система, собирающая данные с сайта mtc travel и сторонних сервисов для формирования и структурирования описания с помощью нейросети.

Прототип итогового описания

Расположение

Отель расположен в центре Петербурга. Гостям открывается вид на Неву, самые красивые петербургские мосты и главные достопримечательности города: Крейсер Аврора, Петропавловская крепость, Эрмитаж.

Удобства

Все номера удобны как для отдыха, так и для работы. В каждом есть ванная комната, телефон, телевизор, ХОЛОДИЛЬНИК.

Услуги

На крыше отеля расположен видовой ресторан премиального уровня La Vue и панорамный ресторан «Беринг» с европейской кухней, а выпить кофе можно в лобби-баре Meeting Point. На территории отеля гости могут посетить: салон красоты и SPA-центр, бассейн и тренажерный зал. Также воспользоваться Конгресс-холлом для проведения конференций, выставок, презентаций любого масштаба.

Прочее

На парковке перед отелем установлена зарядная станция.

Отзывы клиентов

Отель с лучшим видом на Неву. Номера просторные и чистые. Персонал вежливый, учитывает пожелание при заселении. Питание разнообразное.

Сбор и обработка информации

Для формирования описания будет собираться следующая информация:

- Текущее описание
- Удобства в номере
- Услуги отеля
- Расположение и объекты рядом с отелем
- Отзывы клиентов

Далее следует обработка с помощью нейросети. После объединения пунктов у нас должен будет получиться текст в котором указана только нужная информация об отеле.

Структурирование текста

Описание имеет определённую структуру, в виде разделения на абзацы с кратким названием.

План разделения текста:

- Расположение
- Удобства номера
- Услуги отеля
- Для вас на территории работает
- Прочие преимущества
- Отзывы клиентов

Выбор нейросети

Изначально мы планировали использовать нейросеть BERT для написания и разделения текста на абзацы. Но сейчас у нас не имеется необходимого датасета для её обучения. Поэтому было принято решение пока что использовать Большую языковую модель (LMM), а точнее отправлять запросы на Mistral 7B LoRA.

Позже, мы сможем сформировать датасет из результатов работы модели Mistral 7B LoRA и обучить на нём BERT.

Сбор данных

Такие данные как услуги отеля, текущий отзыв и удобства номера можно взять с сайта.

Для того чтобы собрать данные о ближайших объектах необходимо найти адрес и с помощью ObjectManager извлечь ближайшие заведения и достопримечательности.

Данные о отзывах пользователей можно получить с помощью yandex-reviews-parser по id компании полученного из ссылки найденной по названию отеля.

Решение задания второго этапа Сириус.ИИ

Сбор данных с сайта “Ostrovok”

Исследовательский анализ данных

Обучение модели BERT

Парсинг

Мы парсили данные с сайта ostrovok.ru с помощью двух библиотек BeautifulSoup и selenium.

BeautifulSoup использован для извлечения информации с html страницы, когда как selenium для симуляции нажатия кнопки, так как сайт состоит из множества страниц.

Сначало парсер проходится по странице запроса и сохраняет ссылки на отели, затем он проходится уже по страницам отзывов отелей и сохраняет их в файл

Анализ данных

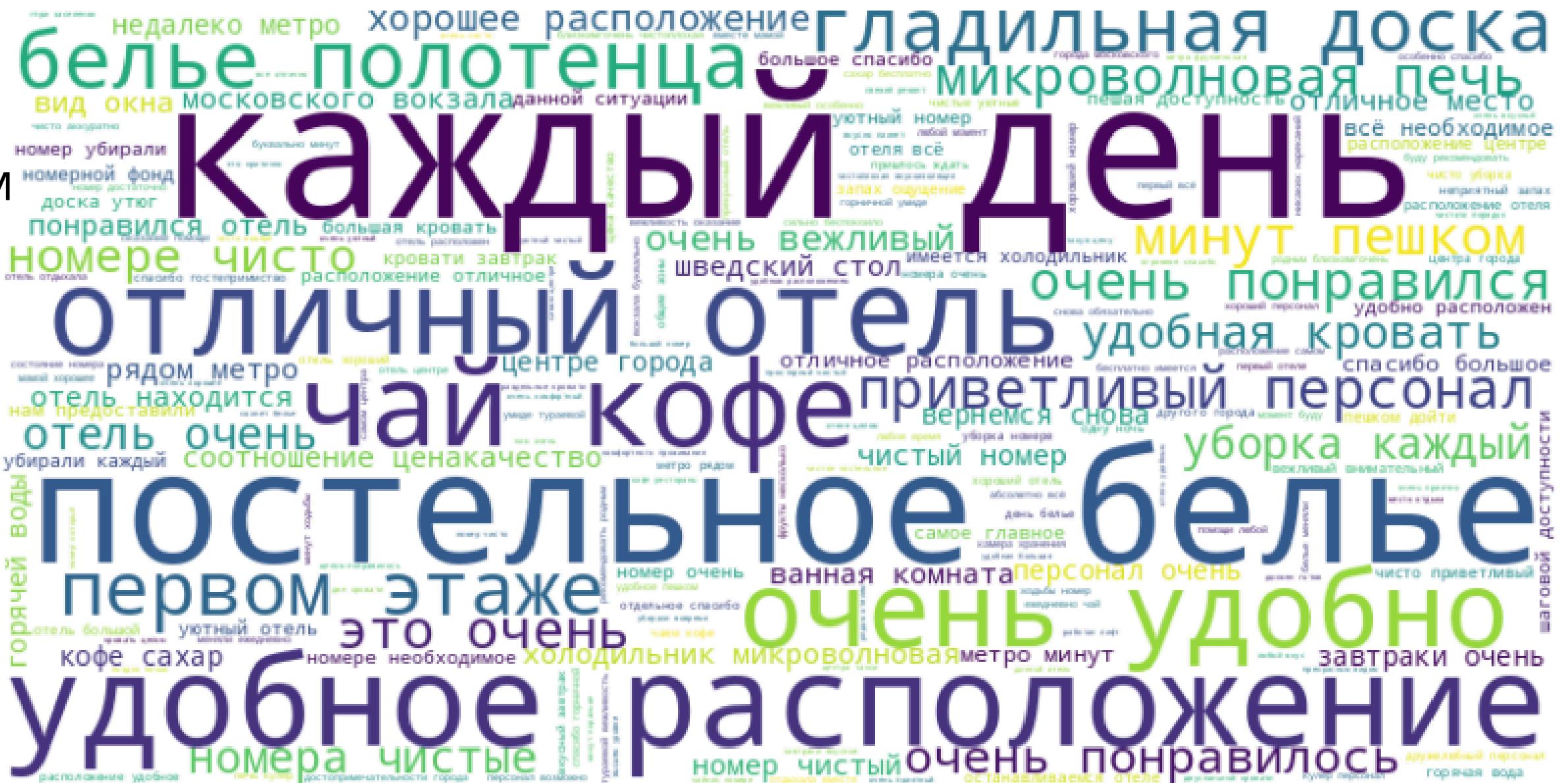
Мы провели анализ длины отзывов и выявили, что наиболее часто встречающиеся тексты имеют длину от 100 до 200 символов. Так же мы обнаружили что большую длину чаще имеют отзывы с критической оценкой, то есть с минимальной или максимальной



Анализ данных

Также мы визуализировали облако слов и сделали вывод что для пользователей очень важно:

- Расположение отеля
- Обслуживание номеров
- Наличие завтрака по утрам
- Удобства в номере



Обучение модели

На основе собранных данных мы обучили модель-классификатор, построенную на логистической регрессии.

Текст отзыва обрабатывается с помощью токенизера, а затем преобразуется в вектор. Позже идёт сопоставление с оценкой отзыва.

Проведённые эксперементы показали, что точность нашей модели составила 90.5%.

Перспектива

- Использование визуальной составляющей в виде значков для увеличения читабельности текста.
- Расширение функционала системы. Не только работа уже с готовыми описаниями с сайта mtc travel и сторонних источников, но и формирование описания, основываясь лишь на изображениях.
- На основе отзывов можно составить список того, что клиенты рекомендуют исправить или доработать в отеле. Это будет видно только для отельера и поможет отелю оставаться "в тренде". Реализовано это будет по подписке.

Спасибо за
внимание!