

академия
больших
данных



mail.ru
group

Персональный гид:

оптимальные маршруты по городу специально для вас

Автор идеи: Сулимов Павел

Ментор: Сулимов Павел

Команда:

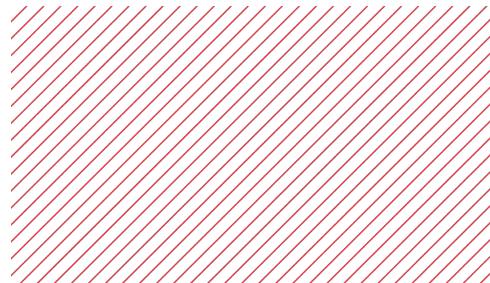
Холматов Темур DS-32

Устиц Юлия ML-31

Жданов Павел ML-32

Чернышенко Алексей ML-31

Балганбаев Сергей ML-32



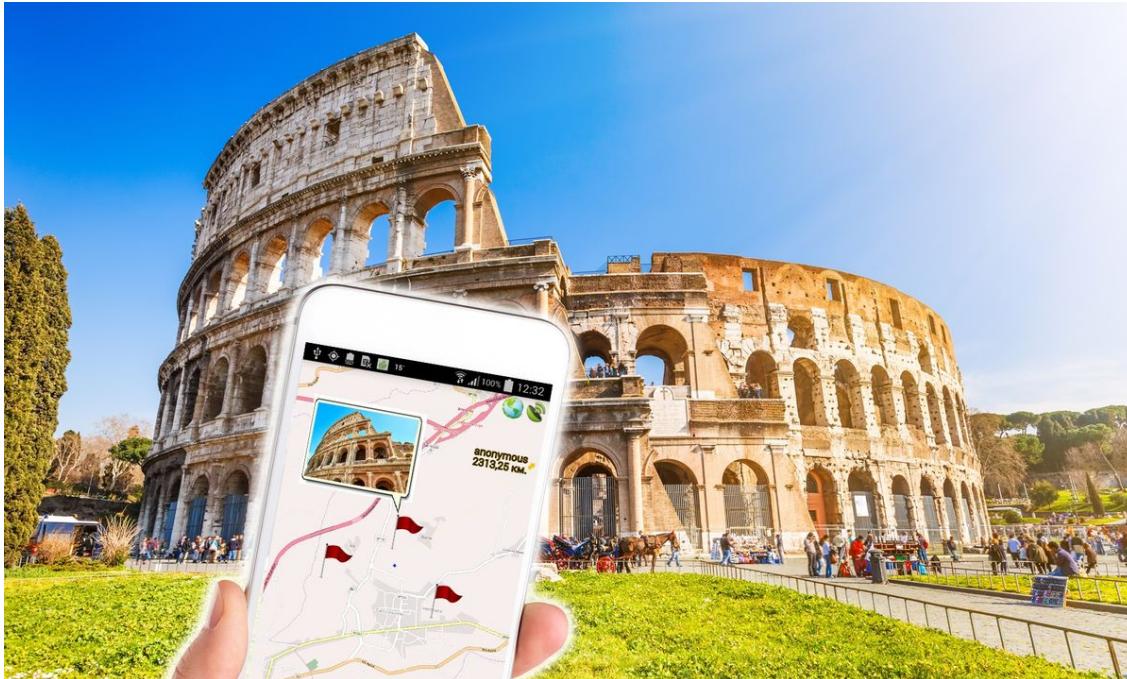
Описание проекта





Описание проекта

Пеший Туристический Маршрут специально для Тебя



- Персонально
- Учет времени
- Типы маршрутов
- Учет доп. требований

♀ Применение (B2C)

для путешественников



когда хочешь прогуляться, но нет
времени планировать

для командировочных

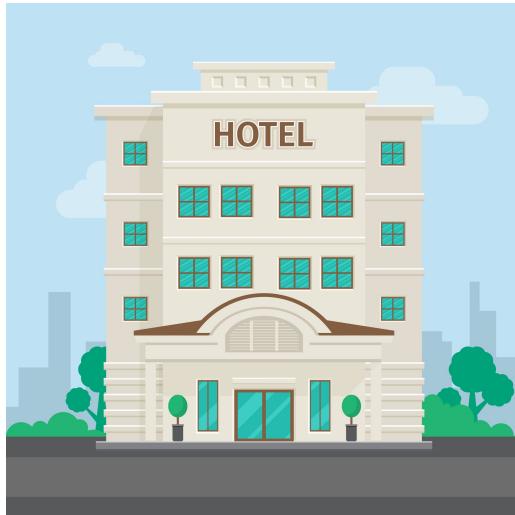


когда осталось время прогуляться в
чужом городе



Применение (B2B)

для турагентств и отелей



повышение лояльности клиентов

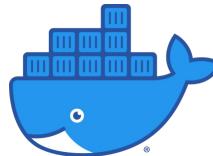
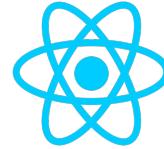
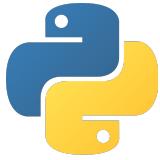
для авиалиний и транспортных компаний



оказание дополнительных услуг туристам

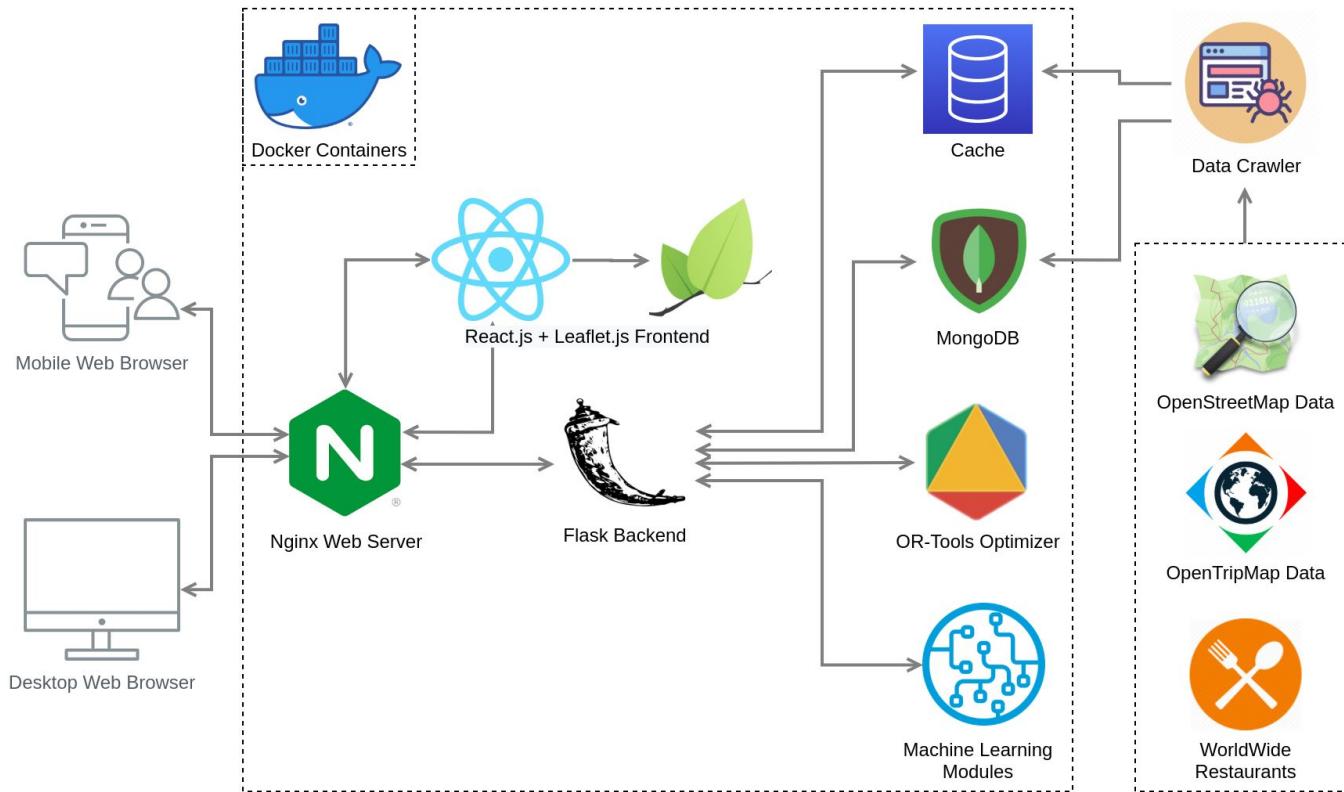


🔧 Технологии



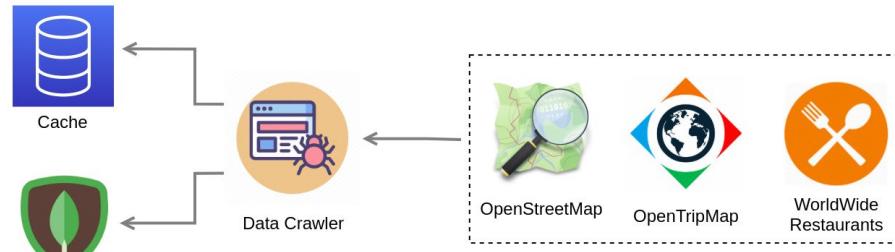
- Python
- Leaflet + React.js
- Flask
- Scikit-learn
- OR-Tools, NetworkX
- MongoDB
- Docker

⚙️ Архитектура



Cloud Данные

- Полученные данные:
 - Дороги, точки и полигоны городов (OpenStreetMap - ODbL)
 - Рейтинги, описание достопримечательностей (OpenTripMap - ODbL)
 - Рейтинги, описание, фото ресторанов/кафе/тп. (Worldwide Restaurants)
- Обработка данных:
 - Упрощение OSM данных, для построения графа
 - Интеграция OSM и WR данных
 - Приведение OTP данных к единому формату
- Сохраненные данные:
 - Cache:
 - Графы городов (Networkx)
 - Пути между соседними вершинами
 - MongoDB:
 - Все узлы графа с информацией о них
 - Точки интереса с кратчайшими расстояниями





↗ Оптимизационный алгоритм

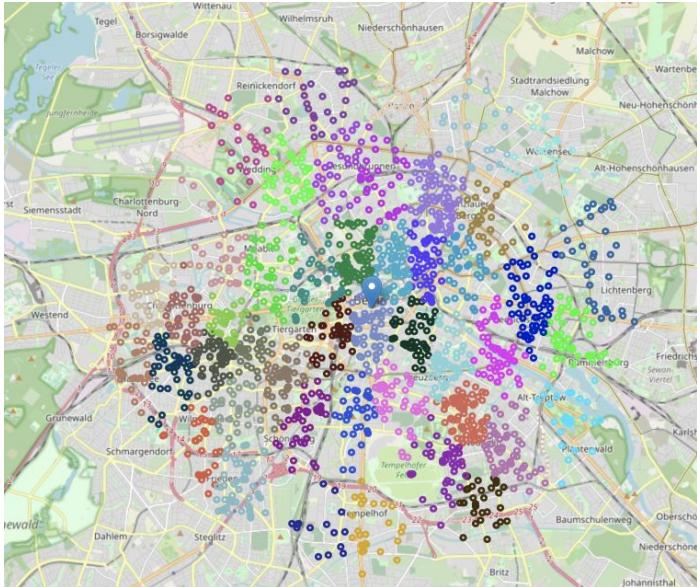
- NetworkX для оптимальных маршрутов между парами точек (+ смотрели OSRM)
- OR-Tools:
 - TSP Solver. Как учесть время, типы точек, рейтинг? RL?
 - Constraint Optimization → Reward Collecting TSP
 - награды - рейтинги достопримечательностей и ресторанов + closeness centrality значения (в частности, для аптек)
 - “Лучше” RL (помогает лишь косвенно)
 - *Vehicle Routing Problem with Time Windows* - для учета времени
- Multi-threads для маршрутов по кластерам



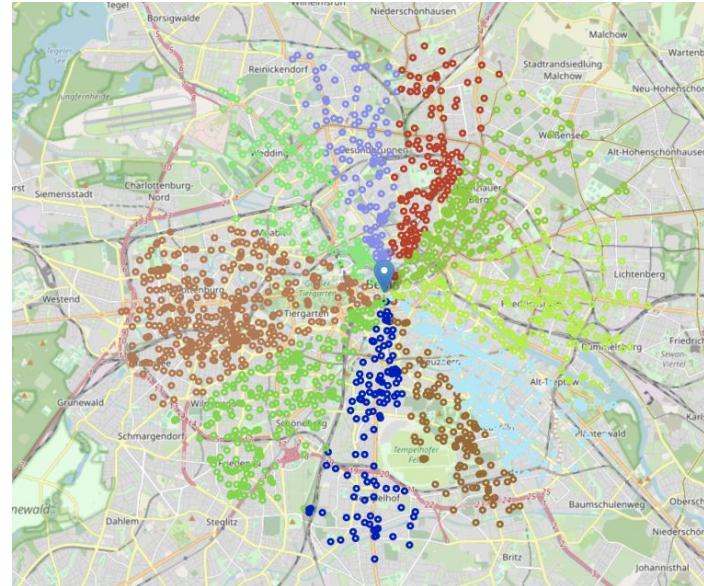
🌐 Кластеризация

Варианты применения	Полученный эффект
Кластеризация по локальности добавочных точек (ресторанов, аптек) и отбор лучших из кластеров	Ускорение работы оптимизационной модели без вреда качеству отобранных точек
Кластеризация всех точек по секторам (по разные стороны от стартовой точки)	Ускорение работы оптимизационной модели и генерация разных вариантами маршрутов

Кластеризация

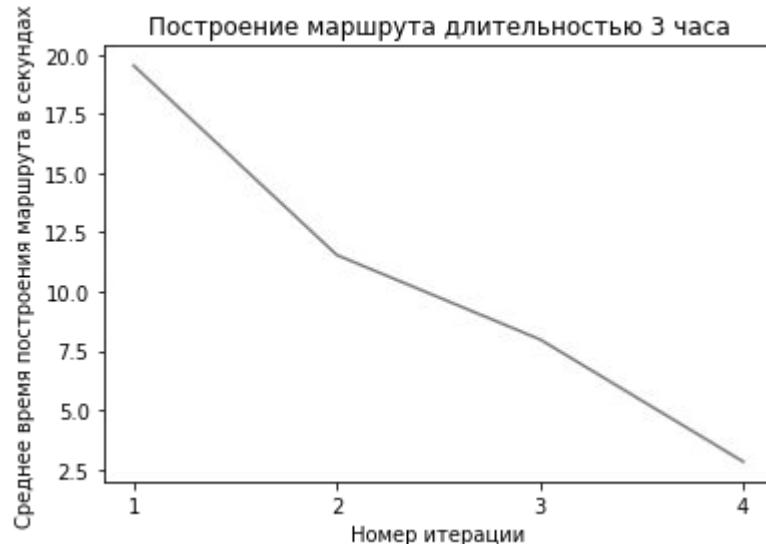


Обычная кластеризация KMeans++
(Вариант 1)



Кластеризация KMeans++ после
удаления точек от стартовой точки
(Вариант 2)

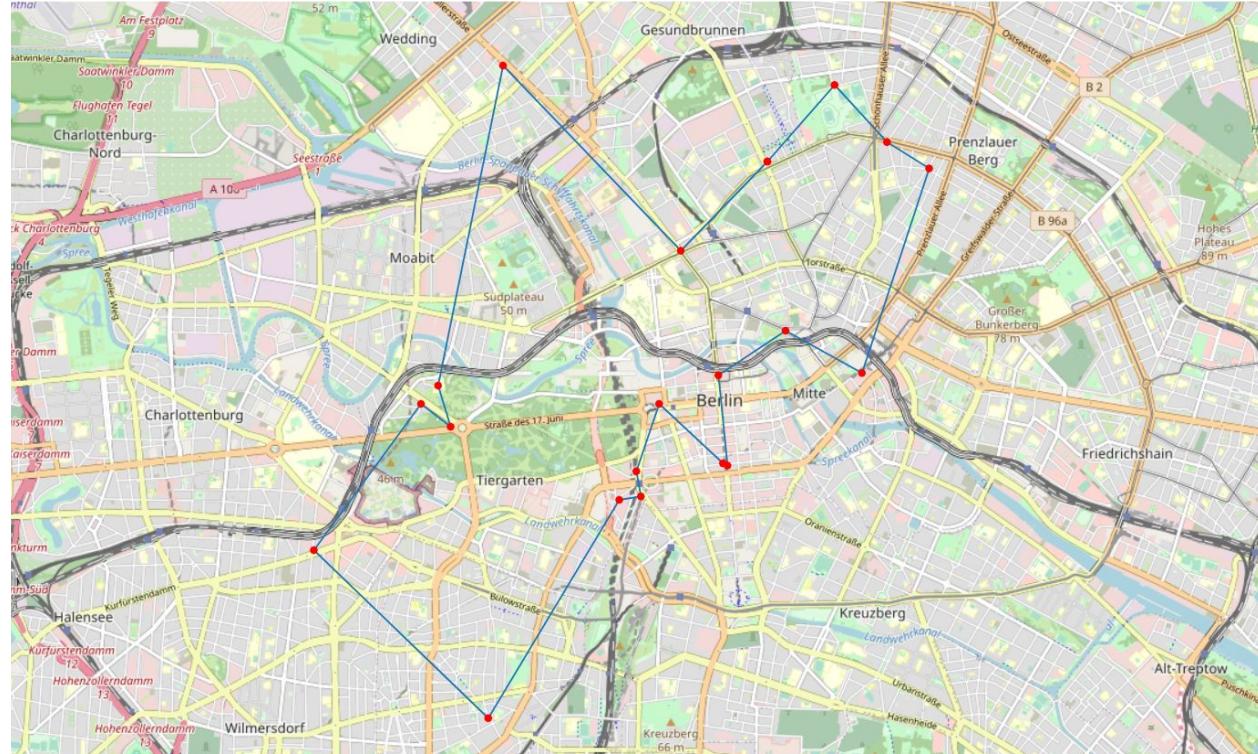
↗ Этапы развития скорости работы



Этапы улучшения скорости:

1. MongoDB + TSP
2. Предрасчет графа города + Constraint Optimization
3. Предрасчет матрицы расстояний + многопоточность и оптимизация параметров модели
4. Упрощения графа + оптимизация параметров кластеризации

↗ Этапы развития продукта



Отзыв пользователя



Аноним

В свой отпуск я решил поехать в Калининград, чтобы получше познакомится с уникальной архитектурой данного города. Однако, в связи со сложившейся ситуацией в мире и по соображениям безопасности, было решено отказаться от услуг гидов, и найти альтернативное решение проблемы. Благодаря приложению мне удалось обойти практически все значимые достопримечательности города, увидеть много живописных мест и получить кучу положительных эмоций. Спасибо ребятам за разработанное приложение!

Демонстрация проекта

262

263

Демонстрация проекта



Персональный гид



Коммерциализация

Freemium

Бесплатно

Длительность прогулки от 30 минут до 6 часов

30 запросов в месяц

4 типа маршрутов

12 дополнительных вариантов заведений

До 10 вариантов построенных маршрутов

Доступна информация по пунктам маршрута

Premium

Месячная/годовая подписка

Всё из Freemium подписки и больше ...

1000 запросов в месяц

Добавление своих пунктов посещения на карте

Указание временных ограничений для заведений

До 20 вариантов построенных маршрутов

Business

Свяжитесь с нами

Все возможности Premium подписки

Доступ к API сервиса

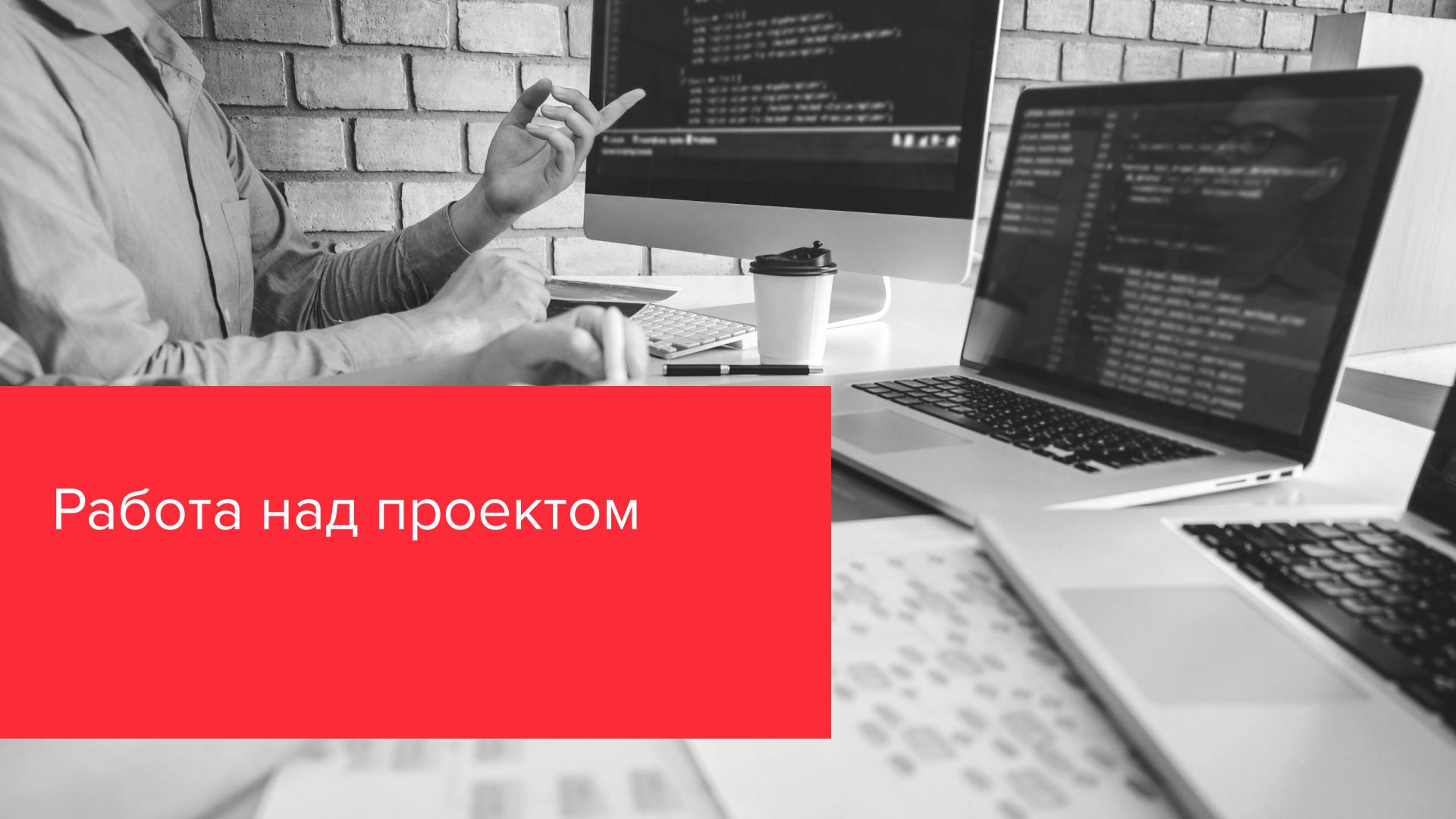
Интеграция карт с маршрутами

50000 запросов в месяц



Х Дальнейшее развитие продукта

- ✓ Учет последовательностей добавочных точек интереса (банкомат должен идти до магазина/ресторана)
- ✓ Добавление своих обязательных точек для посещения (музеи, театры)
- ✓ Регулировка пользователем времени посещения (2 часа на музеи)
- ✓ Обновить визуализацию порядка обхода точек
- ✓ Отображение дополнительных точек не с маршрута (дополнительные аптеки на пути)
- ✓ Разработка мобильного приложения (для определения местоположения, перестроения маршрута, вызова такси к текущей геопозиции)



Работа над проектом



Обновленный Roadmap

	20 окт. - 26 окт.	27 окт. - 02 ноя.	03 ноя. - 09 ноя.	10 ноя. - 16 ноя.	17 ноя. - 23 ноя.	23 ноя. - 30 ноя.	1 дек. - 13 дек.
DevOps							
Темур Павел			Схема пайплайна проекта и структура репозитория		Сбор проекта	Поднять приложение на сервере	
Optimizer							
Алексей	Найти алгоритмы обхода	Протестировать алгоритм	Обойти точки оптимально	Ограничить маршрут по времени	Построить маршрут без возврата Подобрать эвритики	Распараллелить расчеты Построить по кластерам	Добавить переход между кластерами
Темур						Добавить веса к poi	Учесть ограничения на доп. точки
Data							
Юлия	Подобрать критерии оптимизации	Собрать данные osm, рейтинги	Обработать данные Категоризовать тэги	Построить граф из mongo	Рассчитать матрицы расстояний	Построить граф из матрицы Ускорить расчет	Доработать данные Ускорить расчет
Сергей	Найти данные	Собрать данные Найти доп. информацию	Сопоставить найденную информацию				
Темур		Поднять MongoDB					
ML							
Павел	Выбрать ML часть				Кластеризовать poi		
Темур				Изучить применимость RL	Найти альтернативные решения RL		
FronEnd							
Сергей					Вывести маршрут на карту	Добаботать входные параметры Вывести несколько маршрутов	Добавить описание, фото Добавить доп. точки
Темур	Исследовать визуализацию гео-данных	Создать прототип приложения Поднять FrontEnd	Изучить OSRM-Front				
BackEnd							
Павел		Изучить Flask		Реализовать модуль для подключения к базе данных	Реализовать модуль для передачи результатов на Front	Реализовать модуль для кластеризации	Нагрузочное тестирование
Сергей			Изучить OSRM-Back	Изучить технологии описания маршрута			

Команда





Команда



Сулимов Павел

Ментор / Автор идеи



Холматов Темур

DevOps / ML

сборка инфраструктуры
проекта
анализ ML решений задачи



Чернышenko Алексей

Optimizer

разработка оптимизационной
задачи



Устиц Юлия

Data

выгрузка и обработка данных
предварительные расчеты



Жданов Павел

BackEnd / ML

интеграция Front и Back
подключение к MongoDB
реализация кластеризации



Балгамбаев Сергей

FrontEnd / Data

разработка интерфейса
вывод маршрутов на карту
поиск данных

Спасибо за внимание!

<https://walkcreator.ru>



Описание проекта

- + Описание проблемы/задачи (1 слайд с картинкой/схемой; глядя на него даже в полглаза должно быть понятно о чем и для кого проект) (**Персональный/пеший/туристический, Для тех кто ценит время и индивидуальность**)
 - + Польза бизнесу/пользователям от проекта и внедрения

Где применять? Если авторы сами не знают, жюри не догадается за них, дальше рассказывать бесполезно; поставьте себя на место клиента и место разработчика, и оцените себя/попросите друзей

Кто пользователи?

b2b, b2c - разделите варианты применения(**Тур. агентства и отели (повышение лояльности клиентов)/ Путешественник (когда хочется все посмотреть, но нет времени планировать) / Командировочные (когда нашлось время после работы в чужом городе) / Поезда/самолеты (дополнительные услуги)**)

a. Востребованность

Аналогичные решения на рынке, возможность применения бизнес-юнитами Mail.ru Group



Демонстрация проекта

1. Что на входе системы - подготовить примеры для запуска демонстрации жюри с минимальными усилиями (заранее возможность тыкнуть в пример и посмотреть разборчиво результат), уметь ответить на вопрос, можно ли добавить контроль параметров на входе и получить желаемое на выходе
2. Что на выходе проекта получается? (сервис, библиотека, приложение, предложение для интеграции с существующим сервисом)
3. Как оценивается качество технологий (метрика задачи, не лосс!!!) и качество результата проекта (не путать метрику/лосс с оценкой качества и востребованностью продукта!)
4. Примеры где хорошо и где плохо с обоснованием, интерпретация предсказаний модели (если применимо), подготовленное заранее видео работы приложения и вставить в презентацию

Описание проекта

Что вы используете для реализации проекта?

- + Перечень инструментов и технологий, систем и т.п. техническая составляющая проекта (мелко базовый стек и крупно - что-то специфичное для проекта)
- + Данные. Открытые - какая лицензия / свои - как собирали, примеры
- + Выбор моделей. Обоснование, метрики качества, особенности обучения (если применимо), оригинальность решения (новые модели/приемы) **Скорость/соответствие требованиям/Вариативность Максимизация награды (Награда - производная из рейтингов точек интереса и closeness-centrality)**
- + Оригинальность продукта - почему комбинация выбранных технологий может быть полезна, конкретный пример использования / демонстрации

Тур. агентства (повышение лояльности).

Рестораны/музеи (увеличение приоритезации внутри рейтингового диапазона)

Авиалинии, ж/д компании (снижение скидок на рейсы с длительными пересадками).

Коммерциализация: подписка (больше запросов, добавление своих точек с временными ограничениями на посещение (например, музеи), больше альтернативных маршрутов при запросе)

- + Возможность решения существующих недоработок проекта (с примерами) **учет последовательностей добавочных точек интереса (пример ресторан/туалет); добавление обязательных точек; регулировка времени посещения пользователем; более точно визуализировать порядок обхода точек интереса**



🔗 Применение (B2C)



Для путешественников, чтобы
экономить время на планирование



Для командировочных, когда есть
время прогуляться в чужом городе

♀ Применение (B2B)

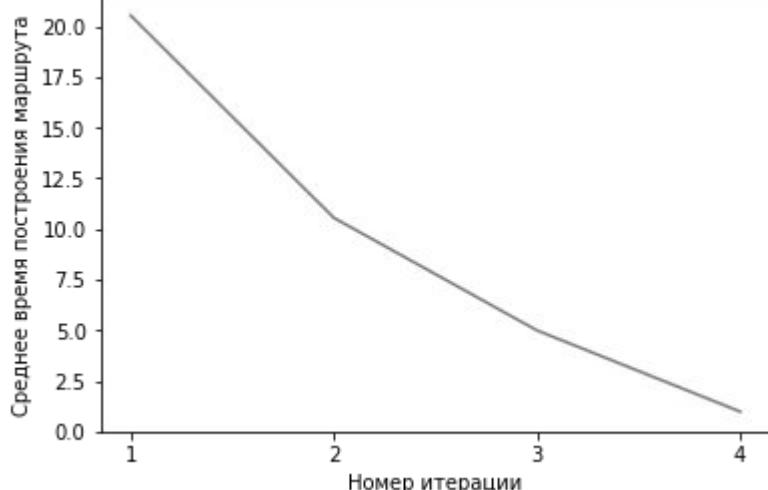


Для турагентств и отелей, чтобы
повысить лояльность клиентов



Для аэропортов и вокзалов, чтобы
оказывать дополнительные услуги
туристам

↗ Метрики качества



1. MongoDB + TSP
2. Предрасчет графа города + Constraint Optimization
3. Предрасчет матрицы расстояний + многопоточность, оптимизация параметров модели
4. Упрощения графа + оптимизация параметров кластеризации



Команда



Холматов Темур
DevOps / ML

создание пайплайна
сбор проекта
поднятие на сервере
изучение RL и аналогов

Устиц Юлия
Data

поиск и выгрузка данных
чистка информации
предварительные расчеты
построение графов



Жданов Павел
BackEnd / ML

организация общения
FrontEnd и BackEnd
реализация модуля
подключения к MongoDB
реализация кластеризации



Чернышенко Алексей
Optimizer

разработка
оптимизационной задачи

Балганбаев Сергей
FrontEnd / Data

разработка интерфейса
вывод маршрутов на карту
поиск данных



Спасибо!

Слайд спасибо с контактами участников

Пояснения к шаблону

В вашем распоряжении есть следующие слайды:

- Титульный слайд
- Слайд №1.1 Стандартный
- Слайд №1.2 Стандартный с подзаголовком
- Слайд №2.1 Разворот справа
- Слайд №2.2 Разворот слева
- Слайд №3.1 Ноутбук справа
- Слайд №3.2 Ноутбук слева
- Слайд №4.1 2 столбца
- Слайд №4.2 2 столбца
- Слайд №5 Контакты
- Слайд №6.1 Отбивка вертикальная
- Слайд №6.2 Пустой слайд

Для акцентов в коде и тексте на слайдах в настройках цвета у вас есть готовая палитра:



HEX #fc2c38
RGB 252, 43, 56
Pantone 032C
CMYK 0, 97, 75, 0



HEX #e5e8e8
RGB 230, 231, 232
Pantone Cool Gray 1C
CMYK 0, 0, 0, 9



HEX #75787b
RGB 117, 120, 123
Pantone Cool Gray 9C
CMYK 0, 0, 0, 73

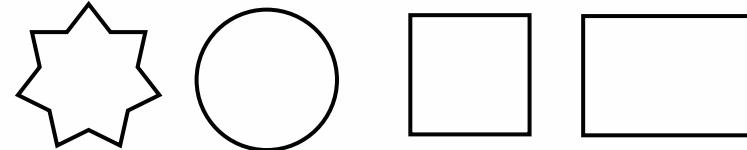
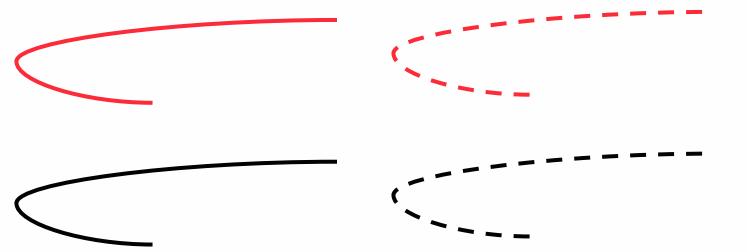


HEX #030303
RGB 3, 3, 3
Pantone Black
CMYK 75, 67, 67, 89

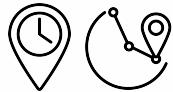
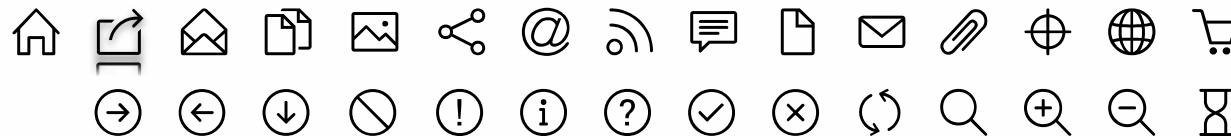
Слайды всех типов находятся в шаблоне, их можно выбрать из выпадающего меню в окне «Создать слайд»

Иконки и элементы

Для создания ориентиров на слайде
используйте иконки из готового набора
или подходящие по цветовой гамме:



Иконки и элементы



- 
3. Интеграция в существующие сервисы / связка с тем, как/кто должен оценивать успех идеи (клиенты/бизнес/технологи)
 - 4.



Логотипы

Логотипы продуктов и сервисов Mail.ru Group

Topic Name Color

Topic Name Color

Topic Name Color

Topic Name Color

Topic Name Shape

Topic Name Shape

Topic Name Shape



Критерии оценки

1. Реализация проекта (соответствие технологий задаче, сложность проекта, дружественный интерфейс, понятный/качественный код, оптимальное решение задачи и т.п.)
2. Презентация проекта (презентационные навыки команды. Логика повествования. Подробный охват спектра работ над проектом)
3. Дополнительные баллы (за уникальность проекта, сложность задачи, уникальность подхода к решению задачи)