

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Волгоградский государственный технический университет»

Факультет Электроники и вычислительной техники
Кафедра Системы автоматизированного проектирования и поискового
конструирования

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к курсовому проекту

по дисциплине Междисциплинарный курсовой проект
на тему _____

Студент Чечеткин Илья Александрович

Группа САПР-1.1п

Руководитель работы (проекта) _____
(подпись, дата) (инициалы и фамилия)

Члены комиссии:

_____ (подпись и дата подписания)	_____ (инициалы и фамилия)
_____ (подпись и дата подписания)	_____ (инициалы и фамилия)
_____ (подпись и дата подписания)	_____ (инициалы и фамилия)

Нормоконтролер _____
(подпись и дата подписания) (инициалы и фамилия)

Волгоград 2016г.

Факультет	Электроники и вычислительной техники
Направление (специальность)	Информационное и программное обеспечение автоматизированных систем
Кафедра	Системы автоматизированного проектирования и поискового конструирования
Дисциплина	Междисциплинарный курсовой проект

Утверждаю
Зав. кафедрой САПР и ПК В. А. Камаев
« » 2016г.

Студент Чечеткин Илья Александрович
(фамилия, имя, отчество)

Группы САПР-1.1п

1. Тема

Утверждена приказом от « » 2016г. №

2. Срок представления работы (проекта) к защите « » 20 г.

3. Постановка задачи

4. Дата выдачи задания « » _____ 2016г.

Руководитель работы (проекта) _____

(подпись, дата) (инициалы и фамилия)

Задание принял к исполнению _____

(подпись, дата) (инициалы и фамилия)

ОГЛАВЛЕНИЕ

1	Введение	4
2	Описание решаемых задач	6
3	Результат анализа и систематизации информации	7
4	Структура магистерской работы	8
5	Описание прототипа	9
6	Заключение	10
7	Приложение	11
7.1	Прототип	11
7.2	Псевдокод алгоритма K-Means	12
7.3	Псевдокод алгоритма Mean Shift	12

1 ВВЕДЕНИЕ

Актуальность. В настоящее время формирование маршрутов в городской среде осуществляется на основе положений, заложенных в городской план развития. Эта информация, как правило, достаточно устаревшая и не учитывает предпочтения жителей. На основе полученных данных о предпочтениях жителей, представленных в виде множества начальных и конечных точек маршрута требуется разработать эффективный метод кластеризации предпочтений жителей города по перемещению.

Цель работы – разработка метода кластеризации предпочтений жителей для минимизации дискомфорта перемещения в городе.

Теоретический этап – рассмотрение информации по существующим алгоритмам; составление теоретической базы проекта и систематизация полученных знаний; составление технического задания.

Практический этап – реализация системы на основе теоретической базы, составленной ранее с использованием технического задания.

Финальный этап – внедрение готового продукта, получение обратной информации и исправление ошибок.

Теоретические задачи:

- разработка алгоритма кластеризации;
- разработка метода учета географических особенностей местности;
- разработка критериев для оценки качества кластеризации.

Практические задачи:

- генерация исходных данных;
- реализация разработанных алгоритмов и методов;
- построение полученных результатов на карте;
- оценка качества кластеризации.

Понятийный аппарат

- **Кластер** – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определенными свойствами.
- **Предпочтение** – пара узлов с определенными координатами и идентификатором пользователя.
- **Node (узел)** – точка с указанными координатами и тегами.
- **Tag (тег)** – пары «ключ – значение».
- **Дискомфорт** – совокупный параметр, определяющий время перемещения из начального узла в конечный.
- **Центроид** – центр тяжести фигуры (геометрический центр).
- **Метрика** – функция, определяющая расстояние в метрическом пространстве.
- **Framework (фреймворк)** – программная платформа, определяющая структуру программной системы; программное обеспечение, облегчающее разработку и объединение разных компонентов большого программного проекта.
- **OpenStreetMap (OSM)** – некоммерческий веб-картографический проект по созданию силами сообщества участников-пользователей Интернета подробной свободной и бесплатной географической карты мира.
- **Project OSRM** – фреймворк для вычисления кратчайших путей в графе дорог. Разработан для использования с картографическим сервисом OSM.

Объект исследования – предпочтения жителей города, выраженные в географических координатах.

Предмет исследования – методы кластеризации предпочтений жителей.

2 ОПИСАНИЕ РЕШАЕМЫХ ЗАДАЧ

В данной работе рассматриваются к решению следующие задачи:

1. разработка механизма генерации исходных данных, которые представляют собой исходные и конечные пункты ежедневных маршрутов;
2. реализация метода кластеризации точек маршрутов, охватывающих все исходные данные;
3. оценка качества кластеризации;
4. отображение результатов метода кластеризации на карте.

Первая задача заключается в создании псевдореалистичных данных для замены отсутствующих реальных на данный момент. Они нужны для работы над последующими задачами как некий приближенный аналог.

Вторая задача заключается в разработке метода кластеризации предпочтений, генерирует оптимальный список кластеров, основываясь на данных об отправных и конечных пунктах маршрутов. Кластеры не должны иметь определенной формы, а количество людей в кластерах должно усредняться, то есть метод подразумевает разделение и слияние кластеров в процессе кластеризации. Так же метод подразумевает использование метрики, основанной на построении маршрутов между точками на карте OSM. На текущем этапе используется алгоритм Mean Shift для кластеризации исходных данных, метрика разработана с использованием фреймворка OSRM. В дальнейшем планируется внедрение метрики маршрутов в работу алгоритма кластеризации, возможно, разработка нового метода кластеризации или метрики.

Третья задача заключается в разработке критериев, по которым можно будет оценить качество проделанной кластеризации. Также предоставить данную информацию пользователю и модулю кластеризации для последующей оптимизации.

Четвертая задача заключается в разработке web-инструмента для отображения и редактирования проделанной кластеризации.

Псевдокод текущий алгоритмов представлен в Приложении (7.2, 7.3).

3 РЕЗУЛЬТАТ АНАЛИЗА И СИСТЕМАТИЗАЦИИ ИНФОРМАЦИИ

Среди различных источников информации мной была выделена следующая литература:

Воронцов К. В. Машинное обучение. <http://www.machinelearning.ru/>
Курс лекций, охватывающий множество современных алгоритмов и проблем классификации и кластеризации. Алгоритмы рассмотрены очень подробно, ко многим приведены псевдокоды. В лекциях предоставлена информация по эффективности, производительности и сложности различных алгоритмов и методов классификации и кластеризации.

Mean Shift Clustering.

http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/TUZEL1/MeanShift.pdf

Статья, описывающая один алгоритмов кластеризации данных: Mean Shift. В ней предоставлена обширная информация по его работе и методах, на которых он основан.

Mean Shift: A Robust Approach Toward Feature Space Analysis.

<https://courses.csail.mit.edu/6.869/handouts/PAMIMeanshift.pdf>

Статья, рассматривающая применение алгоритма Mean Shift при анализе изображений. В ней дана исчерпывающая информация по его работе, применениям, достоинствам и недостаткам. Даны наглядные примеры работы алгоритма в различных условиях и для различных целей. Также приведен довольно большой список источников информации.

D. Allard, G. Guillot. Clustering geostatical data.

http://people.compute.dtu.dk/gigu/article_capetown.pdf

Статья, описывающая методы кластеризации геостатических данных. Рассмотрен вопрос близости объектов, приведены некоторые алгоритмы кластеризации и их сравнение на тестовой выборке, а также результаты работы алгоритмов на разных выборках. Было выделено два критерия проделанной кластеризации: «будущность» (*likelihood*) и «дисперсия» (*variance*). Подробно рассмотрены алгоритмы EM и EC-M.

4 СТРУКТУРА МАГИСТЕРСКОЙ РАБОТЫ

1. Введение

- Актуальность
- Цели, задачи
- Ожидаемый результат

2. Введение в проблему кластеризации

- Анализ предметной области
- Состояние современных исследований
- Требования к методам

3. Метод кластеризации

- Общее описание
 - Схематическое представление
 - Идея метода
- Подробное описание

4. Испытание и обоснование эффективности предлагаемых подходов

- Проектирование ПО
- Методика проведения эксперимента
 - Данные
 - Критерии
 - Методика
- Проведение эксперимента и описание результатов
- Обсуждение результатов
- Интеграция

5. Заключение

6. Список используемой литературы

7. Приложение

5 ОПИСАНИЕ ПРОТОТИПА

Прототип включает в себя два модуля:

- пользовательский интерфейс;
- инструмент для кластеризации данных о предпочтении жителей.

Пользовательский интерфейс представляет из себя карту, на которой будут отображаться данные о предпочтениях и полученные различными методами кластеры. Каждый из кластеров можно будет выбирать и получать по нему подробную информацию: метод кластеризации, оценку качества кластеризации, приписанные кластеру точки.

Инструмент для кластеризации данных – приложение, которое получая на вход данные о предпочтениях жителей будет генерировать набор кластеров и передавать данные в пользовательский интерфейс.

Примерный вид интерфейса и ссылка на прототип в Приложении 7.1.

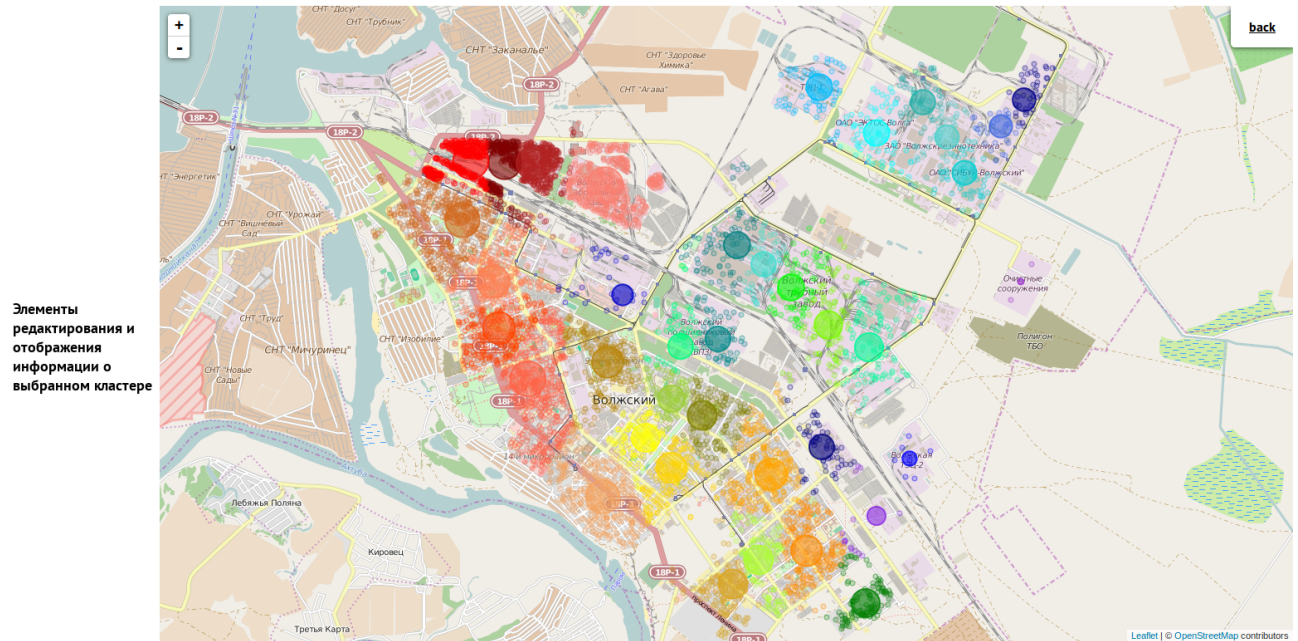
6 ЗАКЛЮЧЕНИЕ

Конечным результатом будет инструмент, который дает возможность на основе данных о предпочтении жителей найти скопления пунктов отправления и прибытия, на основе которых в дальнейшем будет строиться сеть маршрутов городского транспорта.

7 ПРИЛОЖЕНИЕ

7.1 Прототип

Представление Web UI:



Реализованный алгоритм: <https://github.com/vstu-cad-stuff/clustering>

Текущая реализация: <http://vstu-cad-stuff.github.io/clustering/>

7.2 Псевдокод алгоритма K-Means

1. Генерирование начального распределения центроидов
2. итерация = 0
3. **ПОВТОРЯТЬ**
4. Расчет принадлежности всех точек центроидам
5. Расчет нового положения центроидов
6. Сдвиг центроидов
7. итерация += 1
8. **ПОКА** разница между рассчитанным положением и текущим не равна 0 и пока количество итераций не достигло максимума
9. **ВЫВОД** центроиды

7.3 Псевдокод алгоритма Mean Shift

1. Генерирование начального распределения центроидов
2. **ПОВТОРЯТЬ**
3. Определение соседних точек к центроидам
4. Определение среднего веса соседних точек к центроидам
5. Расчет нового положения центроидов
6. Сдвиг центроидов
7. **ПОКА** разница между рассчитанным положением и текущим не будет равна 0
8. **ВЫВОД** центроиды