



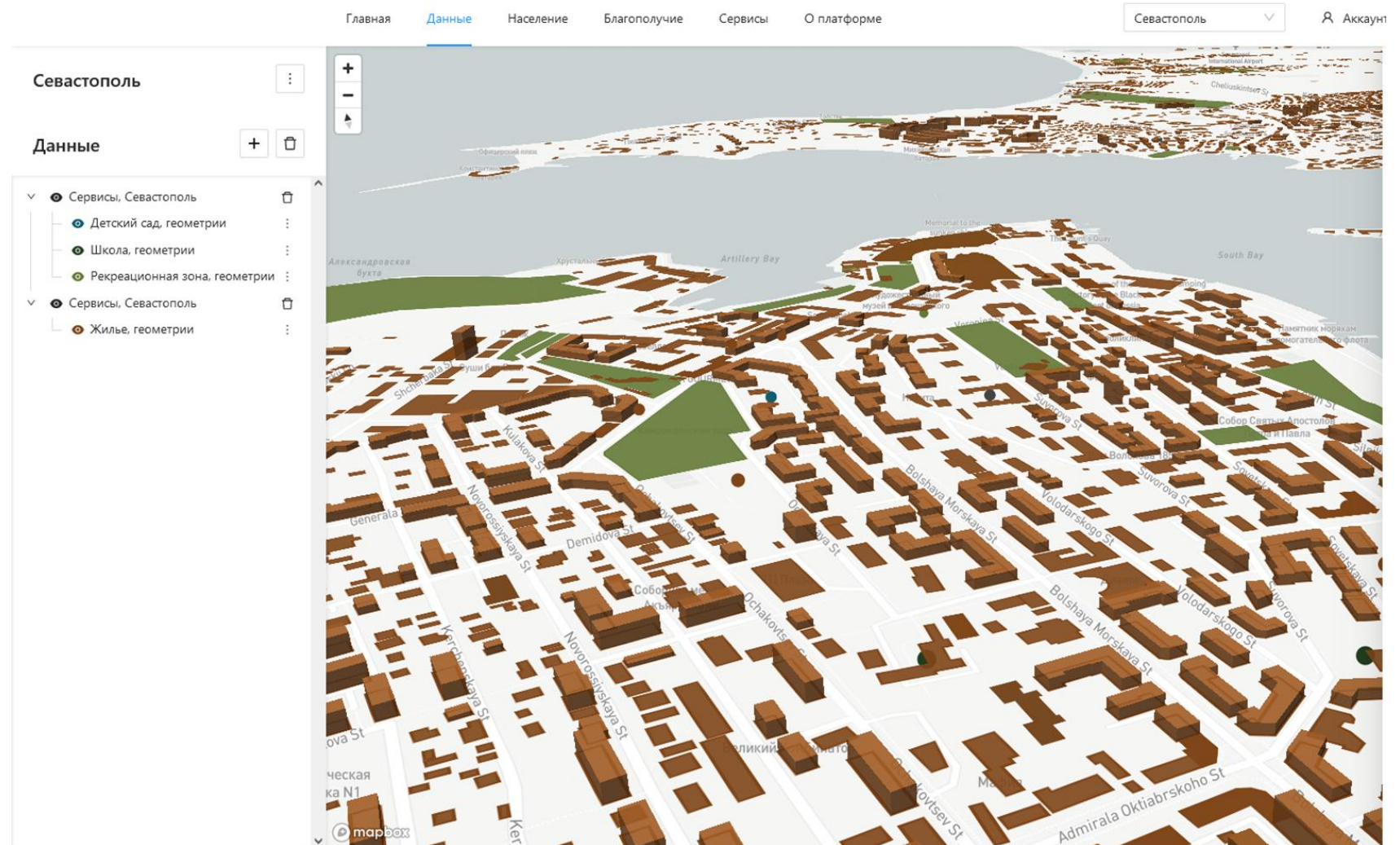
# ІТМО

## **SOIKA на хвосте принесла: собираем живые данные о городе**

**Антонов Александр, аспирант**

2023

- ✓ Цифровые модели городов позволяют нам изучать проходящие в них процессы *in silico*, не влияя на существующую городскую среду
- ✓ Они требовательны к наборам данных
- ✓ Актуальные наборы данных крайне тяжело добыть



## Количественные показатели

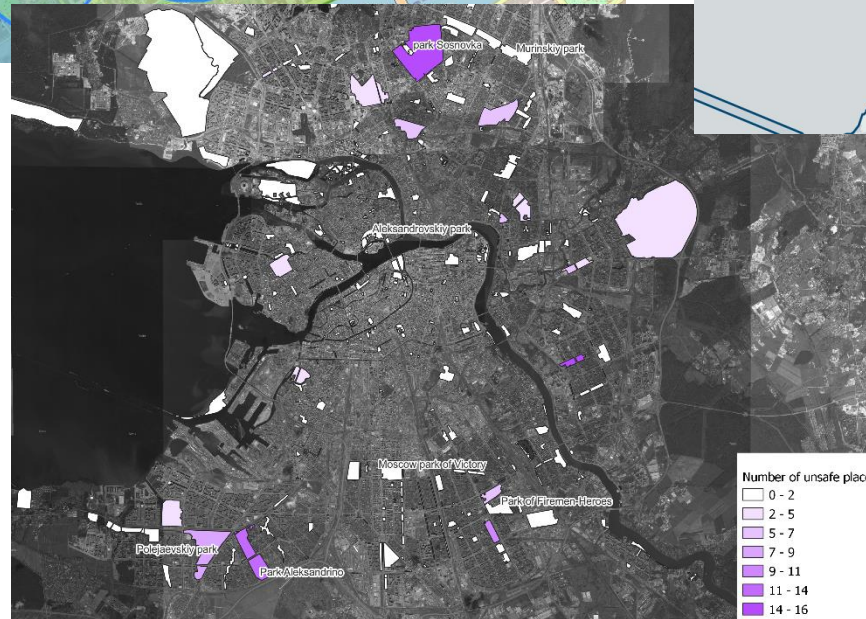
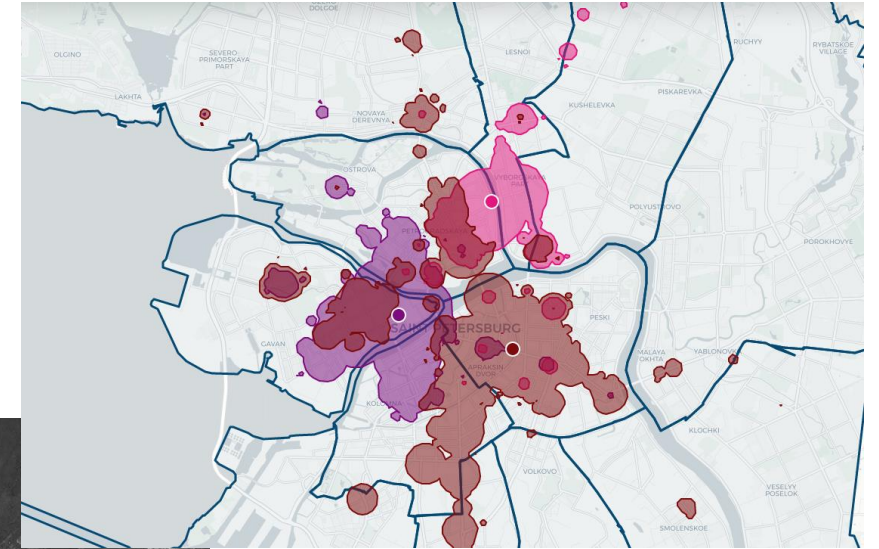
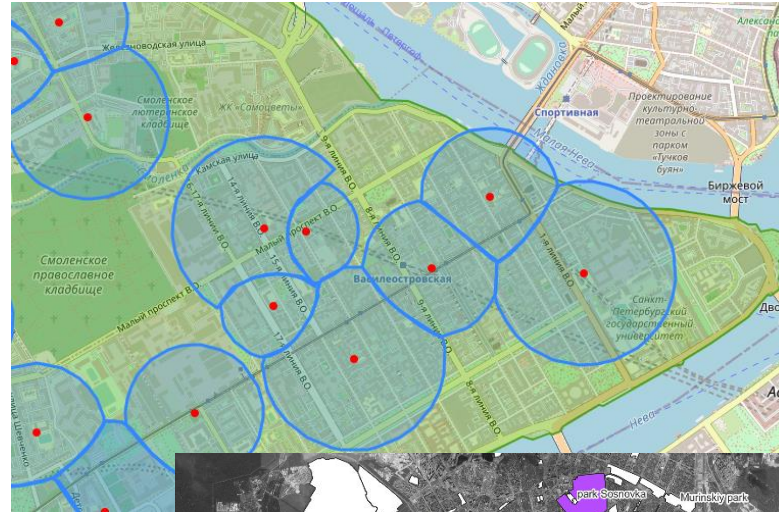
- ✔ Обеспеченность сервисами
- ✔ Доступность городских территорий

## Качественные показатели

- ✔ Износ зданий и инфраструктуры
- ✔ Качество получаемых сервисов и услуг
- ✔ Качество городской среды в целом

## Интегральные показатели / индексы

- ✔ Индекс качества городской среды
- ✔ Индекс инфраструктурной обеспеченности



- ✓ Закрытые источники (по договоренности)
- ✓ Общие и государственные порталы открытых данных (OSM, Реформа ЖКХ)
- ✓ Сенсоры и датчики
- ✓ Неструктурированные спонтанные данные: обращения, комментарии, как вторичный вариант – новости и так далее



- ✔ Жители описывают город по его элементам – с помощью OpenStreetMap
- ✔ Жители оставляют отзывы на сервисы – их можно агрегировать
- ✔ Жители создают обращения в сервисах электронного участия, например Наш Санкт-Петербург – и они уже структурированы
- ✔ Жители описывают свою оценку города в обсуждениях социальных сетей
- ✔ Жители пишут о событиях в городе – и если о них упоминается, значит, они уже достаточно важные

**ДТП и ЧП Санкт-Петербург**  
 -15 °C 4 балла

Грязь не убирается на Косыгина11, к.2, с бго пл 10й парадные 😡😡😡😡😡 нет дворника...  
 15 ноя. в 16:00 Ответить Подѐлиться 3

Статус проблемы	Количество проблем
Решенные проблемы	4946450
На рассмотрении	130141
На модерации	34

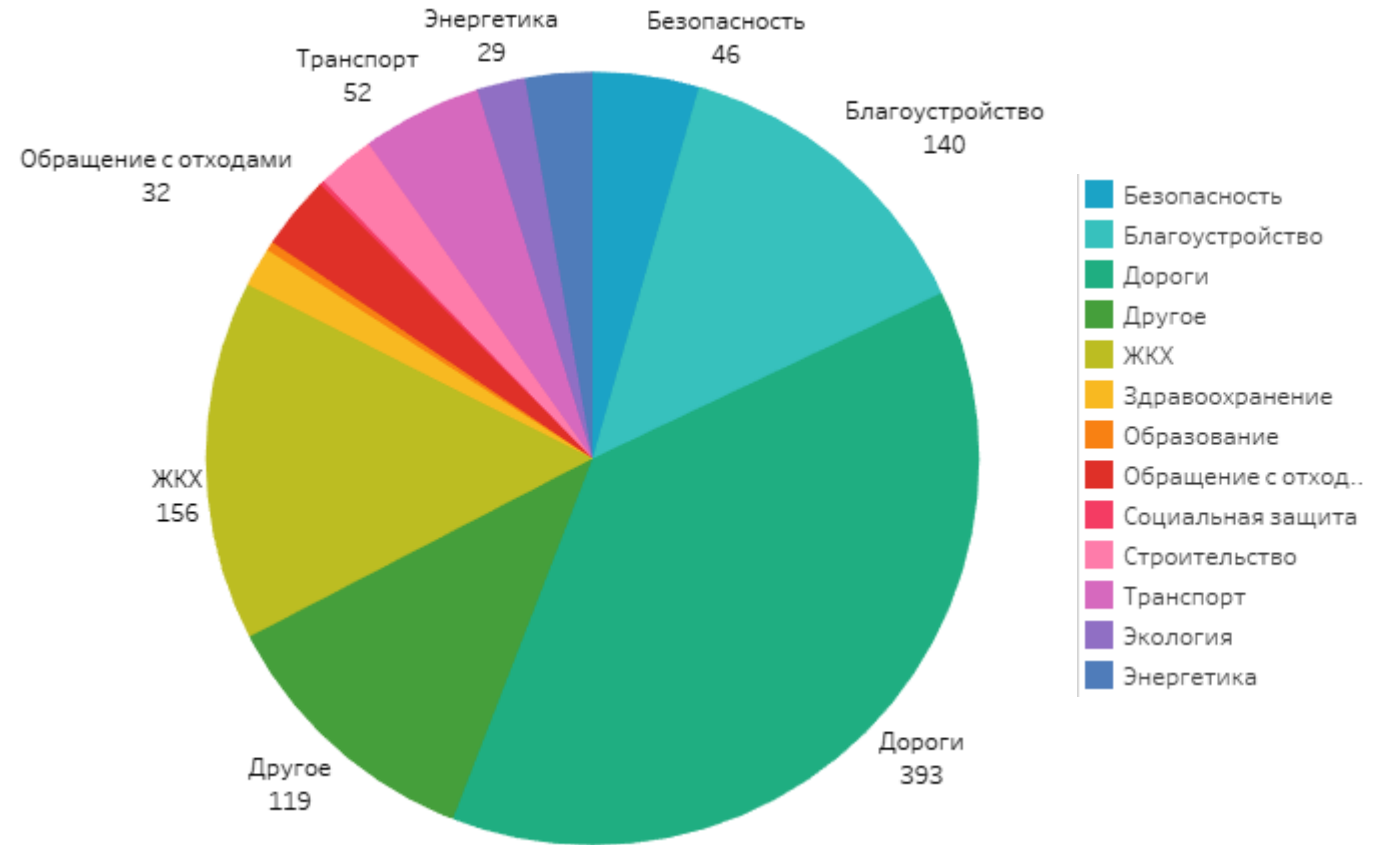
Библиотека SOIKA направлена на автоматизацию задач по структуризации текстов социальных сетей и других источников, описывающих городские процессы.

Основные особенности:

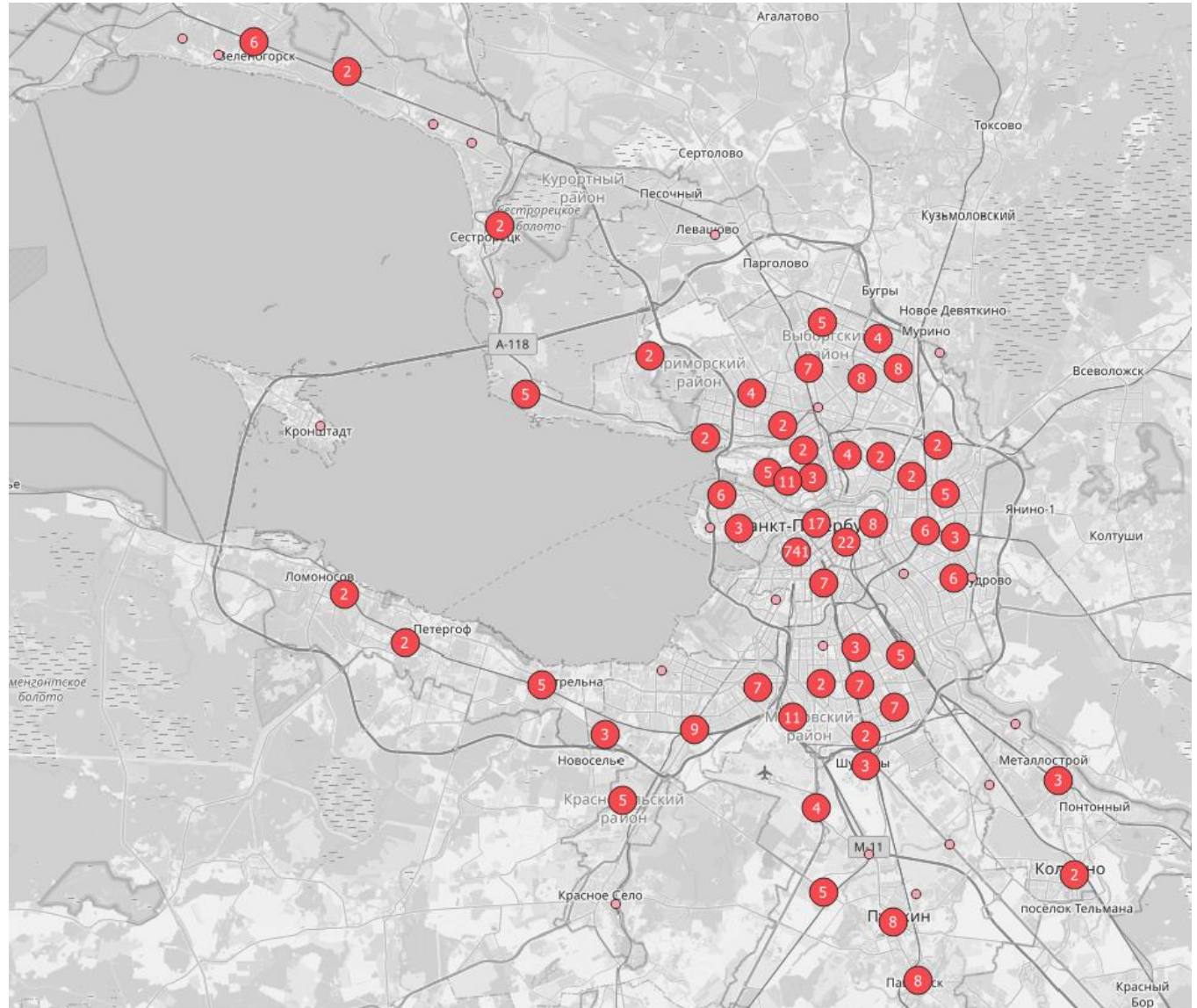
- ✓ Классификация городских функций производится при помощи классификатора, обученного на 90 000 текстах обращений, размеченных по 12 функциям
- ✓ Выявление и геокодирование адресов на основе модели NER в комбинации с другими методами извлечения и обработки адресов
- ✓ Формирование событий на основе семантической кластеризации в пространственной иерархической модели. Итеративный подход позволяет выявлять события по принципу сита на разных пространственных уровнях города, а также определять смысловые связи между ними



- ✓ С точки зрения функциональной классификации город разбит на ряд инфраструктур, поддерживающих тот или иной аспект жизни горожан
- ✓ С ограниченным набором основных городских функций на размеченном наборе обращений можно натренировать классификатор (что мы и сделали)



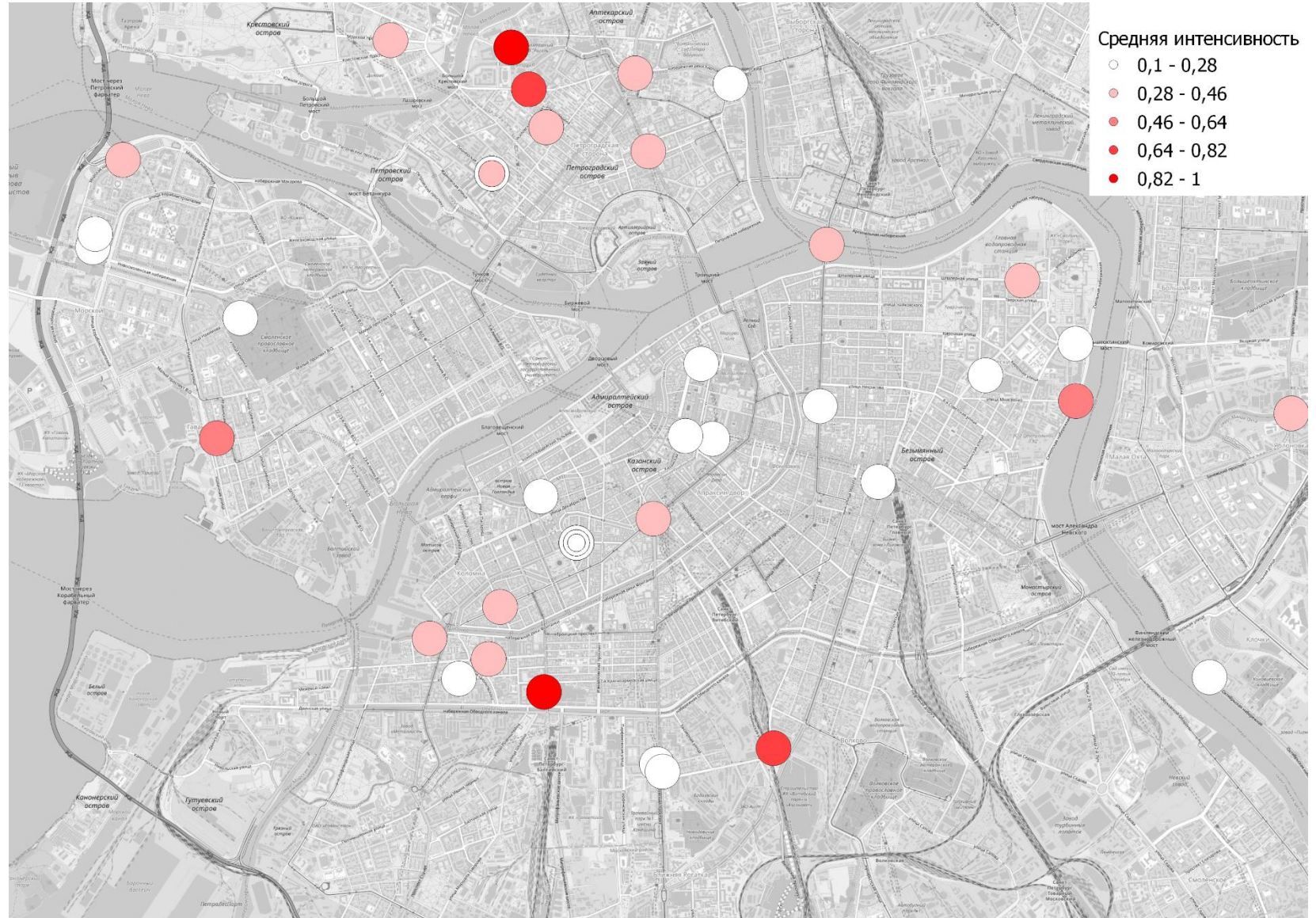
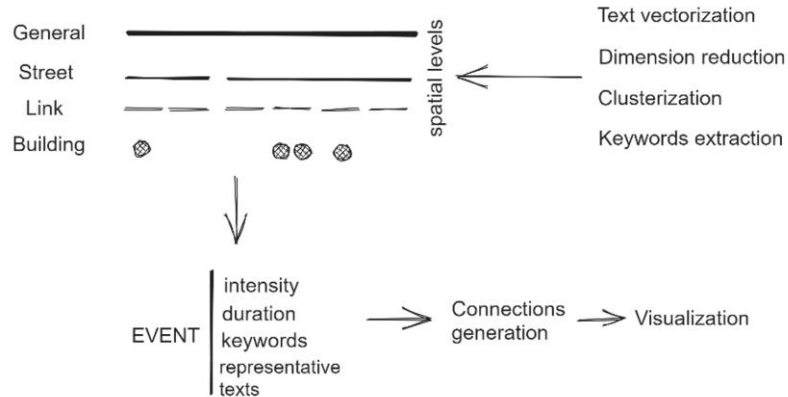
- ✓ Наиболее важной частью структуризации спонтанных данных является их геопарсинг – привязка текстов к пространству за счет определения упоминаемых урбанонимов (названия улиц, организаций, парков и др.) в тексте и их геокодирования
- ✓ Также важно учитывать особенности описания городского пространства жителями



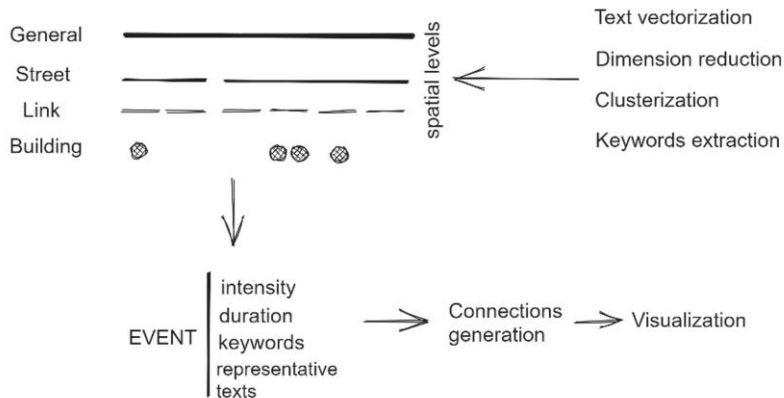


✓ После привязки сообщений к территории и при наличии временных меток появляется возможность выявить набор событий, послуживших триггером для создания дискуссии в социальных сетях

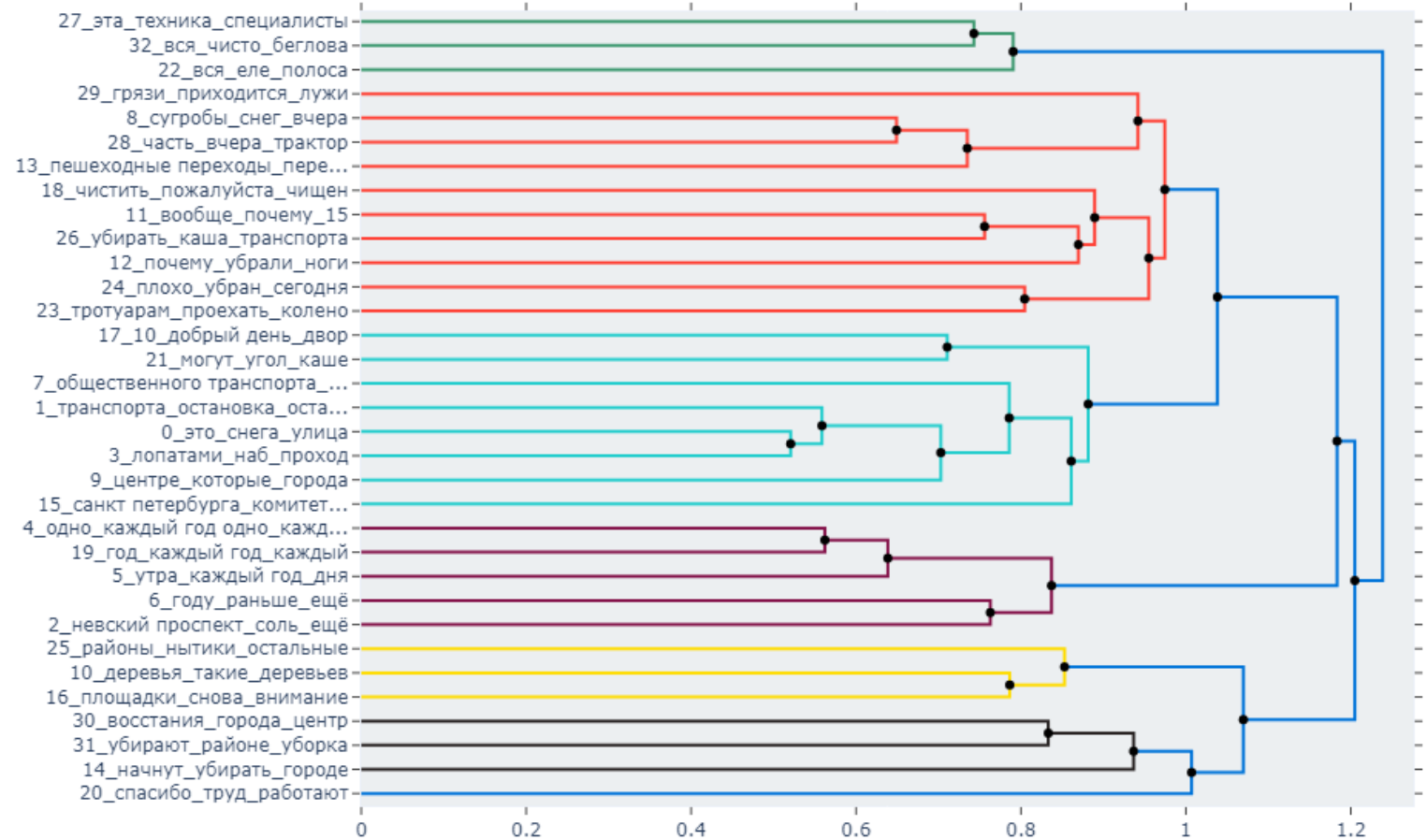
✓ Это можно выполнить с помощью применения семантической кластеризации на уровнях иерархической пространственной модели, по аналогии с ситом



- ✓ После привязки сообщений к территории и при наличии временных меток появляется возможность выявить набор событий, послуживших триггером для создания дискуссии в социальных сетях
- ✓ Это можно выполнить с помощью применения семантической кластеризации на уровнях иерархической пространственной модели, по аналогии с ситом



## Hierarchical Clustering



- ✓ Структурированная пространственная модель
- ✓ Не менее структурированный подход к семантическим параметрам данных
- ✓ «Прослойка» между пользовательскими спонтанными данными и цифровыми моделями городов
- ✓ Возможность моделировать и прогнозировать динамику в результате изменений физической и смысловой составляющей города

