

Создаем карту границ участков избирательных комиссий Москвы из открытых данных

[Обсудить в форуме](#) Комментариев — 8

Эта страница является черновиком статьи.

Шаги для создания карты границ участков

В разных уголках нашей страны, а иногда и целиком по всей стране иногда проводятся выборы разных уровней. По закону, количество избирателей на участок не должно быть больше 3000, поэтому вся страна или область разбивается на участки. Люди проживающие на территории участка голосуют на избирательном участке - месте голосования. В виде открытых геоданных, за редким исключением некоторых регионов, государство данные об участках и их границах не распространяет, границы доступны только в виде описаний. Если местоположения мест голосований еще можно [собрать](#), с участками количество работы по самостоятельной их отрисовке возрастает минимум на порядок. Поэтому имеет смысл попробовать применить автоматизацию.

Эта статья иллюстрирует некоторые, достаточно простые подходы для создания карты участков в границах г. Москвы (исключая присоединенные территории Новой Москвы). Статья может рассматриваться как иллюстрация применения географического анализа для решения более-менее конкретной задачи. Используя данные, ссылки на которые приведены, можно всё воспроизвести самостоятельно.

Забегая вперед хочется сказать, что существующие наборы данных неадекватны для решения этой задачи и результат неудовлетворителен. Но лучше хоть какая-то попытка, чем никакой.

Спасибо Сергею Шпилькину, что еще раз напомнил о том, что это стоит попробовать сделать и Виктору Олайя за помощь с Sextante и быстрый багфикс в нем. Опенсорс FTW!

Содержание

- [1 Основная идея](#)
- [2 Исходные данные](#)
- [3 Используемое ПО](#)
- [4 Получение и подготовка данных](#)
 - [4.1 Границы АТД](#)
 - [4.2 База адресов](#)
- [5 Геообработка](#)
- [6 Автоматизация](#)
- [7 Результат](#)
- [8 Почему получилось плохо](#)

Основная идея

Границы избирательных участков можно восстановить зная причисленные к ним адреса.

Исходные данные

Границы административно-территориального деления г. Москва (источник - [GIS-Lab](#))

Адреса причисленные к избирательным участкам (источник - сервис "Поиск места голосования" [Мосгоризбиркома](#)). Используется выгрузка из ресурса на 29 августа 2013 г.

Картографические данные - [OpenStreetMap](https://www.openstreetmap.org/).

Используемое ПО

Будем использовать только открытое программное обеспечение.

Нам понадобится:

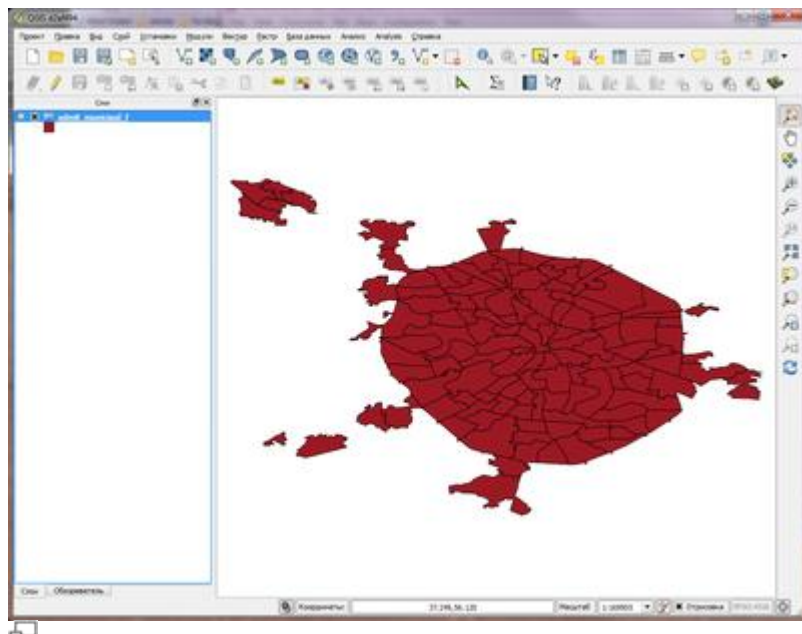
- Python - программирование скриптов для получения данных с МГИКа и пакетного геокодирования.
- QGIS/Sextante - геообработка, визуализация.

Получение и подготовка данных

Границы АТД

Скачаем [стабильную версию](#) внутригородских муниципальных образований Москвы и Санкт-Петербурга. Удалим все что относится к Санкт-Петербургу, так как нас в данном случае он не интересует. Для этого откроем слой данных в QGIS, выделим и удалим объекты относящиеся к СПб.

Нам нужны эти границы, так как участки создаются в границах районов Москвы.



Границы муниципальных образований Москвы (без Новой Москвы)

База адресов

Наша задача - получить все адреса домов и привязанный к ним номер УИКа.

Этап 1 - скачаем всю базу адресов МГИК.

Это можно сделать используя специальный скрипт на Python ([исходник](#)). Скрипт работает очень просто, обходит весь сервис МГИКа по поиску УИКов по адресу и складывает все найденные данные в CSV файл. Для запуска скрипта нужен чистый питон 2.7.x и пакет BeautifulSoup.

Пример фрагмента результата после скачивания:

| ID | UIK_ID | LVL1 | LVL2 | LVL3 | SEARCH_L | LVL2_LINK |
|-------|--------|--------------------|------|---------|---|--|
| 16393 | 413 | верещагина ул. | | | 4 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15635/search |
| 16394 | 413 | верещагина ул. | | | 6 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15636/search |
| 16395 | -1 | верещагина ул. | | | 7 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15637/search |
| 16396 | 413 | верещагина ул. | | | 8 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15638/search |
| 16397 | 68 | верземнека ул. | | | 1 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16534/search |
| 16398 | 68 | верземнека ул. | | | 2 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16535/search |
| 16399 | 68 | верземнека ул. | 2a | | http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16536/search |
| 16400 | 68 | верземнека ул. | | | 3 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16537/search |
| 16401 | 68 | верземнека ул. | | | 5 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16538/search |
| 16402 | 68 | верземнека ул. | | | 6 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16539/search |
| 16403 | 68 | верземнека ул. | | | 7 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16540/search |
| 16404 | 2766 | вернадского просп. | 10 | корп. 1 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19730/search |
| 16405 | 2766 | вернадского просп. | 10 | корп. 2 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19731/search |
| 16406 | 2844 | вернадского просп. | | 100 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19732/search |

Таблица адресов МГИК

В результате присутствуют важные для нас поля:

- UIK_ID - номер УИКА
- LVL1 - уровень 1 адреса: название нас.пункта или улицы
- LVL2 - уровень 2 адреса: название улицы если есть и нас.пункт
- LVL3 - уровень 3 адреса: номер дома

[Скачать исходную таблицу](#) (CSV)

Этап 2 - подготовим адреса

Путем простых операций в табличном процессоре, превратим таблицу в подходящую для геокодирования, для этого нужно сделать чтобы адрес был в одной ячейке, это сделает простой CONCATENATE:

| ID | tik | tik_id | uik | uik_id | LVL1 | LVL2 | LVL3 | SEARCH_L | LVL2_LINK | ADDR | ADDR_V |
|-------|-----|--------|-------|--------|--------------------|------|---------|---|--|-------------------------------|-------------------------------|
| 16393 | 1 | 1 | 16393 | 413 | верещагина ул. | | | 4 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15635/search | верещагина ул. 4 | верещагина ул. 4 |
| 16394 | 1 | 1 | 16394 | 413 | верещагина ул. | | | 6 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15636/search | верещагина ул. 6 | верещагина ул. 6 |
| 16395 | 1 | 1 | 16395 | -1 | верещагина ул. | | | 7 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15637/search | верещагина ул. 7 | верещагина ул. 7 |
| 16396 | 1 | 1 | 16396 | 413 | верещагина ул. | | | 8 http://mosgorizbirkom.ru/precinct/address/1756/search | http://mosgorizbirkom.ru/precinct/address/15638/search | верещагина ул. 8 | верещагина ул. 8 |
| 16397 | 1 | 1 | 16397 | 68 | верземнека ул. | | | 1 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16534/search | верземнека ул. 1 | верземнека ул. 1 |
| 16398 | 1 | 1 | 16398 | 68 | верземнека ул. | | | 2 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16535/search | верземнека ул. 2 | верземнека ул. 2 |
| 16399 | 1 | 1 | 16399 | 68 | верземнека ул. | 2a | | http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16536/search | верземнека ул. 2a | верземнека ул. 2a |
| 16400 | 1 | 1 | 16400 | 68 | верземнека ул. | | | 3 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16537/search | верземнека ул. 3 | верземнека ул. 3 |
| 16401 | 1 | 1 | 16401 | 68 | верземнека ул. | | | 5 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16538/search | верземнека ул. 5 | верземнека ул. 5 |
| 16402 | 1 | 1 | 16402 | 68 | верземнека ул. | | | 6 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16539/search | верземнека ул. 6 | верземнека ул. 6 |
| 16403 | 1 | 1 | 16403 | 68 | верземнека ул. | | | 7 http://mosgorizbirkom.ru/precinct/address/1758/search | http://mosgorizbirkom.ru/precinct/address/16540/search | верземнека ул. 7 | верземнека ул. 7 |
| 16404 | 1 | 1 | 16404 | 2766 | вернадского просп. | 10 | корп. 1 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19730/search | вернадского просп. 10 корп. 1 | вернадского просп. 10 корп. 1 |
| 16405 | 1 | 1 | 16405 | 2766 | вернадского просп. | 10 | корп. 2 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19731/search | вернадского просп. 10 корп. 2 | вернадского просп. 10 корп. 2 |
| 16406 | 1 | 1 | 16406 | 2844 | вернадского просп. | | 100 | http://mosgorizbirkom.ru/precinct/address/1759/search | http://mosgorizbirkom.ru/precinct/address/19732/search | вернадского просп. 100 | вернадского просп. 100 |

Таблица адресов МГИК

Кроме всяких служебных полей, нужных для правильной работы скриптов по геокодированию, добавились важные поля:

- ADDR - слитый адрес
- ADDR_V - результат очистки, поиском-заменой убраны лишние запятые и т.п.

Таблица готова к геокодированию.

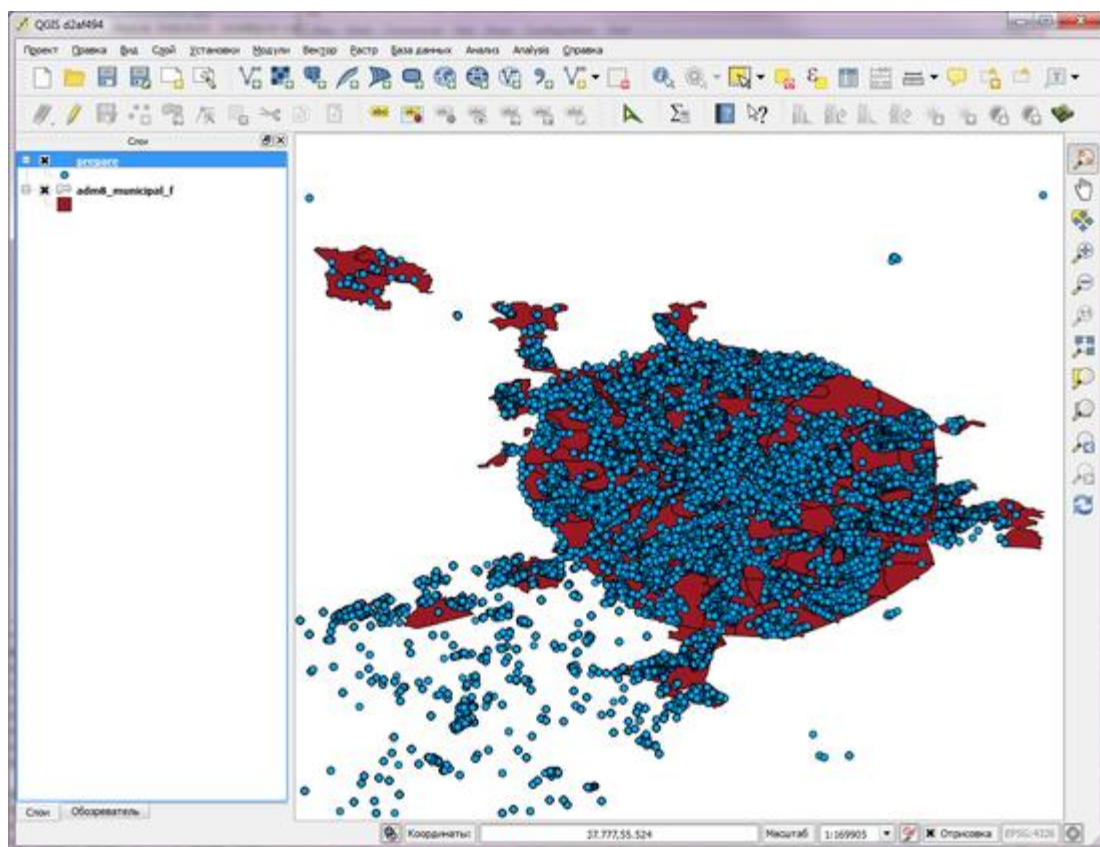
[Скачать обработанную таблицу](#) (CSV)

Этап 3 - геокодируем

Скрипты для геокодирования можно скачать [здесь](#). Это тоже питон использующий бэкэнд OpenStreetMap.ru для геокодирования.

[Скачать результат геокодирования](#)

На карте эти адреса будут выглядеть вот так:



Геокодированные адреса на карте районов

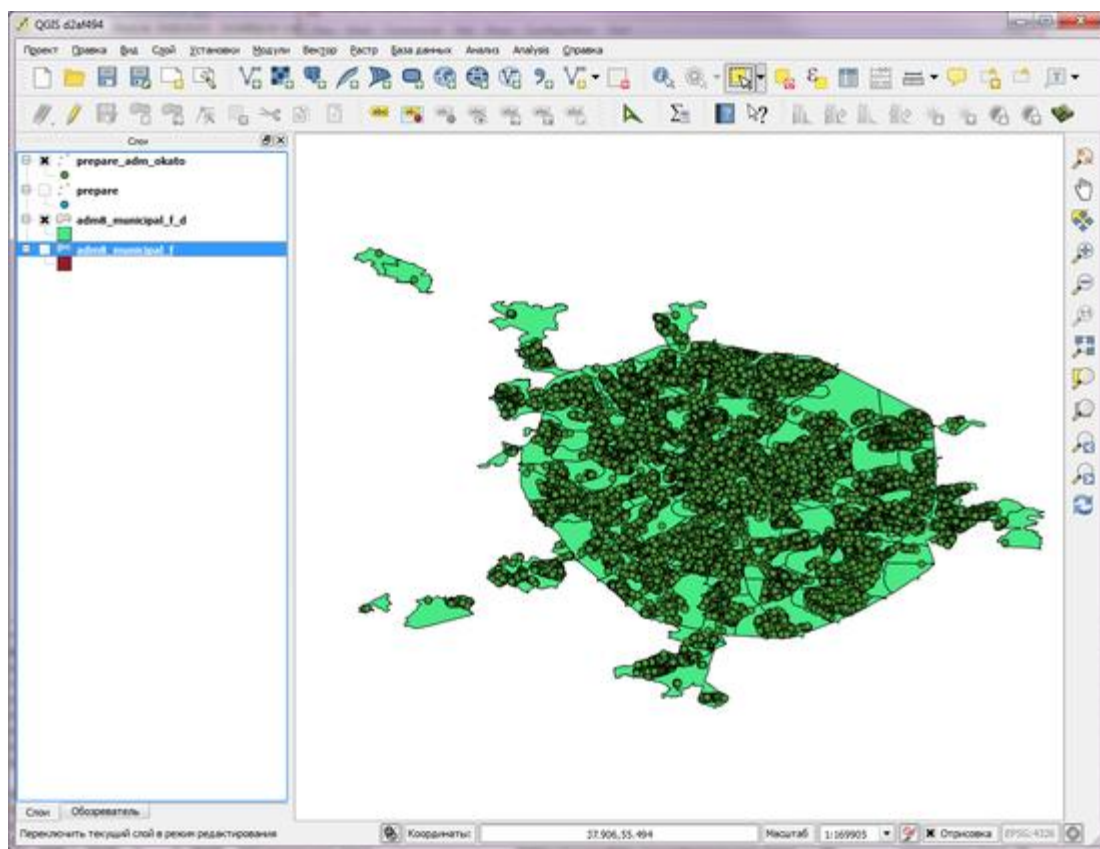
Нам понадобится еще несколько шагов, чтобы подготовить только те данные, которые нам нужны:

Этап 4 - удалим всё, что не в старой Москве

Для этого в QGIS выполним Вектор/Выборка/Пространственная выборка и выделим все точки попадающие в границы нашей Москвы. После выборки - инвертируем ее и удалим все что выделилось (вне границ Москвы).

Этап 5 - удалим все районы Москвы без точек

С помощью того же инструмента проведем и обратную операцию, выбрав и удалив все районы в которые не попало ни одной точки.



Оставшиеся точки и районы

Этап 6 - привяжем точки к районам

У нас каждой точке адреса соответствует УИК, а нам нужно, чтобы еще у них был и район (для чего - увидим позже). Поэтому в QGIS делаем Вектор/Управление данными/Объединение атрибутов по районам. После этой операции, каждая точка получит все атрибуты того полигона района, который ей нужен. Визуально точки не изменятся, просто добавятся новое поле с кодом OKATO из слоёв АД.

Этап 7 - удалим плохие точки Удалим все точки, уровень точки геокодирования которых ниже чем Дом, т.е. эти точки привязались к улице или населенному пункту. Для нашего метода это абсолютно неприемлемо, поэтому мы их лучше удалим. Проще всего это сделать через таблицу атрибутов, где сделать выборку `g_status != "building"` и удалить выделенное.

Из 55 тыс адресов у нас останется около 21 тысячи.

[Скачать очищенный набор геокодированных точек адресов](#) (ESRI Shape)

Этап 8 - разбивка по OKATO Это последний этап подготовки. Теперь у нас есть точки и полигоны и у всех объектов, помимо прочих значений в таблице атрибутов, есть код OKATO, который их объединяет. Разобьем эти два слоя на множество производных слоёв. Каждый из производных будет представлять один код OKATO.

В QGIS: Вектор/Управление данными/Разбить векторный слой Разбиваем по полю OKATO_CODE. Рекомендую сложить результирующие точки и полигоны в отдельные папки, которые предварительно нужно создать. После окончания разбивки, нужно переименовать все файлы так, чтобы в названии остался только код OKATO. В результате количество файлов точек и полигонов должно точно совпадать.

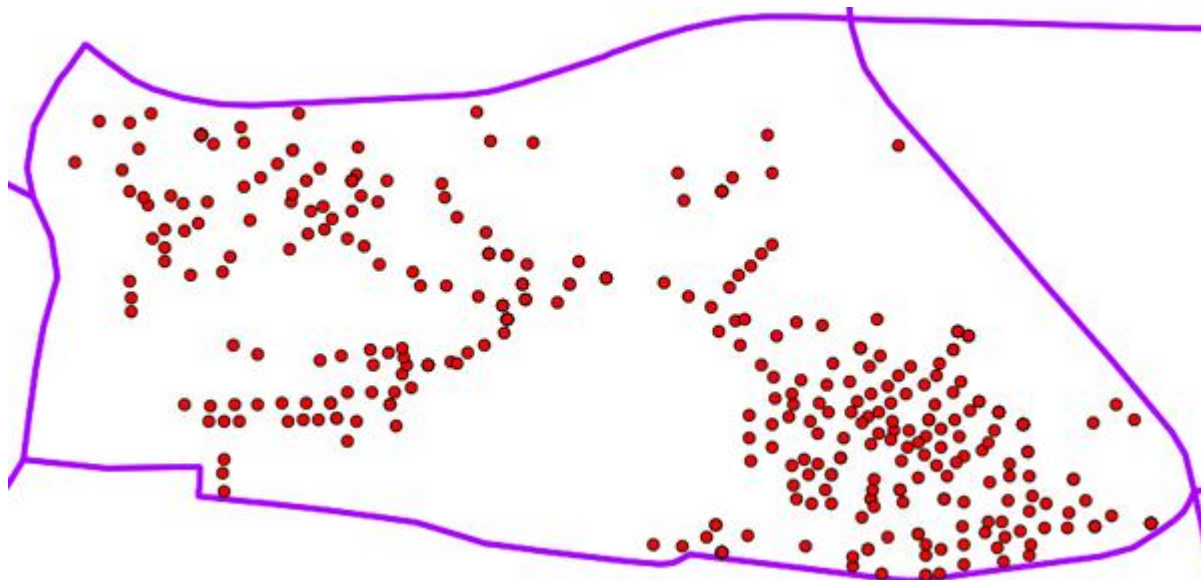
Наши базы адресов и АД готовы, займемся геообработкой.

Геообработка

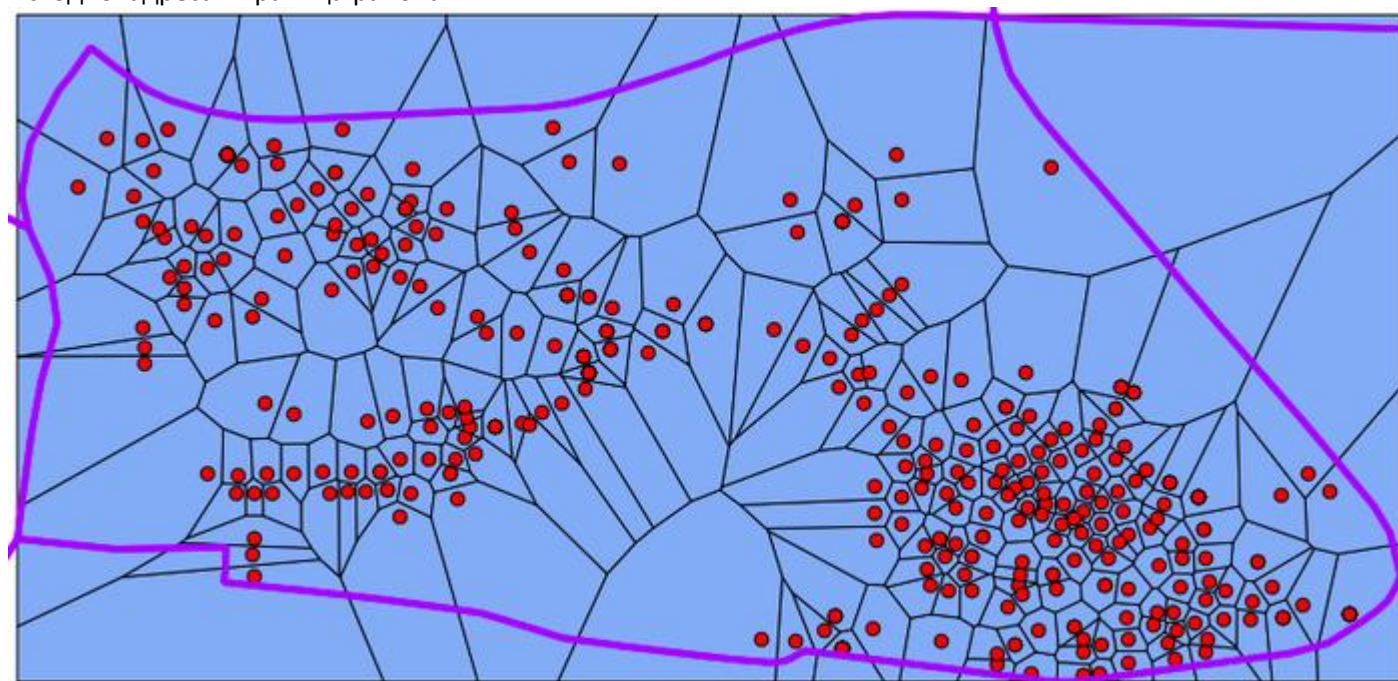
Теперь наша задача следующая, произвести с каждой парой точки (адреса) - полигон (граница района) следующие операции.

1. Рассчитать географический охват полигона
2. Построить полигоны Вороного по точкам в границах рассчитанного охвата
3. Обрезать получившиеся полигоны Вороного по границе района
4. Растворить границы между полигонами с одинаковым номером УИКа.

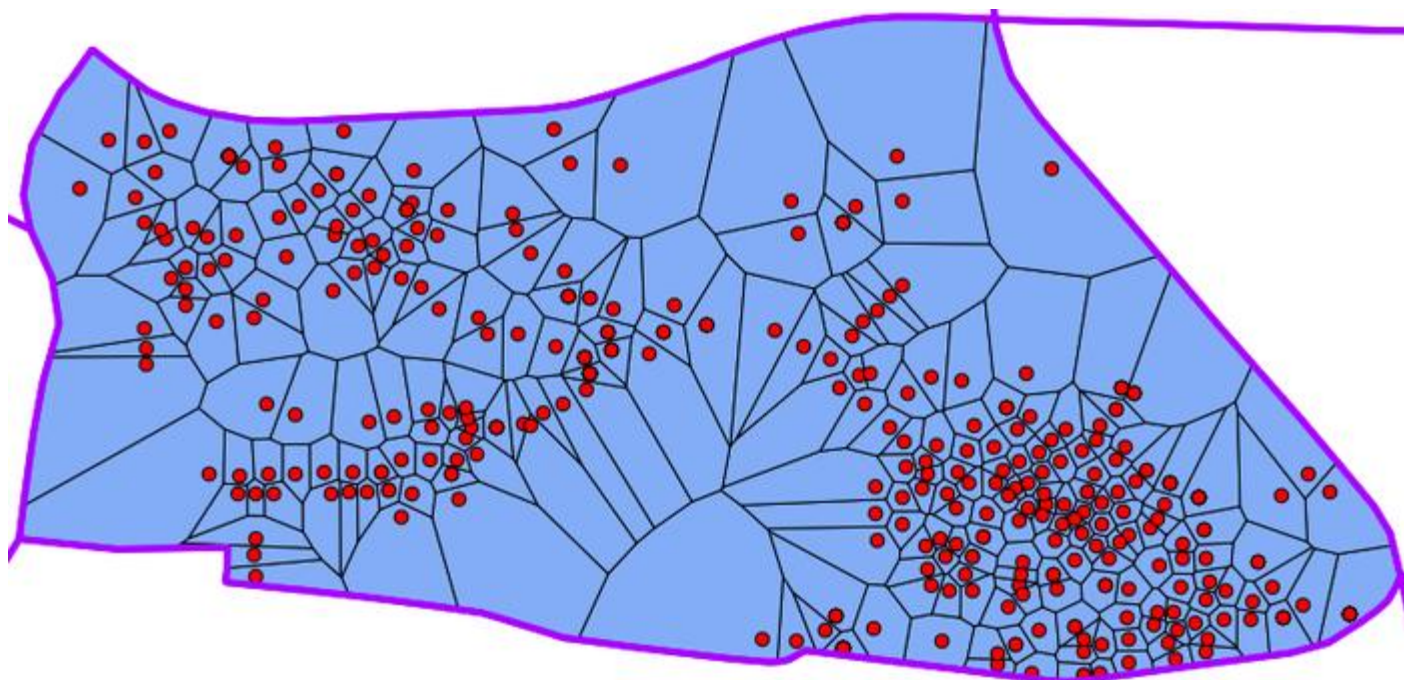
На примере Щукино в картинках:



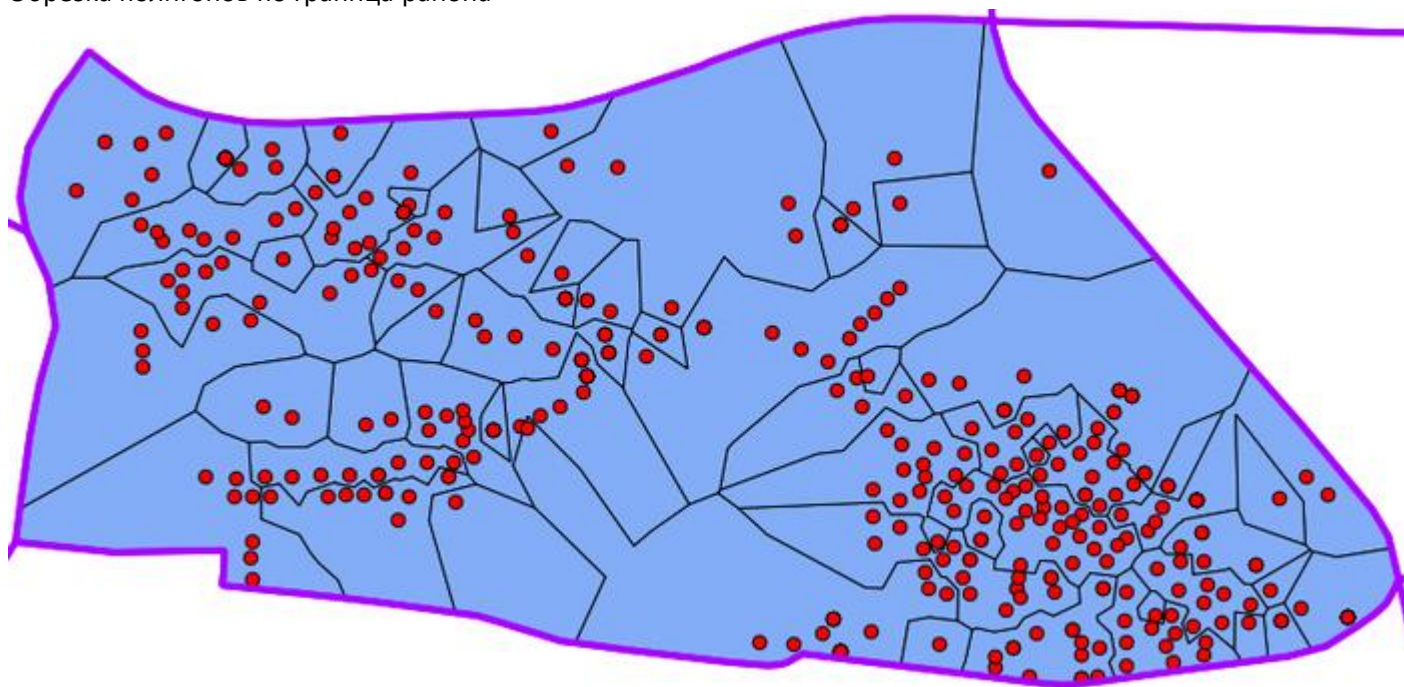
Исходно: адреса и граница района



Полигоны Вороного по адресам в пределах охвата границы района



Обрезка полигонов по граница района

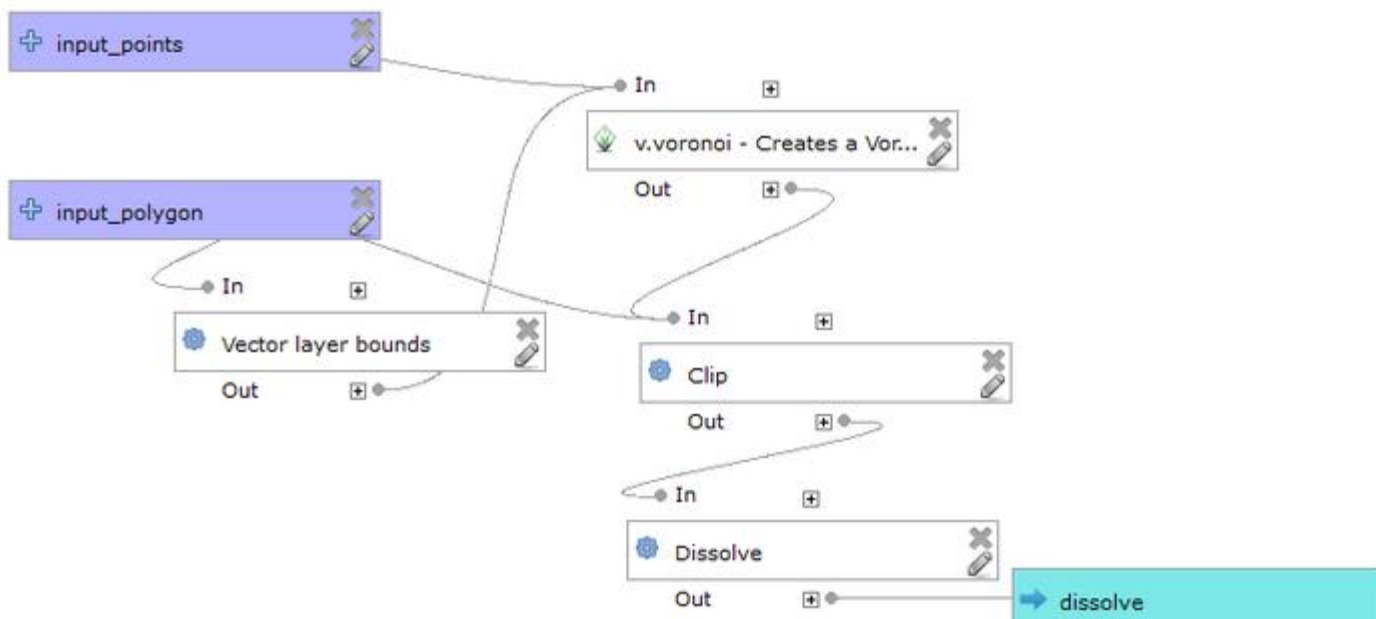


Слияние полигонов относящихся к одному УИКу

Автоматизация

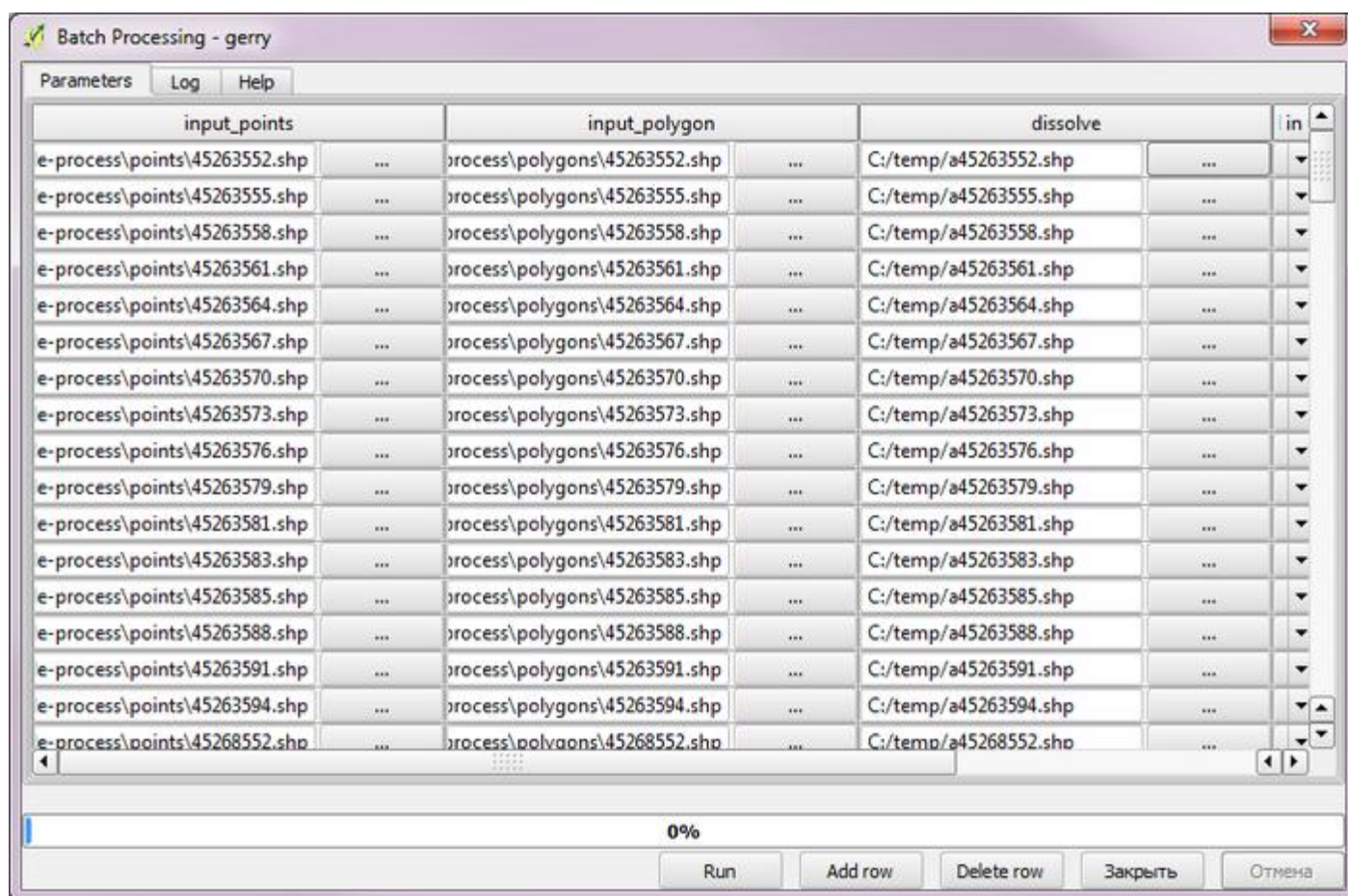
Понятно, что вручную нам это повторять для всех районов не хочется, поэтому будем автоматизировать процесс выше с помощью Sextante. Sextante - это набор модулей для геообработки и среда для удобного объединения этих модулей в последовательности (модели). Он доступен из коробки если у вас одна из последних версий QGIS.

Модель будет выглядеть следующим образом:



Модель геообработки Sextante

Запускать модель будет в пакетном режиме и добавим ей в качестве аргументов все наши точки и полигоны.



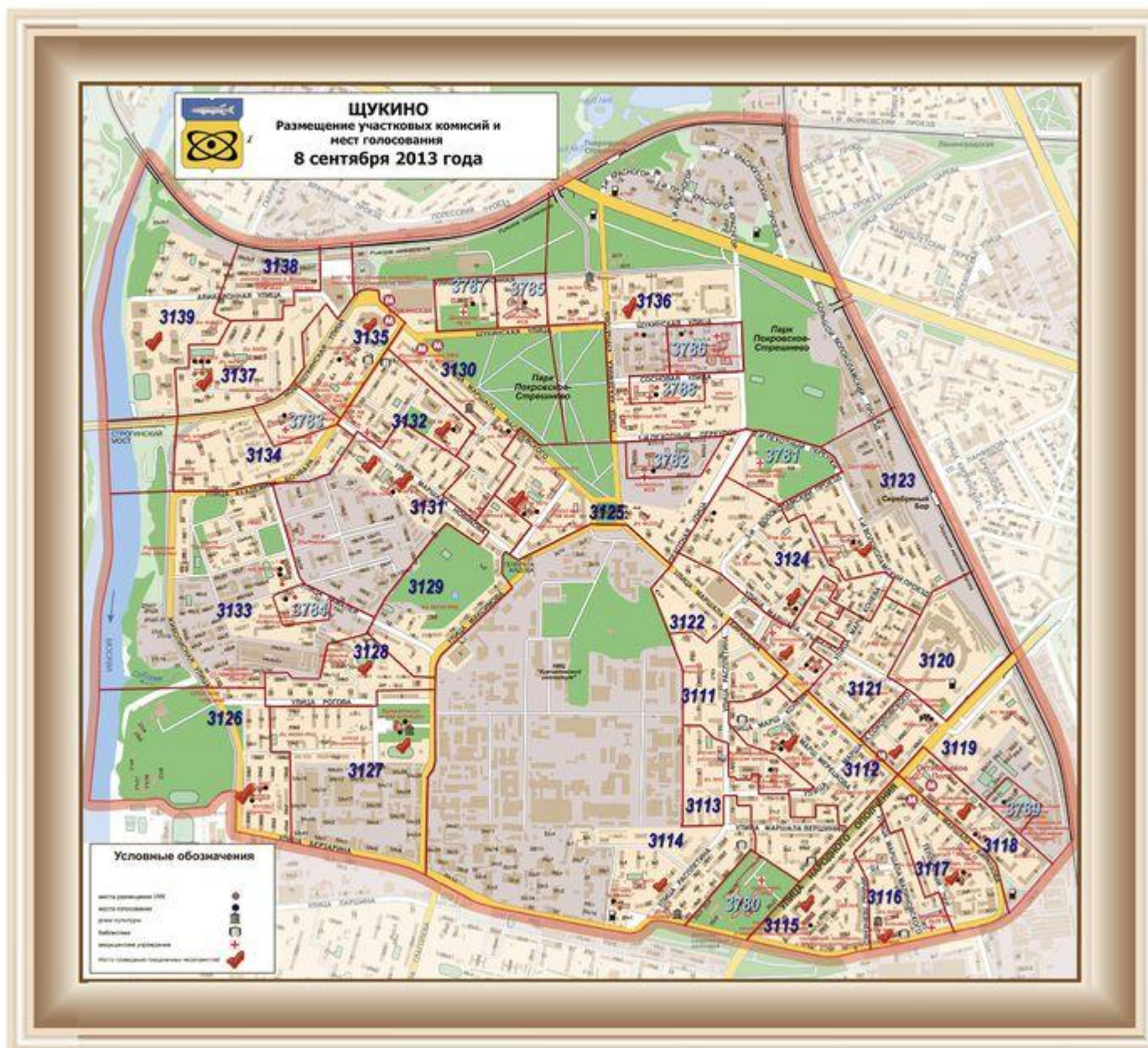
Параметры модели

После выполнения, останется только сделать Вектор/Управление данными/Объединение shape-файлов, чтобы получить единый слой границ участков по г. Москва.

[Скачать готовую модель](#) (Sextante model)

Результат

Границы участков нарисованные управой выглядят следующим образом:



Размещение участковых комиссий и мест голосования [с сайта управы](#)

[Скачать границы расчетных участков](#) (ESRI Shape)

Почему получилось плохо

Результат довольно плохо соответствует подобным картам, которые рисуются вручную и на то есть несколько причин.

1. Это практически полностью автоматический алгоритм не ведающий красивых границ, но зато он математически точен.
2. Качество исходных данных - ниже среднего, потому что геокодирована примерно половина точек. 33% вины на OpenStreetMap - нет геоданных, 33% вины на геокодере - не нашел адреса, хотя они есть, 33% вины на МГИК - плохо структурированные и неправильные адреса, масса отсутствующих домов и адресов которые не приписаны к УИКу. Точные вклады каждого из компонентов в результат оценить сложно, это примерная оценка.
3. Не учтены промзоны, леса, парки, вода.

Очевидно, что улучшить результат можно засчет:

- улучшения карты - нужно продолжать вносить данные в OSM, а попросту - рисовать карту

- улучшения геокодера, есть много ситуаций когда он мог бы сработать лучше
- убеждения МГИК и улучшения охвата описаниями всех домов, которые в участки почему-то не попали.

[Обсудить в форуме](#) Комментариев — 8

Последнее обновление: 2014-05-15 01:54

Дата создания: 04.09.2013

Автор(ы): [Максим Дубинин](#)