

TAGMAP.XYZ

STACKOVERFLOW MAP

Михаил Кольцов

Руководитель: Аркадий Калакуцкий

ЗАДАЧА

~50.000 тегов

~16 миллионов вопросов

$\text{similarity}(\text{tag1}, \text{tag2})$ = количество вопросов, которые
имеют оба тега

ВЗВЕШЕННЫЙ ГРАФ

ЗАДАЧА

- Понять, как применить t -SNE к графам
- Рисовать граф тегов с помощью t -SNE за ≤ 24 часа
- Показывать теги в веб-интерфейсе

T-SNE

- Алгоритм понижения размерности до 2 или 3
- Применяется к точкам в d -мерном пространстве
- Показывает хорошие результаты на практике

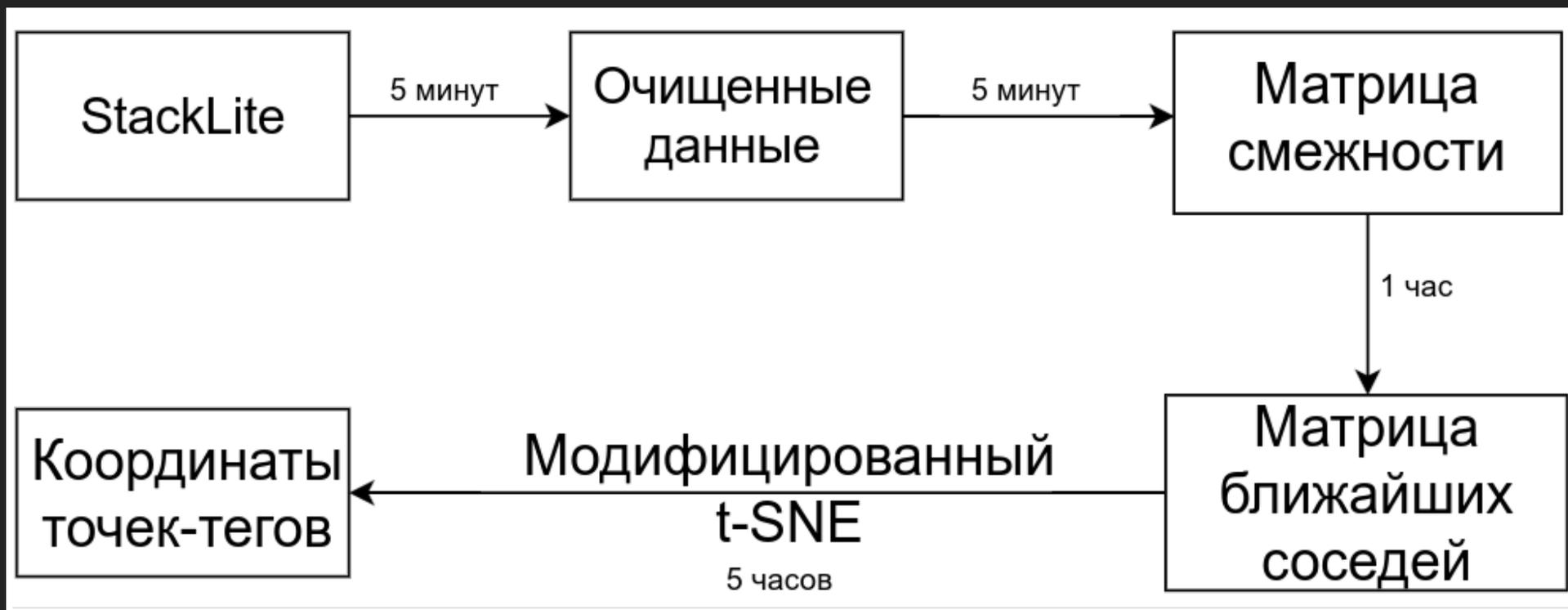
ДААННЫЕ

- .torrent-файлы со снимком базы данных Stack Exchange
- Облегчённая версия на <https://github.com/dgrtwo/StackLite>
- Извлекаем связи между постами и тегами, строим матрицу смежности

T-SNE И ГРАФ

- Вход t-SNE — точки
- От каждой точки берутся только ближайшие соседи
- Давайте запустим алгоритм Дейкстры для каждой вершины графа и сами найдём ближайших соседей
- Модифицировали код
<https://github.com/lvdmaaten/bhtsne>

ОБРАБОТКА ДАННЫХ



Имеем координаты всех точек — нарисуем легко?

НЕТ

КАРТА

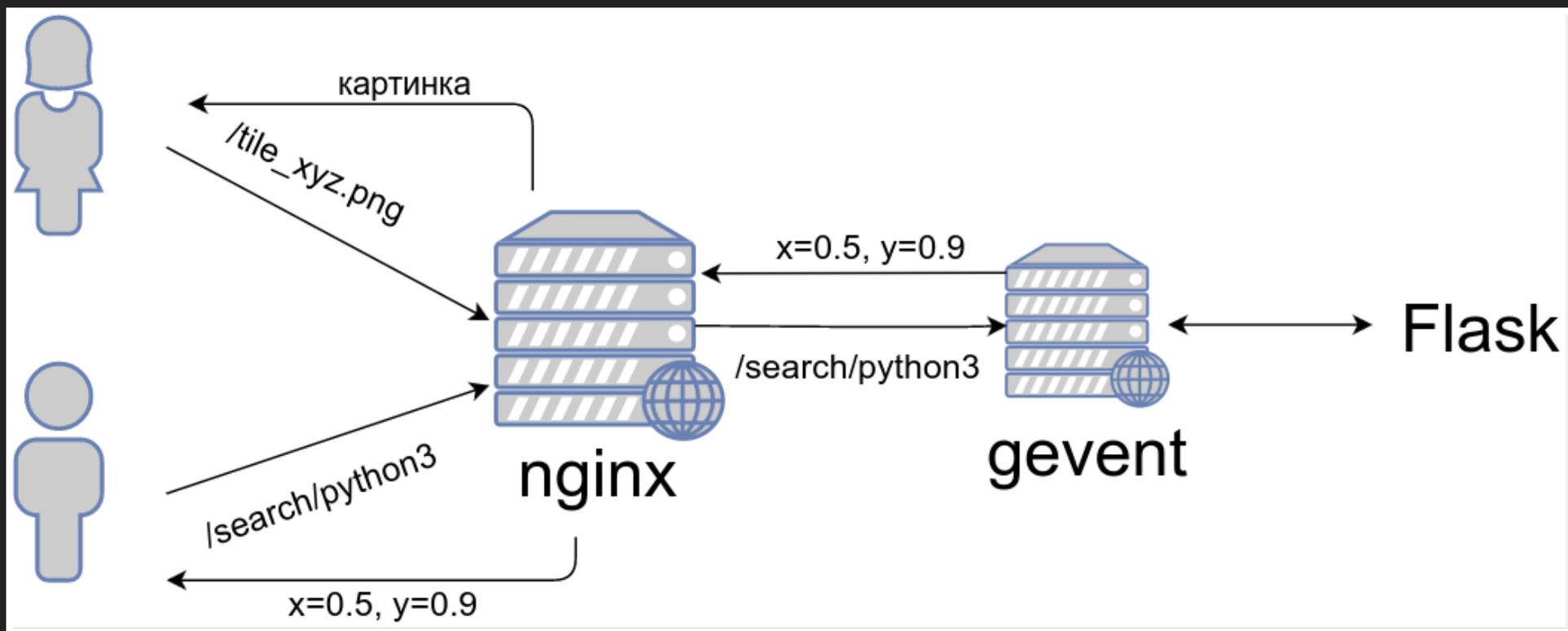
- Посмотрели в Open Street Map устройство настоящей карты
- Картинка делится на тайлы
- На i -м уровне увеличения $2^i \times 2^i$ тайлов
- Тайлы объединяем в группы
- Предподсчитываем всё до седьмого уровня увеличения

БЫСТРОДЕЙСТВИЕ

- Тайлы хочется показывать быстро
- Не только локально, но и во всемирную сеть
- Ещё и интерактив: поиск, ...

Сервер на Flask, стоит за nginx

АРХИТЕКТУРА СЕРВЕРА



ЧЕМУ НАУЧИЛСЯ

- Научился структурировать работу на GitHub
- Зауважал математику визуализаций
- Узнал о возможностях алгоритмов визуализации
- Познал базовые принципы показа карт, проблемы масштабирования на много клиентов
- Настроил сервер, который доступен в Интернете
- ...

Спасибо за внимание!