

Рекомендации навыков с помощью Transformers и графовой БД Neo4j



Артём Шарганов (Rabota.ru)



Кто я и откуда

Rabota.ru

Разработка моделей матчинга и рекомендаций. MLOps.

SAP

Консалтинг в сфере инновационных проектов

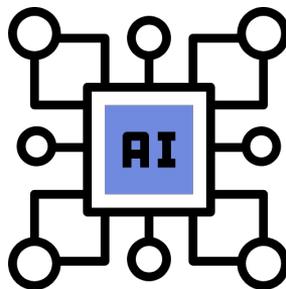


<https://www.linkedin.com/in/artem-sharganov/>

Эволюция Работа.ру



рост числа белых
воротничков



поиск стал сложнее
(появился МЛ)



критерии качества
заполнения стали
выше

Что важно для работодателей?

91%

**Опыт
(года)**

89%

Навыки

84%

**Специальный
опыт**

Иногда о человеке лучше говорят его навыки

Ищем: Инженер в конструкторский отдел

ID	Позиция	Город	Релевантность	...
1	Инженер	Москва	0.9934	
2	Инженер	Москва	0.9875	

Кадастровый инженер

Проектная документация

Проверка технических планов

Знание структуры XML-СХЕМЫ

ФЗ 953, ФЗ 218, Приказ 861

Постановка, снятие с кадастрового учета

Инженер-конструктор

Проектирование металлоконструкций

технология производства

MathCAD

AutoCAD

Умение работать в дедлайне

коммуникабельность

Сервисы, в которых нужны навыки



Зарплаты



Матчинг/скоринг

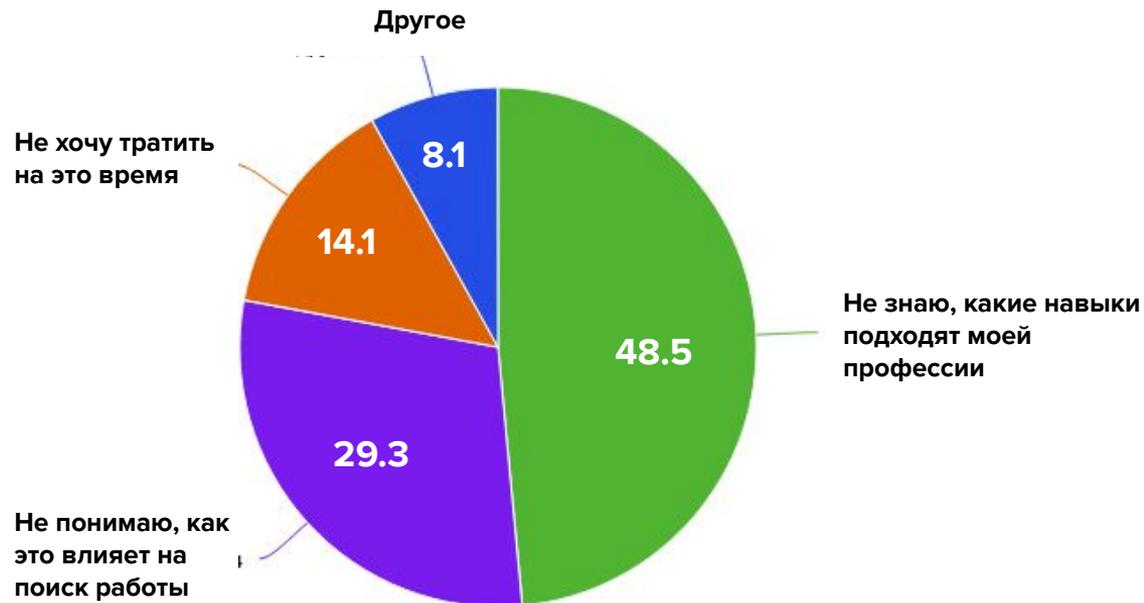


Рекомендации



Аналитика

Почему навыки не заполняют?



Как работает сейчас



Шарганов Артём

Machine Learning Engineer

Опыт работы

Сен 2021 –

Настоящее время

8 месяцев

Работа.ру, Санкт-Петербург [Оставить отзыв](#)

Старший ML-инженер

Дополнительные сведения

Профессиональные навыки и знания

pandas numpy Faust NLP AirFlow Pytorch sklearn Kubernetes Kafka

Natural Language Processing

Профессиональные навыки и знания

Например: менеджмент

CMS vBulletin

показы объектов

марочные отчеты

база TER

ведение протоколов

ЧТО

БУДЕМ

ДЕЛАТЬ

Рекомендация навыков

- Решили сосредоточиться только на предсказании **в момент создания или редактирования резюме**
- Подборка **должна меняться** в зависимости от выбранных **навыков**

Профессиональные навыки и знания

Рекомендовано Вам

Машинное обучение

Computer Vision

TensorFlow

Например: менеджмент

CMS vBulletin

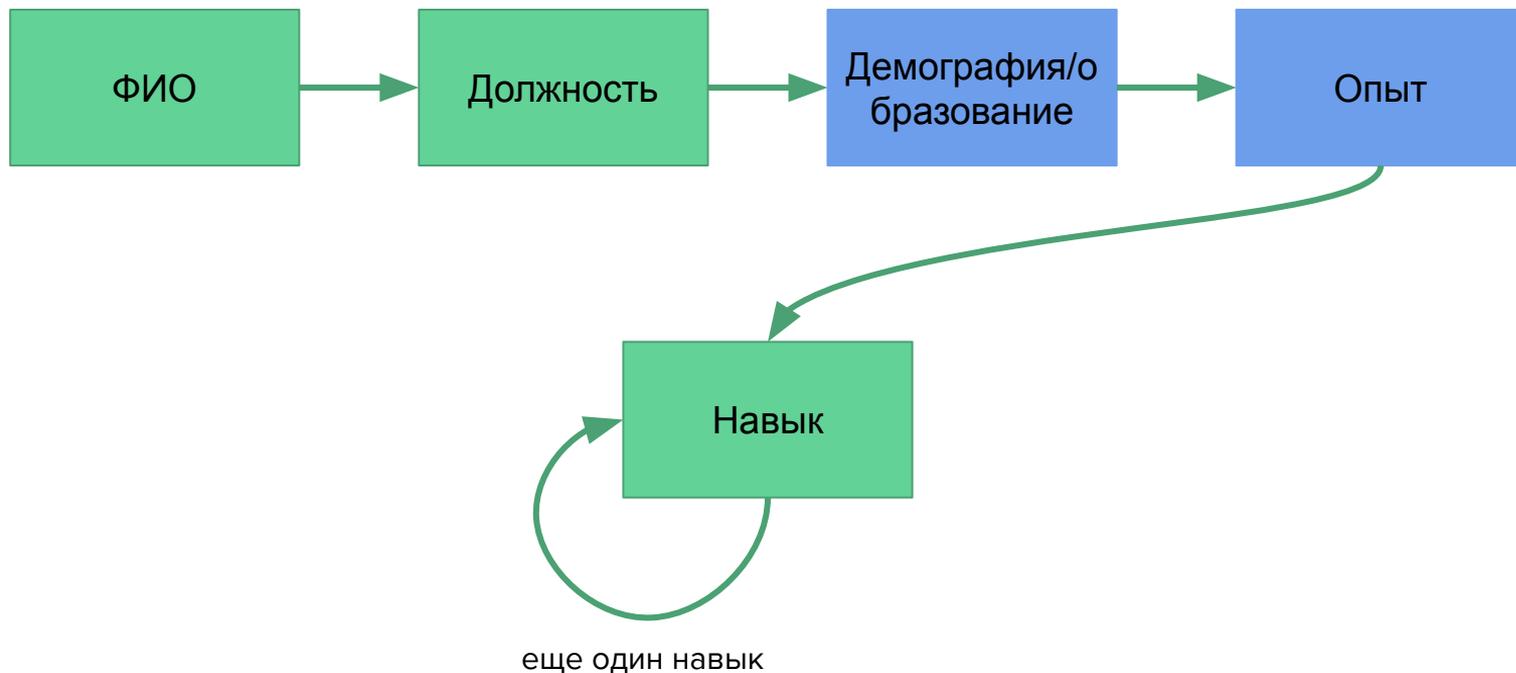
показы объектов

марочные отчеты

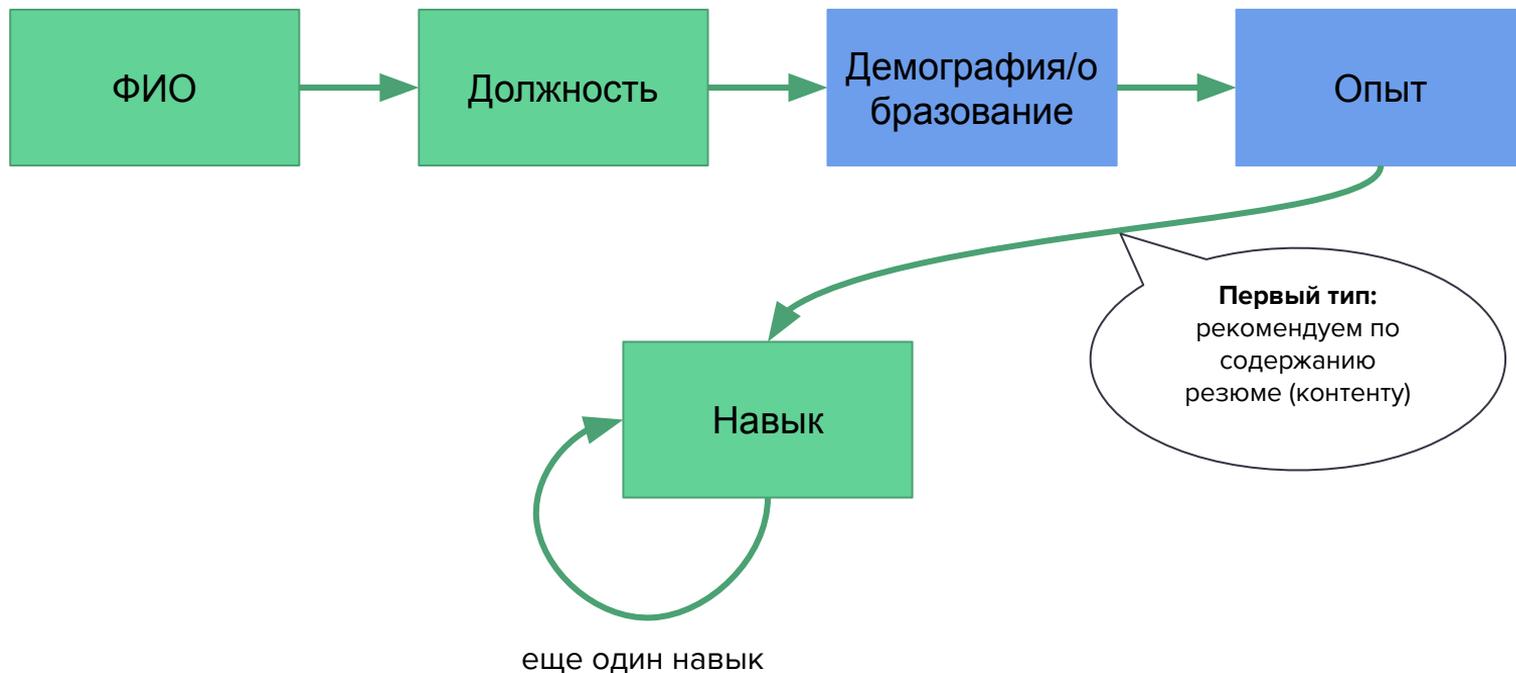
база ТЕР

ведение протоколов

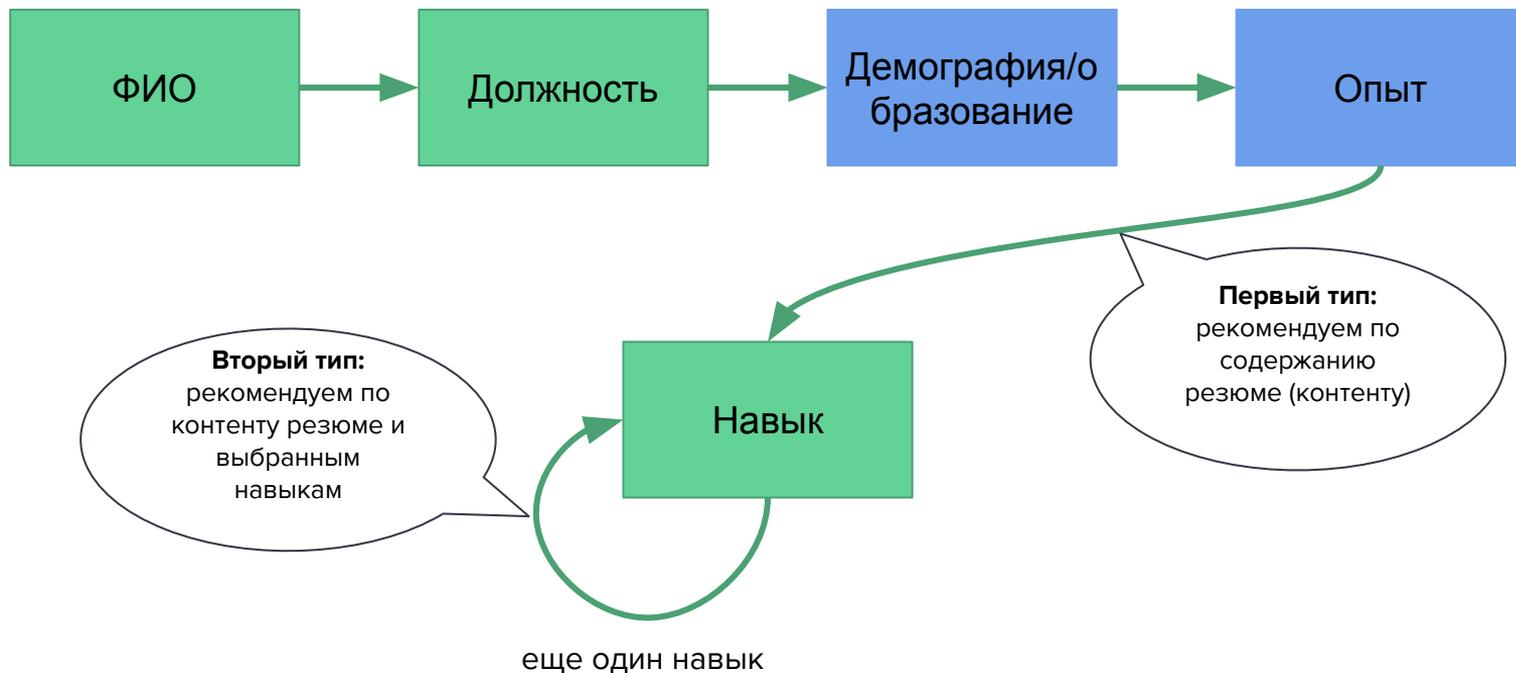
Как выглядит создание резюме



Как выглядит создание резюме

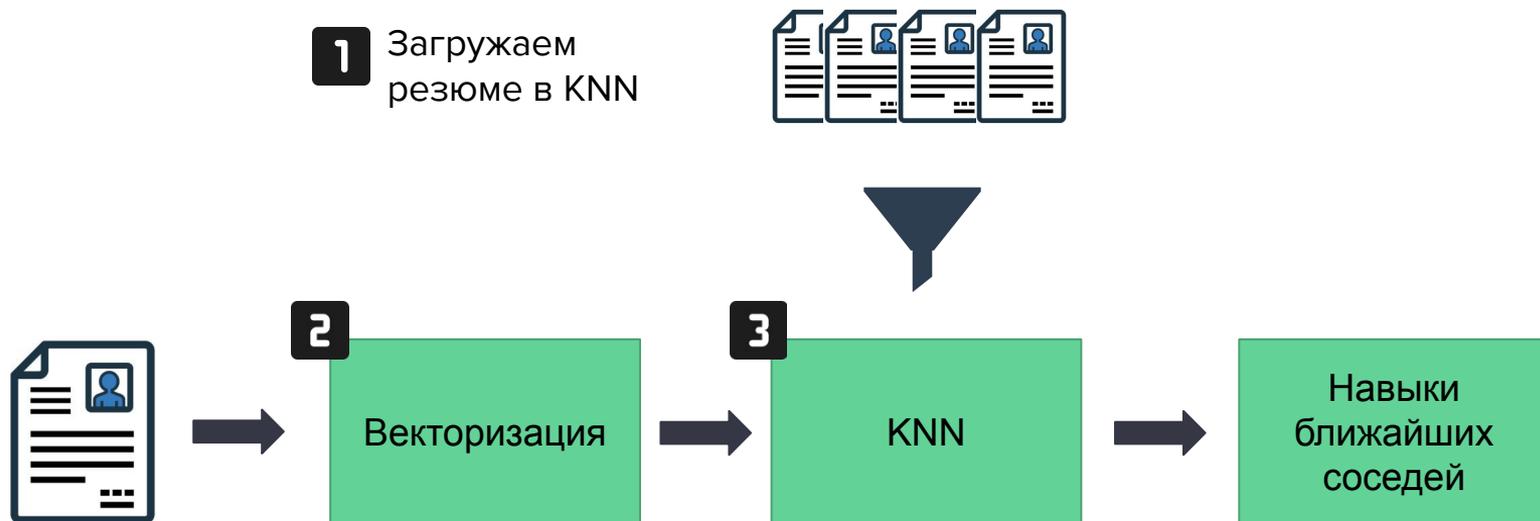


Как выглядит создание резюме



СИСТЕМА РЕКОМЕНДАЦИИ НАВЫКОВ

Первый тип: процесс рекомендаций без навыков

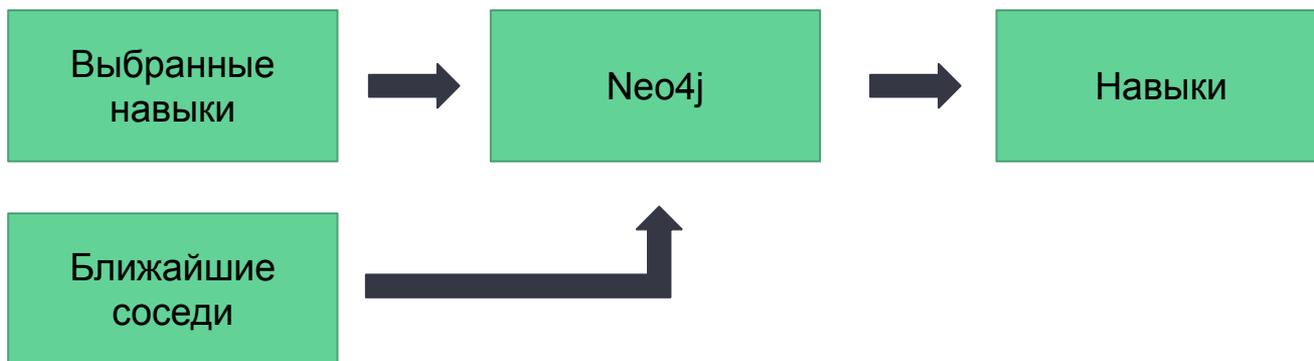


В индексы попадают не все

- фильтры по частотностям
 - берем резюме, у которых больше N навыков
 - берем навыки, которые встречаются больше чем M раз
 - убираем дубликаты навыков
- фильтры по стоп-словам
- модерация на платформе

Второй тип: процесс рекомендации с навыками

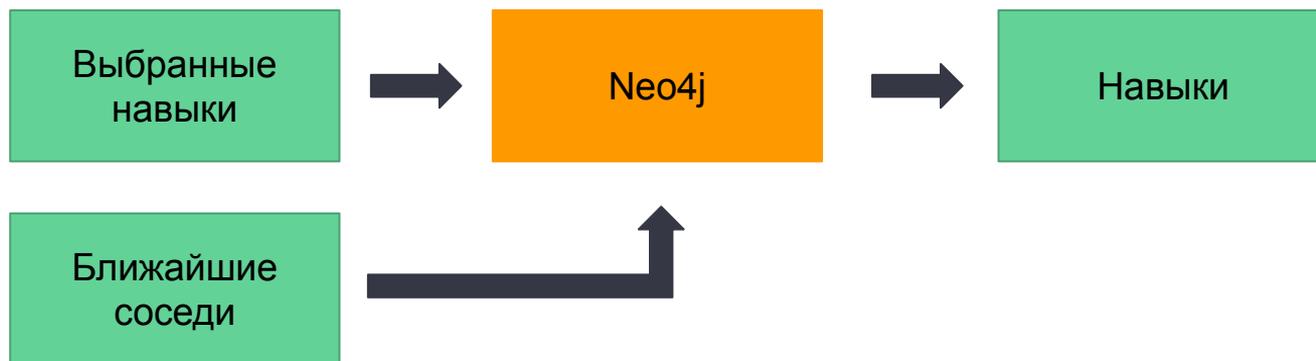
Важно: KNN больше не участвует в основном процессе рекомендации (подбора) навыков



*Предполагаем, что навыки уже существуют

Второй тип: процесс рекомендации с навыками

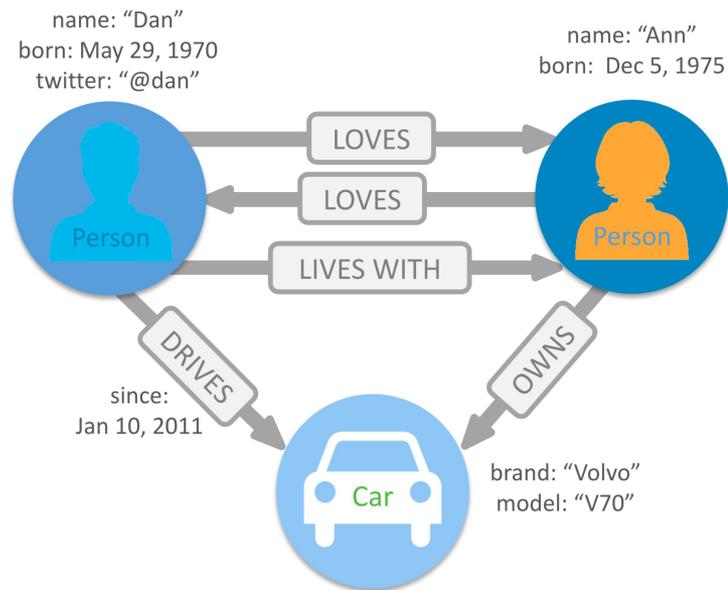
Важно: KNN больше не участвует в основном процессе рекомендации (подбора) навыков



*Предполагаем, что навыки уже существуют

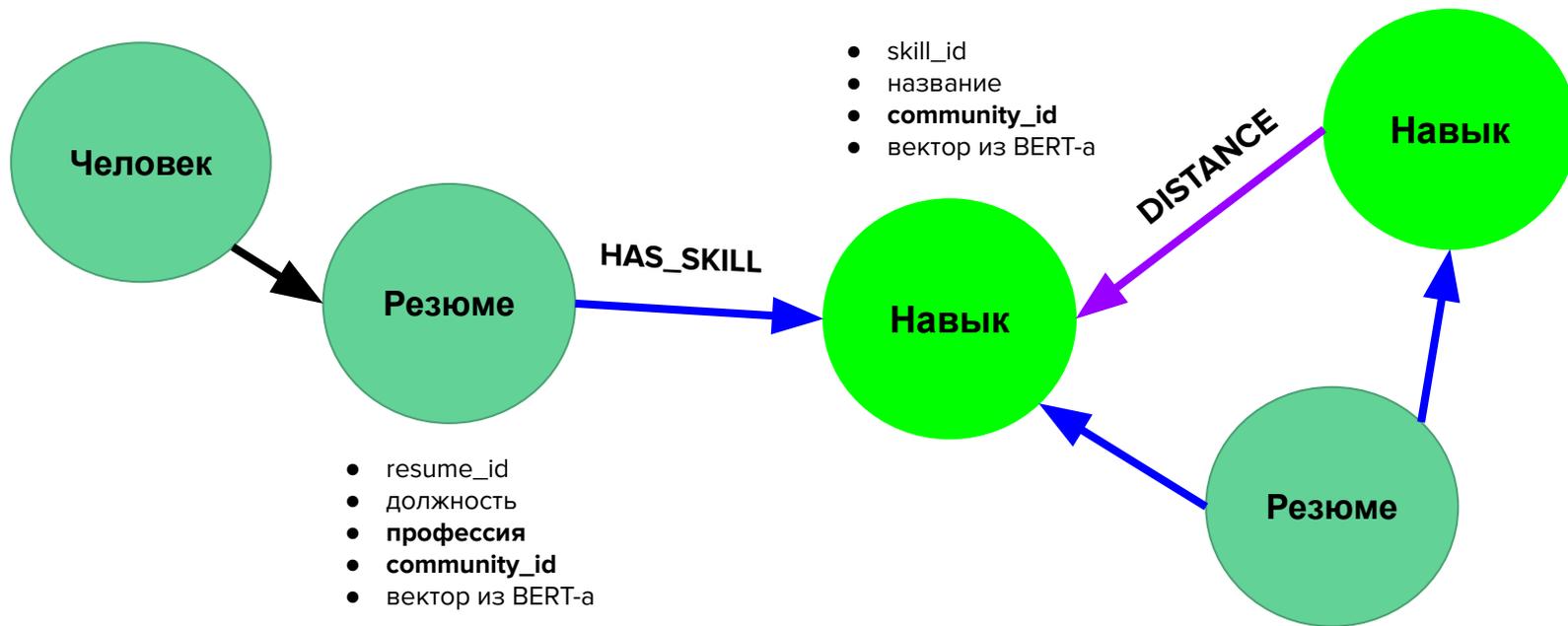
Neo4j - графовая база

- производительная
- иметь свою библиотеку с алгоритмами анализа
- давно уже используется продуктивно у многих компаний

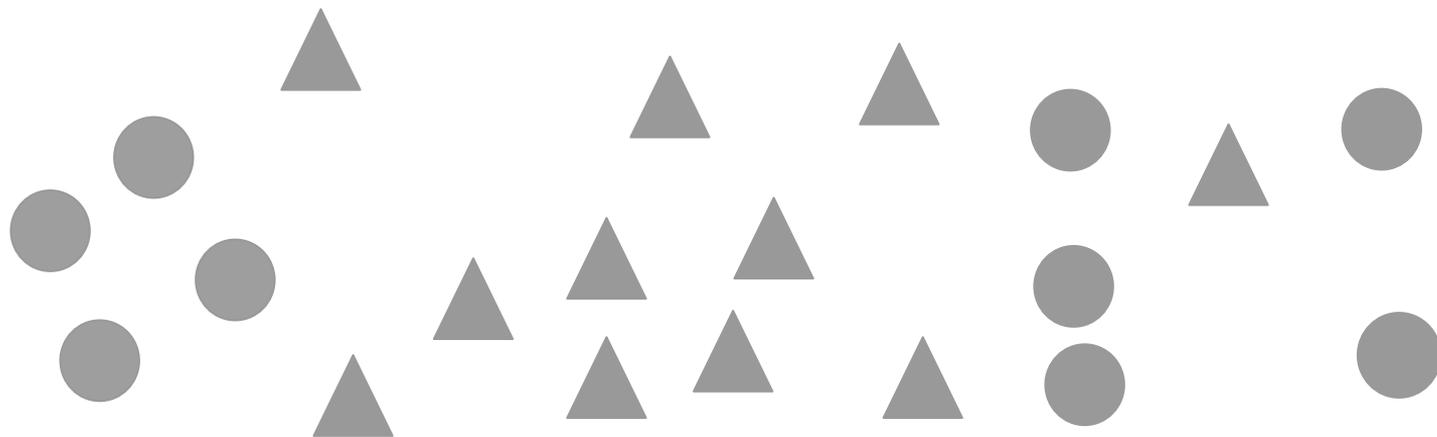


<https://neo4j.com/docs/getting-started/current/graph-database>

Подграф для навыков

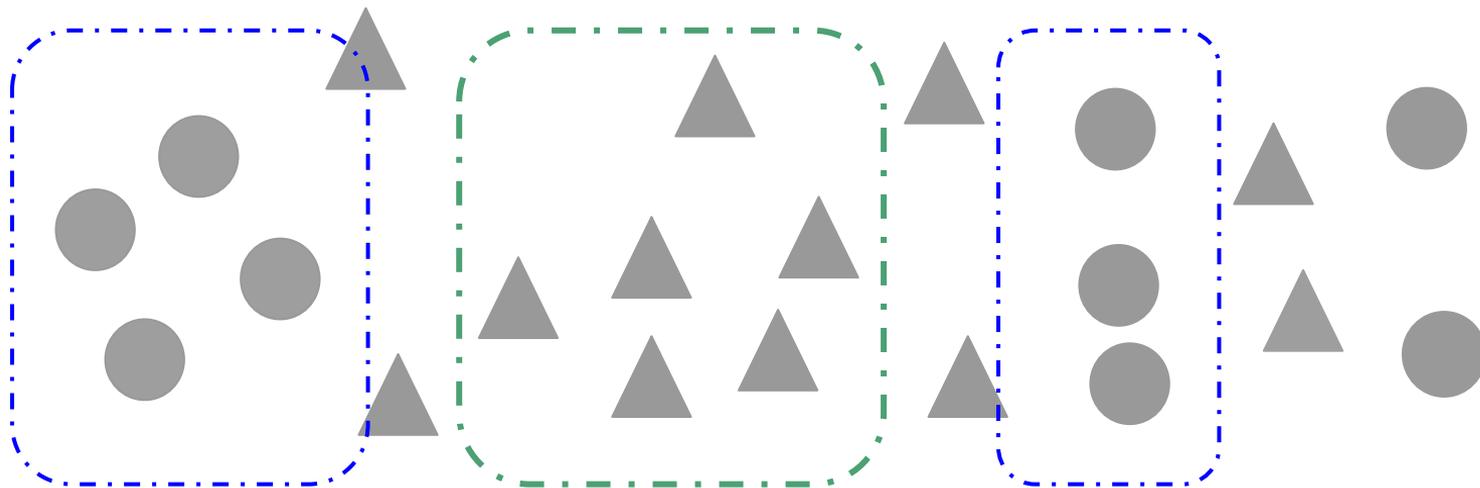


Как это работает



ГДЕ-ТО В ГРАФЕ

Берем соседей из KNN

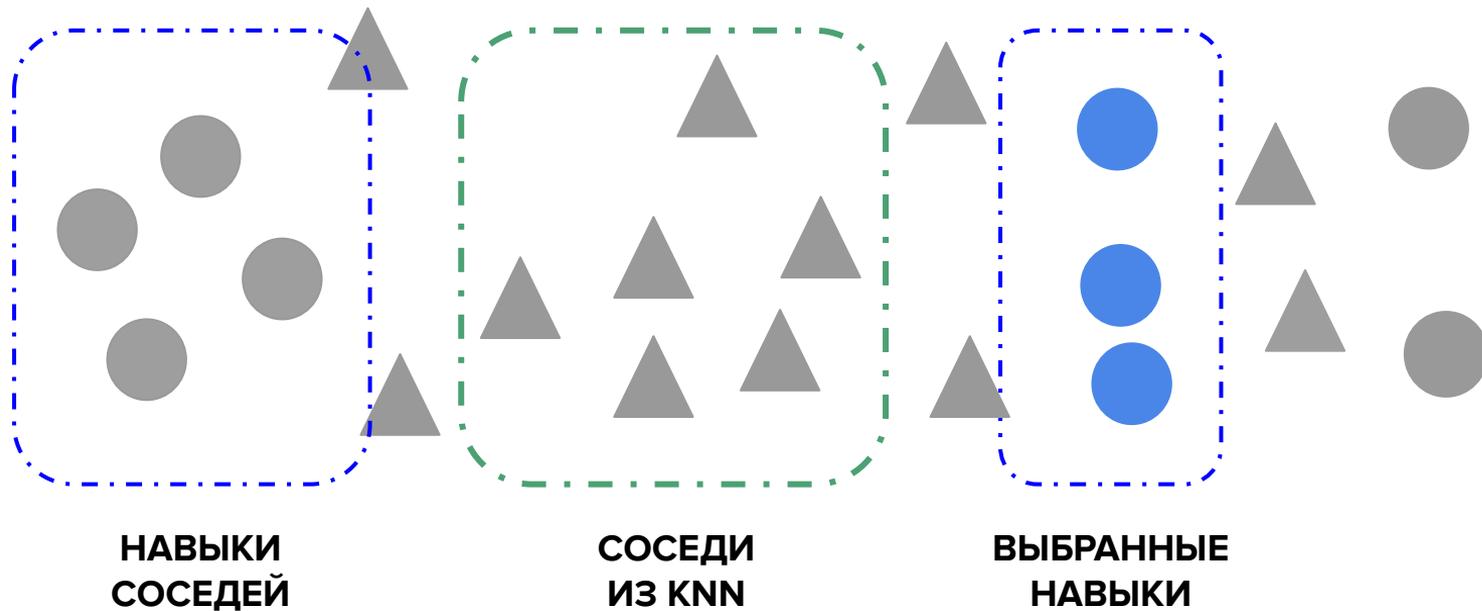


**НАВЫКИ
СОСЕДЕЙ**

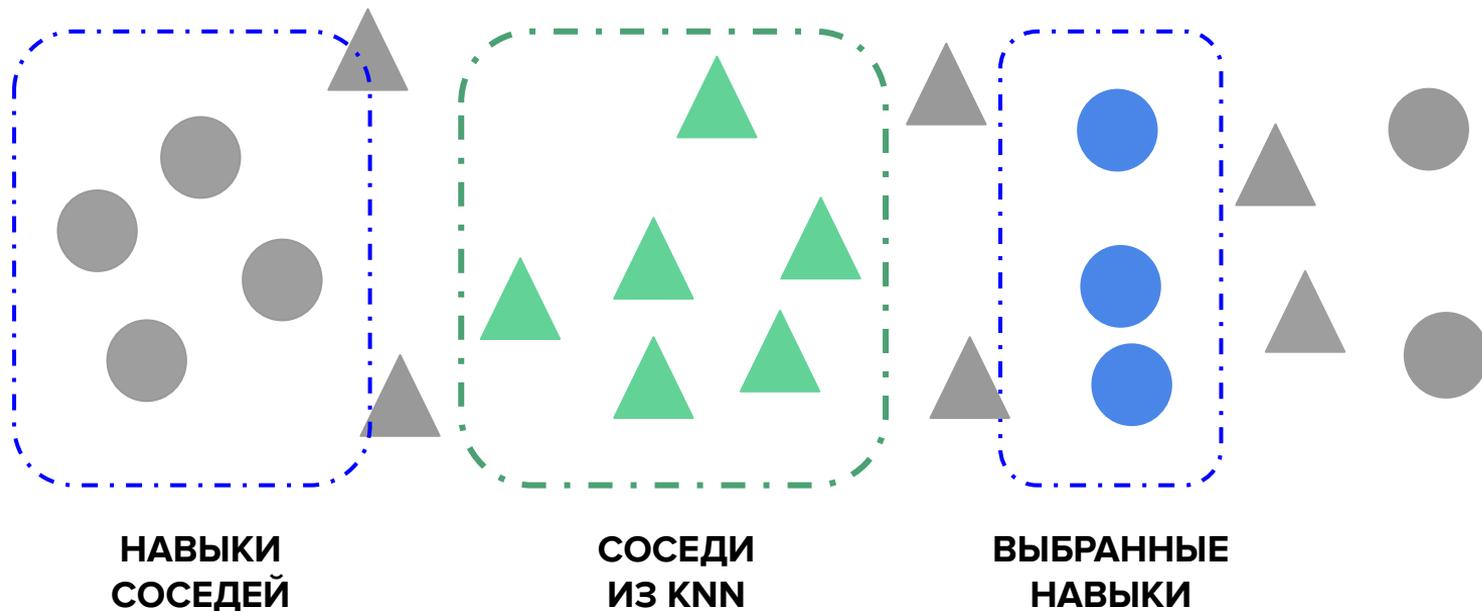
**СОСЕДИ
ИЗ KNN**

**НАВЫКИ
СОСЕДЕЙ**

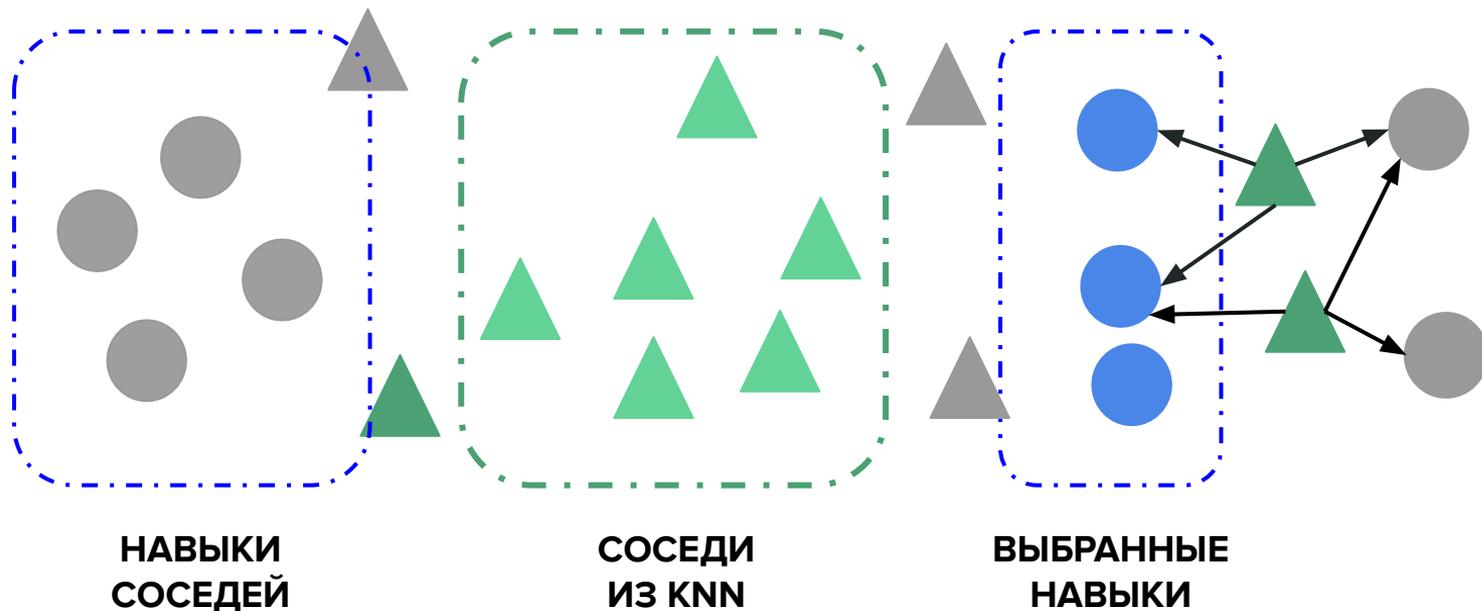
Пользователь выбирает навыки



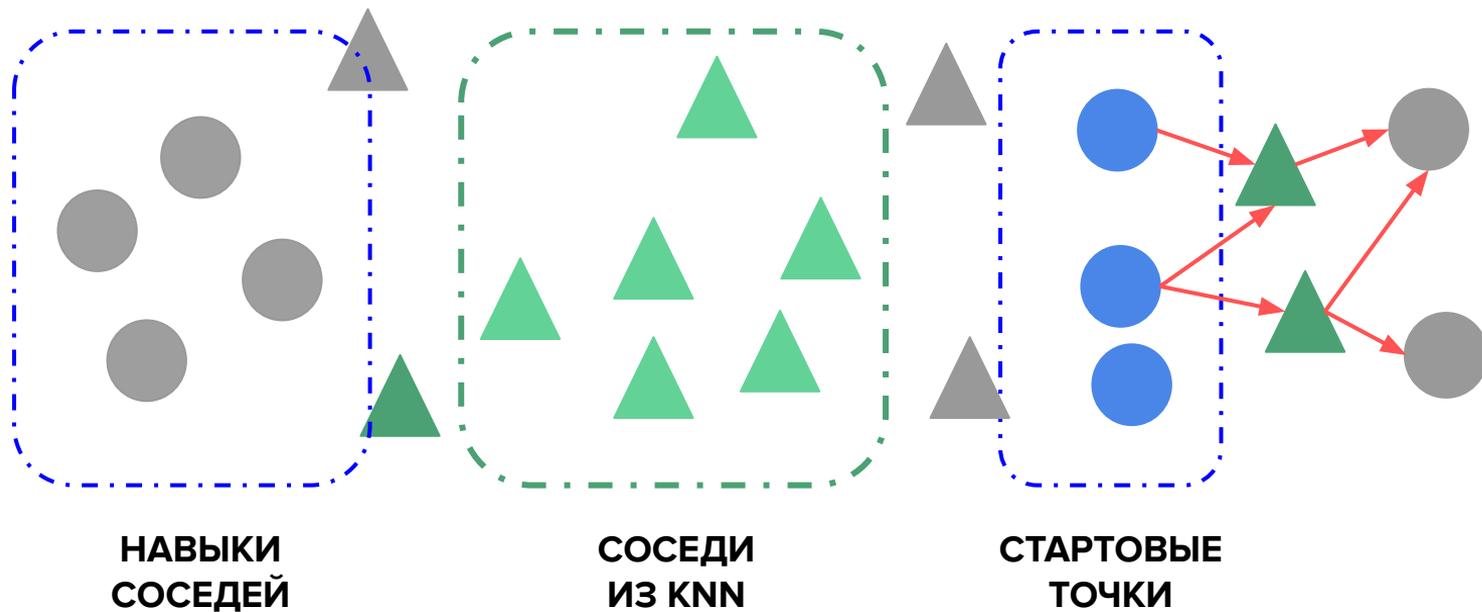
Используем свойства узлов



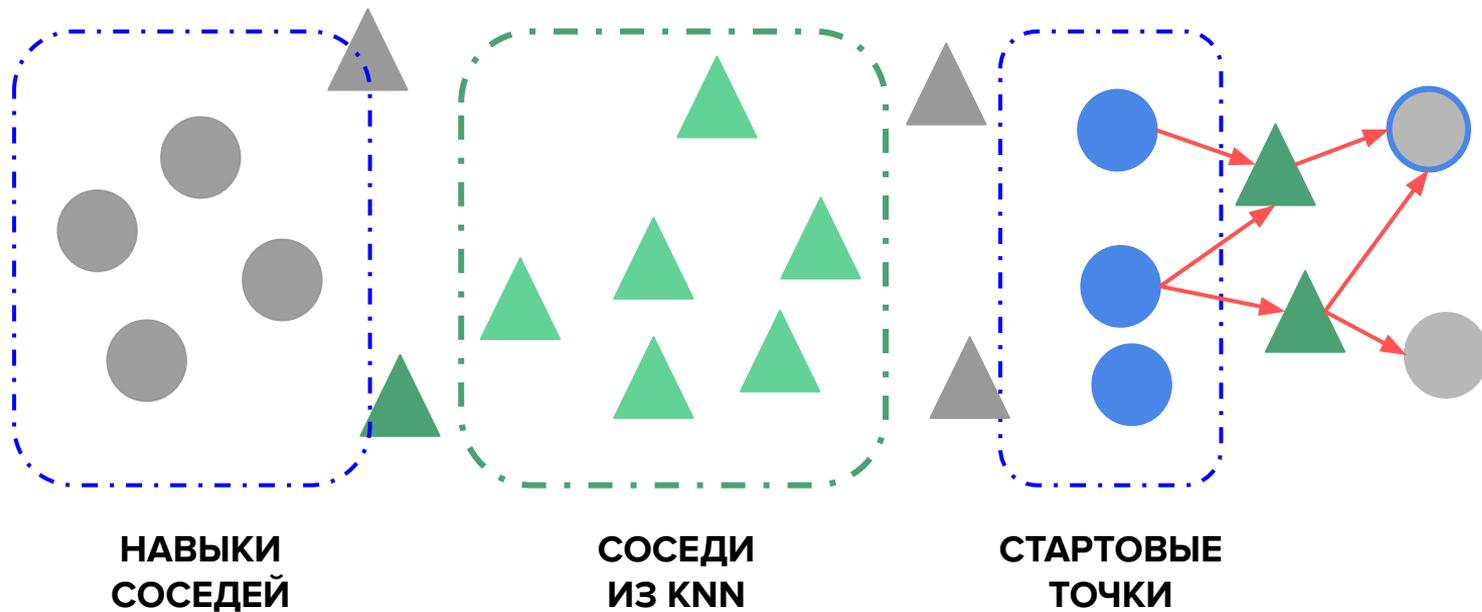
Берем узлы с такими же свойствами



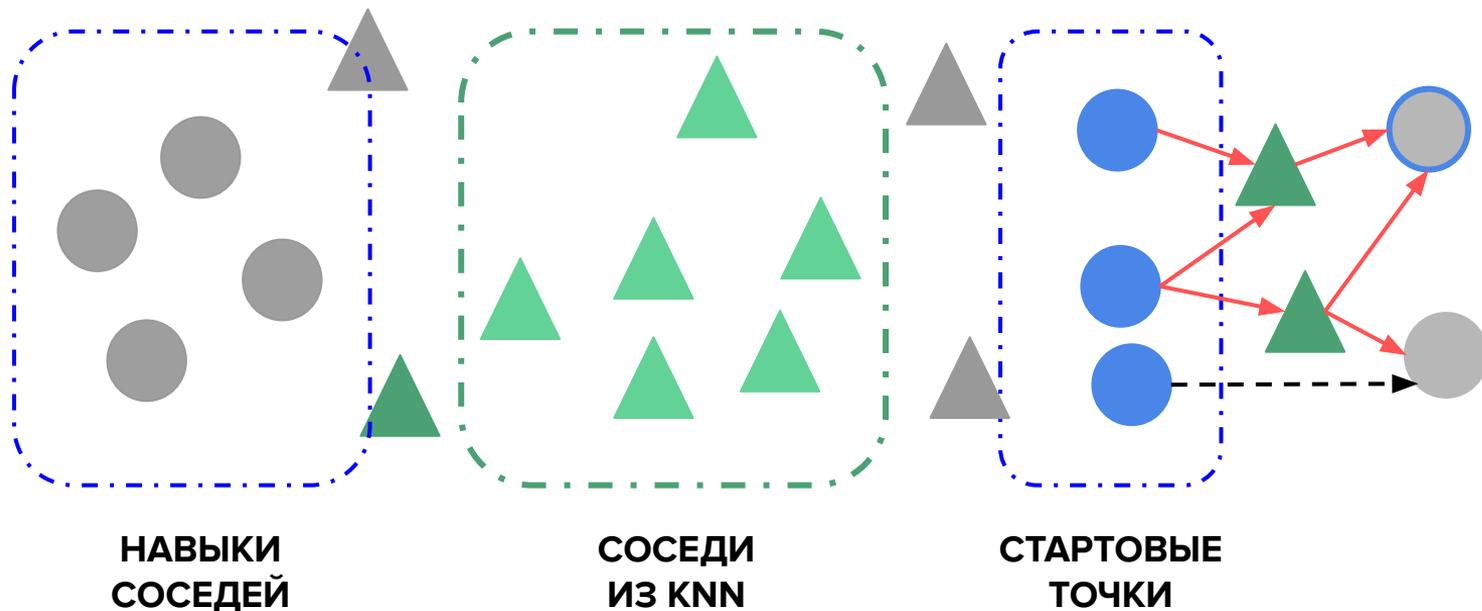
Считаем количество путей



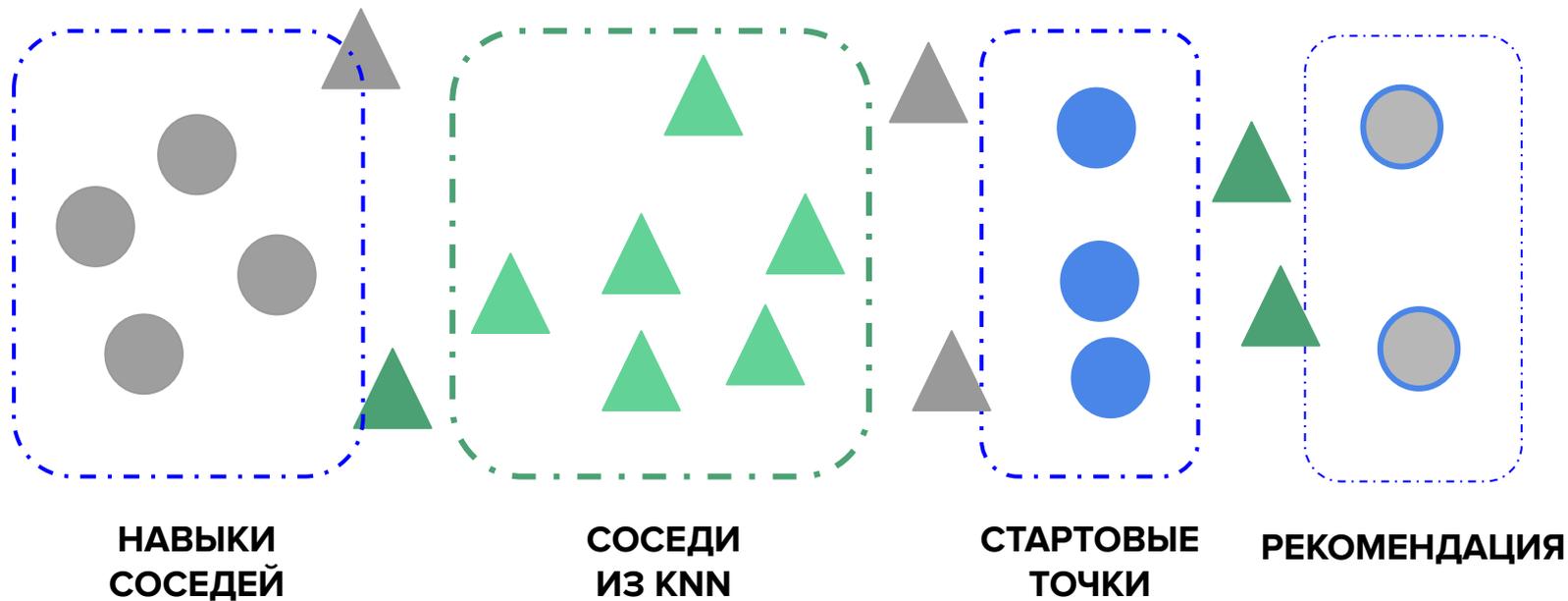
Считаем количество путей



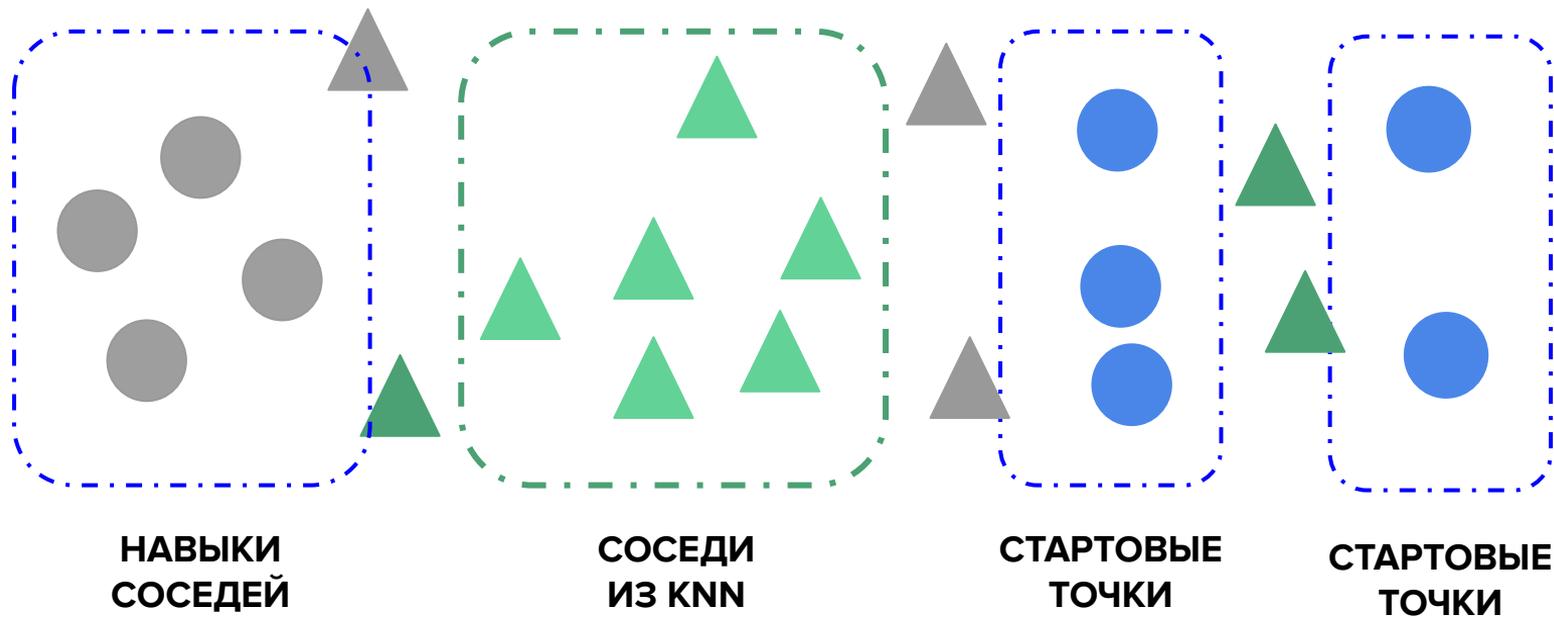
Используем расстояние между скиллами



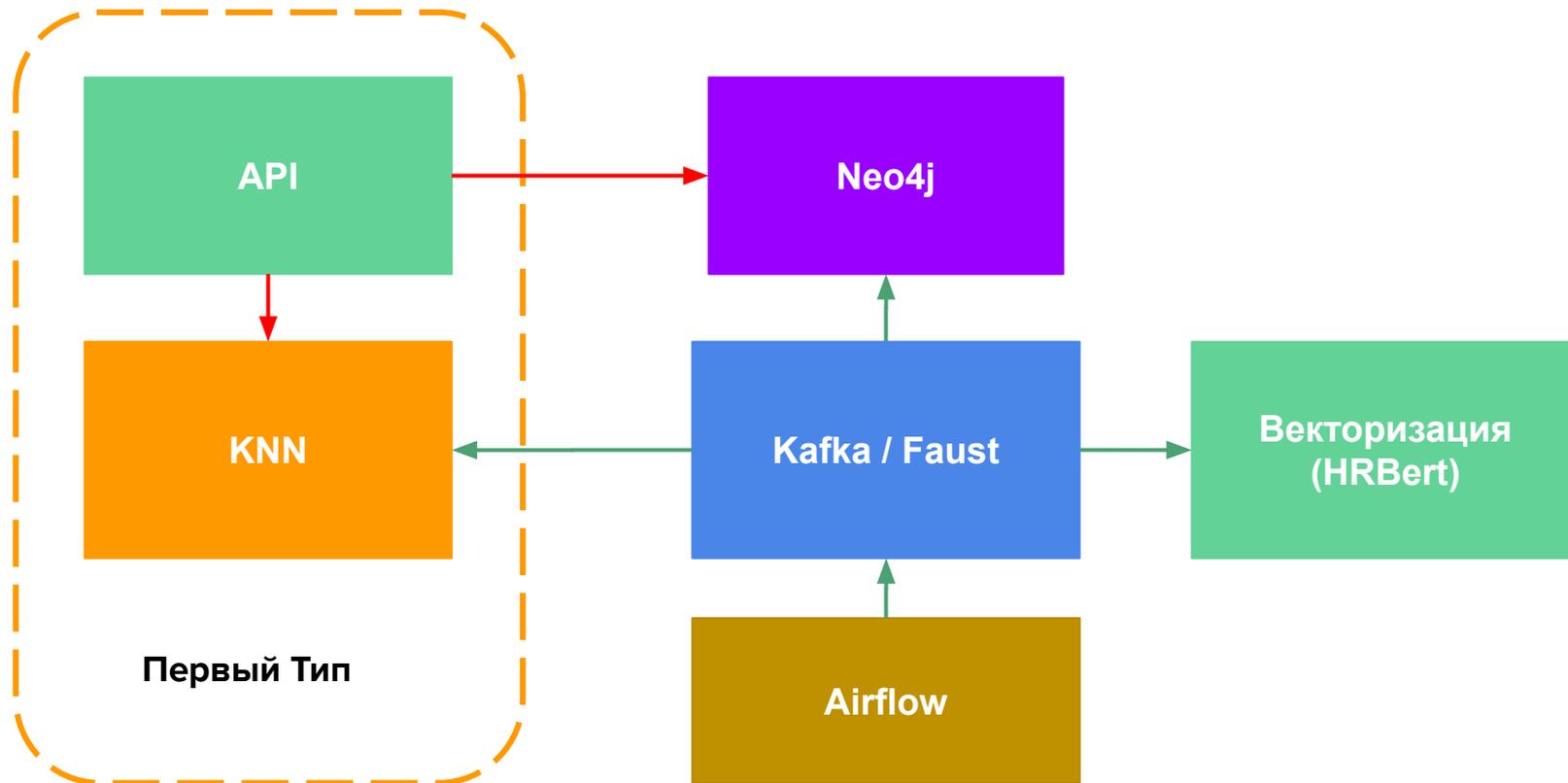
Рекомендации готовы



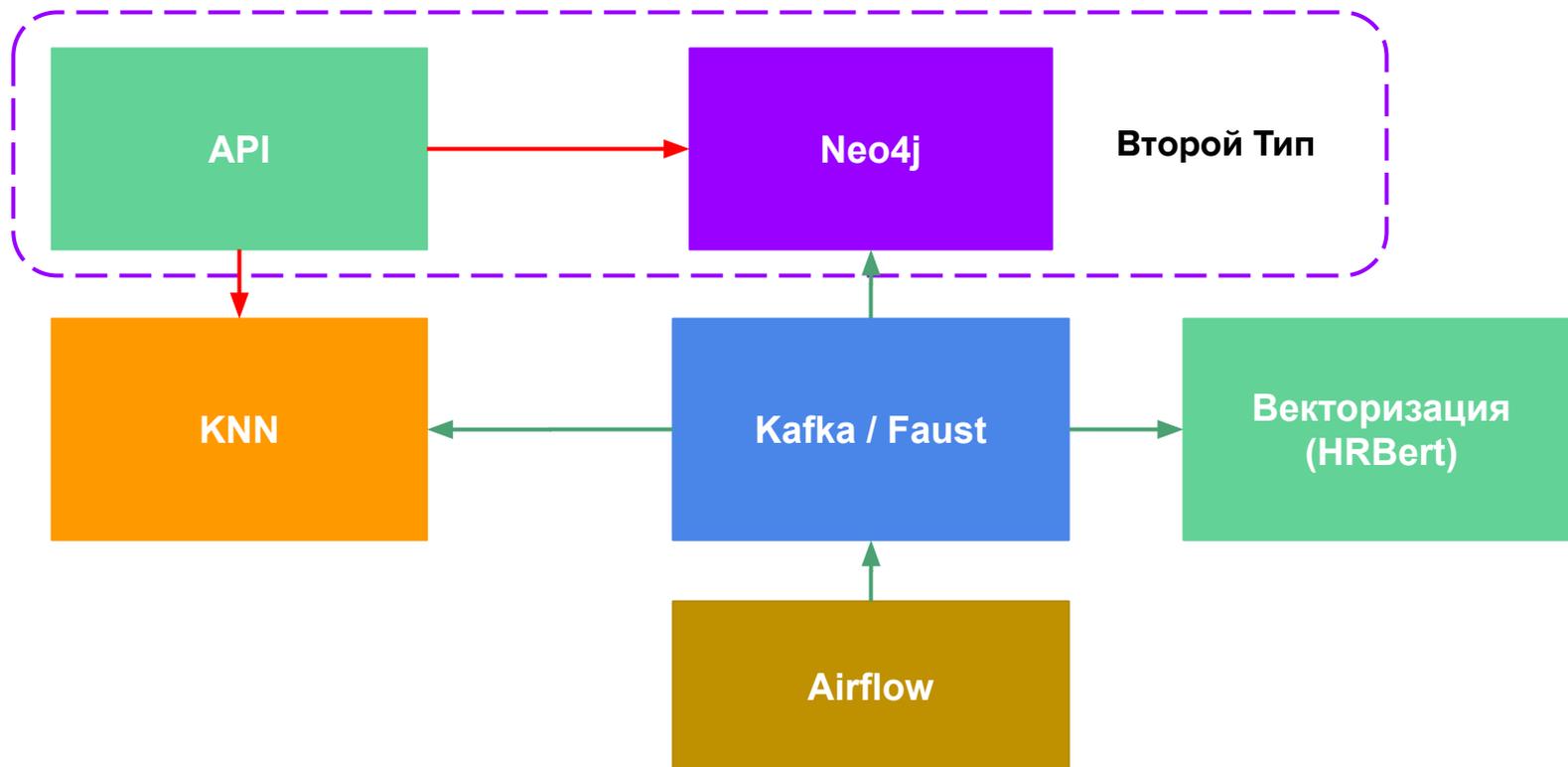
Готовы к следующей итерации



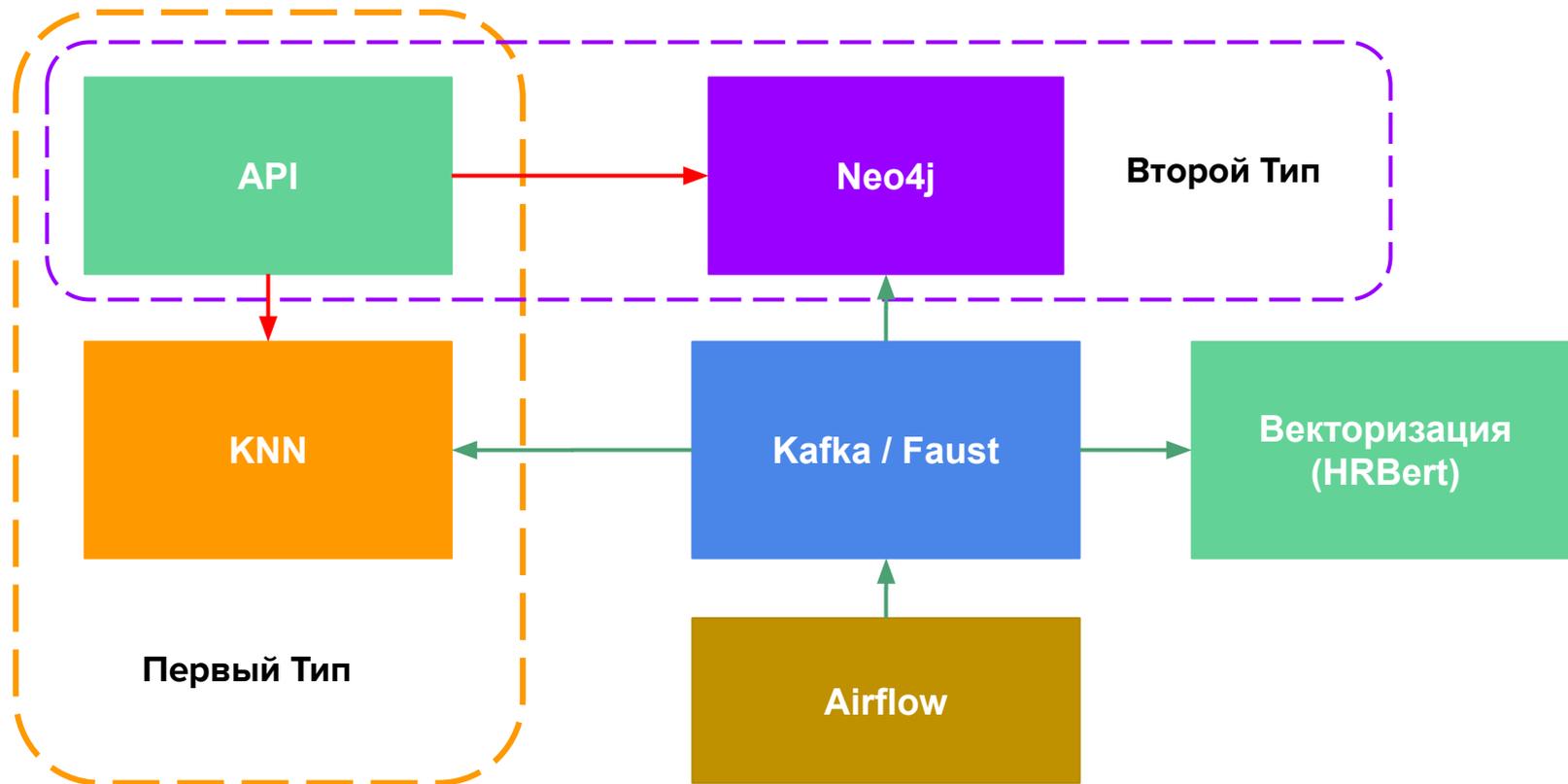
Архитектура



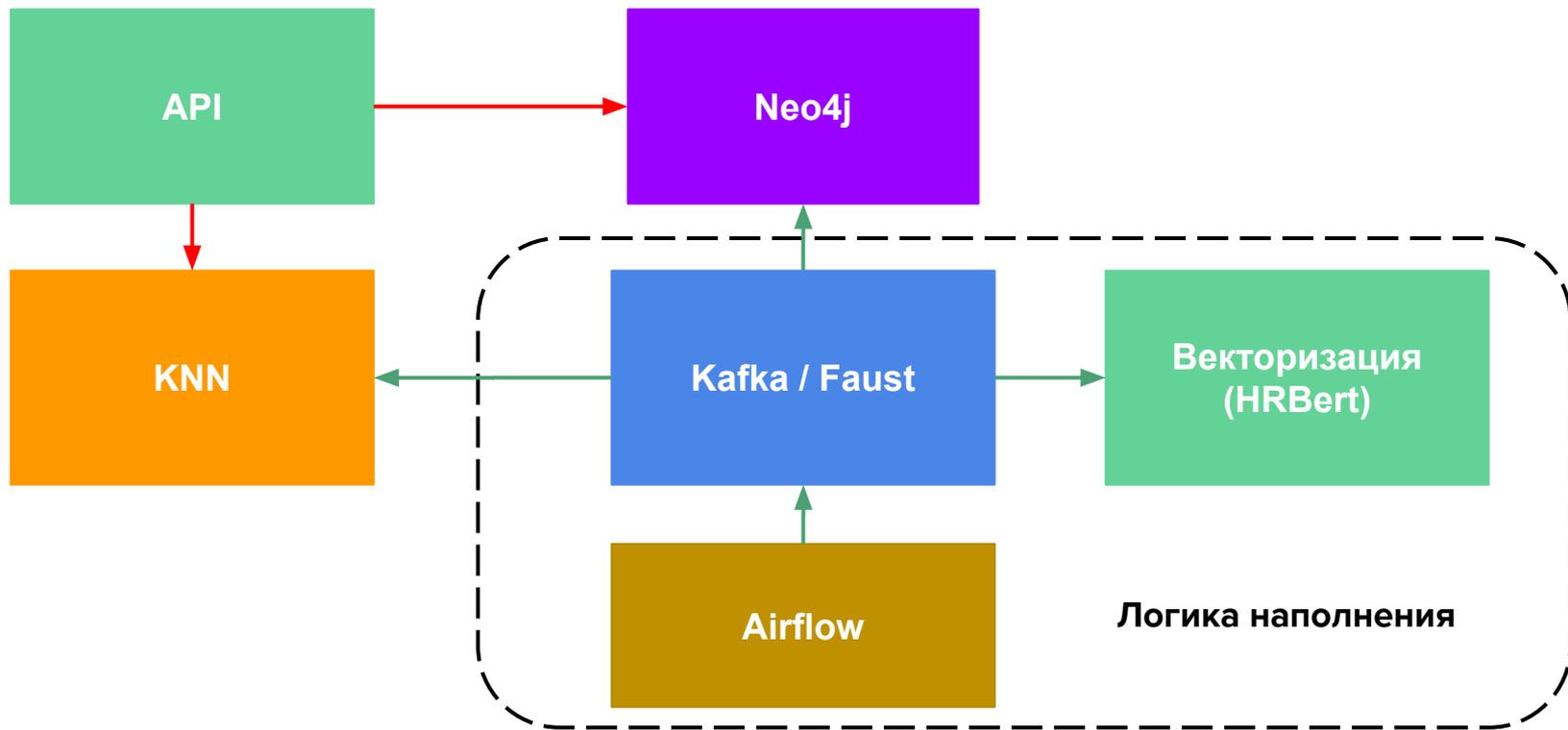
Архитектура



Архитектура



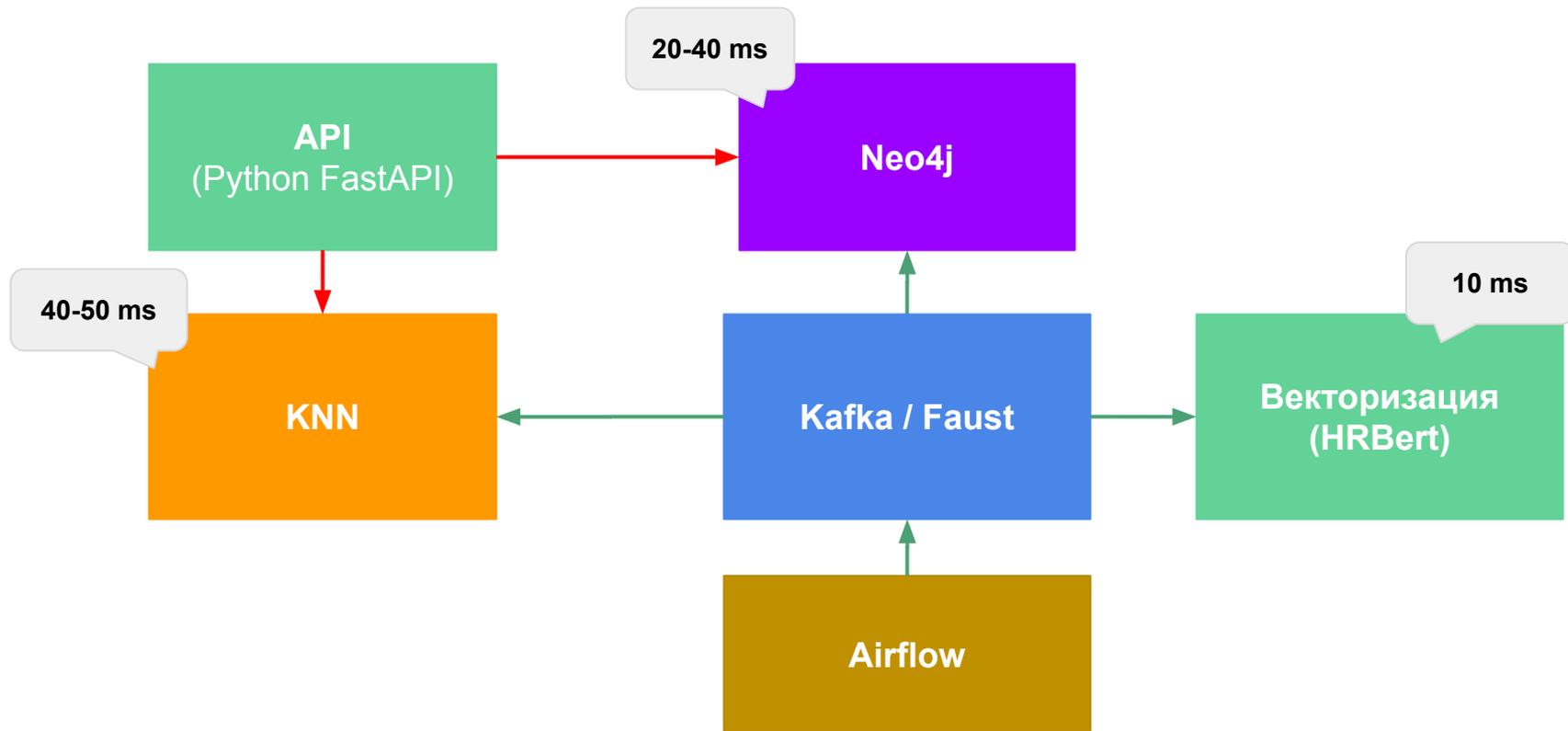
Архитектура



Процесс подготовки графа и KNN

1. Заполнить KNN
 - a. индекс резюме
 - b. индекс навыков
 - c. индекс для профессий
2. Заполнить граф
 - a. узлы резюме
 - b. узлы навыков
 - c. добавляем связи (отношения)
3. Вычисляем профессии
 - a. добавляем метки профессий в узлы резюме
4. Вычисляем Communities в графе
 - a. Запускаем алгоритмы из GDS library
5. Добавить метки схожести по BERT-у для узлов со скиллами

Нагрузка / Производительность



Итоги

- Простота разработки
- Расширяемый подход
 - новые правила
 - новые метки, отношения (отрасль, специализация, зарплата, ...)
 - новые обходы
 - веса на отношения (связи)
- Интерпретируемость
- Высокая производительность
 - изначальный запрос (первый тип) “медленный”
 - последующие запросы (второй тип) работают быстро

Спасибо за внимание!

