

ВКР по теме: анализ рынка труда при помощи Node.js

Кузнецов Вадим Сергеевич

Актуальность

Рынок труда постоянно меняется.

Особенно в области IT.

Почти каждый начинающий специалист не знает, с чего начать на работе, и как её найти.

Около трети выпускников **не работают** по специальности.* Это большая проблема для системы образования

Было бы неплохо узнавать эти тренды хотя бы каждый год, и не по *чутью*, а на натуральных данных.

*по данным росстата

цель

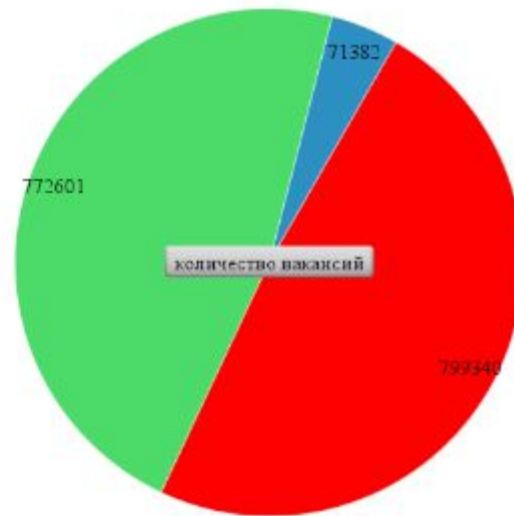
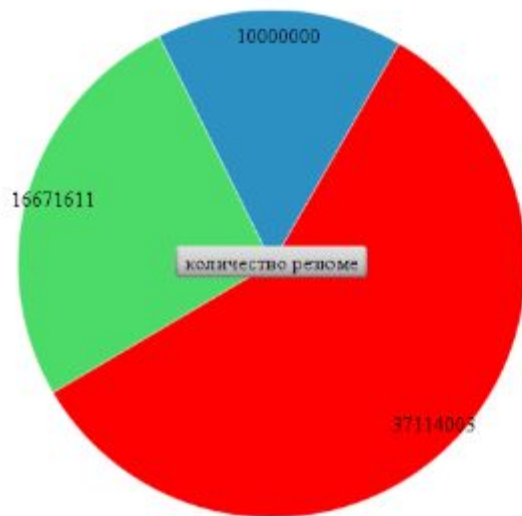
Разработать программу для
автоматического анализа рынка труда

- В свободном доступе
находится около миллиона
открытых вакансий
- обработка их с бумагой и
карандашом заняло бы месяцы
и даже годы
- *раньше это столько и
занимало

задачи

- изучить исторический момент развития рынков труда
- разработать методику изучения рынка труда
- разработать программу
- собрать данные
- обработать данные

■ hh.ru ■ superjob ■ rabota.ru



hh.ru является третьим в мире по популярности сайтом для поиска работы, и **первым** в России

—
hh.ru имеет публичное **API**,
он даёт возможность
загрузить более **800 000**
вакансий

ещё есть **trudvsem.ru....**

—
чем **больше** данных, тем
лучше для исследования.
по итогу работы было
обработано около **100 000**
вакансий

—

Технологии:

Node.js

TypeScript

Docker

PostgreSQL

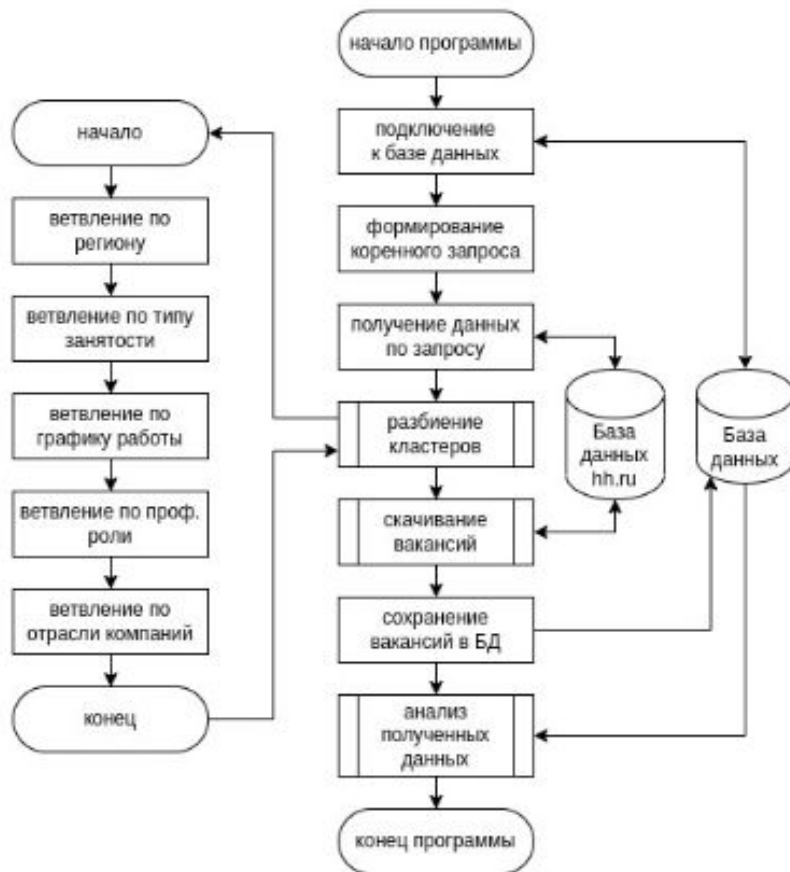
Алгоритм приложения

Скачать как можно больше вакансий (данных)

Взять оттуда полезную информацию (ключевые навыки)

Посчитать мат.статистику...

Нарисовать диаграммы



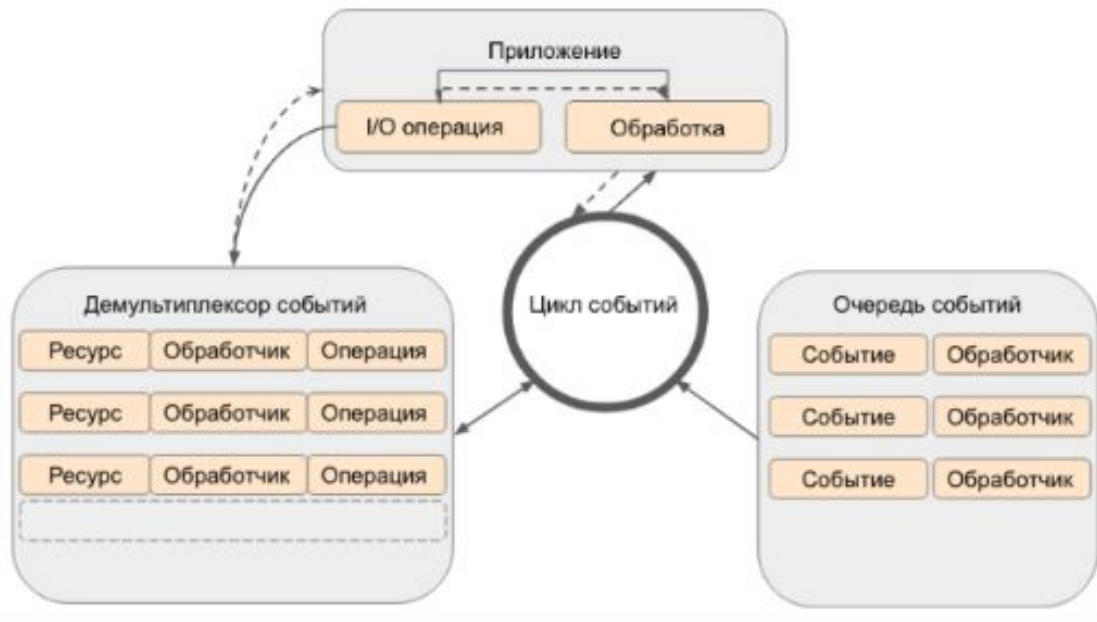
Почему Node.js?

Асинхронные вычисления

Богатая библиотека

Простой Си-подобный
синтаксис

Имелся опыт разработки



База данных и Docker

Было удобно завернуть базу данных в контейнер-систему, чтобы в нужные моменты “поднимать” базу данных, а в ненужные выключать.

```
version: "3.8"

services:
  database:
    image: "postgres"
    ports:
      - 5432:5432
    volumes:
      - database-data:/var/lib/postgresql
    environment:
      - POSTGRES_DB=testdb
      - POSTGRES_USER=testuser
      - POSTGRES_PASSWORD=testpass

volumes:
  database-data:
```

Итоговая база данных
содержит **100** тысяч
вакансий

Интерфейс пользователя

Это приложение в
командной строке

Конечно, программа
рассчитана для
обученного специалиста,
а не для широкого
пользователя.

```
территория поиска вакансий: Москва
✓ формирование коренного запроса...
i Коренной запрос: https://api.hh.ru/vacancies?area=1&specialization=1&clusters=true&per_page=0&page=0
✓ получение данных по запросу...
i всего по данному запросу найдено: 27185 вакансий
i парсинг кластеров...
i начало ветвления...
✓ ветвление по типу занятости...
✓ ветвление графику работы...
✓ ветвление по проф.роли...
✓ ветвление по отрасли компаний...
i количество запросов для получения сокращённых вакансий: 426
i размер чанка: 50
i количество чанков: 9
i установка соединения с базой данных
✓ скачивание вакансий... 9/9
✓ сохранение вакансий в базу данных... 278/278
i поиск вакансий закончен, всего найдено: 97389
Done in 43.65s.
```

Ветвление кластеров

Проблема в том, что hh.ru не дает скачивать больше 2 000 вакансий с запроса.

Пришлось “делить” кластеры до тех пор, пока не будет по 2 000 вакансии в каждом

```
const branchByEmployment = async (
  cluster_items: API.ClusterItem[]
): Promise<API.ClusterItem[]> => {
  const [less_2000_clusters, more_2000_clusters] =
    splitItemsBy2000(cluster_items);

  // branching more_2000_clusters
  const urls = more_2000_clusters.map((item) => item.url);
  const reponses: API.Response[] = await Promise.all(
    urls.map((url) =>
      fetch(url, {
        headers: hh_headers,
      }).then((res) => res.json())
    )
  );

  const branched_clusters: API.Cluster[] = reponses.flatMap(
    (res) => res.clusters
  );

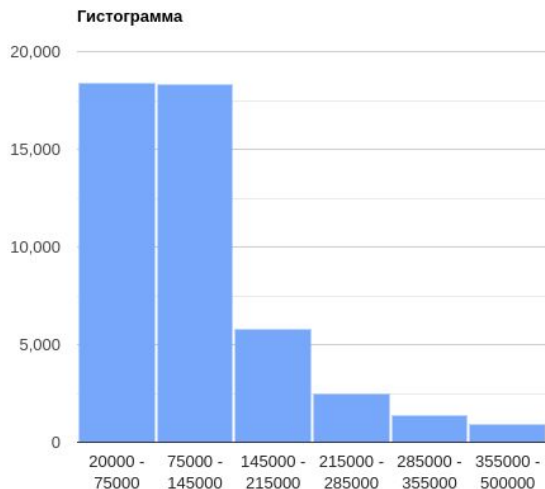
  const experience_items = formatClusters(branched_clusters).experience.items;

  return [...less_2000_clusters, ...experience_items];
};
```

Немного статистики



На основе полученных
данных можно получить
весомый материал для
вычислений
эконометрики и
математической
статистики



```
skills = result_ents
=> {

arr[0],
arr[1],
to_vacancies: parseFloat(
[1] / vacancies_with_keyskills).toFixed(3)

to_key_skills: parseFloat(
[1] / key_skills.length).toFixed(3)

kill) => skill.ratio_to_vacancies >= 0.001)
ll_1, skill_2) =>
count < skill_2.count ? 1 : skill_2.count < skill_1.count ? -1
```

$$M_o = 20000 + 55000 \frac{1}{8} 352 - 0(18352 - 0) + (18352 - 18296) = 74832.681$$

$$Me = 75000 + \frac{7}{0} 00018296 \left(\frac{4}{7} 1752 - 18352 \right) = 95030.881$$

$$s = \sqrt{S^2} = 87367.133$$

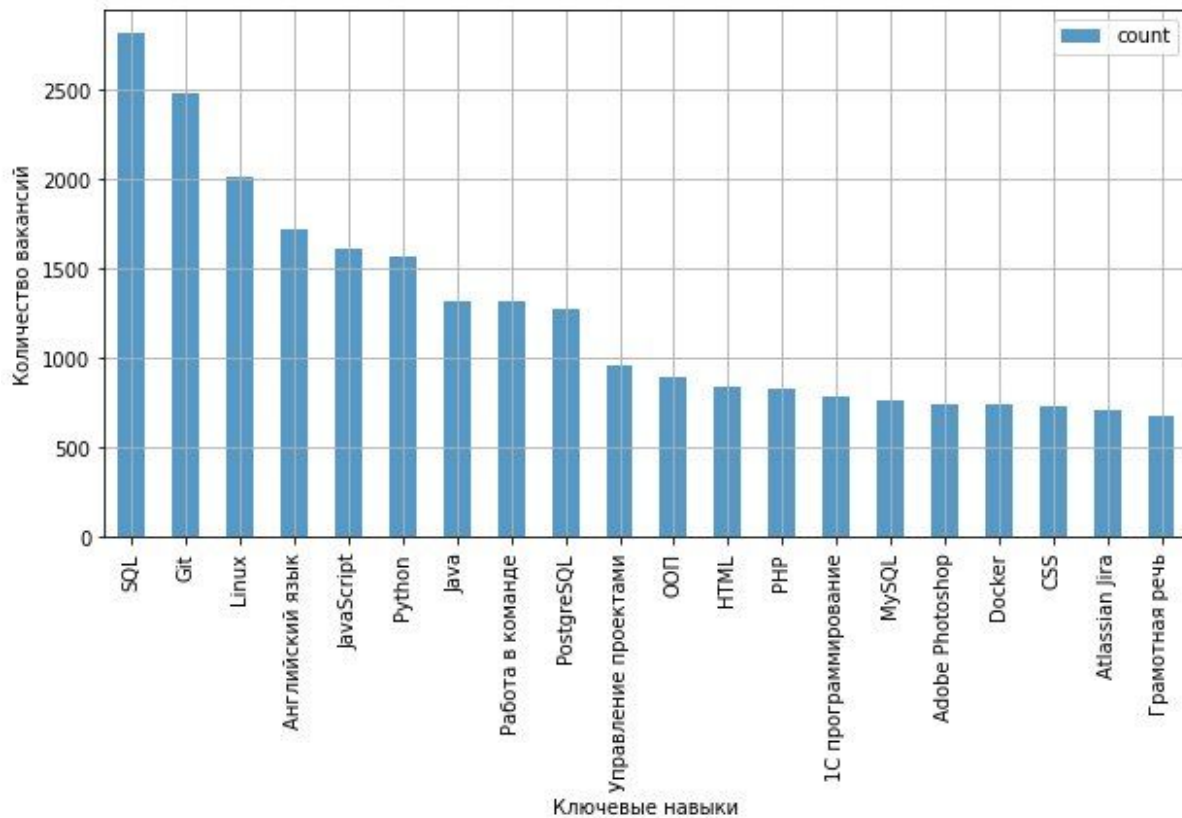
Ключевые навыки в IT:

SQL, Git, Linux,

Английский язык,

JavaScript (+HTML +CSS)

Python,



—
Портрет

специалиста:

Точно знает:

Английский язык, Git, SQL

и базы данных (PostgreSQL / MySQL / MSSQL)

1-2 основных языка: Python / Java / C# / PHP / JavaScript

Для веб-разработчика: JavaScript + HTML + CSS

Желательно: Linux, React / Angular / Vue

Итог

- Была разработана программа для анализа рынка труда
- Составлен портрет работника
- Получены данные для мат. статистики и эконометрики
- Полученные данные уже можно использовать при разработке новых академических курсов