

Федеральное государственное автономное образовательное
учреждение
высшего образования

«Национальный исследовательский университет»

«Высшая школа экономики»

Факультет компьютерных наук
ООП «Прикладная математика и информатика»

Отчёт по проекту

Интерактивная визуализация структуры профессионального
сообщества

Студент: Уржумов Александр Евгеньевич

Группа: БПМИ-154

Руководитель: Кандидат физико-математических наук Яворский Ростислав
Эдуардович

Введение.

Основные цели:

1. Знакомство с JavaScript библиотекой D3.js;
2. Визуализация комментариев с новостного портала;
3. Оформление сайта;

Решаемые задачи:

Задание 1.

1. Создать репозиторий на Github.
2. Написать реферат по библиотеке D3.js.
3. Реализовать пример графа с d3js.

Задание 2.

1. Выбор новостного сайта.
2. Сделать структуру комментариев 10 обсуждений.
3. Составить графы этих структур и визуализировать их с помощью D3.js.
4. Разместить визуализации на сайте.

Задание 3.

1. Сделать меню и конфигурацию для сайта.

Основные инструменты :

1. D3.js.
2. Python
3. JavaScript.
- 4.css.
- 5.html.

Задание 1.

Созданный репозиторий на Github:

https://github.com/AlexUrzhumov/project_visualization

Также был написан реферат про библиотеку d3.js, реферат расположен в репозитории.

Задание 2.

В качестве новостного портала был выбран сайт habrahabr.ru. Работа проделывалась со следующими обсуждениями:

1. <https://habrahabr.ru/post/125999>
2. <https://habrahabr.ru/post/325316>
3. <https://habrahabr.ru/post/324818>
4. <https://habrahabr.ru/post/325236>
5. <https://habrahabr.ru/company/vkontakte/blog/324722>
6. <https://habrahabr.ru/company/mailru/blog/323242>
7. <https://habrahabr.ru/company/ruvds/blog/323330>
8. <https://habrahabr.ru/company/wrike/blog/323550>
9. <https://habrahabr.ru/post/325332>
10. <https://habrahabr.ru/post/324818>

Так как у сайта habrahabr.ru отсутствует готовый api, комментарии добывались парсингом сайта с помощью python библиотек lxml и requests. Полученные комментарии были размешены в графе в виде json словаря

Создание json для визуализации:

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
import requests
from lxml import html
import json
habr_articles = open('articles.txt', 'r')
index = 1
for article in habr_articles:
    response = requests.get(article)
    tree = html.fromstring(response.text)

    path_to_nicks = '//li/div[@class = "comment_body "]/div[@class]/span[@class = "comment-
item__user-info"]/a[@class = "comment-item__username"]/text()'
    path_to_ids = '//li/div[@class = "comment_body "]/div/@rel'
    path_to_parents = '//li/span/@data-parent_id'
    path_to_ratings = '//li/div[@class = "comment_body "]/div/div/div/span[@title]/text()'
    nick_names = tree.xpath(path_to_nicks)
    ids = tree.xpath(path_to_ids)
    parent_ids = tree.xpath(path_to_parents)
    ratings = tree.xpath(path_to_ratings)

    i = 0
    all_comments = dict()
    for elem in ids:
        all_comments[elem] = (nick_names[i], ratings[i], parent_ids[i])
        i += 1
    for elem in all_comments:
        print (elem, ' : ', all_comments[elem])

    i = 0
    nodes = list()
    links = list()
    nodes.append({"id" : "0", "rating" : 20, "name" : "Article"})
    for elem in all_comments:
        x = all_comments[elem][1]
        if x[0] == "-":
            x = "0"
        nodes.append({"id" : elem, "rating" : int(x), "name" : all_comments[elem][0]})
        links.append({"target" : elem, "source" : all_comments[elem][2]})

    graph = {"nodes" : nodes, "links" : links}
    path = 'graph' + str(index) + '.json'
    with open(path, 'w') as output_file:
```

```
json.dump(graph, output_file, sort_keys=False)
index += 1
```

Созданные визуализации располагаются на alex.hse7.ru.

Задание 3.

Для оформления сайта использовались: html, css и javascript.

Была создана титульная страница с меню и кратким описанием. Каждая визуализации расположена в отдельной вкладке, в зависимости от ценности комментария размер и цвет вершины графа меняется для более наглядной визуализации.

Весь код сайта с сопутствующими документами расположен в репозитории на Github.

Заключение.

Данный проект был очень полезен для моего обучения. Я познакомился и разобрался с D3.js и смог реализовать визуализации структуры. Более подробно разобрался с парсингом сайтов, посредством освоения библиотек requests и lxml на Python. У меня появилось примерное понимание способа структурирования данных. Кроме того, я попробовал силы в оформлении сайта. Для этого мне пришлось глубже познакомиться с html, css и javascript.

Данный опыт будет полезен в будущем как во время обучения, так и для применения в других проектах.