

Отчёт по проекту: “Интерактивная визуализация структуры профессионального сообщества”.

Содержание работы.

Визуализация данных с сайта stackoverflow.com.

Результаты.

- файл tags.txt содержит упорядоченный набор более чем 800 тегов, вопросы по которым наиболее популярны на сайте stackoverflow.com, для каждого тега показано число заданных вопросов;
- файл pairs.txt содержит по 25 ближайших (наибольшее количество вопросов в пересечении) тегов для каждого тега из файла tags.txt;
- файл arrows.html – граф-визуализация. Вершины - теги, рёбра - есть между тегами, если имеется достаточно большое количество вопросов в пересечении. Начальными вершинами выбраны 10 наиболее популярных тегов, для каждого из них проведены рёбра к 25 наиболее близким тегам;
- файл arrows2.html – граф-визуализация. Вершины - теги, рёбра - есть между тегами, если имеется достаточно большое количество вопросов в пересечении. Начальными вершинами выбраны 30 наиболее популярных тегов, для каждого из них проведены рёбра к 10 наиболее близким тегам.

Инструменты, использованные при работе над проектом.

- Python + библиотека BeautifulSoup – выкачивание и обработка информации с сайта;
- HTML + JavaScript – визуализации данных.

Навыки, полученные при работе над проектом.

- углубление знаний в разработке веб-страниц в HTML;
- получение знаний в языке программирования JavaScript;
- знакомство с некоторыми видами хранения данных: json, sqlite, csv;
- углубление знаний в языке программирования Python, в частности при работе с файловой системой, а также получение навыков работы с парсером для синтаксического разбора файлов BeautifulSoup.