

AngryBlock

Фильтрация новостей

SPb AU project

Беляев Станислав
Степанов Владимир
Пластинин Виталий

Куратор: Кринкин Михаил

Клиентская часть

- Расширение для Google Chrome:
 - блокирует негативные новости
 - собирает заголовки и посылает на сервер, получая массив оценок
 - обучает систему оценивания, посылая на сервер заголовки с заданной пользователем оценкой

Освоено:

- Основы работы с git
- Javascript + HTML
 - HTTP запросы
 - манипуляции с DOM
 - query selectors
- Написание плагинов для Google Chrome
 - content script
 - context menu
 - local storage

Проблемы

- Новости должны удаляться незаметно для пользователя
- Медленная обработка запросов
- Динамически подгружаемые новости (vk, twitter)
- Проблематичность создания универсальной технологии выборки заголовков из любых сайтов (пришлось делать отдельно для каждого ресурса, ограничились lenta.ru и [politruussia](http://politruussia.ru))

Сервер

- Обучение:
 1. Пересчёт вероятностей для наивного байеса
 2. Вероятности хранятся в sqlite
- Классификация
 3. Разбиваем текст на лексемы (`nltk.RegexpTokenizer`)
 4. Наивный байесовский классификатор для оценки

Познакомился:

- библиотека flask, requests
- библиотека logging
- nltk Tokenizers и NaiveBayesClassifier
- git, GitHub
- Heroku, запуск удалённых приложений на сервере
- SQL: Расширения SQLAlchemy и управление SQL базой из python приложения

Проблемы

- Невозможность хранить локальные файлы на сервере heroku
- Особенности cross-domain запросов

<https://github.com/vitalik239/AngryBlock>