

## 1 Домашнее задание №4

Срок исполнения - 09.11.2019

### Задача №1 "Разработка протокола прикладного уровня"

Необходимо разработать протокол прикладного уровня, который будет соответствовать следующей спецификации:

**Сообщение клиент-сервер:**

- Сначала передается командная строка состоящая из [Команда] [Размер поля полезной нагрузки]
  - Команды - STAT и ENTI, команда STAT возвращает статистические данные, собранные в ходе анализа переданного текста, а команда ENTI возвращает найденные именованные сущности
  - В качестве полезной нагрузки будет передаваться список твитов
  - Нужно посчитать 10 наиболее часто встречающихся слов, 10 наиболее популярных твитов, их авторов и сколько раз они были ретвитнуты, 10 самых популярных авторов. Информацию о странах, в которых пользователи создают контент (твиты) и в которых его потребляют (ретвитят)
  - Датасет - [Tweets](#)
  - Для распознавания именованных сущностей предлагается использовать Stanford CoreNLP [Natural Language Processing Using Stanford's CoreNLP](#)
  - Сервер должен поддерживать несколько одновременных подключений
- После передаются данные, размер которых указан в командной строке
- Данные могут передавать в любом удобном формате - csv, json, xml etc

**Сообщение сервер-клиент:**

- Ответ может передаваться в любом удобном формате - csv, json, xml etc
- Первой строкой передается размер данных ответа

## Задача №2 "Разработка веб сервера"

- Сервер должен поддерживать два endpoint. Первый возвращает статистические данные, собранные в ходе анализа переданного текста, а второй возвращает найденные именованные сущности
- Основной протокол передачи данных http
- В качестве полезной нагрузки будет передаваться список твитов
- Нужно посчитать 10 наиболее часто встречающихся слов, 10 наиболее популярных твитов, их авторов и сколько раз они были ретвитнуты, 10 самых популярных авторов. Информацию о странах, в которых пользователи создают контент (твиты) и в которых его потребляют (ретвитят)
- Датасет - [Tweets](#)
- Для распознавания именованных сущностей предлагается использовать Stanford CoreNLP [Natural Language Processing Using Stanford's CoreNLP](#)
- Сервер должен поддерживать несколько одновременных подключений