Министерство образования и науки РФ

Федеральное государственное образовательное учреждение

высшего профессионального образования

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Кафедра комплексной информационной безопасности электронно-

вычислительных систем (КИБЭВС)

Место прохождения практики:

ТУСУР, каф. КИБЭВС, г.Томск

ОТЧЁТ

по технологической практике

Выполнили:

студентка гр. 722

\_\_\_\_\_\_\_ Мейта М.В.

студент гр. 742

\_\_\_\_\_\_\_ Шиповской В.В.

« »\_\_\_\_\_\_\_\_\_\_\_\_\_\_2015 г.

Руководитель практики от предприятия:

к.т.н. каф. КИБЭВС

\_\_\_\_\_\_\_ Романов А.С.

« »\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2015 г.

Руководитель практики от университета:

старший преподаватель каф. КИБЭВС

\_\_\_\_\_\_\_ Сопов М.А.

« »\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2015 г.

Томск 2013

АННОТАЦИЯ

к технологической практике студентов 3 курса

кафедры КИБЭВС, ТУСУРа, г.Томска

Мейта М.В., Шиповской В.В.

2015, Х стр.

Х ил., библиогр. список – Х наим., Х прил.

Целью технологической практики было начало разработки системы мониторинга настроения людей в социальных сетях для предотвращения массовых беспорядков и экстремизма, то есть такого программного комплекса, который позволит производить семантический анализ текстовых сообщений на предмет негативной эмоциональной окраски. Подобного рода система может быть использована в различных целях, как научно-исследовательских, так и коммерческих.

В рамках данной практической работы была разработана программа для извлечения коротких текстовых сообщений («tweets») по определенному ключевому слову («hashtag») в социальной сети «Twitter», последующего сохранения извлеченной информации в базу данных, а также веб-интерфейс для отображения полученных результатов на основе программной платформы Django версии 1.8.0.

Планируется дальнейшая разработка программного комплекса для семантического анализа высказываний в социальных сетях, а также написание научно-исследовательской статьи в рамках Х международной научно-практической конференции «Электронные средства и системы управления».

СОДЕРЖАНИЕ

[Введение 4](#_Toc429142404)

[1 Кафедра КИБЭВС 5](#_Toc429142405)

[2 Разработка системы мониторинга социальных сетей с целью предотвращения массовых беспорядков и экстремизма 6](#_Toc429142406)

[2.1 Используемые программные средства, обоснование выбора и их описание 7](#_Toc429142407)

[2.1.1 Язык программирования Python 8](#_Toc429142408)

[2.1.2 Программная платформа Django 9](#_Toc429142409)

[2.1.3 Система контроля версий Git и веб-сервис для создания удаленного репозитория GitHub 10](#_Toc429142410)

[2.1.4 Кросплатформенный текстовый редактор Sublime Text 2 11](#_Toc429142411)

[3 Создание шаблона проекта в Django 12](#_Toc429142412)

[4 Проектирование базы данных. Создание моделей 13](#_Toc429142413)

[5 Twitter API. Программа для поиска сообщений по ключевому слову 14](#_Toc429142414)

[Заключение 15](#_Toc429142415)

[Список использованных источников 16](#_Toc429142416)

[Приложение А 17](#_Toc429142417)

# Введение

В период 29 июня 2015 года по 12 июля 2015 года нами была пройдена технологическая практика на кафедре КИБЭВС ТУСУРа. В ходе прохождения практики была проведена следующая работа:

1. изучение предметной области проблемы, существующих разработок в области семантического анализа, инструментов и методов для решения поставленной задачи;
2. написание программы для поиска текстовых сообщений («твитов») на интерпретируемом языке высокого уровня Python версии 2.7.9;
3. проектирование базы данных для хранения полученной выборки сообщений с сопутствующей ей информации об авторах, дате публикации и др.;
4. изучение программной платформы Django, ее установка, а также настройка и создание шаблона проекта для разработки веб-интерфейса системы;
5. разработка веб-интерфейса системы для отображения результатов работы программы, а именно содержимого спроектированной базы данных.

# 1 Кафедра КИБЭВС

Местом прохождения практики была выбрана кафедра ТУСУРа – КИБЭВС ([Кафедра комплексной информационной безопасности электронно-вычислительных систем)](http://kibevs.tusur.ru/).

Кафедра организована в ТУСУР в 1971 году как кафедра «Конструирования и производства электронно-вычислительной аппаратуры» (КиП ЭВА) вскоре переименованной в кафедру «Конструирования электронно-вычислительной аппаратуры» (КЭВА).

21 сентября 1999 г. в связи с открытием новой актуальной специальности 090105 – «Комплексное обеспечение информационной безопасности автоматизированных систем» кафедра КЭВА была переименована в кафедру «Комплексной информационной безопасности электронно-вычислительных систем» (КИБЭВС).

Заведующим кафедрой КИБЭВС на сегодняшний день является ректор ТУСУРа,

Александр Александрович Шелупанов, лауреат премии Правительства Российской Федерации, действительный член Международной Академии наук высшей школы РФ, действительный член Международной Академии информации, Почетный работник высшего профессионального образования РФ, заместитель Председателя Сибирского регионального отделения учебно-методического объединения вузов России по образованию в области информационной безопасности, профессор, доктор технических наук.

С 2008 г. кафедра КИБЭВС входит в состав Института «Системной интеграции и безопасности».

На базе кафедры КИБЭВС ТУСУР в 2002 году организовано «Сибирское региональное отделение учебно-методического объединения Вузов России по образованию в области информационной безопасности.[1]

# 2 Разработка системы мониторинга социальных сетей с целью предотвращения массовых беспорядков и экстремизма

В современном мире набрали огромную популярность такие средства массовой коммуникации, как социальные сети. Подобные информационные площадки являются мощным инструментом маркетинговых, социальных и иных исследований. Также социальные сети становятся источниками распространения экстремизма, социальной, расовой, национальной и религиозной ненависти и вражды. В связи с этим возникает необходимость отслеживания и возможного урегулирования настроений людей в социальных сетях с целью предотвращения массовых беспорядков, нарушений конституционных прав граждан, защиты национальных интересов страны. Данная задача может быть осуществлена при помощи программного комплекса, включающего в себя инструменты для анализа тональности текста сообщений и их классификации по эмоциональной окраске.

Классификация высказываний по эмоциональной окраске может быть проведена различными способами. Условно можно разделить все высказывания на три подмножества: позитивно, негативно или нейтрально окрашенные. Далее встает проблема обучения автоматизированной системы семантическому анализу тональности произвольного высказывания из конечного множества сообщений.

Существует множество разработок в области семантического анализа и создания нейронных сетей, способных решать задачи компьютерной лингвистики. Среди них – разработанная учеными из Стэнфорда нейросеть, способная, по заявлению разработчиков, определять тональность англоязычного текста с точностью 85% [2]. Также есть различные отечественные разработки – как коммерческие, так и доступные для исследования и научной работы. Однако по-прежнему данная тема остается актуальной и интерес к ней активно возрастает.

В качестве первого шага в разработке данной системы была поставлена задача создать программу для извлечения коротких текстовых сообщений («твитов»), публикуемых в социальной сети «Twitter» по ключевому слову («хэштегу») с последующим сохранением в базу данных как самих сообщений, так и сопутствующей им информации (об авторе, дате публикации, локации пользователя и др.). Также разработать веб-интерфейс для отображения полученных данных и продолжить исследование данной предметной области для написания научно-исследовательской статьи в рамках Х Международной научно-практической конференции «Электронные средства и системы управления» [3] и дальнейшей разработки системы.

## 2.1 Используемые программные средства, обоснование выбора и их описание

В качестве инструментов разработки были выбраны:

1. интерпретируемый высокоуровневый язык программирования Python версии 2.7.9;
2. программная платформа Django версии 1.8.0;
3. система контроля версий Git и веб-сервис для создания удаленного репозитория GitHub;
4. кросплатформенный текстовый редактор Sublime Text 2 для написания кода.

## 2.1.1 Язык программирования Python

Python — высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода. Синтаксис ядра Python минималистичен. В то же время стандартная библиотека включает большой объём полезных функций.

Python поддерживает несколько парадигм программирования, в том числе структурное, объектно-ориентированное, функциональное и аспектно-ориентированное. Основные архитектурные черты — динамическая типизация (переменная связывается с типом во время присвоения значения, а не во время объявления переменной), автоматическое управление памятью, полная интроспекция (возможность определять тип и структуру объекта во время выполнения программы), механизм обработки исключений, поддержка многопоточных вычислений и удобные высокоуровневые структуры данных. Код в Python организовывается в функции и классы, которые могут объединяться в модули (они в свою очередь могут быть объединены в пакеты) [4].

Можно выделить следующие преимущества данного языка программирования:

− низкий порог вхождения;

− объектно-ориентированный;

− простота и удобство интуитивно понятного синтаксиса;

− обширный набор различных библиотек;

− метапрограммирование и другие возможности в ООП;

− качественная документация;

− постоянно растущее коммьюнити разработчиков;

− высокая скорость работы.

Кроме всего вышеперечисленного, Twitter имеет множество готовых библиотек с необходимыми API-функциями для различных языков программирования, представленными на официальном сайте в разделе для разработчиков [5]. В том числе среди них представлены и библиотеки для Pyton – например, tweepy [6], которая использовалась в ходе нашей работы.

## 2.1.2 Программная платформа Django

Django — свободный фреймворк (программная платформа) для веб-приложений на языке Python, использующий шаблон проектирования MVC. Проект поддерживается организацией Django Software Foundation.

Сайт на Django строится из одного или нескольких приложений, которые рекомендуется делать отчуждаемыми и подключаемыми. Это одно из существенных архитектурных отличий этого фреймворка от некоторых других (например, Ruby on Rails). Один из основных принципов фреймворка — DRY (англ. Don't repeat yourself)

Также, в отличие от других фреймворков, обработчики URL в Django конфигурируются явно при помощи регулярных выражений, а не выводятся автоматически из структуры моделей контроллеров.

Для работы с базой данных Django использует собственный ORM, в котором модель данных описывается классами Python, и по ней генерируется схема базы данных.

Архитектура Django похожа на «Модель-Представление-Контроллер» (MVC). Контроллер классической модели MVC примерно соответствует уровню, который в Django называется Представление (англ. View), а презентационная логика Представления реализуется в Django уровнем Шаблонов (англ. Template). Из-за этого уровневую архитектуру Django часто называют «Модель-Шаблон-Представление» (MTV).Первоначальная разработка Django, как средства для работы новостных ресурсов, достаточно сильно отразилась на его архитектуре: он предоставляет ряд средств, которые помогают в быстрой разработке веб-сайтов информационного характера. Так, например, разработчику не требуется создавать контроллеры и страницы для административной части сайта, в Django есть встроенное приложение для управления содержимым, которое можно включить в любой сайт, сделанный на Django, и которое может управлять сразу несколькими сайтами на одном сервере. Административное приложение позволяет создавать, изменять и удалять любые объекты наполнения сайта, протоколируя все совершённые действия, и предоставляет интерфейс для управления пользователями и группами (с пообъектным назначением прав) [7].

Преимущества использования Django:

− использование Python в качестве языка программирования, а следовательно, все преимущества данного языка;

− качественная документация;

− встроенный ORM (Object-relational mapper) – технология программирования, которая связывает базы данных с концепциями объектно-ориентированных языков программирования, создавая «виртуальную объектную базу данных»;

− автоматически генерируемый интерфейс администратора;

− поддержка MVT (Model-Template\_View) – паттерн проектирования, предполагающий использование моделей, шаблонов и представлений (view-функций);

− наличие встроенного обработчика ошибок;

− удобство настройки;

− широкое коммьюнити разработчиков;

− высокая скорость работы.

## 2.1.3 Система контроля версий Git и веб-сервис для создания удаленного репозитория GitHub

Аналогично

## 2.1.4 Кросплатформенный текстовый редактор Sublime Text 2

И тут

# 3 Создание шаблона проекта в Django

# 4 Проектирование базы данных. Создание моделей

# 5 Twitter API. Программа для поиска сообщений по ключевому слову

# Заключение

# Список использованных источников

1. Источник - официальный сайт КИБЭВС

http://kibevs.tusur.ru/pages/kafedra/index

1. ссылка на сайт нейросети Стендфорда
2. международная конференция ТУСУР
3. https://www.python.org
4. https://dev.twitter.com/overview/api/twitter-libraries – twitter-API
5. tweepy library - https://github.com/tweepy/tweepy
6. https://www.djangoproject.com

# Приложение А