

Московский государственный технический университет им.Н.Э. Баумана

УТВЕРЖДАЮ
Заведующий кафедрой ИУ-9
_____ (И.П. Иванов)
« ____ » _____ 200__ г.

Техническое задание
на дипломный проект
«Автоматическое установление связей между сообщениями твиттера и
новостными статьями»

Исполнитель: А.И. Выборнов
Группа: ИУ9-111

СОГЛАСОВАНО
Руководитель
квалификационной работы
_____ (Н.В. Лукашевич)
« ____ » _____ 200__ г.

1. Введение.

1.1. Наименование проекта «Автоматическое установление связей между сообщениями твиттера и новостными статьями».

1.2. Исполнитель дипломного проекта – А.И. Выборнов.

2. Основания для разработки.

2.1. Основанием для разработки дипломного проекта является решение кафедры ИУ-9 от «___» _____ 200__ г., протокол № ____.

3. Используемые понятия.

Понятие	Определение
Твиттер	Социальная сеть для публичного обмена короткими (до 140 символов) сообщениями при помощи веб-интерфейса, SMS, средств мгновенного обмена сообщениями или сторонних программ-клиентов.
Твит	Термин сервиса микроблоггинга Twitter, обозначающий сообщение, публикуемую пользователем в его твиттере. Особенностью твита является его длина, которая не может быть больше 140 знаков.
Новость	Оперативное информационное сообщение, которое представляет политический, социальный или экономический интерес для аудитории в своей свежести, то есть сообщение о событиях произошедших недавно или происходящих в данный момент.
Классификация	Это осмысленный порядок вещей, явлений, разделение их на разновидности согласно каким-либо важным признакам.
Тематическая модель	Модель, которая по коллекции текстовых документов определяет, к каким темам относится каждый документ и какие слова (термины) образуют каждую тему.
Тональность текста	Эмоциональная оценка автора по отношению к объектам, речь о которых идет в тексте.
Аннотирование текста	Краткое представление содержания текста в виде аннотации (обзорного реферата).
Точность	Доля релевантных документов выборки по отношению ко всем документам в выборке.
Полнота	Доля релевантных документов в выборке по отношению ко всем релевантным документам коллекции.

4. Цель и назначения разработки.

- 4.1.** Назначение: выявление связи между сообщениями твиттера (твитов) и новостями позволит как расширить информативность твитов, так и обогатить новости. Это позволит лучше решать такие задачи, как построение тематических моделей, классификация сообщений твиттера, определение тональности сообщения, автоматическое аннотирование новостей.
- 4.2.** Цель: исследование и разработка методов автоматического установления связей между сообщениями твиттера и новостными статьями.

5. Требования к программному комплексу.

- 5.1.** Надёжность.
- Система не должна переходить в неопределенное состояние в процессе своей работы, а так же при отсутствии компонентов и необходимых файлов.
- 5.2.** Условия эксплуатации.
- Условия эксплуатации должны соответствовать требованиям по условиям эксплуатации технических средств, выполняющих программу.
- 5.3.** Требования к программной документации.
- Программная документация должна быть выполнена в соответствии с требованиями ЕСПД.
 - На программный пакет должен быть разработан следующий комплект документации:
 - техническое задание (ТЗ);
 - пояснительная записка, включающая руководство пользователя;
 - структурные схемы основных алгоритмов.
- 5.4.** Требования по эргономике и эстетике не предъявляются.
- 5.5.** Требования по маркировке, упаковке и хранению.
- Требования по маркировке, упаковке и хранению должны соответствовать требованиям по маркировке, упаковке и хранению носителей информации, используемых для хранения программы.
- 5.6.** Требования к транспортировке и хранению.
- Требования к транспортировке и хранению должны соответствовать требованиям к транспортировке и хранению носителей информации, используемых для хранения программы.
- 5.7.** Специальные требования.
- Специальные требования не предъявляются.

6. Техничко-экономические требования.

6.1. Требования к расчету технико-экономической эффективности разработки не предъявляются.

7. Стадии и этапы разработки.

7.1. Стадии и этапы разработки приведены в таблице:

Стадии работ	Этапы работ и их содержание	Исполнитель	Сроки исполнения		Результаты работ
			Начало	Конец	
1. Подготовительная	1. Разработка ТЗ	Выборнов	20.09.15	29.09.15	ТЗ утверждено
	2. Сбор информации о существующих разработках		29.09.15	31.12.15	Обзор существующих разработок
	3. Изучение технической литературы		29.09.15	31.12.15	Результаты анализа
2. Основная	1. Разработка и описание алгоритмов		29.09.15	15.03.16	Программная часть проекта
	2. Разработка программной документации		01.03.16	05.04.16	Руководство пользователя
3. Заключительная	1. Оформление пояснительной записки.		06.04.16	10.05.16	Пояснительная записка.
	2. Разработка графической части работы		18.03.16	10.05.16	Презентация

8. Порядок контроля и приемки.

8.1. Приемка работ осуществляется в два этапа:

- предварительная защита перед комиссией представителей кафедры ИУ-9 до 1 июня 2016 г.
- основная защита на Государственной Аттестационной Комиссии в период с 10 по 25 июня 2016 г.

9. Список литературы.

1. W. Guo, H. Li, H. Ji, and M. T. Diab. Linking tweets to news: A framework to enrich short text data in social media. - ACL, pages 239–249, 2013.
2. T. Hoang-Vu, A. Bessa, L. Barbosa and J. Freire. Bridging Vocabularies to Link Tweets and News. - International Workshop on the Web and Databases (WebDB 2014), Snowbird, Utah, US, 2014.
3. J. Sankaranarayanan, H. Samet, B. Teitler, M. Lieberman, J. Sperling. TwitterStand: news in tweets. - 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009, Seattle, Washington.

10. В ходе работы отдельные пункты технического задания могут уточняться, дополняться, исключаться.

Консультант по
конструкторской части
_____ (Н.В. Лукашевич)
«__» _____ 2015 г.

Студент МГТУ
им.Н.Э. Баумана
_____ (А.И. Выборнов)
«__» _____ 2015 г.