

Алина Ильясовна Лозовская,
магистрант 1-го курса
Уральский гуманитарный институт
Уральский федеральный университет

СТИЛИСТИЧЕСКИЙ ПОРТРЕТ ПОЛЬЗОВАТЕЛЯ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»

В тезисах указана область и направление исследования: компьютерная лингвистика, социолингвистический анализ данных. Поставлена проблема взаимосвязи статистического анализа массива данных Big Data и изучения социолингвистических характеристик записей пользователей социальной сети «ВКонтакте».

Ключевые слова: социальные сети, социальные данные, пользовательские данные, анализ социальных сетей, Big Data, компьютерная лингвистика.

Alina Lozovskaya,
1-year master student
Ural Institute of Humanities
Ural Federal University

STYLISTIC PORTRAIT OF A VK SOCIAL MEDIA NETWORK'S USER

The thesis contains the area and direction of the research: computational linguistics and sociolinguistic data analysis. The problem of interrelationship of Big Data statistical analysis and the study of the socio-linguistic characteristics of VK social media network's posts has been raised in this work.

Keywords: social networks, social data, user data, social network analysis, Big Data, computational linguistics.

Анализ социальных данных стремительно набирает популярность благодаря появлению различных социальных сетей (Facebook, Twitter, «ВКонтакте» и другие). Социальные сети становятся

уникальным источником различных открытых данных (визуальных и текстовых), и это открывает беспрецедентные возможности для самостоятельного проведения анализа на стыке лингвистики и социологии.

В настоящей работе осуществляется постановка задачи анализа социальных сетей, а также предложены основные методы анализа данных. Ключевыми для данного исследования являются следующие работы: статья А. Коршунова, И. Белобородова и др., в которой описываются основные компоненты разработанного в ИСН РАН анализа пользовательских данных из социальных сетей с определением различных пользовательских атрибутов [1]; вторая статья Т. В. Батура, Н. С. Копылова и др. посвящена проблеме анализа социальных сетей и описывает авторскую разработку программного комплекса, который позволяет извлекать информацию из социальных сетей и проводить ее обработку [2]. Следовательно, исходя из теоретических данных, изложенных в упомянутых выше статьях, для данной работы важными являются такие методы анализа данных, как статистический анализ, анализ социальных графов, анализ тональности текста для определения отношения пользователей социальных сетей к различным социально значимым событиям, как указывается в статье О. С. Смирновой [3]. Более того, рассмотрена проблема анализа и интерпретации Big Data — набора данных, размер которых превосходит возможности типичных баз данных по хранению, управлению и анализу информации. Для получения такого объема данных в практической части исследования был использован инструмент Streaming API социальной сети «ВКонтакте», который позволяет получать публичные данные из указанной социальной сети по заданным ключевым словам. С помощью данного инструмента был запущен процесс сбора материала: в режиме реального времени индексируется 1 % от записей пользователей, размещенных в открытом доступе. Для каждой записи произведено определение географических атрибутов, указан пол и возраст для отдельного пользователя — автора записи. Произведен статистический анализ 77 707 записей, полученных с помощью запроса, описанного выше, с целью измерения частотности различных языковых явлений, характерных исключительно для интернет-коммуникации.

Рассмотрена корреляционная зависимость силы влияния социальных признаков отдельного пользователя на лексическое наполнение текстовой записи данного пользователя.

Литература

1. Анализ социальных сетей: методы и приложения / А. Коршунов, И. Белобородов, Н. Бузун и др. // Труды ИСП РАН. 2014. № 1 [Электронный ресурс]. URL: <http://cyberleninka.ru/article/n/analiz-sotsialnyhseteymetody-i-prilozheniya> (дата обращения: 15.01.2019).

2. Батура Т. В., Копылова Н. С., Мурзин Ф. А., Проскуряков А. В. Методы анализа данных из социальных сетей // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11. Вып. 3. С. 5–21.

3. Смирнова О. С., Петров А. И., Бабийчук Г. А. Основные методы анализа, используемые при исследовании социальных сетей // Современные информационные технологии и ИТ образование (SITITO'16) : труды XI Международ. науч.-практич. конф. Москва, Россия, 25–26 ноября 2016 г.