

**МЕТОД ВЫЯВЛЕНИЯ ВОЗДЕЙСТВИЯ НА ЧЕЛОВЕКА И СОЦИАЛЬНЫЕ ГРУППЫ
В СОЦИАЛЬНОЙ СЕТИ TWITTER**

А. С. Гончаров, В. М. Саклаков

Научный руководитель: к. т. н., М. А. Иванов

Национальный исследовательский Томский политехнический университет

634050, г. Томск, пр. Ленина, 30

romanov_ky@mail.ru

**METHOD OF DETECTING IMPACT ON HUMAN AND SOCIAL GROUPS
IN THE TWITTER SOCIAL NETWORK**

A. S. Goncharov, V. M. Saklakov

Scientific adviser: c. t. s. M. A. Ivanov

National Research Tomsk Polytechnic University

634050, Tomsk, Lenin Avenue, 30

romanov_ky@mail.ru

***Abstract.** Any communication between individuals and groups is aimed at obtaining the desired result by influencing each other. Modern technologies allow monitoring this influence and analyzing the data obtained. The present work is devoted to development of a graphical tool for data analytics obtained from the social network Twitter: the algorithm of the program and its result is presented. Even the primary results of data analysis show the relevance of using the proposed method of analyzing information. It allows you to cluster the statements of users and visually represents the interconnection of the most popular of them.*

Введение. Любая коммуникация отдельных людей и групп между собой имеет своей целью получение желаемого результата путем воздействия друг на друга. В некоторых источниках [1] такое воздействие называется манипуляцией, однако отмечается, что таковая не обязательно направлена во вред субъекту воздействия. *Объектом анализа* настоящей работы является процесс коммуникации человека и (или) групп людей. Для того, чтобы исследование имело опору на эмпирические данные необходимо задать целевое ограничение: *предметом анализа* является коммуникация человека и (или) групп людей в популярных социальных сетях. В рамках работы для сбора информации выбран Twitter.

Актуальность. В процессе коммуникации в сети объект и субъект воздействия обмениваются сообщениями. Результатом такого обмена становится их действие или бездействие в рамках данного воздействия. При этом по завершении коммуникации процесс воздействия на субъект может продолжаться через тот же или другой объект воздействия. Проблемная ситуация заключается в отсутствии комплексного инструмента сбора и анализа данных, позволяющего автоматизировать процесс выявления и прогнозирования такого воздействия на различных уровнях. Такой инструмент необходимо разрабатывать как междисциплинарный – на стыке экономики, социологии, психологии и информатики. В работах других авторов большое внимание уделяется созданию частных инструментов для бизнеса [2-3], власти [4-5], отраслей наук [6]; отдельно стоит выделить системы сбора и анализа

данных специальных служб [7]. Анализ большого объема данных позволяет собирать, обобщать и обрабатывать статистику сообщений в социальных сетях.

Целью работы на данном этапе является:

1. Разработка комплексной методики выявления массового воздействия на человека и группы людей в сети интернет (преимущественно в социальных сетях).
2. Разработка программного обеспечения, сбора, обобщения и обработки информации, позволяющее в автоматизированном режиме выявлять воздействие на человека и социальные группы и давать оценку его эффективности.

Для достижения поставленных целей необходимо решить следующие задачи:

1. Обзор и кластеризация наиболее востребованных тем в различных социальных сетях.
2. Классификация объектов и субъектов воздействия в сети интернет.
3. Классификация методов воздействия и реакции на воздействие в сети интернет через социально-экономические функции и процессы.
4. Методика анализа связи между объектами в общественном сознании.
5. Программная реализация вышеперечисленных пунктов (сбор, визуализация, анализ).

Данный комплекс целей и задач имеет системный характер и будет достигаться поэтапно. В рамках настоящей работы были частично реализованы задачи обзора наиболее востребованных тем и программная реализация сбора данных их социальной сети Twitter.

Прототип инструмента веб-аналитики. Ранее авторы разработали прототип программного решения для сбора данных в социальной сети Twitter. Собранные 147 публикаций пользователей по теме ЖКХ были вручную обобщены до 17 смысловых высказываний. Данные публикации содержали от 1 до 4 единичных по смыслу сообщения, что при декомпозиции полученных 147 публикаций позволило выделить 174 таких единичных сообщения. Далее был разработан прототип графического инструмента аналитики полученных данных. Опишем алгоритм работы программы:

1. Для чтения информации, хранимой в формате *.xlsx используем библиотеку xlrd;
2. Формируем массив данных на основе ключевого параметра каждого столбца;
3. Создаем класс Node с параметрами: обзор твита, его связи с другими обзорами, лайки, ретвиты;
4. На каждый обзор из массива данных «data» создаем объект;
5. Создаем объект простого графа при помощи библиотеки networkX;
6. Размеры вершин графа определяются на основе суммы лайков и ретвитов;
7. Создаем холст для графа, его вершины и связи между ними;

Результат работы программного кода отражен в виде графа на рисунке 1. Здесь зеленым цветом обозначены его вершины, набравшие меньше десяти тысяч лайков и ретвитов, синие – от десяти до пятидесяти тысяч, красные – свыше пятидесяти тысяч. Ребра графа демонстрируют связи между единичными по смыслу сообщениями пользователей, сделанными в рамках одной публикации.

Вывод. Графическое отображение метода выявления воздействия на группы людей позволило представить гетерогенную информацию в удобном для анализа виде. Даже первичные результаты анализа данных показывают востребованность и актуальность данной тематики. С одной стороны, можно утверждать, что в данном (частном) случае проблемы ЖКХ известны, однако непонятен масштаб явления, его пропорции, а значит и методы работы с ним. При анализе рисунка 1, а также первичных

данных становится понятным: (1) наибольшее количество сообщений (45%) являются демонстрируют связь между критикой ЖКХ и негативной оценкой власти; (2) 25% сообщений являются «информационными», т. е. создают информационный шум; (3) малое количество сообщений, может быть компенсировано большим количеством лайков и ретвитов (например, «Позитивная оценка власти»).

В условиях кризиса избытка информации создание инструмента выявления воздействия на группы людей, а также многомерного анализа полученных данных позволит решать поставленные задачи. В дальнейшем планируется расширить функционал приложения в соответствии с целями и задачами, описанными выше.

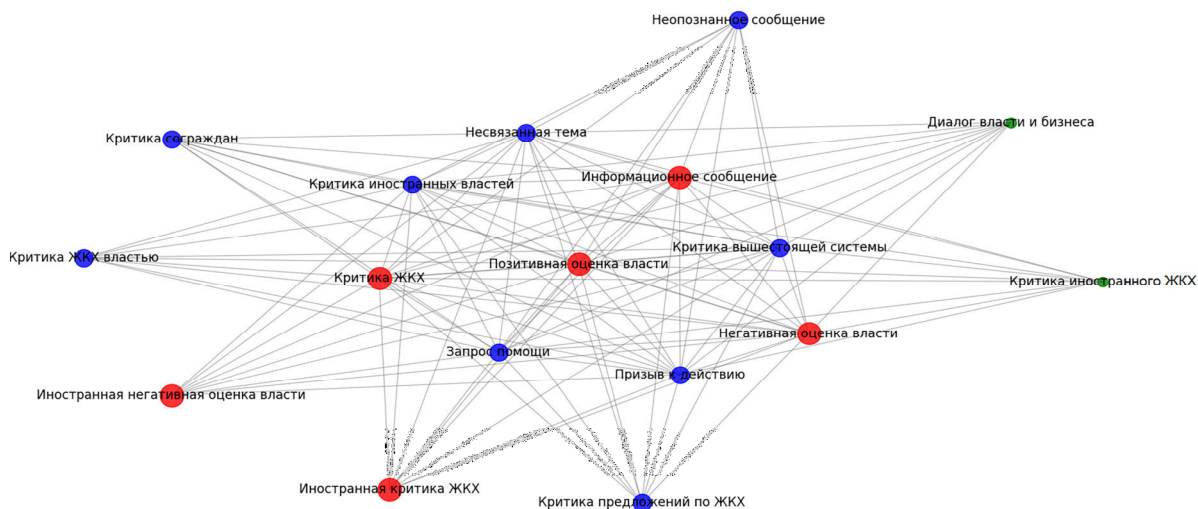


Рис. 1. Частотность запросов по наиболее популярным темам в области ЖКХ

СПИСОК ЛИТЕРАТУРЫ

1. Гагин. Т. В. Голая правда о слепой вере. Сеанс бытовой магии с разоблачением / Т. В. Гагин, С. С. Бородина. – М.: Дрофа-Плюс, 2005. – 448 с.: ил. – (Психология для всех и каждого)
2. Джамак П. Бизнес-анализ больших данных. [Электронный ресурс. – режим доступа: <https://www.ibm.com/developerworks/ru/library/ba-big-data-bi/index.html> (дата обращения 21.02.2018)
3. Аналитический обзор рынка Big Data [Электронный ресурс]. – режим доступа: <https://habrahabr.ru/company/moex/blog/256747/> (дата обращения 10.02.18)
4. Единая межведомственная информационно-статистическая система (ЕМИСС). [Электронный ресурс]. – режим доступа: <https://www.fedstat.ru> (дата обращения 21.02.2018)
5. Портал открытых данных Российской Федерации. [Электронный ресурс]. – режим доступа: <http://data.gov.ru> (дата обращения 21.02.2018)
6. И. Кижнер, М. Террас, М. Румянцев, К. Сычева. Цифровые коллекции музейного фонда РФ как источник данных для гуманитарных исследований: масштабы оцифровки в российских музеях // Информационные технологии в гуманитарных науках: тез. Докл. Науч.-практ. Конф., 18-22 сентября 2017 г. / Сиб. федер. ун-т. – Красноярск: Сиб. федер. ун-т, 2017. с. 35
7. PRISM – Program for Robotics, Intelligents Sensing and Mechatronics (программа разведки США). [Электронный ресурс]. – режим доступа: [https://ru.wikipedia.org/wiki/PRISM_\(программа_разведки\)](https://ru.wikipedia.org/wiki/PRISM_(программа_разведки)) (дата обращения 21.02.2018)