

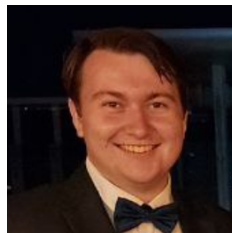
УДК 004.622

## АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ С ПОМОЩЬЮ BIG DATA: ВЫЯВЛЕНИЕ ТРЕНДОВ



**М.В. Реут**

Студент 5 курса, кафедры информатики, инженер-системный программист  
[maksreut1@gmail.com](mailto:maksreut1@gmail.com)



**В.Д. Владымцев**

Ассистент кафедры информатики, инженер-программист ОИАСУ ЦИИР БГУИР  
[v.vladymtsev@bsuir.by](mailto:v.vladymtsev@bsuir.by)



**А.Н. Марков**

Старший преподаватель, заместитель начальника Центра информатизации и инновационных разработок БГУИР  
[a.n.markov@bsuir.by](mailto:a.n.markov@bsuir.by)

### **М.В. Реут**

Студент 5 курса “Информатика и технологии программирования” БГУИР.

### **В.Д. Владымцев**

Ассистент кафедры информатики, инженер-программист ОИАСУ ЦИИР БГУИР

### **А.Н. Марков**

Старший преподаватель кафедры ПОИТ, заместитель начальника Центра информатизации и инновационных разработок Белорусского государственного университета информатики и радиоэлектроники.

**Аннотация.** Использование методов big data для анализа социальных сетей с целью выявления трендов и мнений пользователей. Описывается процесс сбора, обработки и анализа данных, а также основные инструменты и методы, применяемые при работе с большими данными. Рассматривается пример анализа социальной сети Signal и дается оценка этических и правовых аспектов использования таких методов. В заключении подводятся итоги и указывается на значимость анализа социальных сетей в различных сферах деятельности.

**Ключевые слова:** big data, анализ социальных сетей, выявление трендов, сбор данных, обработка данных, инструменты и методы, этические и правовые аспекты, Signal, социальные сети, данные, машинное обучение, алгоритмы, статистический анализ, текстовый анализ, сентимент-анализ, визуализация данных, рекомендательные системы, пользователи, технологии, маркетинг, медиа.

### **Введение.**

Социальные сети играют все более важную роль в нашей повседневной жизни, и с каждым днем становится все сложнее обработать огромное количество информации, создаваемой ежедневно миллионами пользователей. В этом контексте анализ социальных сетей с помощью больших данных становится все более важным, чтобы выявлять тренды и мнения, которые могут помочь нам понять, что происходит в нашем обществе. В данном исследовании мы рассмотрим анализ социальной сети Signal с помощью big data. Представим план, который поможет провести анализ и выявить наиболее популярные темы, настроения пользователей и тренды в использовании данной социальной сети. Начнем со сбора данных из Signal с помощью API, который предоставляет данная социальная сеть, затем обработаем данные и проведем анализ, используя методы машинного обучения и анализа данных. Наконец, мы проанализируем результаты и сделаем выводы на основе данных, полученных из Signal. Анализ социальных сетей с помощью big data может помочь в понимании, как пользователи используют эту социальную сеть, что в свою очередь может помочь в улучшении ее функциональности и предоставлении более полезного опыта для пользователей.

### **Сбор данных в социальных сетях.**

Сбор данных в социальных сетях - это первый и один из самых важных этапов анализа больших данных. В случае с Signal, сбор данных может быть выполнен с помощью API, предоставляемого этой социальной сетью. Первый шаг заключается в определении целевых групп и тем, которые мы хотим проанализировать в Signal. Это может быть сделано, например, путем определения определенных ключевых слов, хештегов или пользователей, которые являются наиболее интересными для исследования. Для того чтобы получить доступ к API необходимо зарегистрироваться на сайте разработчика Signal, получить ключ API и настроить нашу систему для взаимодействия с Signal. После получения доступа к API можно начать сбор данных. Для этого нужно определить параметры, которые нужно получить, например, текст сообщений, даты публикации, количество лайков и репостов и т.д. После сбора данных необходимо их обработать, чтобы избавиться от ненужных данных, провести их очистку и предварительный анализ. Для этого можно использовать специализированные программные инструменты, такие как Python, R и другие. После обработки данных необходимо их сохранить в базу данных или файловую систему, чтобы иметь возможность их использовать в будущем для анализа.

Что касается других мессенджеров и социальных сетей, то Twitter и Facebook продолжают предоставлять доступ к своим API разработчикам, но в свете недавних скандалов с утечкой персональных данных и давлением со стороны правительств на защиту конфиденциальности, они стали более осторожными и регулируют доступ к своим данным. Однако в связи с последними новостями Twitter планирует запустить платный доступ к своему API и стоимость базового доступа к программному обеспечению составит 100\$ в месяц. WhatsApp, с другой стороны, ограничила доступ к своему API в 2021 году, чтобы защитить конфиденциальность своих пользователей и предотвратить распространение мошенничества и спама.

### **Обработка данных.**

После сбора данных необходимо провести их очистку, чтобы удалить шумы и ненужную информацию, такую как рекламные сообщения, боты и т.д.

### **Анализ данных.**

Анализ данных социальных сетей включает в себя несколько шагов. На первом этапе необходимо определить насколько популярна тема, которую исследуем, и выявляем тренды в качестве упоминаний. Затем необходимо анализировать пользователей, выделяя группы с общими интересами и оценивая их влияние. На третьем этапе необходимо анализировать содержимое, выделяя наиболее часто используемые слова и фразы, а также оценивая тональность текстов. После этого используется визуализация данных для наглядного представления результатов анализа.

### **Сравнительный анализ Signal с другими мессенджерами.**

Каждый мессенджер и социальная сеть имеют различные политики конфиденциальности и подходы к обработке и использованию пользовательских данных в научных исследованиях. Некоторые из них предоставляют открытые API (Application Programming Interface), которые разрешают ученым получать доступ к определенным данным, в то время как другие ограничивают доступ к данным и обрабатывают их сами. Например, Twitter предоставляет доступ к своим данным через открытое API, что позволяет исследователям получать данные о твитах, пользовательских профилях и тенденциях. Twitter также получает прибыль от продажи данных пользователей, в том числе информации об их интересах и поведении в социальной сети.

Компания предоставляет доступ к этим данным рекламодателям, которые используют их для создания более эффективных рекламных кампаний. В то же время, Twitter также заявляет, что защищает конфиденциальность данных пользователей и позволяет им контролировать свою конфиденциальность через настройки приватности в своих аккаунтах. Однако, подобные политики использования данных часто становятся объектом обсуждения с точки зрения этики и законности. Facebook также предоставляет открытый доступ к своим данным, но только после

прохождения процедуры аутентификации и получения соответствующих разрешений. Однако, Facebook была критикована за то, что она не защищает свои данные пользователя, в результате чего был создан крупный скандал в 2018 году. WhatsApp, в отличие от Facebook, не позволяет исследователям получить доступ к данным своих пользователей через открытые API и, как следствие, ученые обычно используют альтернативные методы для сбора данных. Signal, с другой стороны, считается одним из наиболее безопасных мессенджеров, так как вся коммуникация между пользователями зашифрована, и компания не продает данные своих пользователей.

Features	DMME	Whatsapp	Telegram	Viber	WeChat	Signal
Military-grade encryption	✓	✗	✗	✗	✗	✓
Encrypted File Transfer	✓	✗	✗	✓	✗	✓
No phone number required	✓	✗	✗	✗	✗	✗
Self-Destructed Messages	✓	✗	✓	✓	✗	✗
Local stored data (Encrypted on your phone & not stored on servers)	✓	✗	✗	✗	✗	✗
Peer-to-peer payments	✓	✗	✓	✗	✓	✗
Public / Private Key (email=pubkey / psw=privkey)	✓	✗	✓	✓	✗	✓

Рисунок 1. Сравнение безопасности мессенджеров

### Выявление трендов и мнений.

Для выявления трендов и мнений в сообществе Signal можно использовать методы анализа тональности текстов, тематического моделирования и анализа социальной сети. Методы анализа тональности позволят определить, какие слова и выражения используются в сообщениях с положительной, отрицательной или нейтральной окраской. Тематическое моделирование поможет выделить ключевые темы и подтемы, обсуждаемые в сообществе Signal. Анализ социальной сети позволит определить наиболее активных пользователей и группы, а также выявить связи между ними.

### Визуализация результатов.

Для наглядной визуализации результатов анализа можно использовать различные инструменты, такие как графики, диаграммы и облака слов. Графики и диаграммы помогут показать изменение количество сообщений и активности пользователей во времени, а также распределение сообщений по тональности и темам. Облака слов позволят выделить наиболее употребляемые слова и выражения в сообщениях.

### Интерпретация результатов.

Полученные результаты анализа социальной сети Signal помогут выявить ключевые тренды и мнения, обсуждаемые в сообществе, а также определить наиболее активных пользователей и группы. Эта информация может быть полезна для маркетинговых

исследований, анализа общественного мнения и разработки стратегий коммуникации в социальных сетях.

#### **Заключение.**

В заключении можно отметить, что анализ социальных сетей с помощью Big Data является важным инструментом для выявления трендов и мнений. В случае социальной сети Signal, были рассмотрены процесс сбора, обработки и анализа данных, показав, что при правильном подходе и использовании соответствующих инструментов, можно получить ценные знания о поведении пользователей и их мнениях. Однако, необходимо учитывать, что сбор и анализ данных социальных сетей связаны с этическими и правовыми вопросами, такими как конфиденциальность и защита личных данных. Поэтому, при использовании таких методов необходимо соблюдать правила и законы, чтобы избежать негативных последствий.

Тем не менее анализ социальных сетей остается важным инструментом для многих областей, от маркетинга и бизнеса до научных исследований и политической аналитики. И при правильном использовании этого инструмента можно получить ценную информацию, которая поможет принимать более обоснованные решения и добиваться успеха в различных областях деятельности.

#### **Список литературы**

- [1] Вайгенд Андреас. Big Data: революция социальных данных.
- [2] Джозел Грас. Data Science: Наука о данных с нуля.
- [3] Ольга Базалева. Мастерство визуализации данных
- [4] Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- [5] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.

## **SOCIAL MEDIA ANALYSIS USING BIG DATA: IDENTIFYING TRENDS**

**M.V. Reut**

*Student of BSUIR, software engineer*

**V. D. Vladymtsev**

*Assistant of the Department of Computer Science, Software Engineer of DIACS CIIR BSUIR*

**A.N. Markov**

*Senior lecturer of the department, Deputy head of the Center for Informatization and Innovative Developments*

*Center for Informatization and Development of the Belarusian University of State Informatics and Radioelectronics, Republic of Belarus*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*

*E-mail: maksreut1@gmail.com, v.vladymtsev@bsuir.by, a.n.markov@bsuir.by*

**Abstract.** Using big data methods to analyze social networks in order to identify trends and user opinions. It describes the process of data collection, processing and analysis, as well as the main tools and methods used when working with big data. An example of Signal social network analysis is considered and an assessment of the ethical and legal aspects of using such methods is given. In conclusion, the results are summarized and the importance of the analysis of social networks in various fields of activity is indicated.

**Keywords:** big data, social network analysis, trend detection, data collection, data processing, tools and methods, ethical and legal aspects, Signal, Social networks, Data, machine learning, algorithms, statistical analysis, text analysis, sentiment analysis, data visualization, recommendation systems, users, technologies, marketing, media.