

DOI: 10.37943/AITU.2020.12.13.010**UDC: 004****M. Ryspayeva**

PhD student in Information technologies and robotics
marya.rys1@mail.ru, orcid.org/0000-0001-5055-4149
A. Baytursynov Kostanay State University, Kazakhstan

O. Salykova

PhD, Associate professor of the Software Department
solga0603@mail.ru, orcid.org/0000-0002-8681-4552
A. Baytursynov Kostanay State University, Kazakhstan

DETERMINATION OF PARAMETERS AND THEIR RELATIONSHIPS IN SOCIAL NETWORK ACCOUNTS

Abstract: The article provides an overview of citizens' participation in social networks according to the results of 2018 in the Republic of Kazakhstan in comparison with the data of the Statistics Committee. From year to year, the main goals of Internet use by residents of the country are participation in social networks, posting information and instant messaging, which show an increasing interest in social networks. In turn, the article's purpose is to determine the parameters that indicate real and fake accounts on social networks and the distribution of parameters in groups. According to open information on social networks and their APIs, three groups of the parameters were selected: account information, completeness of the account, and friends. The groups of parameters are given by the example of using the Graph API Explorer of the Facebook. The relationships between the parameters were analyzed, and their effect on the decisions of the algorithms was determined. The hypothesis is outlined that, it is possible to more accurately determine the ownership of accounts on the basis of friends and family relationships. Often, an account contains bot behaviour; however, it belongs to a person. The algorithms must identify the actions of these exceptions. The researchers suggest that the special influence on the algorithm results within the "Friends" parameters can be defined with a detailed scan of family relationships.

Key words: Social networks parameters, fake accounts, data relationships, big data, social networks.

Рыспаева М.

Докторант специальности «Информационные технологии и робототехника»
marya.rys1@mail.ru, orcid.org/0000-0001-5055-4149
Костанайский государственный университет им. А. Байтурсынова, Казахстан

Салыкова О.

PhD, ассоциированный профессор кафедры «Программное обеспечение»
solga0603@mail.ru, orcid.org/0000-0002-8681-4552
Костанайский государственный университет им. А. Байтурсынова, Казахстан

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ И ИХ ЗАВИСИМОСТЕЙ В АККАУНТАХ СОЦИАЛЬНЫХ СЕТЕЙ

Аннотация

В статье приведена статистика использования социальных сетей по итогам 2018 года на территории Республики Казахстан в сравнении с данными Комитета статистики. Из года в год основными целями использования интернета гражданами страны является участие в социальных сетях, размещение информации и обмен мгновенными сообщениями, что показывает все увеличивающийся интерес к социальным сетям. В свою очередь, целью статьи является определение параметров, указывающих на реальность аккаунта. Отобраны три группы параметров: информация об аккаунте, наполненность аккаунта и друзья. Согласно открытой информации социальных сетей и их API, каждая группа включает в себя ряд параметров, указывающих, что аккаунт принадлежит реальному или фейковому пользователю. Пример указанных групп параметров приведены на примере использования Graph API Explorer социальной сети Facebook. Дополнительно, проанализированы зависимости между параметрами и как они могут повлиять на результаты алгоритма. В статье предложена гипотеза – на основе друзей и выявленных родственных отношений, возможно более точно определить принадлежность аккаунтов к реальному человеку и результаты проверки гипотезы.

Ключевые слова: Параметры социальных сетей, фейковые аккаунты, зависимости параметров, большие данные, социальные сети.

Введение

В рамках цифровизации экономики Республики Казахстан согласно Посланию Главы государства народу от 2 сентября 2019 года, страна взяла курс на развитие и внедрение новейших технологий, такие как 5G, «Умные города», Big Data, блокчейн, цифровые активы, новые цифровые финансовые инструменты на государственном рынке.

На сегодняшний день казахстанские компании все чаще обращаются к Big Data, поскольку сбор данных (Data Mining – DM) позволяет решать проблемы компаний: улучшать маркетинговую, рекламную и другие стратегии. Например, мобильные операторы Beeline, Kcell, Altel, анализируют отток и приток клиентов, их лояльность и финансовые возможности только с помощью данных о звонках, размерах пополнения баланса, длительности звонков и т.д. [1].

Наравне с мобильными операторами, также и кредитные организации занимаются DM и их главной проблемой является определение личности. Эта проблема касается и социальных сетей с тысячами фейковых аккаунтов, которые являются активными и ведут свою деятельность в различных целях, в том числе и незаконных. Как же определить является ли аккаунт настоящим и принадлежит ли он реальному человеку. Какие параметры являются

сигналами о том, что аккаунт принадлежит боту? В рамках написания исследовательской работы проводится научное исследование для определения открытых данных, характеристик и параметров из социальных сетей, отвечающее на выше поставленный вопрос. В [2] были проанализированы четыре социальных сети, пользующиеся популярностью на территории Казахстана, а именно «ВКонтакте», «Одноклассники», «Facebook» и «Instagram».

Основная часть

Согласно докладу о состоянии медиа-сферы по итогам 2018 года, опубликованному Министерством информации и общественного развития Республики Казахстан в государстве насчитывается более 1,7 млн. активных пользователей «ВКонтакте», более 1 млн. в «Instagram», около 0,5 млн. – «Facebook». Несмотря на это, более 10% населения не пользуются социальными сетями и не зарегистрированы [4]. Около 50% опрошенных социологического исследования считают Интернет основным источником информации, при этом 32,76% респондентов постоянно находятся в режиме online. С другой стороны, Комитет статистики Министерства национальной экономики Республики Казахстан выделяет такие основные цели использования сети Интернет членами домашних хозяйств (всех типов жилых помещений, за исключением проживающих в общих коммунальных квартирах, общежитиях, домах-интернатах для престарелых и инвалидов, детских домах, тюрьмах, гостиницах) [5, 6]:

1. размещение информации или мгновенный обмен сообщениями;
2. участие в социальных сетях;
3. скачивание фильмов, изображений, музыки, просмотр видео, прослушивание музыки, игра или скачивание игр;
4. получение информации о товарах и услугах;
5. отправка и получение электронной почты.

Первые две цели подразумевают регистрацию и размещение информации в социальных сетях, которые используются для общения, ведения бизнеса, анализа активной аудитории для дальнейшего маркетинга, различных исследований. В 2016 году наблюдается высокий рост пользователей, обменивающихся мгновенными сообщениями, показатель увеличился с 25,7% в 2015 году до 75,7% в 2019 году (диаграмма 1). В то же время, в социальных сетях участие пользователей выросло с 56% в 2015 г. до 70,7% в 2019 г.

В таблице 1 приведено процентное содержание по каждой цели использования сети Интернет членами домашних хозяйств согласно Комитету статистики Министерства национальной экономики Республики, Казахстан за 2015-2019 гг.

Таблица 1. Основные цели использования сети Интернет членами домашних хозяйств за 2015-2019 гг.

Цель	Год	2015, %	2016, %	2017, %	2018, %	2019, %
Размещение информации или мгновенный обмен сообщениями		25,7	63,5	72,2	75,4	75,7
Участие в социальных сетях		56	66,6	65,7	69,7	70,7
Скачивание фильмов, изображений, музыки, просмотр видео, прослушивание музыки, игра или скачивание игр		52,4	60,4	61,2	63,3	63,3
Получение информации о товарах и услугах		31,4	34,7	35,1	36,8	37,1
Отправка и получение электронной почты		37,8	34,7	31,1	32,6	33,7

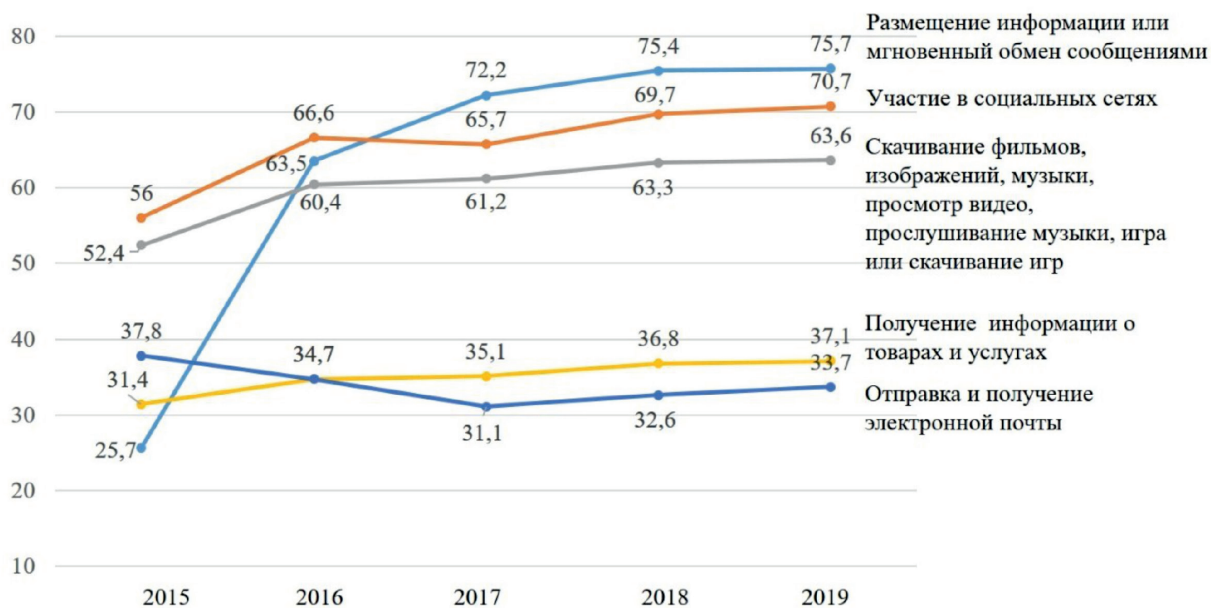


Диаграмма 1. Основные цели использования сети Интернет членами домашних хозяйств

Данная статистика показывает, что население Республики Казахстан активно пользуется социальными сетями и процентное соотношение действий пользователей из года в год увеличивается. В связи с этим анализ социальных сетей является актуальным направлением в информационно-телекоммуникационных технологиях, требующий современный подход к изучению, а именно анализ больших данных с применением методов машинного обучения. На первичном этапе написания диссертации стоит задача проанализировать параметры социальных сетей, которые будут использованы в дальнейшей работе.

Интерфейс всех социальных сетей содержит одинаковый и дополнительный ряд параметров для заполнения пользователей, которые необходимы для лучшего поиска друзей, общения, знакомства и т.д. При заполнении информации о себе настоящим человеком наблюдается определенный алгоритм действий, а при принадлежности аккаунта боту – другой. При разработке алгоритма, определяющего принадлежность аккаунта боту или человеку, необходимо учесть характеристики и принять решение.

Проанализированы группы параметров, указывающих на фейковые аккаунты в социальных сетях, а именно информация об аккаунте, наполненность аккаунта, друзья.

Информация об аккаунте

Информация о пользователей социальной сети включает в себя такие параметры, как: Фамилия Имя, место проживания и рождения, статус – Холост / Женат / Замужем, школа, место учебы и работы, телефон, адрес электронной почты и т.д. При этом современные инструменты позволяют создавать такие алгоритмы, которые могут автоматически заполнять перечисленные параметры, что невозможно определить реальность пользователя [7].

Алгоритмы машинного обучения дают возможность анализировать различные параметры. Зачастую на аватар реальный пользователь размещает личную фотографию, но при создании бота, фотография скачивается с просторов Интернета или различные абстракции, и выдается за собственное. Аккаунты с одинаковыми фотографиями у различных пользователей является одним из сигналов бота [4].

Дата создания аккаунта также является одним из параметров, например, если аккаунт был создан несколько лет назад, или несколько месяцев назад, то в первом случае решение будет склоняться к реальному человеку, а во втором – к боту.

Существуют дополнительные поля, такие как «О себе», «О жизни», «Любимая музыка», «Любимые фильмы», «Любимые книги» и др, которые не заполняются для ботов, чтобы не тратить время на данные поля.

В таблице 2 указан список параметров, относящиеся к информации об акканту, а также их описание.

Таблица 2. Параметры групп информации об акканте

№ п/п	Параметр	Описание параметра
1	id	Уникальный идентификатор учетной записи владельца аккаунта
2	about	Краткое описание о владельце аккаунта
3	birthday	День рождения человека
4	education	Образование владельца аккаунта
5	favorite_athletes	Любимые спортсмены
6	favorite_teams	Любимые команды
7	first_name	Имя человека
8	gender	Пол человека: male, female
9	hometown	Название родного города
10	languages	Наименование языка
11	last_name	Фамилия человека
12	link	Ссылка на аккаунт
13	location	Идентификатор местоположения
14	quotes	Цитаты
15	relationship_status	Статус отношений: Married, Single, Divorced
16	significant_other	Значимый человек
17	sports	Вид спорта, которым увлекается владелец аккаунта
18	work	Место работы
19	books	Книги, которые прочитал пользователь
20	movies	Список фильмов, которые нравятся пользователю
21	music	Понравившиеся музыкальные группы и певцы
22	picture	Миниатюра аккаунта пользователя
23	television	Понравившиеся телевизионные программы
24	groups	Группы, на которые подписан пользователь
25	likes	Люди, которые поставили лайки пользователю

К примеру в социальной сети Facebook указанные параметры в таблице 2 предоставляются через API для каждого пользователя, имеющего открытый аккаунт (рисунок 1).

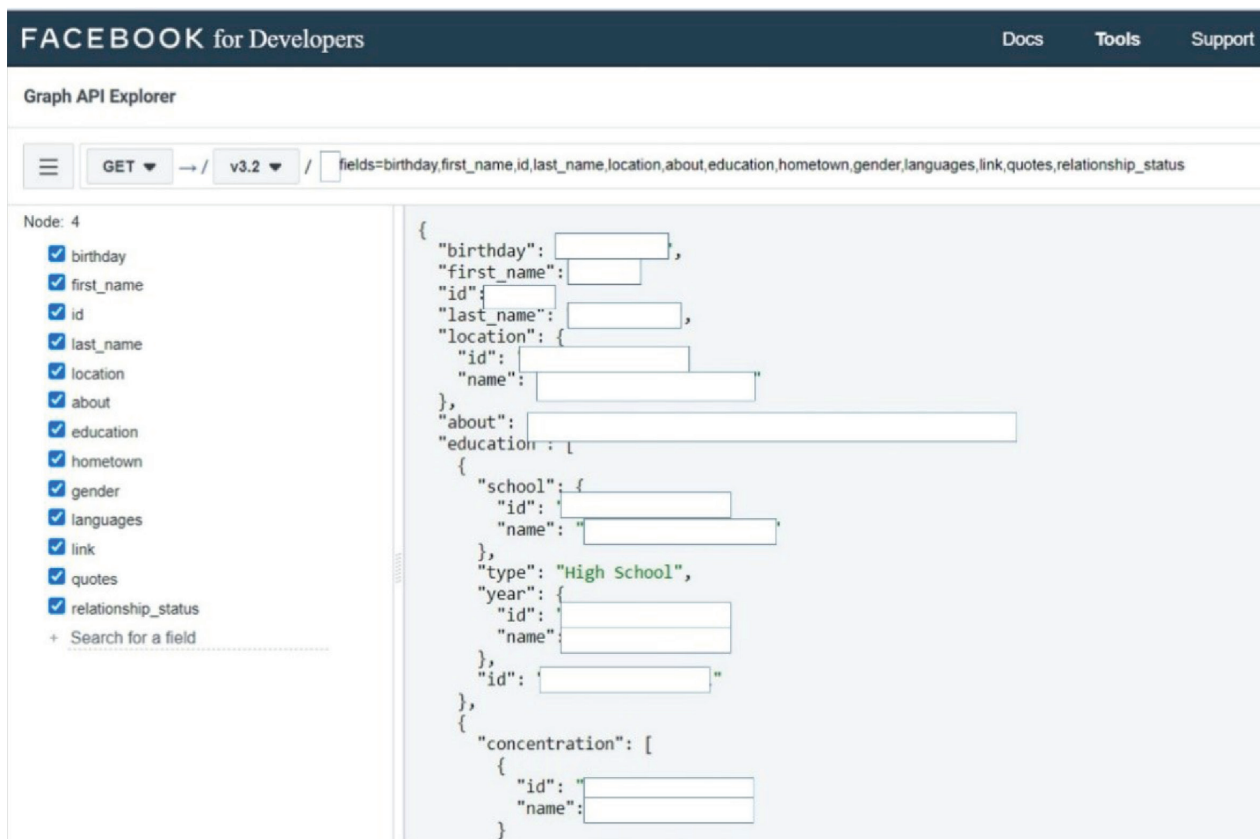


Рис. 1. Открытые параметры социальной сети через API

Наполненность аккаунта

К наполненности аккаунта относятся тематика постов, количество и дата постов. Большинство людей в основном публикуют посты о своей жизни, о себе, загружают личные фотографии, делятся различными эмоциями со своими подписчиками и друзьями. Большинство фейковых аккаунтов размещают информацию узкой направленности, например, о политике, о каких-то проблемах, организуя информационную войну и атаку на реальных пользователей.

Высокая периодичность постов, также является признаком бота. К примеру, аккаунт, созданный 5 лет назад и с 10 постами за весь период существования, начинает массово размещать ссылки на сторонние источники, репосты одних и тех же СМИ без личных комментариев, публиковать информацию узкой тематики и т.д., то алгоритм машинного обучения будет классифицировать аккаунт как фейковый из-за аномального поведения. Время публикации поста – боты размещают публикации каждые 2-3 часа, в точно определенное время, это означает что на данном аккаунте работает сервис автоматической публикации, так называемый постер. Параметры – недавняя дата создания и высокая периодичность публикации – являются основными признаками ботов, а также одинаковое количество постов, которое каждый день размещается [8].

К параметрам постов относятся изображение к посту, ссылка поста, настройки конфиденциальности, репосты, лайки, комментарии и реакции к постам.

Немаловажными параметрами также являются параметры альбомов пользователя, а именно: количество альбомов и фотографий в них, комментарии, лайки и реакции к альбому, описание и миниатюры к альбому. К примеру, в социальной сети Facebook данные параметры отображаются с помощью Graph API Explorer и токена приложения (Рисунок 2).

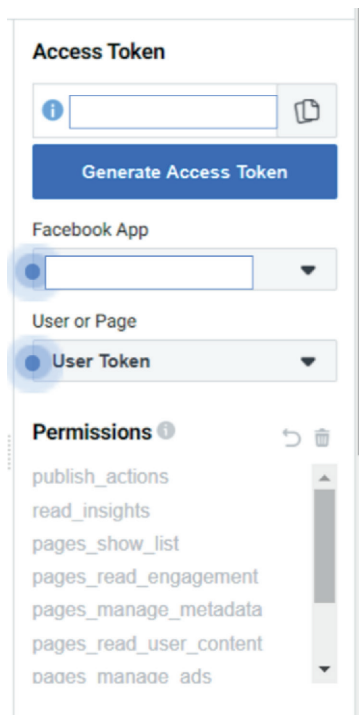


Рис. 2. Access Token API

Друзья

В основном, в первую очередь, с психологической точки зрения, человек является личностью и социальным существом, это значит для социальных сетей, что каждый человек имеет в качестве друзей других людей, таких как родных и родственников, знакомых, коллег, одноклассников, сокурсников и т.д. [9].

Алгоритм машинного обучения обязательно должен анализировать параметры друзей.

Параметры, характеризующие реального человека:

1. Родственные связи. Несколько друзей в аккаунте имеют одинаковую фамилию с анализируемым аккаунтом. Люди, обладающие одинаковой фамилией, предполагается, что являются родственниками. Алгоритм определяет родственные связи, характерные практически для каждого человека и гражданина страны. Выявление родственников в виде графов строится на каждого человека и людей, окружающих его (диаграмма 2).

Возможны случаи, когда алгоритм определяет людей с одинаковой фамилией и отмечает их родственниками, но на самом деле они не являются таковыми. После этого, алгоритм должен выявить на основе друзей и контактов, родственные связи каждой отдельной семьи, это делается на анализе всех отобранных аккаунтов-однофамильцев и их друзей. Часто происходит так, что родственники одной семьи добавляются друзьями в социальных сетях для поддержания общения, образуя собственную ячейку. А другая половина однофамильцев никак не связаны ни через одного человека, напрашивается вывод, что это на самом деле другая семья, только с одинаковой фамилией [10].

2. Поле «Married», указана фамилия человека, совпадающая с фамилией владельца аккаунта. При этом, анализируется аккаунт, указанного в поле «Married» по всем вышеперечисленным параметрам (наличие фото, дата создания аккаунта, периодичность и тематика постов и т.д.).

3. Пол друзей. После анализа тысячи аккаунтов, выявлена связь между полами друзей у аккаунта. Человека окружают с детства женщины и мужчины в лице матери и отца, мужа и жены, детей, друзей и подруг и т.д. Предположим, что у аккаунта добавлены 200 человек в качестве друзей, примерно 50% из них мужчины, а остальные 50% – женщины, то вывод –

это реальный человек. При наличии большинства друзей только представителей мужского пола и если владелец аккаунта женского пола, то вывод – бот, созданный для различных маркетинговых или рекламных целей, и также наоборот.

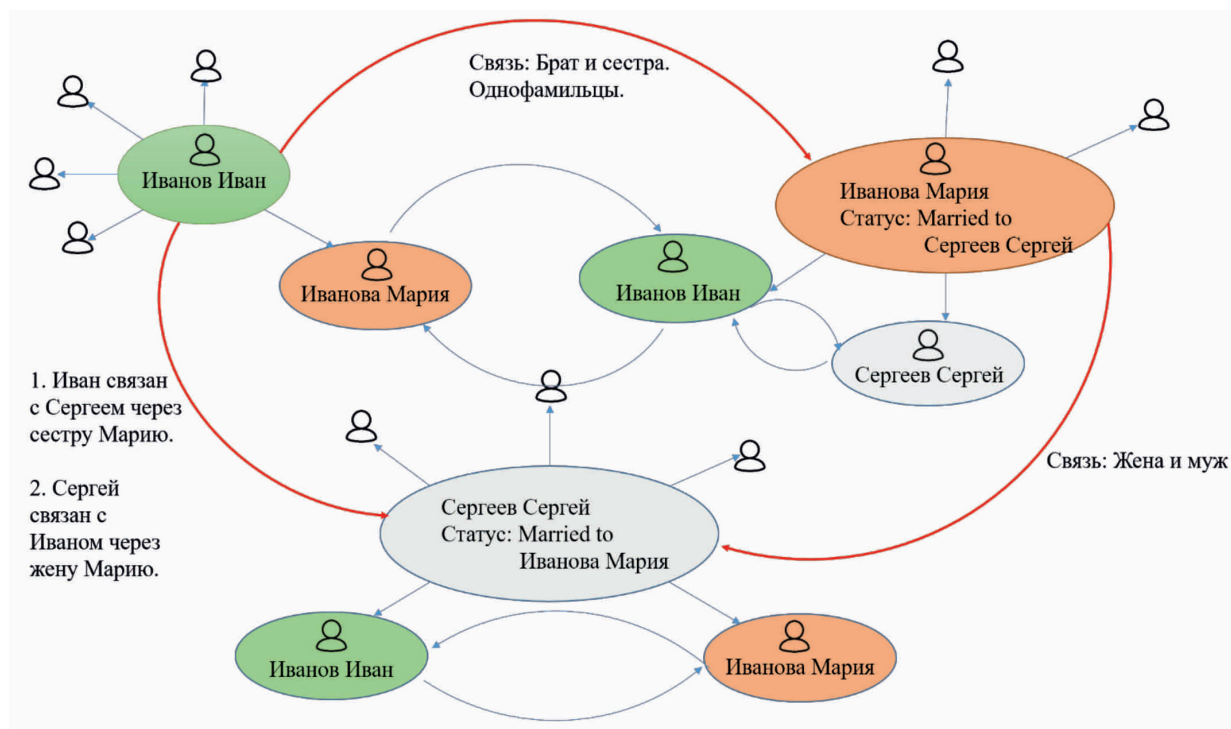


Диаграмма 2. Родственные связи между друзьями

Совокупность признаков, сигнализирующих о фейковом аккаунте (бот):

1. После удаления давно существующего аккаунта, создается другой аккаунт с одинаковым поведением удаленного аккаунта, так называемый клон. Подобные клоны создаются, когда человек хочет скрыть свою личность и свои намерения.

2. Соотношение даты создания аккаунта, количества постов и количества друзей. Много постов, недавняя дата создания, отсутствие родственных связей с такими же фамилиями, друзья в основном мужчины или только женщины в зависимости от пола самого аккаунта.

3. Частота публикаций и репостов, их тематика, время, тональность. Например, аккаунт публикует регулярно посты о политике, в основном посты несут негативный эффект, подрывающий стабильность страны [11].

Заключение

Проанализированы параметры аккаунтов пользователей с помощью API социальных сетей, на примере Facebook. Параметры были разделены на три группы: содержание персональной информации, наполненность аккаунта и наличие друзей. Авторы приводят пример, когда параметры являются более характерными для ботов и реального человека. В случаях, если параметры, характерные для ботов встречаются в аккаунтах настоящего владельца аккаунта, необходимо обучать алгоритм для выявления подобных исключений.

В ходе исследования были сделаны выводы, что значимыми параметрами, выявляющие фейковые аккаунты является наличие родственных связей (построение родственных связей), друзья и их гендерная принадлежность, а также анализ поведения аккаунта.

Анализ аккаунтов социальных сетей не ограничивается анализом только одной социальной сети, так как это зависит от популярности социальных сетей, целей и причин использования.

При написании статьи не рассматривалось поведение аккаунта: участие в группах и сообществах, понравившийся контент, анализ текста и фотографии в постах. Язык аккаунта также может быть разным – казахским, русским или английским, это не влияет на предложенный выше анализ, так как рассматривается интерфейс социальных сетей и его заполнение.

Литература

1. Рыспаева М.К., Салыкова О.С., Иванова И.В. «Уровень развития Data Science в Республике Казахстан» // Наука сегодня: глобальные вызовы и механизмы развития: материалы международной научно-практической конференции. – Вологда: 29 апреля 2020. – С. 54-57
2. Ryspayeva M., Salykova O., Ivanova I. Review of Data Science development in the Republic of Kazakhstan // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ, АВТОМАТИЗАЦИИ И МЕХАТРОНИКЕ: Сборник научных трудов 2-й Международной научно-технической конференции (30 апреля 2020 года). Юго-Зап. гос. ун-т. Курск.; 30 апреля 2020 года. – С. 7-11.
3. Bonchi, Francesco & Castillo, Carlos & Gionis, Aristides & Jaimes, Alejandro. (2011). Social Network Analysis and Mining for Business Applications. ACM TIST. 2. 22. 10.1145/1961189.1961194.
4. Медиа-сфера Казахстана – 2018, (2018). Министерство информации и общественного развития Республики Казахстан URL: <https://qogam.gov.kz/ru/kategorii/rezultaty-provedennyh-nauchnyh-konsaltingovyh-socialnyh-analiticheskikh-issledovaniy> (дата обращения: 16.02.2020).
5. Приказ Председателя Комитета по статистике Министерства национальной экономики Республики Казахстан от 21 октября 2016 года №246(2016). 'Об утверждении Методики по статистике информационно-коммуникационных технологий'
6. Министерство национальной экономики Республики Казахстан Комитет по статистике (2020). 'Основные цели использования сети Интернет членами домашних хозяйств' URL: <https://stat.gov.kz/> (дата обращения: 16.02.2020).
7. Pollet, T.V., Roberts, S.G.B., Dunbar, R.I.M. (2011) 'Use of social network sites and instant messaging does not lead to increased offline social network size, or to emotionally closer relationships with offline network members. Cyberpsychology, Behavior, and Social Networking, vol.14, pp. 253-258
8. Алгоритм определения ботов и офферов во ВКонтакте (2018) Хабр URL: <https://habr.com/ru/post/413855/> (дата обращения: 17.02.2020).
9. Чем отличается бот от человека? (2012) VC.RU URL: <https://vc.ru/flood/271-chem-otlichaetsya-bot-ot-cheloveka> (дата обращения: 17.02.2020).
10. Roberts S.B.G., Dunbar R.I.M. (2011) 'The costs of family and friends: an 18-month longitudinal study of relationship maintenance and decay. *Evol. Hum. Behav.*, 32 (2011), pp. 186-197
11. Фейковые аккаунты в соцсетях: что это и как их вычислить? // TexTerra URL: <https://texterra.ru/blog/kak-opredelit-feykovye-akkaunty-v-sotsialnykh-setyakh.html> (дата обращения: 17.02.2020).

References

1. Ryspayeva, M., Salykova O., & Ivanova I. (2020, April 29). Development level of Data Science in the Republic of Kazakhstan. *Science today: global challenges and development mechanisms: proceedings of an international scientific-practical conference.* – Vologda, 54-57
2. Ryspayeva, M., Salykova O., & Ivanova I. (2020, April 30). Review of Data Science development in the Republic of Kazakhstan. *Information Technologies in Management, Automation and Mechatronics: Collection of scientific papers of the 2nd International Scientific and Technical Conference*, Kursk. Southwestern State University, 7-11.
3. Bonchi, Francesco & Castillo, Carlos & Gionis, Aristides & Jaimes, Alejandro. (2011). *Social Network Analysis and Mining for Business Applications.* ACM TIST. (2)22, <https://doi:10.1145/1961189.1961194>.

4. Media Sphere of Kazakhstan – 2018, (2018). Ministry of Information and Social Development of the Republic of Kazakhstan. <https://qogam.gov.kz/ru/kategorii/rezultaty-provedennyh-nauchnyh-konsaltingovyh-socialnyh-analiticheskikh-issledovaniy> (date: 16.02.2020).
5. Order of the Chairman of the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan, No. 246 (2016, October 21). On approval of the Methodology for Information and Communication Technology Statistics.
6. Ministry of National Economy of the Republic of Kazakhstan Committee on Statistics (2020). 'The main goals of using the Internet by household members. <https://stat.gov.kz/> (date: 16.02.2020).
7. Pollet, T.V., Roberts, S.G.B., & Dunbar, R.I.M. (2011). Use of social network sites and instant messaging does not lead to increased offline social network size, or to emotionally closer relationships with offline network members. *Cyberpsychology, Behavior, and Social Networking*, 14, 253-258
8. Algorithm for determining bots and offers on VKontakte. (2018). <https://habr.com/ru/post/413855/> (date: 17.02.2020).
9. What is the difference between a bot and a person? (2012). <https://vc.ru/flood/271-chem-otlichaetsya-bot-ot-cheloveka> (date: 17.02.2020).
10. Roberts S.B.G., Dunbar R.I.M. (2011). The costs of family and friends: an 18-month longitudinal study of relationship maintenance and decay. *Evol. Hum. Behav.*, 32, 186-197.
11. Fake accounts in social networks: what is it and how to determine them? (2015). TexTerra <https://texterra.ru/blog/kak-opredelit-feykovye-akkaunty-v-sotsialnykh-setyakh.html>.