МРНТИ 20.53.19
УДК 81.93.29

https://doi.org/10.51889/2021-4.1728-7901.14

*Ш.Ж. Мусиралиева[1], Р.Қ. Оспанов[1], Қ.М. Шалабаев[1], Б.Ә. Саттар[1]**

[1]әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан
**e-mail: bibizere98@gmail.com*

TWITTER ӘЛЕУМЕТТІК ЖЕЛІСІНДЕГІ ДІНИ БАҒЫТТАҒЫ ПАРАҚШАЛАРДАН МӘТІНДЕРДІ ЖИНАУ ӘДІСІ

Аңдатпа

Қазіргі уақытта әлеуметтік желілер көптеген ақпараттың алаңы болып табылады. Пайдалы ақпаратты қоспағанда, әлеуметтік желілер заңсыз әрекеттер үшін ыңғайлы алаңға айналды. Әлеуметтік желілерде қауіп-қатерлерді анықтау және талдау жүйелерінің болмауына байланысты көбінесе күдікті белсенділік көлеңкеде қалады. Пайдаланушылар әлеуметтік желілерде жалған профильдерді құруы көптеген күдікті әрекеттердің пайда болуына әсер етеді. Зерттеу нәтижелеріне сәйкес жалған профильдерді әлеуметтік боттар немесе қылмыскерлер жалған жаңалықтарды насихаттау, ұлтаралық араздықты қоздыру немесе басқа ақпаратты тарату үшін қолдана алады. Мақалада пайдаланушы профиліндегі ақпаратты талдауға арналған тәсілдерге қысқаша шолу берілген. Зерттеу барысында Twitter әлеуметтік желісі қолданылды. Ашық ақпарат көзін барлау құралы қарастырылды. Әлеуметтік желінің бағдарламалау интерфейсіне қол жеткізу арқылы бастапқы деректерді алу әдістері көрсетілді.

Түйін сөздер: әлеуметтік желілер, белсенділік, анықтау жүйелері, қауіптерді талдау, пайдаланушы профилі.

Аннотация

Ш.Ж. Мусиралиева[1], Р.Қ. Оспанов[1], Қ.М. Шалабаев[1], Б.Ә. Саттар[1]
[1] Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

МЕТОД СБОРА ТЕКСТОВ С РЕЛИГИОЗНЫХ СТРАНИЦ В СОЦИАЛЬНОЙ СЕТИ TWITTER

В настоящее время социальные сети являются площадкой большого количества информации. За исключением полезной информации социальные сети стали удобной платформой для противозаконных действий. Зачастую подозрительная активность остается в тени ввиду отсутствия систем обнаружения и анализа угроз в социальных сетях. Создание пользователями поддельных профилей в социальных сетях влияет на появление множества подозрительных действий. Согласно результатам исследования, поддельные профили могут использоваться социальными ботами или преступниками для пропаганды поддельных новостей, разжигания межнациональной розни или распространения другой информации. В статье представлен краткий обзор подходов для анализа информации из профиля пользователя. В ходе исследования была использована социальная сеть Twitter. Рассматривался инструмент разведки открытого источника информации. Были продемонстрированы методы получения исходных данных путем доступа к интерфейсу программирования социальной сети.

Ключевые слова: социальные сети, активность, системы обнаружения, анализ угроз, профиль пользователя.

Abstract

METHOD OF COLLECTING TEXTS FROM RELIGIOUS PAGES ON TWITTER SOCIAL NETWORK

Mussiraliyeva Sh.Zh.[1], Ospanov R.K.[1], Shalabaev K.M.[1], Sattar B.A.[1]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan

Currently, social networks are a platform for a large amount of information. With the exception of useful information, social networks have become a convenient platform for illegal actions. Suspicious activity often remains in the shadows due to the lack of threat detection and analysis systems in social networks. The creation of fake profiles by users on social networks affects the appearance of many suspicious actions. According to the results of the study, fake profiles can be used by social bots or criminals to promote fake news, incite ethnic hatred or spread other information. The article provides a brief overview of approaches for analyzing information from a user profile. During the research, the social network Twitter was used. An open source intelligence tool was considered. Methods of obtaining initial data by accessing the programming interface of a social network were demonstrated.

Keywords: social networks, activity, detection systems, threat analysis, user profile.

Кіріспе

Әлеуметтік медиа сайттарын пайдаланушылар қалдыратын "сандық іздер" әлеуметтік ғылымдар саласындағы зерттеушілерге, сондай-ақ белсенді онлайн-пайдаланушылардың көпшілігін құрайтын маркетинг пен жастармен жұмыс істеу саясатына жаңа мүмкіндіктер ашады. Алайда, пайдаланушылар туралы деректерді жинау және талдау кезінде ескеру қажет маңызды фактор - олардың сауалнамаларында қамтылған ақпараттың толық болмауы болып табылады. Жынысы, жасы және тұрғылықты жері туралы ақпараттың болмауы, алынған желінің бұрмалануына әкелуі мүмкін және оны дұрыс түсіндіруге кедергі келтіруі мүмкін.

Әдебиеттерге шолу

ТМД-да ең көп қолданылатын әлеуметтік желі – 100 миллион белсенді қолданушысы бар ВКонтакте әлеуметтік желісі [1]. Әлеуметтік медиа әлемнің назарын өзінің миллиондаған қолданушыларымен байланыс орнату және байланыс орнату қабілетімен аударды. Әлеуметтік медианың әлеуетін көбіне ақпаратсыз пайдаланушылардан құпия ақпаратты шығаратын немесе әлеуметтік медианы заңсыз әрекеттер алаңы ретінде пайдаланатын шабуылдаушылар пайдаланады. Заңсыз әрекеттер үшін жасырындықты жасаудың ең кең таралған тәсілдерінің бірі - жалған профильдерді пайдалану, мұнда зиянды пайдаланушылар профильдегі жалған немесе нақты адамдар ретінде көрінеді. Зерттеулер саясаткерлердің, танымал адамдардың және бұқаралық ақпарат құралдарының парақтарындағы жалған аккаунттардың күдікті әрекеттері туралы хабарлайды [2].

Зерттеу нәтижелеріне сәйкес жалған профильдерді әлеуметтік боттар немесе қылмыскерлер жалған жаңалықтарды насихаттау, ұлтаралық араздықты қоздыру немесе басқа ақпаратты тарату үшін қолдана алады [3].

2012 жылы Facebook әлеуметтік желіс өз платформасында теріс қолданушылықты, соның ішінде жалған жаңалықтар, жек көрінішті сөздер, сенсация мен поляризация және басқа да ақпараттардың орналастырылғанын байқады [4]. Орын алған жағдай қауіпті әрі жалған ақпаратты таратуды алдын алудың жаңа әдістерін құрауды талап етті.

2015 жылы Facebook ай сайынғы белсенді қолданушыларының 14 миллионға жуығы қажет емес, веб-сайттың қызмет шарттарын бұза отырып жасалған зиянды, жалған аккаунттарды ұсынады деп бағалады [5].

Қазіргі жағдайда екі міндетті атап өтуге болады, олардың шешімін әлеуметтік желідегі белгілі бір параметрлерді талдау арқылы табуға болады. Пайдаланушылардың нақты тобындағы көшбасшыларды анықтау және профильдерді анықтау міндеттері.

Сілтемелерді анықтайтын және профильдерді тексеретін жүйені дамыту - қазіргі кездегі ең күрделі міндеттердің бірі.

Көбіне әлеуметтік желілерде іс-әрекетті талдау жүйесі қолданылады, соның арқасында пайдаланушы келісімін сол немесе басқа жолмен бұзған пайдаланушыларды анықтауға болады [6]. Күдікті әрекет Captcha тексеру механизмінің пайда болуына әкеледі.

Әлеуметтік желілердің мүмкіндіктері [7, 8] әр-түрлі қылмыстық іс-әрекеттерді жасау үшін пайдаланылған түрлі жағдайлар мысалдары қарастырылған:

- экстремистік немесе ұлтшыл идеяларды насихаттаушы қолданушыларды іздеу және қарым қатынас орнату, терроризм, есірткіні қолданушылар және басқа да девиантты көзқарасты қолданушылармен топтық байланыс орнату.

- зиян келтіруге қабілетті патогендік ақпаратты тарату, кәмелет жасқа толмағандарға арналған экстремистік бағыттағы ақпараттар, есірткіні насихаттайтын материалдарды тарату. Сонымен қатар деструктивті контенттің таралу жылдамдығы жоғары және көп жағдайда шын өмірдегі қарым қатынаспен тең болады, ал әлеуметтік желілерде жеке ақпаратты көрсету теріс ақпаратты мақсатты түрде тарату мүмкіндігін береді.

Әдістер мен құралдар

Орындалған жұмыста Twitter әлеуметтік желісінен жазылған хабарламаларды талдау үрдісі көрсетілген. Көрсетілген нұсқаулық Twitter қосымшасын құрастырып Python бағдарламалау тілінде арнайы сценарий жазу мүмкіндігін сипаттайды [2]. Құралған қосымша Twitter әлеуметтік желісінің API (Application programming interface) бағдарламалу интерфейсімен жұмыс істеуге арналған Tweepy кітапханасын қолданады. Twitter бағдарламалау интерфейсі тіркеу жазбасын басқарып әлеуметтік

медиа арқылы маңызды мәліметтерді алу мүмкіндігін береді. Талдау жүргізу үшін, хабарламаға қатысты келесі атрибуттар таңдалып алынды:

1. Хабарлама идентификаторы.
2. Хабарлама жасалу уақыты.
3. Хабарлама мәтіні.

Діни бағыттағы парақшалардан мәтіндерді алу үшін төменде көрсетілген құралдар қолданылды:

1. Twitter әлеуметтік желісінің бағдарламалау интерфейсі:
https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api
2. Python 3.6 бағдарламалау тілі.
3. Tweepy 3.10 бағдарламалау тілінің кітапханасы.

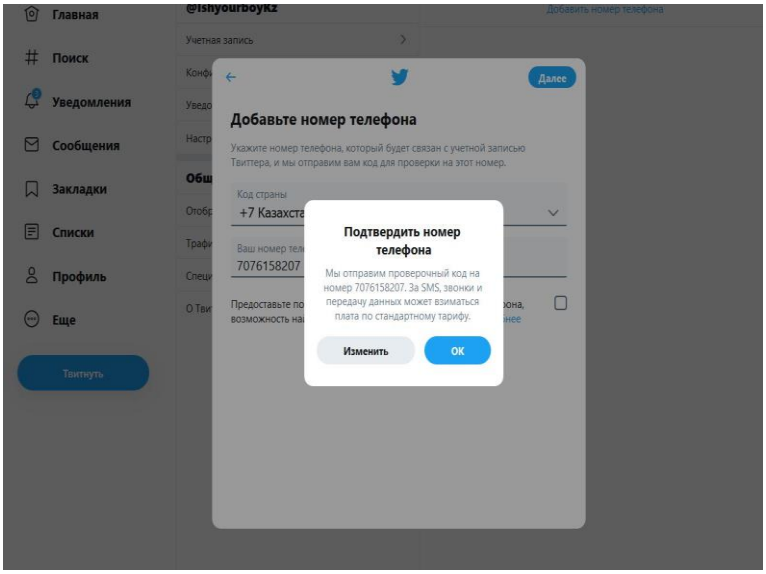
Tweepy кітапханасына сұраныстарды орындау үшін 1 - кестеде көрсетілген параметрлер қолданылды.

Кесте 1. Twitter API параметрлері

Параметр аты	*Параметр сипаттамасы*
bearer_token	*Twitter API рұқсат алушыға берілетін токен*
consumer_key	*Twitter API қолданушысының кілті*
consumer_secret	*Құпия параметр*
access_token	*Рұқсат токені*
access_token_secret	*Twitter API – не рұқсат токенінің құпиясы*

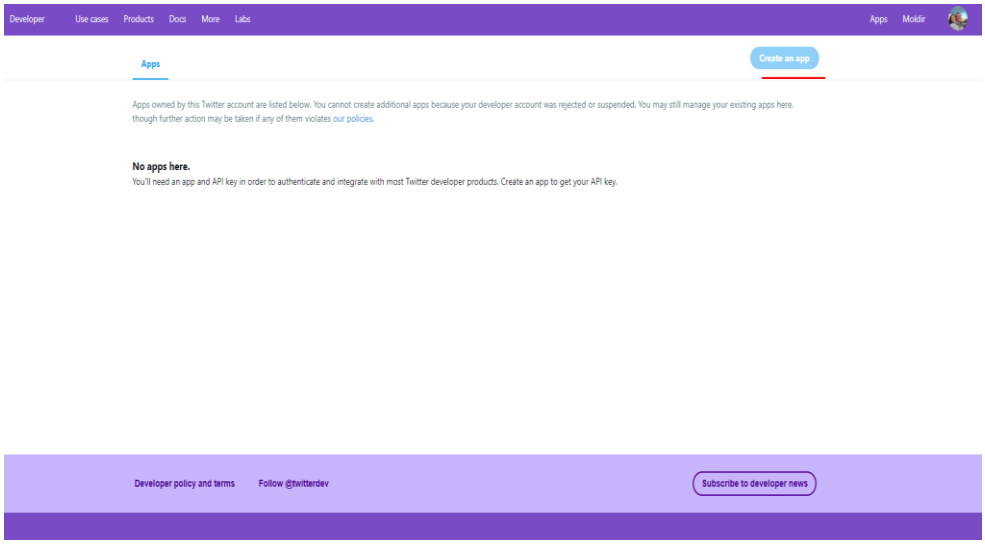
Twitter қосымшасын әзірлеуге қажетті әрекеттер

Қосымшаны тіркеу бетінде қосымшаны құрып кілт және API интерфейсіне рұқсат алу токенін алу қажет. Токен Twitter қосымшасының аутентификациясы үшін қажет. 1 суретте талаптар бөлімінде көрсетілгендей, токенды алу үшін расталған телефон нөмірі қажет болады.



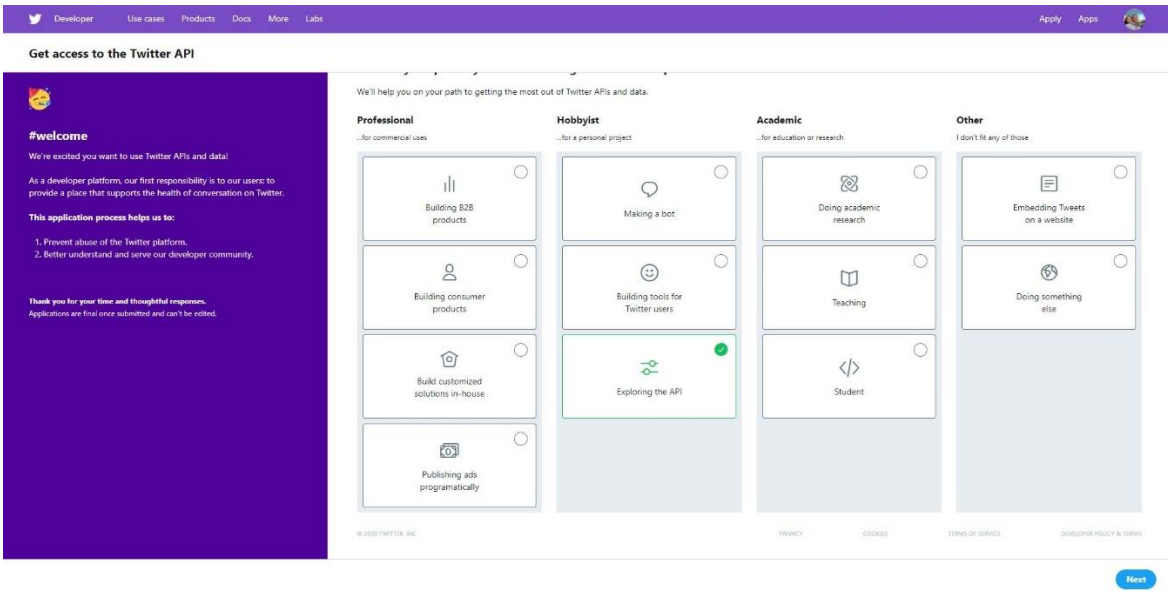
Сурет 1. Нөмірді растау терезесі

Аутентификацияны өту өшін Twitter тіркеу жазбасының деректерін еңгізу қажет және Create new app батырмасын басу қажет. Қосымашын құру терезесі 2- суретте көрсетілген.



Сурет 2. Қосымшаны құру терезесі

3 -суретте көрсетілгендей экранда қосымшаны әзірлеу парақшасы пайда болады.



Сурет 3. Қосымшаны әзірлеу парақшасы

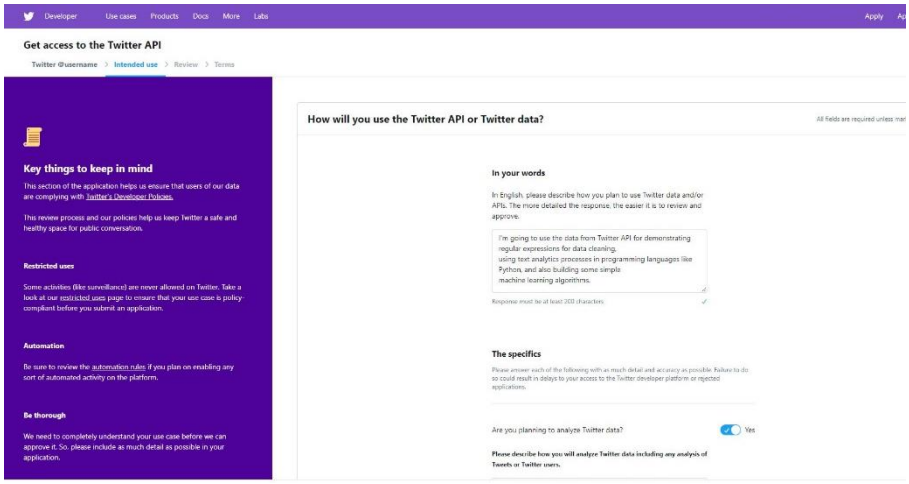
Жоғарыда көрсетілген әрекеттерді орындағаннан кейін Details парақшасын бағыттау орындалады. 4- суретте көрсетілген Details парақшасында қосымша жайлы жалпы ақпараттар көрсетілген.

Қосымшаға қолдану құқығын реттеу баптау үшін Details бетінен Permissions бетіне өту қажет.

Бастапқы баптаулар бойынша, Twitter қосымшаға оқу және жазу құқығын береді. Басқа жағдайда, қолдану рұқсатын өзімізге өзгертуге болады. Мұндай баптаулар қосымшаға сіздің атыңыздан хабарламалар жіберуге мүмкіндік береді.

Келесі қадам қосымшаның кілттері мен токенін генерациялау болып табылады. Ол үшін Keys and Access Tokens бетіне өту керек. Мұнда Consumer Key и Consumer Secret тізімін табуға болады. Сонымен қатар осы бетте Access Token и Access Token Secret кілттерін генерациялауға болады. Бұл кілттер мен токендер Twitter қосымшасының аутентификациясы үшін қажет.

Мәтіндік хабарламаларды жинақтау үшін Twitter әлеуметтік желісінен алынған кездейсоқ түрде таңдалған @ISLAMSEMYA парақшасы таңдалып алынды. Бастапқы талдау жасау үшін 10-08-2020 мен 08-05-2021 аралығында осы парақшада жасалған хабарламалар алынды, алынған хабарламалардың жалпы саны 3341 хабарлама болды.



Сурет 4. Details парақшасы

Алынған хабарламалардан діни бағыттағы кілттік сөздердің кездесу жиілігі 2 – кестеде көрсетілгендей талданды.

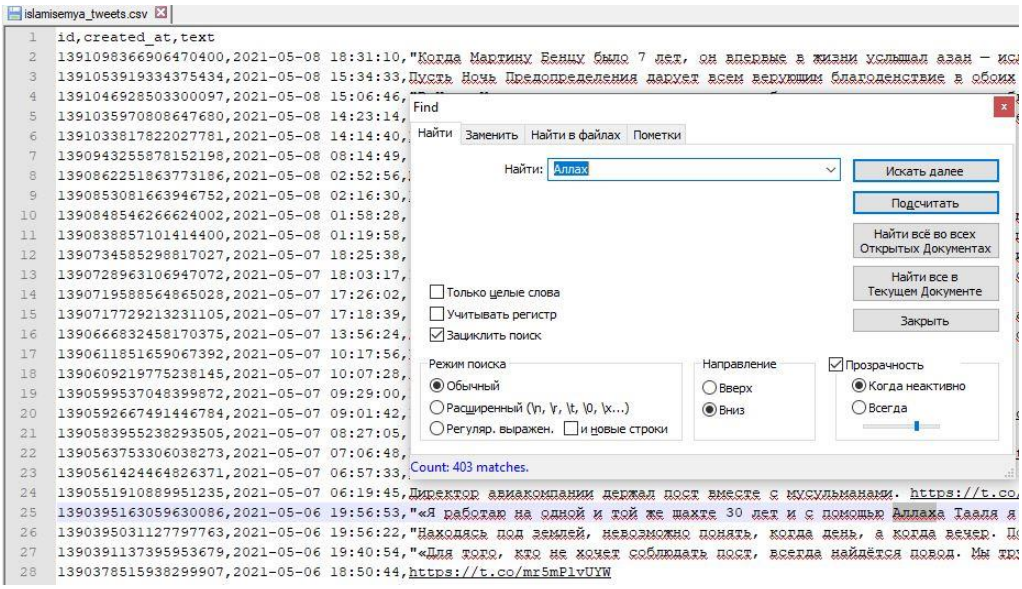
Кесте 2. Діни бағыттағы кілттік сөздердің кездесу жиілігі

Кілттік сөз (Key word)	*Кездесу жиілігі (Count)*
Мусульман	*718*
Аллах	*403*
Коран	*239*
Пророк	*140*

Ашық кодты орындау нәтижесінде, қолданылған код төменде көрсетілген ақпаратты қамтитын csv форматындағы файлды қалыптастырады:

1. Хабарлама идентификаторы.
2. Хабарламаны құру уақыты.
3. Хабарлама мәтіні.

Төменде көрсетілген 5-ші суретте ашық кодты орындау нәтижесінде қалыптасатын есептік файлдың данасы көрсетілген.



Сурет 5. Есептік файл нәтижесі

Зерттеу барысында алынған нәтижелер ең алдымен Twitter API бағдарламалау интерфейсіне рұқсат алу қиындықтары, әлеуметтік желінің геолокацияға байланысты қауіпсіздік саясатындағы шектеулерге байланысты екенін көрсетті. Ашық парақшалардағы хабарламаларды алу бағдарламалау тілінің қосымша кітапханаларын қолдану қажеттілігін тудырады.

Қорытынды

Орындалған зерттеу аясында Twitter парақшасындағы ашық атрибуттарды талдау әдісі қарастырылды. Ашық ақпарат көзін барлау құралы қарастырылды. Әлеуметтік желінің бағдарламалау интерфейсіне қол жеткізу арқылы бастапқы деректерді алу әдістері қарастырылған. Алынған нәтижелер құралдың дұрыстығын және нәтижелердің сенімділігін растайды. Әдісті әлеуметтік желілерді бақылау және кибер-криминалистика жүйесінің элементтерінің бірі ретінде пайдалануға болады. Бұл әдіс пен құралдар машиналық оқыту әдістеріне арналған корпусты қалыптастыруға, және нақты деректер (датасеттер) жиынтығын жинауға пайдалы болуы мүмкін.

Берілген мақала Қазақстан Республикасының цифрлық даму, инновациялар және аэроғарыш өнеркәсібі министрлігінің тапсырысы бойынша ғарыштық қызмет және ақпараттық қауіпсіздік саласындағы қолданбалы ғылыми зерттеулер бағытындағы "Мәтіндегі экстремистік бағытты анықтау үшін веб-ресурстардағы семантикалық талдау модельдерін, алгоритмдерін құрастыру және кибер-криминалистика құрал-жабдықтарын әзірлеу" жобасы аясында жазылды, ЖТН AP06851248.

Пайдаланылған әдебиеттер тізімі:

1 Craig Smith. VK Statistics and Facts. // [Электронный ресурс] / URL: https://expandedramblings.com/index.php/vk-statistics-facts/ (дата обращения: 01.07.2020).

2 Cresci S., Di Pietro R., Petrocchi M., Spognardi A., Tesconi M., Fame for sale: Efficient detection of fake Twitter followers. [Текст]. / Decision Support Systems – 2015.

3 Ferrara E., Varol O., Davis C., Menczer F., and Flammini A., The Rise of Social Bots. [Текст]. / Communications of the ACM – 2016.

4 Facebook shares drop on news of fake accounts. // [Электронный ресурс] / URL: https://www.cbc.ca/news/technology/facebook-shares-drop-on-news-of-fake-accounts-1.1177067 (дата обращения: 01.07.2020).

5 Ripeanu, K. Beznosov, and E. Santos-Neto, "Thwart-ing fake osn accounts by predicting their victims, inProceedings of the 8th ACM Workshop on Artificial Intelligence and Security. [Текст]. /ACM – 2015.

6 Чесноков В.О., Применение алгоритма выделения сообществ в информационном противоборстве в социальных сетях. [Текст]. / Вопросы кибербезопасности No1(19) – 2017.

7 Левшиц Н. Г. Участились случаи мошенничества в социальных сетях при сборе средств на лечение. – М., 2015. // [Электронный ресурс] / URL: http://echo.msk.ru/blog/nlevshits/1519362- echo/ (дата обращения: 10.04.16)

8 Соколова А. А. Средний ущерб от мошенников в социальных сетях». – М., 2015. // [Электронный ресурс] / URL: http://rusbase.com/news/internet-fraud-statistics/ (дата обращения: 10.04.16).

References:

1 Craig Smith. VK Statistics and Facts. // [Elektronnyj resurs] / URL: https://expandedramblings.com/index.php/vk-statistics-facts/.

2 Cresci S., Di Pietro R., Petrocchi M., Spognardi A., Tesconi M., Fame for sale: Efficient detection of fake Twitter followers. [Tekst]. / Decision Support Systems – 2015.

3 Ferrara E., Varol O., Davis C., Menczer F., and Flammini A., The Rise of Social Bots. [Tekst]. / Communications of the ACM – 2016.

4 Facebook shares drop on news of fake accounts. // [Elektronnyj resurs] / URL: https://www.cbc.ca/news/technology/facebook-shares-drop-on-news-of-fake-accounts-1.1177067 .

5 Ripeanu, K. Beznosov, and E. Santos-Neto, "Thwart-ing fake osn accounts by predicting their victims, inProceedings of the 8th ACM Workshop on Artificial Intelligence and Security. [Tekst]. /ACM – 2015.

6 Chesnokov V.O., (2017) Primenenie algoritma vydelenija soobshhestv v informacionnom protivoborstve v social'nyh setjah [Application of the algorithm for identifying communities in information warfare in social networks]. [Tekst]. Voprosy kiberbezopasnosti No1(19). (In Russian)

7 Levshic N. G. (2015) Uchastilis' sluchai moshennichestva v social'nyh setjah pri sbore sredstv na lechenie More frequent cases of fraud in social networks when collecting funds for treatment. M., [Elektronnyj resurs] / URL: http://echo.msk.ru/blog/nlevshits/1519362- echo/.(In Russian)