

[4] Зотиков В.И., Наумкина Т.С., Грядунова Н.В., Сидоренко В.С., Наумкин В.В. Зернобобовые культуры – важный фактор устойчивого экологически ориентированного сельского хозяйства // Зернобобовые и крупяные культуры. - 2016. - №1.- С.6-13.

[5] Германцева Н.И. Нут - культура больших возможностей // Теоретические и прикладные аспекты современной науки. - 2014. - № 4-1. - С. 50-53.

[6] Тарасенко Н.А., Баранова З.А., Быкова Н.С., Третьякова Н.Р. Использование пищевых волокон в функциональных кондитерских изделиях // Успехи современного естествознания. – 2016. – № 11. – С. 86-90;

[7] Химический состав российских пищевых продуктов: Справочник / Под ред. член-корр. МАИ, проф. И. М. Скурихина и академика РАМН, проф. В. А. Тутельяна. - Х46 М.: ДеЛи принт, 2002. - 236 с.

[8] Дайрашева С.Т., Мулдабекова Б.Ж., Рустемова А.Ж. «Кондитер өндірісінің технологиясы» пәнінен зертханалық жұмыстардың әдістемелері. Тамақ өндірісінің жоғары оқу орындарының студенттеріне арналған оқу әдістемелік құралы. Алматы, АТУ, 2012. – 36 б.

[9] Мулдабекова Б.Ж., Исакова Г.К., Бек Р.Б. Нан, макарон және кондитер өндірісі технокимиялық бақылау пәніне арналған лабораториялық оқу құралы. -Алматы: «Полиграфия-сервис и К°», 2016. – 114 б.

Мухарбек А.С., Мулдабекова Б.Ж., Исакова Г.К., Тенгельбаева А.А.

Исследование влияния нутовой муки на качество сахарного печенья

Резюме. Использование нового нетрадиционного сырья с целью повышения ценности мучных кондитерских изделий является одним из основных научно-практических направлений. Таким образом, проведены исследования для изготовления мучных кондитерских изделий и обогащения их состава белками, витаминами и минеральными веществами. В этом аспекте использовалось самое ценное зернобобовые культуры. Создана рецептура приготовления мучных кондитерских изделий. Исследованы органолептические и физико-химические показатели печенья с нутовой мукой.

Ключевые слова: мучные кондитерские изделия, нутовая мука, пищевая ценность, химический состав, пшеничная мука.

ӘОЖ 81.93.29

K. Shalabayev, K. Alipbay, M. Bolatbek, Sh. Mussiraliyeva

(Al-Farabi Kazakh National University, Almaty, Kazakhstan)

E-mail: mussiraliyevash@gmail.com

DEFINITION AND CLASSIFICATION OF EXTREMIST TEXTS IN VKONTAKTE SOCIAL NETWORK

Abstract. Online extremists in social networks pose a new threat to society. In this paper, the authors identify the content of the extremist grouping of incoming text using machine-learning methods, analyzing the content of extremist texts in the Kazakh and Russian languages, analyzing the previously defined groups in the social network Vkontakte. An algorithm for obtaining social network indicators has been developed. For classification, a corpus consisting of social network texts was created, and the analysis was performed using classification methods: Bayesian, support vector machine, logistic regression, etc.

Keywords. Extremism, trustful qualifier Bayes, Methods of Machine Learning, classification of Texts.

К. Шалабаев, К. Әліпбай, М. Болатбек, Ш. Мусиралиева

(Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан Республикасы)

E-mail: mussiraliyevash@gmail.com

ВКОНТАКТЕ ӘЛЕУМЕТТІК ЖЕЛІСІНДЕГІ ЭКСТРЕМИСТІК МӘТІНДЕРДІ АНЫҚТАУ ЖӘНЕ ЖІКТЕУ

Аннотация. Әлеуметтік желілердегі онлайн-экстремистер қоғам үшін жаңа қауіп болып табылады. Бұл жұмыста авторлар қазақ және орыс тіліндегі экстремистік мәтіндердің мазмұнын анықтау барысында Вконтакте әлеуметтік желісіндегі алдын ала анықталған топтарға талдау жасай отырып, машиналық оқыту әдістер көмегімен кіріс мәтіннің экстремистік топқа жату дәлдігін анықтайды. Әлеуметтік желідегі көрсеткіштерді алу алгоритмі құрылды. Жіктеу үшін әлеуметтік желіден алынған мәтіндерден тұратын корпус құрылып, аңқау Байес жіктеуіші, тірек векторлар машинасы, логистикалық регрессия, жіктеу ағаштары және кездейсоқ орман жіктеуіші сияқты машиналық оқыту әдістері пайдаланылды.

Түйінді сөздер. Экстремизм, аңқау Байес жіктеуіші, машиналық оқыту әдістері, мәтінді жіктеу.

КІРІСПЕ

Соңғы жылдарда әлеуметтік желі қолданушыларына қысым көрсету, жаңа мүшелерді тарту және зорлық-зомбылықты күшейту үшін құрылған онлайн-экстремистік топтардың саны мен көлемінің ұлғаюы байқалуда. Бұл топтарға Ирак және Сирия Ислам мемлекеті (ISIS) [1] ұлтшылдар және ұлтшылдардың жақтаушылары [2] сияқты, сондай-ақ, адамдарға қорқыту жасайтын кибер-содырлар жатады [3]. Лаңкестік топтардың қоғам қауіпсіздігіне қауіп төндіруі ерекше алаңдаушылық тудырады. ISIS (ИГИЛ) секілді террористік топтардың қауіп қатты болғаны соншалықты, АҚШ-тың экс-президенті Барак Обама «Құрама Штаттар осыдан былай серіктестермен бірге ИГИЛ-дің жек көрушілік насихатына, әсіресе желіде қарсыласуына қарсы әрекет жасауға өз үлесімізді қосамыз» деген болатын [4]. ИГИЛ-дің ғаламторда болуы жеке тұлғаларды радикалдандыруға және террор актілерін жасауға ынталандыруы мүмкін деген болжам бар [5].

Әлеуметтік желі жақында желідегі экстремистермен белсенді күресу үшін шаралар қабылдауға кірісті. Мысалы, ИГИЛ өз насихатын тарататын басты орынға айналған Твиттер ИГИЛ-ге өте агрессивті жауап беруде. 2016 жылдың тамыз айында Twitter 360,000-нан астам ИГИЛ-дің тіркеу-жазбаларын жауып тастаған және оның лаңкестікке байланысты тіркеу-жазбаларын күнделікті тоқтатылуы, 2015 жылдан бастап 80 пайызға жетті. Twitter экстремистік аккаунттарды анықтауда, бірінші кезекте оның пайдаланушылардың есептеріне негізделеді, бірақ ол бұл есептерді толтыру үшін арнайы спамға қарсы құралдарды қолдануға кірісті. Бұл құралдар Твиттерде лаңкестікке жол бермеу үшін тоқтатылған тіркеу-жазбалары үштен бірінен астамын автоматты түрде анықтауға көмектесті [6].

Twitter сияқты әлеуметтік желілердің күштері ИГИЛ секілді онлайн-экстремистік топтардың мүмкіндіктерін шектеуде тиімді. Дегенмен, барлық экстремистік қолданушылар тоқтатылмайды және тоқтатылғаннан кейін олар үнемі әлеуметтік желіге қайта оралады. Сонымен қатар, экстремистік топтардың қатерін азайтудағы табыстың көп бөлігі әлеуметтік желілердің өзара әрекеттесуіне негізделген. Мысалы, Твиттерде экстремистік аккаунттар жайында пайдаланушылардың есептерін қарауға арналған арнайы топтар бар [7]. Алайда, экстремистік пайдаланушылар басқа әлеуметтік желілерге көшіп жатса, бұл желілерді басқаратын компаниялар онлайн-экстремистермен күресу үшін осыншама күш жұмсауға немесе ресурстар бөлетіндігіне еш кепілдік бермейді. Сондықтан, билік органдарына әлеуметтік желі операторларының ынтымақтастығына сүйенбейтін онлайн-экстремистермен күресуге және кез келген әлеуметтік желіге қолдануға болатын мүмкіндіктер жиынтығы қажет.

Әдебиеттерге шолу

Лаңкестік топтардың желілеріндегі экстремистік бағыттағы ұйымдардың іс-қимылы мен мақсаттарын талдау үшін айтарлықтай зерттеулер жүргізілді. Ғалымдар [8] әлеуметтік желілердегі мәліметтерге интеллектуалды талдау, зерттеу тәсілдерін жүргізеді. Келесі ғылыми еңбекте [9] лаңкестік желілердегі жасырын мәтінді анықтау тәсілдері қарастырылған. Ғалымдардың бірігіп жасаған еңбегінде [10] ғаламтордағы жихадқа зерттеулер келтірілген және ұлттық қауіпсіздіктің киберлық инфрақұрылымы [11] жан-жақты зерттелген. Дегенмен, веб-форумдардың мазмұнын талдауға арналған зерттеулер жоқтың қасы. Келесі зерттеуде [12] ағылшын және араб тілдеріндегі сөйлесуді талдау үшін энтропияға арналған генетикалық алгоритммен (EWGA) синтаксистік, стилистикалық ерекшеліктер мен олардың үйлесімділігін зерттеген. Көптеген зерттеушілердің еңбектері ұйымдардың белгілі бір оқиғаларға және т.б. қатысты жалпы жұртшылық пікірлерін талдауға арналған.

Ғалымдардың [13-15] еңбектерінде жекеленген хабарламаларды талдау үшін, сондай-ақ тұтастай құжатта және құжаттар жинақтарында пікір қорытындыларын шығару үшін пікір білдіруге және сенімділікті талдау әдістеріне және тілдік үлгілерге терең шолу жасалған. Берілген жұмыстың мақсаты дискурстың автоматтандырылған синтаксистік, семантикалық және статистикалық талдауына (блогтар, форумдар және әлеуметтік желілердегі пікірталастар), стилі мен тақырыбына қарамастан, әлеуметтік шиеленістің болуын анықтау болып табылады.

Әлеуметтік желідегі көрсеткіштерді алу алгоритмі

Әлеуметтік желінің API функцияларын бағдарламада пайдалану үшін алдымен жүйеге тіркелуіміз керек.

Біздің бағдарламада қолданылатын қосымша түрі - Standalone. ВКонтакте API функцияларын кез-келген Standalone қосымшалардан қолдану үшін, OAuth 2.0 хаттамасына негізделген клиентті авторизациялау механизмін өтіміз керек. Клиент кез келген веб-браузерді қолдану мүмкіндігі бар Desktop / мобильді қосымша болуы мүмкін.

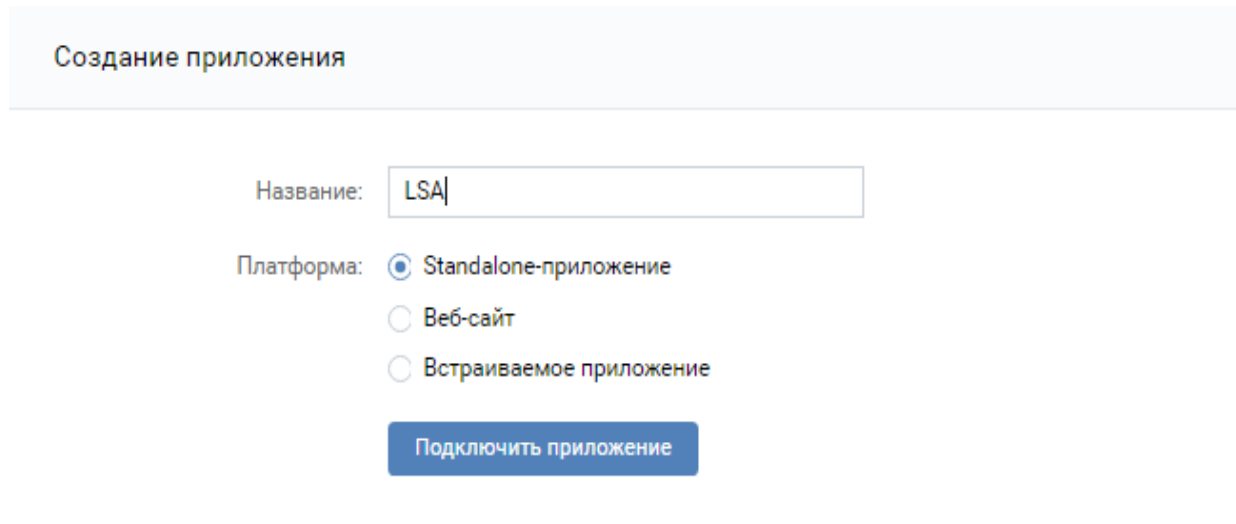
Қосымшаны тіркеу үрдісі үш қадамнан тұрады:

• Технические науки

1. Вконтакте developers желісінде аутентификация үрдісін өту.
2. Қолданушыларға қосымша мәліметтерін қолдануға рұқсат беру.
3. API қолдану үшін access_token кілтін қосымшада қолдану.

Вконтакте developers желісінде аутентификации үрдісін өту.

Standalone-қосымшасын құру үшін Вконтакте developers парақшасына өтеміз. Standalone-қосымшасын құру терезесін 1-суреттен көруге болады.



Создание приложения

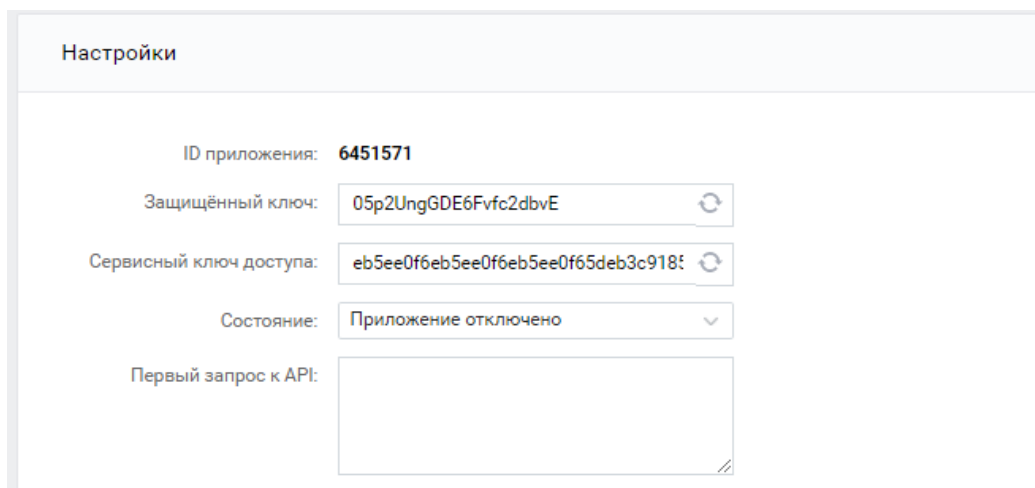
Название: LSA

Платформа: Standalone-приложение
 Веб-сайт
 Встраиваемое приложение

Подключить приложение

1-сурет. Standalone қосымшасын құру терезесі

Қосымшаны авторизациялау үрдісі кезінде API қолдану рұқсат кілті access_token-ді ашпау керек. Сондықтан бағдарлама кодының ортасында құпия кілті көрсетуден сақтану керек, себебі қылмыскерлер оны бағдарламаны декомпиляциялау арқылы ала алады. Қосымша баптауларын 2-суреттен көруге болады.



Настройки

ID приложения: 6451571

Защищённый ключ: 05p2UngGDE6Fvfc2dbvE

Сервисный ключ доступа: eb5ee0f6eb5ee0f6eb5ee0f65deb3c918:

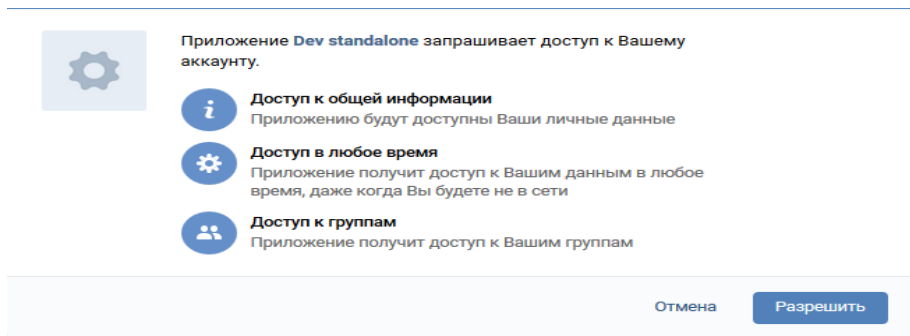
Состояние: Приложение отключено

Первый запрос к API:

2-сурет. Қосымша баптаулары

Қатынау құқықтарын беру

Қосымшаның сәтті авторизациясынан кейін, пайдаланушы браузері авторизация диалогын ашқан кезде көрсетілген REDIRECT_URI мекенжайына қайта бағытталады. Бұл жағдайда access_token API және басқа параметрлерге қатынау кілті сілтеме URL фрагментінде жіберіледі. Қатынау құқықтары терезесін 3-суреттен көруге болады.



3-сурет. Қатынау құқықтары терезесі

Пайдаланушы кіру немесе құпия сөзді өзгертсе немесе қол жеткізу параметрлерінде бағдарламаны жойған кезде access_token сұрауы қажет. Сонымен қатар, қайтарылған параметрлердің ішінен user_id - әлеуметтік желіде авторизацияланған пайдаланушының идентификаторы көрсетіледі.

Авторизация қатесі орын алғанда, бұл қате туралы ақпарат GET параметрлері ретінде REDIRECT_URI-ге беріледі:

http: // REDIRECT_URI? error = access_denied & error_description = The + user + or + authorization + server + denied + the + request.

Сәтті авторизациядан кейін API-ға сұраулар жасай аламыз.

Vkontakte API әдістерін шақыру үшін HTTPS протоколы арқылы көрсетілген URL мекенжайына POST немесе GET сұрауын жасау қажет:

https://api.vk.com/method/METHOD_NAME?PARAMETERS& access_token = ACCESS_TOKEN

METHOD_NAME - API функцияларының тізімінен әдіс атаулары,

PARAMETERS - API сәйкес әдісінің параметрі,

ACCESS_TOKEN – қосымшаны дұрыс авторизациялауы нәтижесінде алынған кілт.

Мысалы:

https://api.vk.com/method/getProfiles?uid=66748& access_token = 533bacf01e11f55b536a565b57531ac114461ae8736d6506a3

Мұндай сұраудың жауабы JSON форматында қабылданады.

API параметрлерін қолдану

Параметрлер GET және POST әдістерін пайдалануы мүмкін. Егер үлкен көлемдегі деректерді (2 килобайтты) аударсаңыз, POST әдісін пайдаланған жөн.

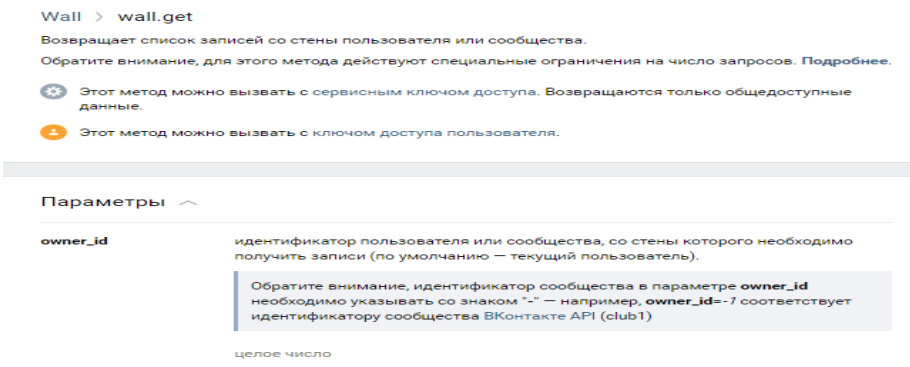
Әрбір әдіске қолдау көрсетілетін параметрлердің өз жиынтығы бар, бірақ барлық әдістер қабылдайтын параметрлер бар:

lang - әртүрлі деректер қайтарылатын тілді, мысалы, елдер мен қалалардың атауларын анықтайды;

v - пайдаланылатын API нұсқасын көрсетеді, мысалы, 4.95;

https - 1 - фотосуреттерге және басқа медиаларға https сілтемелерін алуға мүмкіндік береді. 0 - әдістері http сілтемелерін қайтарады.

Жалпы топтардың қабырғасынан деректерді алу үшін wall.get әдісі қолданылады. Wall.get әдісі терезесін 4-суреттен көруге болады.



4-сурет. Wall.get әдісінің терезесі

Вконтакте әлеуметтік желісіндегі топтардың деректерін алу кодының үзіндісі:

```
import requests
import json
from datetime import datetime
from time import sleep
def write_json(data):
    with open('New1.json','w',encoding='utf8') as file:
        json.dump(data, file, indent=2, ensure_ascii=False)
def main():
    start = datetime.now()
    group_id = '-30666517'
    tokenzi = '16b837e316b837e316b837e34616de002a116b816b837e34d1cb7902b7237d5c464f907'
    offset = 0
    date_x = 1516947754
    all_posts = []
    while True:
        sleep(1)
        r=requests.get('https://api.vk.com/method/wall.get?&v=5.71',
params={'owner_id':group_id,'access_token': tokenzi,'count': 100, 'offset':offset})
        posts = r.json()['response']['items']
        all_posts.extend(posts)
        oldest_post_date = posts[-1]['date']
        offset+=100
        print(offset)
        if oldest_post_date < date_x:
            break
        write_json(r.json())
if __name__ == '__main__':
    main()
```

JSON форматта қайтарылатын деректер түрі:

```
"response": {
  "count": 11572,
  "items": [{
    "id": 28508,
    "from_id": -90689018,
    "owner_id": -90689018,
    "date": 1525257849,
    "marked_as_ads": 0,
    "post_type": "post",
    "text": "Хазреті Әлидің дәрежесі \n \nБір күні Сахабалар хазреті Әлидің мәртебесін білу үшін
Пайғамбарымыздан былай деп сұрады: \n- Ей Расулаллах!",
    "attachments": [{
      "comments": {
        "count": 0,},
      "likes": {
        "count": 13,
        "user_likes": 0,},
      "reposts": {
        "count": 0,
        "user_reposted": 0},
      "views": {
        "count": 274}
    },.....}
}
```

"Вконтакте" әлеуметтік желісіндегі экстремистік және діни топтарға талдау жүргізілді. 07.11.2018-29.01.2019 аралығында "Вконтакте" желісінде 170-ке жуық топтар талқыланды. Олардың ішінде 25 топтар бағдарламада қолданылды:

1. Alhamdulillah | Ислам
2. Allahu Akbar
3. Анти Вахабизм
4. Ислам және отбасы
5. Дінім Ислам
6. KZ Islam
7. Ислам діні
8. Ислам Жұмаққа жол



Жіктеу нәтижесінде берілген кіріс корпустағы мәтіндердің экстремистік топқа жату дәлдігі 1-кестеде келтірілген.

1-кесте. Кіріс мәтіндерін жіктеу нәтижелері

Машиналық оқыту әдісі	Дәлдік
Тірек векторлар машинасы	0,95
АңқауБайес жіктеуіші	0,96
Логистикалық регрессия	0,99
Жіктеу ағаштары	0,92
Кездейсоқ орман жіктеуіші	0,37

Кіріс мәтіндердің экстремистік топқа жату дәлдігімен қатар ең жиі кездесетін терминдер тізімі 2-кестеде көрсетілген.

2-кесте. Қазақ тіліндегі экстремистік мәтіндерде жиі кездесетін терминдер

Термин	Жиілігі	Термин	Жиілігі
Мұхаммад	20	саяфиттер	4
Аллах	18	Тағала	11
Джихад	8	уәссәләм	7
Уәссәләту	7	Алла	16
Ислам	9	бауырлар	8
кәпірлер	4	ахи	4

Қорытынды

Берілген жұмыста қазақ және орыс тіліндегі мәтіндерге талдау жасалып, оларға машиналық оқыту әдістері көмегімен жіктеу жүргізілді.

Мәтіндер корпусын жинау үшін Вконтакте әлеуметтік желісі қолданылды.

Әлеуметтік желідегі көрсеткіштерді алу алгоритмі құрылды және жүйеге сәтті тіркелінді. Жіктеуіштердің көмегімен ең жоғары дәлдікті көрсететін жіктеуіштер анықталды.

ӘДЕБИЕТ

- [1] Berger J. and Jonathan Morgan. The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter. The Brookings Project on US Relations with the Islamic World, 3:20, 2015.
- [2] Amar Toor. Twitter may be cracking down on ISIS, but white nationalists are still thriving. <http://www.theverge.com/2016/9/5/12798196/twitter-nazi-white-nationalist-isis-study>, September 2015. Accessed: 2016-10-13.
- [3] Kristen V. Brown. Twitter just permanently suspended Milo Yiannopoulos, the internet's biggest troll. <http://fusion.net/story/327536/milo-yiannopoulos-nero-permanently-banned-from-twitter/>, July 2016. Accessed: 2016-10-13.
- [4] Scott Higham. Why the Islamic State leaves tech companies torn between free speech and security. https://www.washingtonpost.com/world/national-security/islamic-states-embrace-of-social-media-puts-tech-companies-in-a-bind/2015/07/15/0e5624c4-169c-11e5-89f3-61410da94eb1_story.html?kmap=1, July 2015. Accessed: 2016-10-13.
- [5] Russell Brandom. President Obama says Orlando killer was inspired by online extremism. <http://www.theverge.com/2016/6/13/11922034/orlando-attack-barack-obama-briefing-isis-internet-terrorism>, July 2016. Accessed: 2016-10-13.
- [6] Twitter. An update on our efforts to combat violent extremism. <https://blog.twitter.com/2016/an-update-on-our-efforts-to-combat-violent-extremism>, August 2016. Accessed: 2016-10-13.
- [7] Yasmeen Abutaleb. Twitter suspended 360,000 accounts for 'promotion of terrorism'. <http://www.reuters.com/article/us-twitter-terrorism-dUSKCN10T1ST>, August 2016. Accessed: 2016-10-13.
- [8] Memon N., Hu J., Hicks D., Chen H., "Social Network Data Mining: Research Questions, Techniques, and Applications", I: Data Mining for Social Network Data. red. / N.h Memon ; J. Xu ; D. Hicks ; H. Chen. Vol. 12 Springer Publishing Company, 2010, pp.1-7.
- [9] Memon N., Larsen H., Hicks D. and Harkiolakis N., "Detecting Hidden Hierarchy in Terrorist Networks: Some Case Studies", Intelligence and Security Informatics, Springer, pp. 477-489, 2008.
- [10] Chen H., Chung W., Qin J., Reid E., Sageman M., Weimann G., "Uncovering the dark Web: A case study of Jihad on the Web" in Journal of the American Society for Information Science and Technology, vol. 59 Issue 8, June 2008.
- [11] Raghu T. S., Chen H., "Cyberinfrastructure for homeland security: Advances in information sharing, data mining, and collaboration systems" in Journal: Decision Support Systems - DSS, vol. 43, no. 4, pp. 1321-1323, 2007.
- [12] Abbasi A., Chen H., Salem A., "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums", in ACM Transactions on Information Systems (TOIS), vol. 26, Issue 3, June 2008
- [13] Bing Liu. "Sentiment Analysis: A Multifaceted Problem." Invited contribution to IEEE Intelligent Systems, 2010.
- [14] Bollen J., Mao H., and Zeng X.-J. Twitter mood predicts the stock market // Technical Report arXiv:1010.3003, CoRR, 2010. Http: <http://arxiv.org/pdf/1010.3003v1>
- [15] Pang B. and Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval, v.2 n.1-2, January, 2008 - pp.1-135.
- [16] Moreau, L., Groth, P., Miles, S., Vazquez-Salceda, J., Ibbotson, J., Jiang, S., Munroe, S., Rana, O., Schreiber, A., Tan, V., Varga, L.: The provenance of electronic data. Commun. ACM 51(4), 52–58 (2008)
- [17] Groth, P., Jiang, S., Miles, S., Munroe, S., Tan, V., Tsasakou, S., Moreau, L.: An Architecture for Provenance Systems. Technical report, University of Southampton (2006)
- [18] Ukustov, S., S. Fomenkov and A. Polovinkin, 2013. of the Russian Fundamental Research Fund. Wrong metrics of success for open source software
- [19] Witten, L.H., Frank, E., Hall, M.A.: Data mining: practical machine learning tools and techniques, Amsterdam, p. 560 (2005)
- [20] Foltz, P. W. 1996. Latent Semantic Analysis for text-based research. Behavior Research Methods, Instruments and Computers. 28 197-202.
- [21] Groth, P., Tan, V., Munroe, S., Jiang, S., Miles, S., Moreau, L.: Process Documentation Recording Protocol. Technical report, University of Southampton (2006)

Шалабаев К., Алипбай К., Булатбек М., Мусиралиева Ш.

Определение и классификация экстремистских текстов в социальной сети Вконтакте

Резюме. Онлайн экстремисты в социальных сетях представляют новую угрозу для общества. В данной работе авторы идентифицируют содержание экстремистской группировки входящего текста с помощью методов машинного обучения, анализируя содержание экстремистских текстов на казахском и русском языках, анализируя заранее определенные группы в социальной сети Вконтакте. Разработан алгоритм получения индикаторов социальных сетей. Для классификации было создан корпус, состоящий из текстов социальных сетей, и выполнен анализ с использованием методов классификации: Байеса, машины опорных векторов, логистической регрессии и др.

Ключевые слова. Экстремизм, доверчивые классификатор Байес, методы машинного обучения, классификация текстов.