ҒТАХР 81.93.29

10.51889/2959-5894.2023.84.4.016

С.Т. Мамбетов[1], Е.Е. Бегимбаева[1,2], А.К. Хикметов[3],*
С.К. Джолдасбаев[3], Г.Н. Казбекова[4]

[1]Әль-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан
[2]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы қ., Қазақстан
[3]Халықаралық Ақпараттық технологиялар университеті, Алматы қ., Қазақстан
[4]Қожа Ахмет Яссауи атындағы Халықаралық Қазақ-Түрік университеті, Түркістан қ., Қазақстан
[*]e-mail: mambetov.saken@gmail.com

ХАКЕРЛІК ИНТЕРНЕТ-ФОРУМДАРДЫ СЕНТИМЕНТ-ТАЛДАУ

Аңдатпа

Бұл мақала хакерлік интернет форумдар контекстінде сентимент талдау әдістеріне шолу мен зерттеуді ұсынады. Хакерлік форумдар компьютерлік қауіпсіздік туралы ақпарат пен тәжірибе алмасуға арналған алаң болып табылады және мұндай форумдардағы хабарламалардың эмоционалды бояуын талдау хакерлік қауымдастықтар мен олардың қызметін зерттеу үшін пайдалы түсініктер бере алады. Мақалада әртүрлі әдістер мен тәсілдер қарастырылады. Қарастырылған әдістерге сентимент сөздігі арқылы, машиналық оқыту, эмоцияларды талдау, сондай-ақ осы тәсілдерді біріктіретін гибридті модельдер жатады. Әр әдістің принциптері мен әдістемелері, сондай-ақ олардың артықшылықтары мен шектеулері егжей-тегжейлі сипатталған. Мақалада хакерлік форумдарға сентимент талдауын қолдану кезінде алынған нәтижелер мен қорытындылар талқыланады. Зерттеу нәтижелері хакерлер қауымдастығындағы эмоционалды компонентті жақсырақ түсінуге, трендтер мен көңіл-күйлерді анықтауға және ықтимал қауіпті әрекеттерді немесе тенденцияларды анықтауға мүмкіндік береді. Мақаланың қорытындысында жалпы қорытындылар шығарылып, хакерлік интернет форумдарды талдау саласындағы одан әрі зерттеулердің ықтимал бағыттары көрсетіледі. Сондай-ақ қауіпсіздікті арттыру және киберқылмысқа қарсы іс-қимыл мақсатында талдау нәтижелерін пайдалану бойынша ұсыныстар беріледі.

Түйін сөздер: сентимент талдау, интернет форумдар, киберқауіпсіздік, даркнет.

С.Т. Мамбетов[1], Е.Е. Бегимбаева[1,2], А.К. Хикметов[3], С.К. Джолдасбаев[3], Г.Н. Казбекова[4]
[1]Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан
[2] Казахский национальный исследовательский технический университет им. К. И. Сатпаева, г. Алматы, Казахстан
[3]Международный университет информационных технологий, г. Алматы, Казахстан
[4]Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан

СЕНТИМЕНТ-АНАЛИЗ ХАКЕРСКИХ ИНТЕРНЕТ-ФОРУМОВ

Аннотация

В статье представлен обзор и исследование методов анализа настроений в контексте взлома интернет-форумов. Хакерские форумы – это форум для обмена информацией и опытом в области компьютерной безопасности, и анализ эмоциональной окраски сообщений на таких форумах может дать полезную информацию об изучении хакерских сообществ и их деятельности. В статье рассматриваются различные методы и подходы. Рассматриваемые методы включают словарь настроений, машинное обучение, анализ настроений, а также гибридные модели, объединяющие эти подходы. Подробно описаны принципы и методологии каждого метода, а также их преимущества и ограничения. Кроме того, в статье обсуждаются результаты и выводы, полученные при применении анализа настроений к хакерским форумам. Результаты исследования позволяют лучше понять эмоциональную составляющую в хакерском сообществе, выявить тенденции и настроения, выявить

потенциально опасные действия или тенденции. В конце статьи делаются общие выводы и указываются возможные направления дальнейших исследований в области анализа хакерских интернет-форумов. Также даны рекомендации по использованию результатов анализа для повышения безопасности и борьбы с киберпреступностью.

Ключевые слова: сентимент анализ, интернет форумы, кибербезопасность, даркнет.

S.T. Mambetov [1], Ye.Ye. Begimbayeva[1,2], A.K. Khikmetov[3], S.K. Joldasbayev[3], G.N. Kazbekova[4]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2] Satpayev Kazakh National Research Technical University, Almaty, Kazakhstan
[3]International University of Information Technologies, Almaty, Kazakhstan
[4]Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan

SENTIMENT ANALYSIS OF HACKER INTERNET FORUMS

Abstract

This article presents an overview and study of sentiment analysis techniques in the context of hacking internet forums. Hacker forums are a forum for exchanging information and experiences about computer security, and analyzing the emotional color of messages in such forums can provide useful insights into the study of hacker communities and their activities. Various methods and approaches are considered in the article. Methods considered include sentiment dictionary, machine learning, sentiment analysis, as well as hybrid models that combine these approaches. The principles and methodologies of each method, as well as their advantages and limitations, are described in detail. In addition, the article discusses the results and conclusions obtained when applying sentiment analysis to hacking forums. The results of the research allow us to better understand the emotional component in the hacker community, identify trends and moods, and identify potentially dangerous activities or trends. At the end of the article, general conclusions are drawn and possible directions of further research in the field of analysis of hacker Internet forums are indicated. Recommendations for using the analysis results to increase security and combat cybercrime are also provided.

Keywords: sentimental analysis, internet forums, cybersecurity, darknet.

Кіріспе

Қазіргі уақытта хакерлік ақпараттық қауіпсіздік саласындағы ең өзекті және маңызды мәселелердің бірі болып табылады. Хакерлер жүйелерге еніп, оларды бұзу арқылы бизнес пен жеке адамдарға зиянын тигізуде. Бұл мәселемен тиімді күресу үшін хакерлердің не ойлайтынын, не сезінетінін және олар қандай стратегияларды қолданатынын түсіну қажет. Хакерлер бір-бірімен тәжірибе алмасып, әр түрлі жүйелердің осалдықтарын және сол жүйелерге қалай қауіп төндіруге болатынын талқылайтын хакерлік интернет форумдардағы пікірлерді сентимент талдау қажет болады.

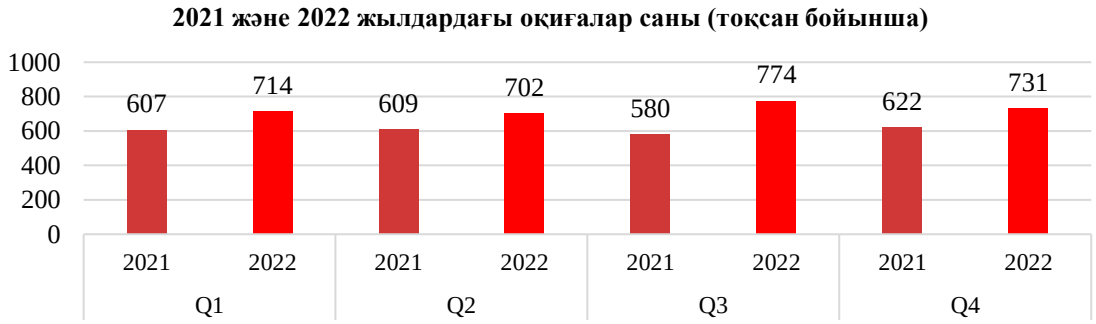
Хакерлік интернет-форумдарда жазылған мәтіндердің эмоционалды бояуын талдаудың күшті құралы сентимент талдау болып табылады. Бұл әдіс хакерлік туралы оң (positive), теріс (negative) және бейтарап (neutral) пікірлерді анықтауға, сондай-ақ хакерлікке қатысты әртүрлі тақырыптардың кілті мен көзқарасын сипаттауға мүмкіндік береді.

Соңғы жылдары кез-келген ақпараттық жүйелерге, жеке тұлғаларға хакерлік шабуылдар көбею үстінде. Postive Technologies интернет сайтының 2022 жылғы қорытынды мәліметіне сүйене отырып, жалпы шабуыл оқиғаларының саны 1 суретте келтірілген [1].

Бұл жұмыста хакерлік интернет-форумдардың сентимент талдауы және оны хакерлікке байланысты мәтіндердің эмоционалды бояуын талдау үшін қолдану жолдары қарастырылады. Негізгі мақсатқа жету үшін қолдануға болатын Машиналық оқыту мен табиғи тілді өңдеудің әртүрлі әдістері, сөздікке негізделген әдісі қарастырыла отырып, ақпараттық қауіпсіздік және кибершабуылдардың алдын алу саласында сентимент талдаудың нәтижелерін қолдану мүмкіндіктері де талқыланады.

Хакерлік интернет-форумдарды сентимент талдау қиын болуы мүмкін, өйткені бұл форумдар көбінесе слэнгтер мен жаргондарды пайдаланумен қатар балағат сөздерді де қамтиды. Алайда, егер біз осындай форумдардағы талқылаудың негізгі тақырыптарын ескере отырып, қатысушылардың көңіл-күйін анықтауға болады.

Мысалы, хакерлік форумдарда ақпарат қауіпсіздігі, хакерлік, осалдық тақырыптары жиі талқыланады, сонымен қатар бағдарламалауға, желілерге және т. б. қатысты техникалық мәселелер талқыланады. Осы форумдардағы хабарламаларды талдайтын болсақ, хакерлер қауымдастығындағы қауіпсіздік қатерлері қаншалықты маңызды деп саналатыны және қазіргі уақытта қандай мәселелер ең өзекті болып табылатыны туралы қорытынды жасауға болады.



Сурет 1. 2021 және 2022 жылдардағы оқиғалар (шабуылдар) саны

Хакерлік форумдарда интернет цензурасына, сөз бостандығына, құпиялылық құқықтарына және т.б. қатысты саяси және әлеуметтік мәселелер талқылануы да мүмкін. Алайда, хакерлік форумдарда арандату, әзіл-қалжың немесе қоғамдастықтың реакциясын тексеру мақсатында хабарламалар жиі кездесетінін де атап өткен жөн. Сондықтан мұндай форумдардың ерекшеліктерін ескере отырып, контекстті ескеру және хабарламаларды бағалау маңызды.

Әдебиеттерге шолу

Сентимент талдау табиғи тілді өңдеудегі маңызды мәселелердің бірі болып табылады және қазіргі уақытта оны шешудің көптеген әдістері мен тәсілдері бар. Хакерлік интернет-форумдар контекстінде бұл тапсырманың өзіндік ерекшеліктері бар, өйткені мұндай форумдардағы мәтіндер жоғары дәрежеде ерекшелік пен күрделілікке ие болуы мүмкін. Осы мәселелермен айналысқан авторлардың жұмыстарына қысқаша тоқтала кетсек. Көптеген авторлар киберқауіпсіздікке қатысты мәтіндерді сентимент талдау үшін қолданылатын машиналық оқытудың әртүрлі әдістерін зерттеді. Талдау барысында авторлар әртүрлі дерек көздерінен, соның ішінде әлеуметтік желілер мен форумдардан алынған деректерді пайдаланды және Naive Bayes, Support Vector Machines (SVM) және Random Forest сияқты әртүрлі алгоритмдер арқылы қарастырды. Нәтижесінде, SVM мен Random Forest мәтіндерді оң, теріс және бейтарап деп жіктеу үшін ең жақсы дәлдікті көрсеткенін көрсетті [2].

Келесі жұмыста да авторлар хакерлік форумдарда көңіл-күйді талдауды машиналық оқытуға негізделген әдісті ұсына отырып, форумдардағы негізгі талқылау тақырыптарын бөліп көрсетеді және хабарламалардың кілтін анықтау үшін Naive Bayes сияқты жіктеу алгоритмдерін пайдаланған. Зерттеуде хакерлік форумдардағы жазбалардың көпшілігі қауіпсіздік пен хакерлік тақырыптармен байланысты теріс түске ие болып табылады [3].

Келесі мақалада авторлар хакерлік форумдарды сентимент талдауға машиналық оқыту әдістері және мұғаліммен оқытуды қолданатын аралас тәсіл ұсынады. Авторлар ұсынған әдіс хабарламалардың кілтін дәл анықтауға және талқылаудың мәнін, мәтінін ескеруге мүмкіндік береді деп көрсетеді [4].

Авторлар хакерлер тілінің лексикалық ерекшеліктерін талдап, оларды көңіл-күй мен эмоционалды бояумен байланыстыратын жұмысты атап өтуі маңызды. Зерттеу көрсеткендей, хакерлер өз хабарламаларында жаргон сөздерді, қысқартылған сөздерді және жағымсыз лексиканы жиі қолданатыны көрсетілген. Бұл наразылық пен қанағаттанбауды көрсетуі мүмкін [5].

Мақалада авторлар хакерлік форумдардағы хабарламаларды талдау үшін қолданылатын сентимент талдаудың негізгі әдістерін қарастырады. Олар сондай-ақ хакерлік форумдарды сентимент талдауға бағытталған әртүрлі зерттеулерге шолу жасайды және олардың әдістемесі мен нәтижелерін сипаттайды [6].

Тағы бір авторлар хакерлік форумдарды сентимент талдау үшін тереңдетіп оқытуды қолдануды ұсынады. Олар форумдардағы хабарламалардың кілтін талдау үшін пайдаланылуы мүмкін әртүрлі нейрондық желі архитектураларын сипаттап көрсетеді [7].

Келесі авторлар сентимент талдауды ғана емес, сонымен қатар хакерлік форумдардағы жазбаларды талдау үшін пайдаланылуы мүмкін деректерді талдаудың басқа әдістерін де қарастыра отырып, форумдарда талқыланатын әртүрлі тақырыптарды жан-жақты талдай отырып, қауіпсіздік осалдықтарын анықтау үшін деректерді талдауды қалай пайдалануға болатынын сипаттайды [8].

Кейбір авторлар киберқауіпсіздік контекстінде хакерлік форумдардағы хабарламаларды талдау үшін сентимент талдаудың әртүрлі талдау әдістерін, соның ішінде кибершабуылдар мен басқа да қауіпсіздік қауіптерін анықтау үшін деректерді талдауды қалай пайдалануға болатынын талқылайды [9].

Хакерлік форум хабарламаларын сентимент талдау үшін авторлар табиғи тілді өңдеудің әртүрлі әдістерін қолданды, соның ішінде көңіл-күйді талдау, тақырыпты модельдеу және желілік талдауға назар аударады. Зерттеу нәтижелері хакерлік форумдардағы киберқылмыс әрекеттері туралы құнды ақпарат беретіндігіне және киберқылмыспен күресу стратегияларын әзірлеу үшін пайдалы болуы мүмкіндігіне көз жеткізеді [10].

Twitter хабарламаларынан хакерлік пен киберқауіпсіздікке қатысты хабарламаларды сентимент талдау үшін зерттеуші табиғи тілді өңдеу әдістерін қолдануды ұсынады. Олар сөздікке негізделген кілттерді талдау әдістерін және SVM және Multinomial Naive Bayes сияқты Машиналық оқыту әдістерін қолдана отырып, нәтижесінде Support Vector Machines басқа әдістерге қарағанда жақсы нәтиже көрсеткенін атап көрсетті [11].

Киберқауіпсіздікке қатысты мәтіндерді сентимент талдауда машиналық оқыту әдістерін қолдануды зерттей отырып, деректерді әлеуметтік желілер мен форумдардан ала отырып, сөздіктерге негізделген мәтіндердің кілтіне талдау жасады. Зерттеу нәтижесінде сөздікке негізделген әдіске қарағанда, машиналық оқытуға негізделген әдістер жоғары дәлдік көрсеткен [12].

Онлайн форумдардағы мәтіндердің кілтін талдауға бағыттап, авторлар мәтіндердің эмоционалды бояуын анықтау үшін әртүрлі талдау әдістерін, атап айтқанда сөздердің жиілігін талдау арқылы және сөздікке негізделген кілттерді талдауды қолданған. Бұл зерттеулер нәтижесінде сөздікке негізделген сентимент талдау жоғары нәтиже көрсеткен [13].

Тағы бір зерттеу пікірлерді талдау саласында қолданылатын әртүрлі сентимент талдау әдістеріне шолу жасап, авторлар сөздікке негізделген кілтті талдау сынды классикалық әдістерге қоса, тереңдетіп оқыту және нейрондық желілерді қамтитын заманауи әдістерді қарастырған [14]. Келесі жұмыста авторлар жаңалықтар мақалаларынан алынған мәліметтер бойынша киберқауіпсіздік саласындағы мәтіндерді сентимент зерттеу жүргізген. Жиналған мәліметтерді талдауда сөздік негізге сүйеніп те, машиналық оқыту әдістерін де қолданған [15].

Келесі жұмыста авторлар ағылшын және орыс тілдеріндегі пікірлерді машиналық талдау үшін қолжетімді әр түрлі сөздіктерді қарастыратын аналитикалық талдау жасай келе, ағылшын тіліне арналған SentiWordNet, MPQ, LWC және орыс тіліне арналған AFINN, RuSentiLex, SentiRuEval-2015 сияқты танымал сөздіктерге шолу жасайды. Авторлар осы сезім сөздіктерінің өнімділігін сентимент талдау тапсырмалары үшін дәлдік, қамту және пайдалану мүмкіндігі тұрғысынан бағалап, осы сөздіктерді пайдаланудағы қиындықтарды көрсете отырып, оларды шешу жолдарын да ұсынады. Зерттеу ағылшын және орыс тілдерінде сезімдерді талдауға және пікірлерді машиналық талдауға қызығушылық танытатын зерттеушілер мен тәжірибешілер үшін пайдалы ресурс болып табылады. Ол әртүрлі сезім сөздіктерінің күшті және әлсіз жақтарын түсінеді және сентимент талдауының өнімділігін

жақсарту бойынша ұсыныстар береді [16]. Аталған авторлардың зерттеу нәтижесіне қарай отырып, сөздікке негізделген әдіске қарағанда, машиналық оқыту әдістері, табиғи тілді өңдеу әдістері жоғары нәтиже беретініне көз жеткізуге болады.

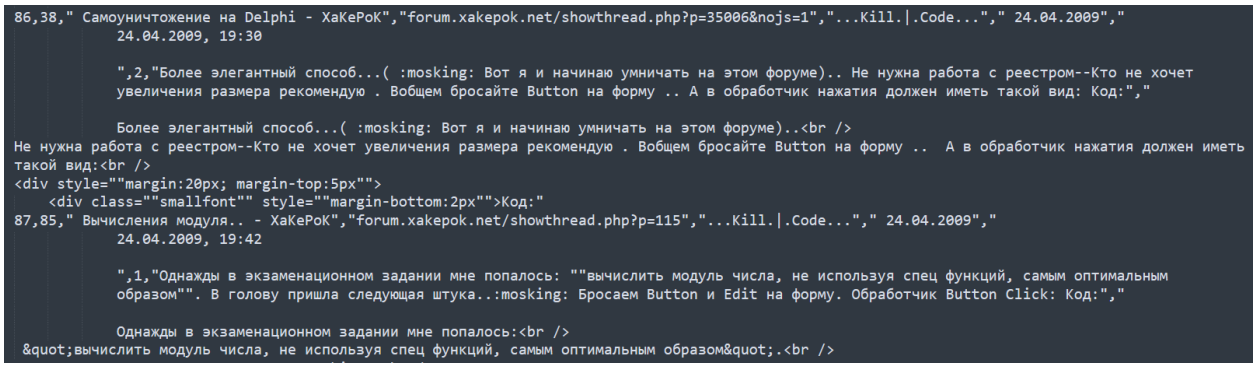
Материалдар мен әдістер

Хакерлік интернет-форумдарды сентимент талдау әртүрлі әдістер мен тәсілдерді қолдана отырып жүргізілуі мүмкін. Ең кең таралған тәсілдердің бірі-нейрондық желілер, SVM (Support Vector Machine) және Naive Bayes сияқты машиналық оқыту әдістерін қолдану. Бұл әдістер мәтіндерді олардың тоналдылығына (оң, теріс немесе бейтарап) қарай жіктеуге, сондай-ақ мәтіннің эмоционалды түсін анықтауға мүмкіндік береді. Тағы бір тәсіл - сөздіктерге негізделген кілтті талдау әдістерін қолдану болып табылады. Бұл әдістер мәтіннің тоналдылығын анықтау үшін белгілі кілті бар (оң, теріс немесе бейтарап) сөздік қорын пайдаланады. Ең көп таралған сөздіктердің бірі - ағылшын тіліндегі SentiWordNet сөздігі, ал орыс тілінде RuSentiLex сөздігі болып табылады.

Сонымен қатар, хакерлік интернет-форумдарға сентимент талдауын жүргізу үшін сөз жиілігін талдау және тақырыптық модельдеу сияқты статистикалық талдау әдістерін қолдануға болады. Жиілікті талдау мәтіндерде жиі кездесетін сөздерді және олардың жиілігін анықтауға мүмкіндік береді, ал тақырыптық модельдеу форумда талқыланатын ең маңызды тақырыптарды бөліп көрсетуге мүмкіндік береді.

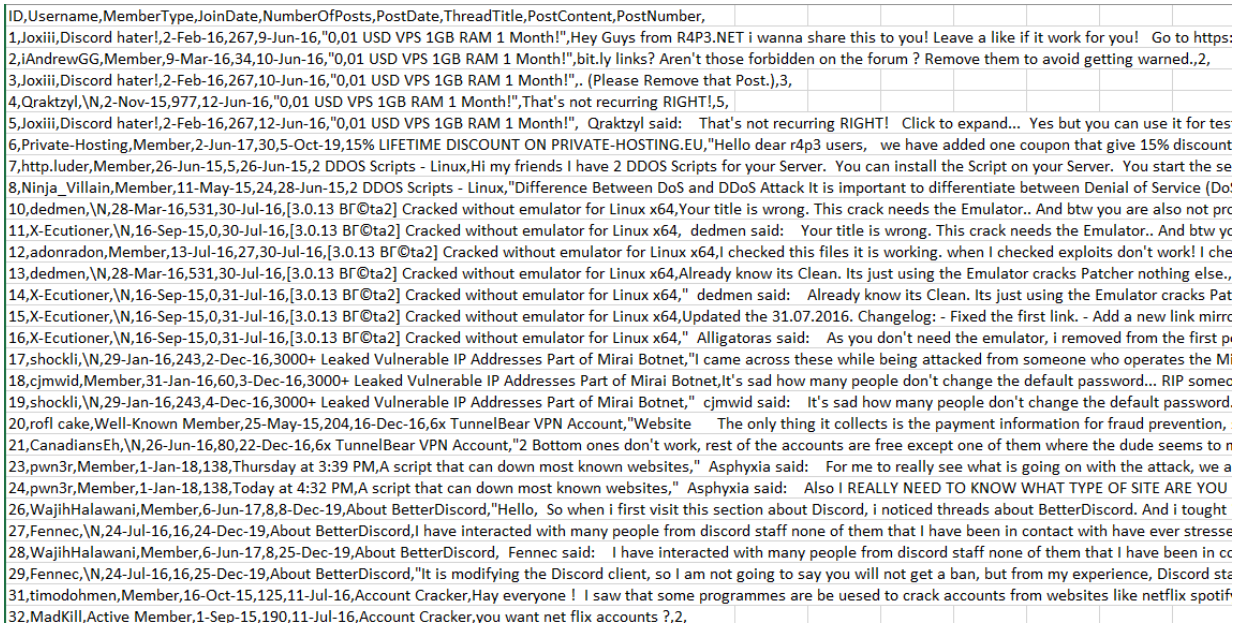
Хакерлік интернет-форумдарға сентимент талдауын жүргізу үшін форум пайдаланушылары жазған мәтіндерден тұратын жеткілікті үлкен деректер жиынтығын жинау қажет. Ол үшін әртүрлі әдістерді қолдануға болады, мысалы, веб-скрапинг немесе форумдардың API (Application Programming Interface).

Хакерлік интернет-форумдарды талдау әдістері мен әдістемелері нақты тапсырмаға және қол жетімді деректерге байланысты. Алайда, Машиналық оқыту әдістерін қолдану және сөздікке негізделген кілттерді талдау хакерлік интернет-форумдардағы мәтіндерді сентимент талдаудың ең кең таралған және тиімді тәсілдері болып табылады. Кез-келген деректерді сентимент талдаудың ең бірінші қадамы деректер жиынтығы болып табылады. Интернет форумдардан python тілінің BS4 + Requests кітапханалары көмегімен парсинг арқылы жиналған орыс тілді деректер базасы 2-суретте көрсетілген. Ал 3-суретте ағылшын тілді хакерлік интернет форумдардан жиналған дерек көрсетілген [17].



Сурет 2. Орыс тілді хакерлік форумнан алынған деректер кескіні

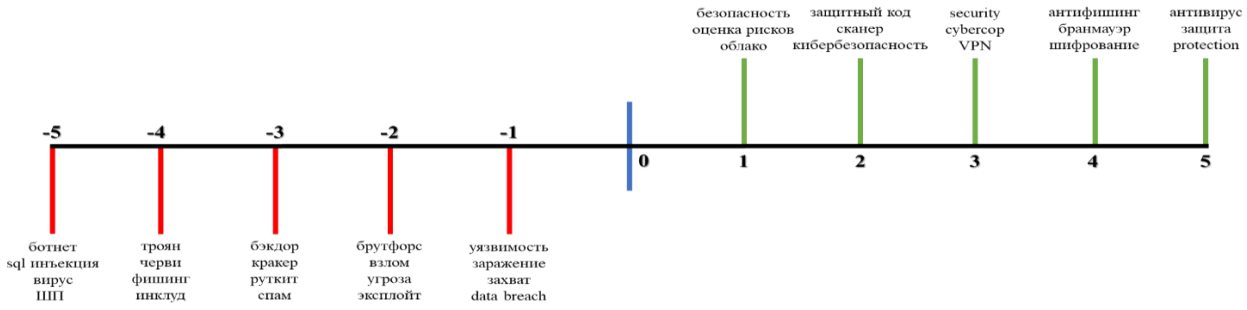
Алынған деректер қатарына автордың ID, Username, MemberType, JoinDate, NumberOfPosts, PostDate, ThreadTitle, PostContent, PostNumber сынды семантикалық белгілер бойынша іріктелді.



ID,Username,MemberType,JoinDate,NumberOfPosts,PostDate,ThreadTitle,PostContent,PostNumber,
1,Joxiii,Discord hater!,2-Feb-16,267,9-Jun-16,"0,01 USD VPS 1GB RAM 1 Month!",Hey Guys from R4P3.NET i wanna share this to you! Leave a like if it work for you! Go to https:
2,iAndrewGG,Member,9-Mar-16,34,10-Jun-16,"0,01 USD VPS 1GB RAM 1 Month!",bit.ly links? Aren't those forbidden on the forum ? Remove them to avoid getting warned.,2,
3,Joxiii,Discord hater!,2-Feb-16,267,10-Jun-16,"0,01 USD VPS 1GB RAM 1 Month!",. (Please Remove that Post.),3,
4,Qraktzyl,\N,2-Nov-15,977,12-Jun-16,"0,01 USD VPS 1GB RAM 1 Month!",That's not recurring RIGHT!,5,
5,Joxiii,Discord hater!,2-Feb-16,267,12-Jun-16,"0,01 USD VPS 1GB RAM 1 Month!", Qraktzyl said: That's not recurring RIGHT! Click to expand... Yes but you can use it for tes
6,Private-Hosting,Member,2-Jun-17,30,5-Oct-19,15% LIFETIME DISCOUNT ON PRIVATE-HOSTING.EU,"Hello dear r4p3 users, we have added one coupon that give 15% discount
7,http.luder,Member,26-Jun-15,5,26-Jun-15,2 DDOS Scripts - Linux,Hi my friends I have 2 DDOS Scripts for your Server. You can install the Script on your Server. You start the se
8,Ninja_Villain,Member,11-May-15,24,28-Jun-15,2 DDOS Scripts - Linux,"Difference Between DoS and DDoS Attack It is important to differentiate between Denial of Service (Do
10,dedmen,\N,28-Mar-16,531,30-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64,Your title is wrong. This crack needs the Emulator.. And btw you are also not pro
11,X-Ecutioner,\N,16-Sep-15,0,30-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64, dedmen said: Your title is wrong. This crack needs the Emulator.. And btw yo
12,adonradon,Member,13-Jul-16,27,30-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64,I checked this files it is working. when I checked exploits don't work! I che
13,dedmen,\N,28-Mar-16,531,30-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64,Already know its Clean. Its just using the Emulator cracks Patcher nothing else.,
14,X-Ecutioner,\N,16-Sep-15,0,31-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64," dedmen said: Already know its Clean. Its just using the Emulator cracks Pat
15,X-Ecutioner,\N,16-Sep-15,0,31-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64,Updated the 31.07.2016. Changelog: - Fixed the first link. - Add a new link mirro
16,X-Ecutioner,\N,16-Sep-15,0,31-Jul-16,[3.0.13 BГ©ta2] Cracked without emulator for Linux x64," Alligatoras said: As you don't need the emulator, i removed from the first p
17,shockli,\N,29-Jan-16,243,2-Dec-16,3000+ Leaked Vulnerable IP Addresses Part of Mirai Botnet,"I came across these while being attacked from someone who operates the Mi
18,cjmwid,Member,31-Jan-16,60,3-Dec-16,3000+ Leaked Vulnerable IP Addresses Part of Mirai Botnet,It's sad how many people don't change the default password... RIP somec
19,shockli,\N,29-Jan-16,243,4-Dec-16,3000+ Leaked Vulnerable IP Addresses Part of Mirai Botnet," cjmwid said: It's sad how many people don't change the default password.
20,rofl cake,Well-Known Member,25-May-15,204,16-Dec-16,6x TunnelBear VPN Account,"Website The only thing it collects is the payment information for fraud prevention,
21,CanadiansEh,\N,26-Jun-16,80,22-Dec-16,6x TunnelBear VPN Account,"2 Bottom ones don't work, rest of the accounts are free except one of them where the dude seems to n
23,pwn3r,Member,1-Jan-18,138,Thursday at 3:39 PM,A script that can down most known websites," Asphyxia said: For me to really see what is going on with the attack, we a
24,pwn3r,Member,1-Jan-18,138,Today at 4:32 PM,A script that can down most known websites," Asphyxia said: Also I REALLY NEED TO KNOW WHAT TYPE OF SITE ARE YOU
26,WajihHalawani,Member,6-Jun-17,8,8-Dec-19,About BetterDiscord,"Hello, So when i first visit this section about Discord, i noticed threads about BetterDiscord. And i tought
27,Fennec,\N,24-Jul-16,16,24-Dec-19,About BetterDiscord,I have interacted with many people from discord staff none of them that I have been in contact with have ever stresse
28,WajihHalawani,Member,6-Jun-17,8,25-Dec-19,About BetterDiscord, Fennec said: I have interacted with many people from discord staff none of them that I have been in co
29,Fennec,\N,24-Jul-16,16,25-Dec-19,About BetterDiscord,"It is modifying the Discord client, so I am not going to say you will not get a ban, but from my experience, Discord sta
31,timodohmen,Member,16-Oct-15,125,11-Jul-16,Account Cracker,Hay everyone ! I saw that some programmes are be uesed to crack accounts from websites like netflix spotif
32,MadKill,Active Member,1-Sep-15,190,11-Jul-16,Account Cracker,you want net flix accounts ?,2,

Сурет 3. Форумдардан жиналған ағылшын тілді деректер үлгісі

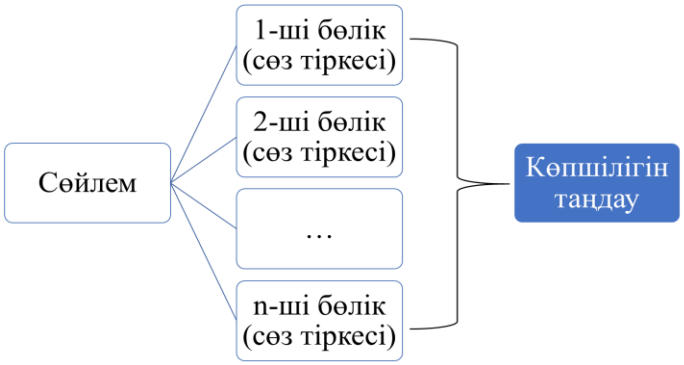
Жұмыс барысында 5 түрлі ағылшын және орыс тілді хакерлік интернет форумдардан деректер жиналды. Сол деректер негізінде сентимент талдау үшін хакерлік терминологияға қатысты сөздер базасы жасалынды. Осы база бойынша -5 – тен +5 – ке дейін қауіптілік деңгейі бойынша пікірлер бөлініп қарастырылды. Сөздікте теріс мәнге шабуылдар тізімі, осалдықтар, қауіп көздері, құпия сөзді іріктеу, алаяқтыққа қатысты сөздерден тұрады, ал оң мәнге жоғарыда келтірілген қауіптерге қарсы қорғанысқа қатысты терминологиядан тұрады. Екеуіне де қатысы жоқ сөздер бейтарап сөздер тізіміне енгізіледі. Оларға 0-мәні беріледі. 4 суретте ақпараттық және киберқауіпсіздікке қатысты терминологияның оң, теріс және бейтарап түрге бөлінуі кескіні келтірілген. Мәндері қауіп түрлері мен олардан қорғану деңгейі бойынша алынған. Суретте барлық терминдер көрсетілмеген. Осындай сөздер жиынтығы орыс және ағылшын тілінде жасалынған.



Сурет 4. Сөздердің жіктелу формасы

Жоғарыда көрсетілген жіктелу формасы бойынша әр пікірді талдау арқылы, жалпы пікірдің дұрыстығына немесе бұрыстығына көз жеткізуге болады. Яғни әрбір сөзге сандық мән беріліп, сол мәндер бойынша арифметикалық амалдар көмегімен пікірді талдайтын боламыз.

Біз ұсынып отырған сентимент сөздікті талдау сөз тіркестерін қарастырып талдауға мүмкіндік береді. Белгілі бір сөйлемді бірнеше бөлікке бөліп, әрбір бөлікті жекелей қарастырып, аталмыш бөлікте айтылып отырған ойдың оң, теріс немесе бейтараптығына көз жеткізу арқылы, толық сөйлемнің эмоционалды бояуын анықтаймыз. Оның логикасы 5 суретте кескінделген.



Сурет 5. Сөйлемді бөліктер арқылы жіктеу формасы

Келтірілген сөздік талдау арқылы жиналған деректер негізінде форум пайдаланушыларының ой-пікірлерін талдап қарайық.

Нәтижелер мен талқылау

Хакерлік форумдардағы сентимент талдаудың нақты нәтижелері көптеген факторларға, соның ішінде таңдалған әдістерге, деректердің көлемі мен сапасына, сондай-ақ форумның өзіне де байланысты болып табылады. Осы ұсынған сөздік әдіске бірнеше мысал келтірейік (мысалдар xss.is және Probiv.one хакерлік форумдарынан алынған):

1. “There are plenty of WordPress eCommerce plugins in the market. But not all of them have the right set of features for your use-case. For example, some eCommerce plugins are made for selling digital goods like eBooks, photos, music, etc. Others are better suited for selling physical products that need shipping”

- *"There are plenty of WordPress eCommerce plugins in the market"* - WordPress-те көптеген электрондық коммерция плагиндері туралы айтылады. Таңдаудың әртүрлілігімен байланысты бейтарап немесе позитивті тоналдылықты болжауға болады.

- *"not all of them have the right set of features for your use-case"* - барлық плагиндерде белгілі бір пайдалану жағдайына сәйкес функциялар жиынтығы жоқ деген мәлімдеме айтылған. Кейбір плагиндердің шектеулеріне байланысты теріс немесе бейтарап тоналдылықты болжауға болады.

- *"some eCommerce plugins are made for selling digital goods"* - сандық тауарларды сатуға арналған плагиндер туралы айтылған. Функционалдылықтың әртүрлілігіне байланысты бейтарап немесе позитивті тоналдылықты болжауға болады.

-*"Others are better suited for selling physical products"* - физикалық тауарларды сатуға қолайлы плагиндер туралы айтылған. Плагиндердің нақты қажеттіліктерге бейімделуіне байланысты бейтарап немесе позитивті тоналдылықты болжауға болады.

Тұтастай алғанда, мәтінде WordPress-тегі электрондық коммерция плагиндерінің әртүрлілігімен және олардың кейбіреулерінің нақты талаптарға бейімделуіне байланысты бейтарап және оң бояуы бар деп есептеуге болады.

2. "DDOS Scripts - Linux, Hi my friends I have 2 DDOS Scripts for your Server. You can install the Script on your Server. You start the server with: perl kill.pl [IP] [Port] [65500] [Time] or: perl ssh.pl [IP] [Port] [65500] [Time] http://www.filedropper.com/kill http://www.filedropper.com/ssh Have FUN,1,"

- *"DDOS Scripts - Linux"* - бұл бөлік теріс және заңсыз әрекеттермен байланысты DDoS шабуылының сценарийлерін сипаттайды, сондықтан теріс кілтті болжауға болады.

-*"Hi my friends I have 2 DDOS Scripts for your Server* - бұл бөлікте де серверге DDoS шабуылдарын жүргізуге арналған екі сценарийдің болуы туралы ақпарат айтылады. Мұндай мәлімдемелер теріс контексте де қарастырылуы мүмкін, өйткені олар зиянды және заңсыз әрекеттермен байланысты.

-*"You can install the Script on your Server"* - бұл мәлімдеме сценарийлерді серверге орнатуды ұсынады, оны DDoS шабуылдарымен байланысы болғандықтан теріс тұрғыдан да қарастыруға болады.

-*"Have FUN"* - бұл сөз тіркесі сарказм немесе ирония ретінде қабылдануы мүмкін, өйткені ол зиянды әрекеттерді жасауға шақырады.

Жоғарыда аталған факторларды ескере отырып, бұл мәтіннің теріс немесе тіпті дұшпандық кілті бар деп болжауға болады, өйткені онда DDoS сценарийлері туралы ақпарат бар және оларды қолдануға шақырады.

3. “I shall find your IP, then find you!” – осы пікірге сөздік жіктеу бойынша сентимент талдау жасайық:

I	*бейтарап*	*0*
shall	*бейтарап*	*0*
find	*теріс*	*-1*
your	*алдыңғы сөзге сүйене отырып теріс қатарға жатқызуға болады*	*-1*
IP	*теріс*	*-3*
then	*бейтарап*	*0*
you	*бейтарап*	*0*
0 + 0 + (-1) + (-1) + (-3) + 0 + 0 = -5		

Хабарламадағы әрбір сөзге оң, теріс немесе бейтарап екенін анықтап, шыққан өрнекті арифметикалық амалмен есептегенде, жалпы хабарлама теріс пікірге жататынына көз жеткіздік.

Дәл осы пікірді бөліктерге бөліп қарастырайық:

- *"I shall find your IP"* - әңгімелесушінің IP мекенжайын табу туралы қауіп болып табылады. Жеке құпиялылықты бұзумен және ықтимал теріс ниеттермен байланысты теріс тоналдылық деп болжауға болады.

- *"then find you"* - әңгімелесушіні табу және қайда екенін анықтау қаупі. Теріс ниеттермен және ықтимал зиянды әрекеттермен байланысты теріс тоналдылыққа ие деп болжауға болады.

Жалпы мәтінде әңгімелесушіге қатысты жағымсыз элементтер мен қауіп көздері бар, яғни бұл осы хабарламаның теріс тоналдылығын көрсетеді.

Жоғарыда келтірілген мысалдар негізінде сөздік арқылы сентимент талдаудың қаншалықты дұрыстығына көз жеткіздік. Хакерлік форумдарды сентимент талдау - форумдардағы пікірталастар мен пікірлердің эмоционалды түсін зерттеудің маңызды құралы болып табылады. Оны пайдаланушы хабарламаларының кілтін талдау үшін, сондай-ақ форум қатысушыларының қызығушылығын тудыратын тақырыптар мен мәселелерді анықтау үшін пайдалануға болады. Дегенмен, хакерлік форумдарда сентимент талдау жүргізу бірқатар қиындықтарға тап болуы мүмкін. Біріншіден, хакерлік форумдарда нақты терминдер мен жаргондар қолданылуы мүмкін, бұл мәтіннің кілтін жіктеуді қиындатады. Екіншіден, көптеген хакерлік форум пайдаланушылары бүркеншік аттар (псевдонимы) қолдана алады, бұл бір пайдаланушыдан екіншісіне хабарлама кілтінің өзгеруін бақылауды қиындатуы мүмкін. Сонымен қатар, форумдарда сентимент талдауды қиындататын хакерлік форумдарда заңсыз әрекеттер талқылануы мүмкін екенін ескеру маңызды. Кейбір қатысушылар өздерінің нақты тондарын жасыруға тырысуы мүмкін немесе анықтамау үшін көп мағыналы сөз тіркестерін қолдануы да мүмкін. Хакерлік форумдар пікірлерін талдау арқылы киберқауіпсіздіктің, кез-келген ақпараттық жүйенің осал тұстарын және DDoS немесе т.б. шабуылдардың алдын алуға дейінгі мүмкіндік береді.

Жалпы, хакерлік форумдарға сентимент талдау жүргізу құнды нәтижелер әкелетін күрделі және қызықты зерттеу болып табылады. Алайда, осы форумдардың ерекшеліктерін ескеру және нақты нәтижелерге қол жеткізу үшін тиісті әдістер мен тәсілдерді қолдану маңызды.

Қорытынды

Хакерлік форумдарды сентимент талдау технологияға, киберқауіпсіздікке және IT индустриясына, сонымен қатар басқа да тақырыптарға қатысты әртүрлі мәселелерді қарастыратын қауымдастық мүшелерінің пікірлері мен көңіл-күйлерін түсінудің маңызды құралы болып табылады. Алайда, талдау нақты терминологияны, стандартты емес байланыс ережелерін және аралас хабарламалардың болуы сияқты хакерлік форумдардың ерекшеліктерін ескеруді талап етеді.

Хакерлер қауымдастығының ерекшеліктерін ескеретін талдау жүргізу үшін қолайлы әдіс пен әдіснаманы таңдау маңызды. Мысалы, Машиналық оқыту мен терең оқыту әдістерін қолдану нәтижелердің дәлдігін жақсартуға көмектеседі, сонымен қатар модельді оқыту үшін көптеген деректерді қажет етеді. Сонымен қатар зерттеу жұмыста біз ұсынып отырған сөздік арқылы талдауда, әрбір хакерлік форумдарда жазылған пікірлердің оң-терістігіне, не бейтараптығына көз жеткізуге көмектеседі.

Сондай-ақ, сентимент талдауының нәтижелері қауымдастық мүшелерінің көңіл-күйі мен пікірлерін толық көрсете алмайтынын атап өткен жөн, өйткені барлық хабарламалар айқын тоналды түске ие бола алмайды немесе екіұшты болуы мүмкін. Алайда, талдау жүргізу жалпы тенденцияларды анықтауға және қатысушылардың негізгі тақырыптарға қатысты көңіл-күйін түсінуге көмектеседі. Осы жағдайда ой-пікірлерді талдауға сөздік әдісін қолданған дұрыс. Өйткені форумдарда әртүрлі жаргондар, слэнгтер пайданылады. Ал сіз оны алдын-ала өз сөздігіңізде белгілі бір топқа жатқызып аласыз.

Қорыта келгенде зерттеу жұмысымызда әртүрлі сентимент талдау әдістеріне тоқталып, сөздік арқылы талдауды хакерлік форумдардың хабарламаларында айтылған ой-пікірлердің талдауын ұсындық және сөздік арқылы талдауға мысалдар келтірілді.

Жалпы, хакерлік форумдарды сентимент талдау хакерлер қауымдастығындағы пікірлер мен көңіл-күйлерді талдаудың пайдалы құралы бола алады және IT индустриясы мен киберқауіпсіздікке қатысты тенденциялар мен мәселелерді анықтауға көмектеседі. Алдағы жұмыстарда ақпараттық қауіпсіздік саласының осалдықтары мен қауіптерін жоғарыда көрсетілген сентимент әдістерді пайдаланып, хакерлік форумдардың ой-пікірлерінен қарастыратын боламыз.

Пайдаланылған әдебиеттер тізімі:

1 Электронды ресурс (Дата обращения: 21/05/2023). URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2022/

2 Said, A. M., Almohri, H. A., & Alnuaimi, O. A. 2020. «Sentiment analysis of cybersecurity related text using machine learning techniques». In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE. https://doi.org/10.1109/ICCCI49829.2020.9159315

3 Sánchez-González, J. D., Sánchez-Yáñez, J. M., Martínez-Rojas, A., & Martínez-Trinidad, J. F. 2016. «Analyzing sentiment in hacker forums». International Journal of Software Science and Computational Intelligence, 8(3). https://doi.org/10.1504/IJSSCI.2016.077315

4 Chen, H., Zhu, J., Chen, H., Wang, L., & Xie, J. 2015. «Mining Hacker Forums: Sentiment Analysis and Semi-Supervised Learning». IEEE Transactions on Information Forensics and Security, 10(6), 1254-1264. https://doi.org/10.1109/TIFS.2015.2403904

5 Moyer, D. D., Lee, R. K., Kim, K. H., & Miner, J. M. 2006. «Hacker's language: A sentiment analysis of the discourse of the computer underground». Computers in Human Behavior, 22(2), 200-213. https://doi.org/10.1016/j.chb.2004.07.009

6 Garcia-Salicetti, M. A., Montesi, A., Spolaor, R., & Conti, M. 2020. «Sentiment Analysis in Hacker Forums: A Comprehensive Review». IEEE Communications Surveys and Tutorials, 22(4), 2404-2434. https://doi.org/10.1109/COMST.2020.2980856

7 Hussain, S., Aslam, S., Malik, M. H., & Qureshi, M. A. 2020. «Sentiment Analysis of Hacker Forum Data using Deep Learning Techniques». In 2020 4th International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE. https://doi.org/10.1109/ICCAIS49205.2020.9149354

8 Alshammari, T., Lu, W., Rigby, J. M., & Jaatun, M. G. 2016. «Analyzing hacker forums: A systematic literature review». Computers & Security, 59, 185-201. https://doi.org/10.1016/j.cose.2016.02.002

9 Sharma, N., Chauhan, N., Bansal, A., & Kumar, S. 2021. «A Review on Sentiment Analysis in Cybersecurity». International Journal of Advanced Science and Technology, 30(1), 1-11. https://doi.org/10.22146/ijast.6187

10 Al-Azawei, M. N., Al-Enezi, R. A., Tawfiq, W. A., & Alfaraj, A. H. 2020. «Mining Hacker Forums: A Study of Cybercriminal Environments». IEEE Access, 8, 89926-89937. https://doi.org/10.1109/ACCESS.2020.2980453

11 Rahman, M. T., Islam, M. R., Hoque, M. R., & Islam, M. R. 2020. «Sentiment Analysis of Hacking and Cybersecurity-Related Tweets». In 2020 2nd International Conference on Computer and Communication Engineering (ICCCE) (pp. 255-260). IEEE. https://doi.org/10.1109/ICCCE49051.2020.9200770

12 Liang, H., Yang, C., & Wang, H. 2020. «Cybersecurity sentiment analysis using machine learning». IEEE Access, 8, 54484-54494. https://doi.org/10.1109/ACCESS.2020.2981284

13 Khoo, C. S., & Gillam, L. 2011. «Sentiment analysis in online forums». The Journal of Systems and Software, 84 (12), 2099-2108. https://doi.org/10.1016/j.jss.2011.06.043

14 Thakur, R., Singh, V. K., & Gupta, P. 2018. «A Review on Sentiment Analysis and Opinion Mining». In 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-6). IEEE. https://doi.org/10.1109/ICRAIE.2018.8594166.

15 Olteanu, A.-M., Stoian, M., & Vrabie, C. 2020. «Sentiment analysis of cybersecurity news». In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 207-212). IEEE. https://doi.org/10.1109/ICCP51668.2020.9217864

16 Karpov, A., Kotelnikov, E., Nikolenko, S., & Koltsova, O. 2018. «Modern sentiment lexicons for opinion mining in English and Russian (analytical survey)». Computational Linguistics and Intellectual Technologies, 17(26), 91-111. https://doi.org/10.28995/2079-8136-2018-26-91-111

17 Мамбетов С.Т., Бегимбаева Е.Е., Хикметов А.К., & Джолдасбаев О.К. Тақырыптық интернет ресурстардан деректерді алу алгоритмін әзірлеу. Вестник Национальной инженерной академии Республики Казахстан. 2023. № 2 (88) https://doi.org/10.47533/2023.1606-146X.6

References:

1 Elektrondy resurs (Data obrashcheniya: 21/05/2023). URL:https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2022/ (in Russian)

2 Said, A. M., Almohri, H. A., & Alnuaimi, O. A. 2020. «Sentiment analysis of cybersecurity related text using machine learning techniques». In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE. https://doi.org/10.1109/ICCCI49829.2020.9159315

3 Sánchez-González, J. D., Sánchez-Yáñez, J. M., Martínez-Rojas, A., & Martínez-Trinidad, J. F. 2016. «Analyzing sentiment in hacker forums». International Journal of Software Science and Computational Intelligence, 8(3). https://doi.org/10.1504/IJSSCI.2016.077315

4 Chen, H., Zhu, J., Chen, H., Wang, L., & Xie, J. 2015. «Mining Hacker Forums: Sentiment Analysis and Semi-Supervised Learning». IEEE Transactions on Information Forensics and Security, 10(6), 1254-1264. https://doi.org/10.1109/TIFS.2015.2403904

5 Moyer, D. D., Lee, R. K., Kim, K. H., & Miner, J. M. 2006. «Hacker's language: A sentiment analysis of the discourse of the computer underground». Computers in Human Behavior, 22(2), 200-213. https://doi.org/10.1016/j.chb.2004.07.009

6 Garcia-Salicetti, M. A., Montesi, A., Spolaor, R., & Conti, M. 2020. «Sentiment Analysis in Hacker Forums: A Comprehensive Review». IEEE Communications Surveys and Tutorials, 22(4), 2404-2434. https://doi.org/10.1109/COMST.2020.2980856

7 Hussain, S., Aslam, S., Malik, M. H., & Qureshi, M. A. 2020. «Sentiment Analysis of Hacker Forum Data using Deep Learning Techniques». In 2020 4th International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE. https://doi.org/10.1109/ICCAIS49205.2020.9149354

8 Alshammari, T., Lu, W., Rigby, J. M., & Jaatun, M. G. 2016. «Analyzing hacker forums: A systematic literature review». Computers & Security, 59, 185-201. https://doi.org/10.1016/j.cose.2016.02.002

9 Sharma, N., Chauhan, N., Bansal, A., & Kumar, S. 2021. «A Review on Sentiment Analysis in Cybersecurity». International Journal of Advanced Science and Technology, 30(1), 1-11. https://doi.org/10.22146/ijast.6187

10 Al-Azawei, M. N., Al-Enezi, R. A., Tawfiq, W. A., & Alfaraj, A. H. 2020. «Mining Hacker Forums: A Study of Cybercriminal Environments». IEEE Access, 8, 89926-89937. https://doi.org/10.1109/ACCESS.2020.2980453

11 Rahman, M. T., Islam, M. R., Hoque, M. R., & Islam, M. R. 2020. «Sentiment Analysis of Hacking and Cybersecurity-Related Tweets». In 2020 2nd International Conference on Computer and Communication Engineering (ICCCE) (pp. 255-260). IEEE. https://doi.org/10.1109/ICCCE49051.2020.9200770

12 Liang, H., Yang, C., & Wang, H. 2020. «Cybersecurity sentiment analysis using machine learning». IEEE Access, 8, 54484-54494. https://doi.org/10.1109/ACCESS.2020.2981284

13 Khoo, C. S., & Gillam, L. 2011. «Sentiment analysis in online forums». The Journal of Systems and Software, 84 (12), 2099-2108. https://doi.org/10.1016/j.jss.2011.06.043

14 Thakur, R., Singh, V. K., & Gupta, P. 2018. «A Review on Sentiment Analysis and Opinion Mining». In 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-6). IEEE. https://doi.org/10.1109/ICRAIE.2018.8594166.

15 Olteanu, A.-M., Stoian, M., & Vrabie, C. 2020. «Sentiment analysis of cybersecurity news». In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 207-212). IEEE. https://doi.org/10.1109/ICCP51668.2020.9217864

16 Karpov, A., Kotelnikov, E., Nikolenko, S., & Koltsova, O. 2018. «Modern sentiment lexicons for opinion mining in English and Russian (analytical survey)». Computational Linguistics and Intellectual Technologies, 17(26), 91-111. https://doi.org/10.28995/2079-8136-2018-26-91-111

17 Mambetov S.T., Begimbayeva Ye.Ye., Khikmetov A.K., Joldasbayev S.K. (2023) Takyryptyk internet resurstardan derekterdi alu algoritmin azirleu. [Development of an algorithm for obtaining data from thematic internet resources]. KR UEA habarshysy, №2(88) https://doi.org/10.47533/2023.1606-146X.6 (in Kazakh)