

ӘОЖ – 004.89

Д.А. Абдраманов, С.Ж. Куракбаева, Ж.Д. Изтаев, П.А. Қожабекова

магистрант, М.Әуезов атындағы ОҚУ, Шымкент, Қазақстан
т.ғ.к., доцент, М.Әуезов атындағы ОҚУ, Шымкент, Қазақстан
п.ғ.к., доцент, М.Әуезов атындағы ОҚУ, Шымкент, Қазақстан
т.ғ.к., доцент, М.Әуезов атындағы ОҚУ, Шымкент, Қазақстан

***Корреспондент авторы:** Zhalgasbek71@mail.ru

ВЕБ-КОНТЕНТТЕГІ ҚЫЛМЫСТЫҚ МӘТІНДЕРДІҢ ЕРЕКШЕЛІКТЕРІН ТАЛДАУ

Түйін

Интернеттің негізі - веб-контент, бұл әртүрлі онлайн платформалар арқылы жасалған және таратылатын мәтіндер, суреттер, бейне, аудио элементтерінің жинағы. Ол ойлауымызға, шешім қабылдауымызға және пікір қалыптастыруымызға әсерін тигізеді. Осы орайда веб-контенттің тез таралуымен интернеттің пайдасымен қатар белгілі бір қауіп-қатерлері де бар және әлеуметтік мәселелер туғызады. Сол мәселелердің бірі - веб-контентте қылмыстық мәтіндердің болуы. Ол қылмыстар, заңсыз әрекеттер, зорлық-зомбылық, заңсыз сауда және басқа да келеңсіз құбыл ыстар туралы ақпаратты қамтиды.

2020 жылы жаппай онлайнға көшу нәтижесінде киберқылмыстардың саны 2019 жылмен салыстырғанда 358%-ға, 2021 жылы онлайн қылмыстар саны алдыңғы жылмен салыстырғанда 125%-ға өсті. Әлемде әр сағат сайын 97 адам деректердің сыртқа шығуының құрбаны болды. 2022 жылдың бастапқы жартыжылдығында 236,1 миллион төлемдік бағдарлама шабуылы тіркелді және әрбір 5-ші ғаламтор пайдаланушысы зиянды сілтемесі бар электрондық пошта хабарламаларын ашты. Сол жылы Meta мобильді пайдаланушылардың Facebook жүйесіне кіру тіркелгі деректерін ұрлауға бағытталған 400-ден астам зиянды iOS және Android қолданбаларын тапты. Осылайша киберқылмыс, алаяқтық, хакерлік шабуылдар және басқа да қылмыс түрлері қоғам үшін күрделі мәселеге айналды.

Кілттік сөздер: Веб-контент, Семантика, iOS, Android, Facebook, Instagram, Twitter, машиналық оқыту.

Кіріспе

Қазіргі қоғамда ғаламтор желісі өміріміздің барлық салаларында басты рөл атқарады. Интернетке кең қол жетімділік пен цифрлық технологиялардың дамуының арқасында адамдар санаулы секундтарда нақты уақыт режимінде деректерге емін-еркін қол жеткізуге және қарым-қатынас жасауға мүмкіндік алды.

Қылмыстарды жоспарлау және оған шақыру, жалған ақпаратпен бөлісу сияқты қылмыстық мәтіндер желілік ортадағы қауіпсіздікке қауіп төндіреді. Мұндай криминалды мәтіндерді анықтау және жіктеу желідегі қылмыспен күрестің құрамдас бөлігіне айналууда. Желіде қолжетімді ақпарат көлемінің ұлғаюына және Интернетке қатысты заңға қарсы әрекеттер көбеюіне байланысты қылмыстық мәтіндерді автоматты түрде анықтау және саралаудың тиімді әдістері мен тәсілдерін әзірлеу қажет[1].

Веб-контенттегі қылмыстық мәтіндердің ерекшеліктерін талдау

Қылмыстық мәтіндер қылмыстар, жасау әдістері, оларды ұйымдастыру, алаяқтық схемалары және басқа да заңсыз әрекеттер туралы мәліметтерді қамтиды. Мұндай мәтіндер оларды тануға және жіктеуге көмектесетін белгілі бір тілдік ерекшеліктер мен стилистикаға ие болып келеді. Олардың ерекшеліктерін талдау осындай мазмұндарды қарапайым мәтіндерден ерекшелендіретін белгілер мен белгілерді анықтауға мүмкіндік береді.

Қылмыстық мәтіндердің тағы бір ерекшелігі олардың мазмұнында болады. Оларда қылмыстар, қылмыскерлер, олардың мотивтері, жасау әдістері мен зардаптары туралы мәліметтер болуы мүмкін. Бұған қылмыстардың сипаттамасы, алаяқтық схемалары, заңсыз мазмұнды тарату және т.б. кіреді. Қылмыстық мәтіндерде қоқан-лоққылар, зорлық-зомбылық, балағат сөздер және пайдаланушылар үшін зиянды және қорлауы мүмкін басқа мазмұн болуы

мүмкін. Бұған қоса, веб-мазмұндағы қылмыстық мәтіндерді табу және жою қиын болуы мүмкін. Қылмыскерлер өз материалдарын жасыру және тарату үшін әртүрлі әдістер мен тактикаларды қолданады. Құқық қорғау органдары мен веб-сайт әкімшілері анықтамауы үшін олар анонимді тіркелгілерді, ақпаратты шифрлауды, жасырын сілтемелерді және басқа әдістерді пайдаланады.

Қылмыстық мәтіндердің белгілерін анықтауға арналған жұмыстарда негізінде төменде веб-контенттегі қылмыстық мәтіндердің негізгі ерекшеліктері анықталды[2]:

1. Арнайы лексика және терминология: Веб-контенттегі қылмыс мәтіндері көбінесе қылмысқа, заң бұзушылықтарға, құқық қорғау органдарына және сот ісін жүргізуге қатысты арнайы лексика мен терминологияны, жаргондарды және аббревиатураларды қамтиды. Бұл терминдер әртүрлі қылмыс түрлеріне, қылмыстық әрекеттің әдістеріне және құқықтық ұғымдарға қатысты қолданылуы мүмкін. Бұл өрістен тыс адамдар үшін белгісіз немесе түсіну қиын болуы мүмкін бірегей өрнектерді, қысқартуларды немесе кодтық сөздерді қамтуы мүмкін.

2. Стыльдік белгілер: Қылмыстық мәтіндердің басқа мәтін түрлерінен ерекшеленетін өзіндік стильдік белгісі болуы мүмкін. Бұл дөрекі және жанама сөздерді қолдануды, шиеленісті және драмалық атмосфераны құруды және сарказм, ирония немесе қара юмор элементтерін қолдануды қамтуы мүмкін.

3. Эмоционалды бояу: веб-мазмұндағы қылмыстық мәтіндер оқырмандарда күшті эмоционалдық реакцияларды тудыруы мүмкін. Олар зорлық-зомбылық, қылмыстар немесе қорқыныш, реніш немесе эмпатия сезімін тудыратын құрбандардың сипаттамаларын қамтуы мүмкін.

4. Арнайы пішім мен құрылым: Веб-контенттегі қылмыстық мәтіндер арнайы пішім мен құрылымға ие болуы мүмкін. Мысалы, оларда қылмыстардың егжей-тегжейлі сипаттамасы, айғақтар, заңды құжаттар немесе қылмыстық оқиғалар туралы жаңалықтар болуы мүмкін. Оларды әңгімелер, баяндамалар немесе аналитикалық мақалалар түрінде де ұйымдастыруға болады.

5. Манипуляция және алдау: веб-мазмұндағы қылмыстық мәтіндер оқырмандарды манипуляциялау және алдау үшін пайдаланылуы мүмкін. Оларда белгілі бір имидж жасау немесе аудиторияның пікіріне әсер ету мақсатында жалған ақпарат, жалған дәлелдер немесе жасырын ниет болуы мүмкін.

Қылмыстық мәтіндерді манипуляциялау қаржылық шығындар, қауіпсіздікке қауіп төндіруі немесе қылмыс деңгейінің артуы сияқты ауыр зардаптарға әкелуі мүмкін. Сондықтан манипуляция мен алаяқтық кең таралған онлайн ортада қырағы болу және ақпаратты сыни тұрғыдан бағалау маңызды[3].

Веб-контенттегі қылмыстық мәтіндердің осы ерекшеліктерін талдау ақпараттың осы түрін жақсы түсінуге және өңдеуге, пайдаланушылардың қауіпсіздігі мен қорғалуын қамтамасыз етуге және онлайн ортадағы қылмыстық әрекеттерге тиімді қарсы тұруға көмектеседі. Одан бөлек қылмыстық әрекетке байланысты веб-мазмұн құпиялылық пен құқық қорғау органдарының тыңдауынан немесе бақылауынан қорғау үшін жиі кодталады және шифрланады. Бұл хабарлар мен деректерді қорғау үшін арнайы бағдарламалар мен алгоритмдерді пайдалануды қамтуы мүмкін.

Мұндай мәтіндер әдетте анонимді түрде немесе бүркеншік аттар арқылы жарияланады. Қылмыстық қауымдастықтың мүшелері өздерінің жеке басын сақтап қалуға және құқық қорғау органдарының анықтауына жол бермеуге тырысады. Олар әрекеттерін анонимді сақтау үшін анонимді браузерлерді, виртуалды жеке желілерді немесе VPN (Virtual Private Network) және басқа құралдарды пайдалана алады. Қылмыстық мәтіндердің ерекшеліктері мен белгілері үнемі өзгеріп отырады және өзгермелі желілік ортаға бейімделеді. Сондықтан веб-контенттегі қылмыстық мәтіндерді талдау мен жіктеудің жаңа әдістері мен құралдарын жасауға

бағытталған зерттеулер өзекті болып табылады.

Қолданыстағы әдістердің әлсіз жақтарын анықтау

Веб-контенттегі қылмыстық мәтіндерді талдаудың қолданыстағы әдістерінің әлсіз жақтарын анықтау жіктеудің тиімділігі мен дәлдігін арттырудың маңызды зор. Кейбір типтік әлсіздіктерге мыналар жатады[4]:

1. Дәлдік жеткіліксіз: кейбір әдістер қылмыстық мәтіндерді анықтауда шектеулі дәлдікке ие болуы мүмкін. Бұл мәтінді алдын ала өңдеудің жеткіліксіздігінен, толық емес мүмкіндіктер жиынынан немесе үлгіні оқытуға арналған шектеулі деректер жиынынан туындауы мүмкін. Нәтижесінде жалған оң немесе жалған теріс нәтижелердің жоғары деңгейі орын алуы мүмкін.
2. Тілдік тәуелділік: Кейбір әдістер тілге тән болуы мүмкін және басқа тілдердегі мәтіндермен тиімді жұмыс істемеуі мүмкін. Бұл көптілді немесе халықаралық мәтіндерді талдау кезінде қиындықтар тудыруы мүмкін.
3. Стандартты емес мәтіндерді өңдеу жоқ: Қолданыстағы әдістер әдетте жаңалықтар мақалалары немесе блогтар сияқты стандартты мәтін пішімдерін талдау үшін әзірленеді. Дегенмен, веб-мазмұндағы қылмыстық мәтіндер түсініктемелер, әлеуметтік желідегі жазбалар немесе шифрлық мәтіндер сияқты стандартты емес пішімдерді қамтуы мүмкін. Бұл оларды тиімді талдау мен жіктеуді қиындатады.
4. Айналуға қарсылық: Қылмыстық элементтер қолданыстағы талдау жүйелерін айналып өту үшін үнемі бейімделіп, жаңа әдістерді дамытады. Бұл синтаксистік немесе семантикалық трюктерді қолдануды, шу туралы ақпаратты енгізуді немесе жаңа мәтін пішімдерін пайдалануды қамтуы мүмкін. Демек, қолданыстағы әдістер мұндай манипуляцияларға осал болып, тиімділігін жоғалтуы мүмкін.
5. Мәтінмәндік талдаудағы шектеулер: Кейбір әдістер қылмыстық мәтіндерді жіктеу кезінде мәтіннің контекстік ерекшеліктерін жеткілікті түрде ескермеуі мүмкін. Мәтінмәндік шынайы табиғатын анықтауда маңызды рөл атқаруы мүмкін, сондықтан оны жеткілікті түрде ескермеу дұрыс емес нәтижелерге әкелуі мүмкін.
6. Семантикалық ақпаратты өңдеу мүмкіндігінің шектеулілігі: Морфологиялық талдаудың кейбір әдістері тек мағыналық мазмұнына жеткілікті мән бермей, мәтіннің формалды және грамматикалық ерекшеліктерін талдаумен шектелуі мүмкін. Дегенмен, қылмыстық мәтіндерде жасырынмағыналар, екіұштылықтар және тереңірек түсінуді қажет ететін арнайы сөздік қолдану болуы мүмкін.
7. Мәтінмәндік ақпаратты елемей: көршілес сөздер, сөйлемдер немесе құжаттар сияқты мәтінмәндік факторлар мәтінді жіктеуге айтарлықтай әсер етуі мүмкін. Бірақ кейбір әдістер олардың арасындағы байланыстар ментәуелділіктерді есепке алмай, жеке сөздерді немесе сөз тіркестерін талдаумен ғана шектелуі мүмкін. Бұл қылмыстық мәтіндердің контекстік ерекшеліктерін жеткіліксіз түсінуге және түсіндіруге әкелуі мүмкін [5].

8. Әлеуметтік-мәдени контексттің ескерілмеуі: веб-контенттегі қылмыстық мәтіндер белгілі бір топтармен, қауымдастықтармен немесе аймақтармен байланысты нақты әлеуметтік-мәдени сипаттамаларға ие болуы мүмкін. Бұл мүмкіндіктер қылмыстық әрекетпен байланысты болуы мүмкін белгілі бір сөздік, қысқартулар немесе техникалық терминдерді пайдалануға әсер етуі мүмкін. Осындай әлеуметтік-мәдени аспектілерді есепке алу қылмыстық мәтіндердің жіктелуін жақсартуға көмектеседі.

Веб-контенттегі қылмыстық мәтіндер үнемі дамып отырады, ал қылмыстық схемалар мен тактика уақыт өте өзгеруі мүмкін. Сондықтан морфологиялық талдау әдістері икемді және қылмыстық мәтіндердің жаңа түрлеріне және қылмыстық әрекеттің өзгеретін үлгілеріне бейімделуге қабілетті болуы керек.

Бұл әлсіз жақтарды анықтау зерттеушілерге осы шектеулерді ескеретін және еңсеретін модификацияланған морфологиялық талдау әдістерін жасауға, веб-контенттегі қылмыстық

мәтіндерді жіктеудің тиімділігі мен дәлдігін арттыруға мүмкіндік береді. Осы әлсіз жақтарды анықтау және түсіну зерттеушілерге осы аспектілерді ескеретін және қолданыстағы тәсілдердің шектеулерін еңсеретін модификацияланған морфологиялық талдау әдістерін жасауға мүмкіндік береді.

Веб-контент пайдаланушылардың және жалпы қоғамның қауіпсіздігіне қауіп төндіретін ақпаратты, оның ішінде қылмыстық мәтіндерді таратудың негізгі алаңына айналды. Мәтіндерді жіктеуге арналған морфологиялық талдау әдістерін өзгертуге бағытталған зерттеу онлайн ортадағы қылмыспен күресу және пайдаланушылардың қауіпсіздігін қамтамасыз ету контекстінде өте өзекті болып табылады. Себебі қылмыстық мәтіндерді дәл анықтау және жіктеу мыналарға мүмкіндік береді:

- ықтимал қауіпті жағдайлар туралы ескерту және қылмыстар туралы ақпарат беру арқылы желідегі пайдаланушылардың қауіпсіздігін арттыру;
- Желілік ортада қылмыстық ақпараттың таралуын жедел анықтау және алдын алу;
- Құқық қорғау органдары мен киберқауіпсіздік мамандарына қылмыстық мәтіндерді автоматты түрде өңдеу және талдау құралдарымен қамтамасыз ету арқылы олардың тиімділігін арттыру[6].

Қорытындылай келе, веб-контенттегі қылмыстық мәтіндер назар аударуды және шешуді қажет ететін күрделі мәселе болып табылады. Олардың әртүрлілігі, таралуы және мазмұны оларды анықтау мен күресуді қиындатады. Дегенмен, қазіргі заманғы машиналық оқыту әдістерін және NLP технологияларын қолданудың арқасында веб-контентте қылмыстық мәтіндердің таралуын анықтау және алдын алу, интернет пайдаланушыларының қауіпсіздігі мен қорғауын қамтамасыз етудің тиімді жүйелерін құруға болады.

Әдебиеттер тізімі

1. Гаужаева В. А., Прокофьева Е. В., Прокофьева О. Ю. Преступность в сети Интернет: криминологические характеристики // Вестник экономической безопасности. — 2019. — № 4. — С. 111–114.
2. Бородкина Т. Н., Павлюк А. В. Киберпреступления: понятие, содержание и меры противодействия // Социально-политические науки. — 2018. — № 1. — С. 135–137.
3. Болушевская И. Н. Корпусное исследование лингвистических особенностей дискурса криминальных интернет-новостей на примере новостей о похищении (на материале английского языка) // Филологические науки. Вопросы теории и практики. — 2019. — Т. 12, вып. 10. — С. 184–188.
4. Барахнин В. Б., Федотов А. М., Бакиева А. М., Бакиев М. Н., Тажибаева С. Ж. Алгоритмы генерации и стемматизации словоформ казахского языка // Вестник Казахского национального технического университета. — 2017. — № 3. — С. 123–130.
5. Bird S., Klein E., Loper E. Natural Language Processing with Python. — Sebastopol: O'Reilly Media, 2018. — 479 p.
6. Литвинова Т. А. Возможности компьютерной лингвистики для решения задач диагностирования личности по тексту // Вестник Воронежского государственного университета. Серия: Филология. Журналистика. — 2015. — № 3. — С. 37–41.

References

1. Gauzhaeva V. A., Prokof'eva E. V., Prokof'eva O. Ju. Prestupnost' v seti Internet: kriminologicheskie harakteristiki // Vestnik jekonomicheskoy bezopasnosti. — 2019. — № 4. — S. 111–114.
2. Borodkina T. N., Pavljuk A. V. Kiberprestuplenija: ponjatie, sodержanie i mery protivodejstvija // Social'no-politicheskie nauki. — 2018. — № 1. — S. 135–137.
3. Bolushevskaja I. N. Korpusnoe issledovanie lingvisticheskikh osobennostej diskursa kriminal'nyh internet-novostej na primere novostej o pohishhenii (na materiale anglijskogo jazyka) //

Filologicheskie nauki. Voprosy teorii i praktiki. — 2019. — Т. 12, vyp. 10. — С. 184–188.

4. Barahnin V. B., Fedotov A. M., Bakieva A. M., Bakiev M. N., Tazhibaeva S. Zh. Algoritmy generacii i stemmatizacii slovoform kazahskogo jazyka // Vestnik Kazahskogo nacional'nogo tehniceskogo universiteta. — 2017. — № 3. — С. 123–130.

5. Bird S., Klein E., Loper E. Natural Language Processing with Python. — Sebastopol: O'Reilly Media, 2018. — 479 p.

6. Litvinova T. A. Vozmozhnosti komp'juternoj lingvistiki dlja reshenija zadach diagnostirovanija lichnosti po tekstu // Vestnik Voronezhskogo gosudarstvennogo universiteta. Serija: Filologija. Zhurnalistika. — 2015. — № 3. — С. 37–41.

Д.А. Абдраманов, С.Ж. Куракбаева, Ж.Д. Изтаев*, П.А. Кожабекова

магистрант, ЮКУ им. М. Ауэзова, Шымкент, Казахстан

к.т.н., доцент, ЮКУ им. М. Ауэзова, Шымкент, Казахстан

к.п.н., доцент, ЮКУ им. М. Ауэзова, Шымкент, Казахстан

к.т.н., доцент, ЮКУ им. М. Ауэзова, Шымкент, Казахстан

*Автор для корреспонденции: Zhalgasbek71@mail.ru

АНАЛИЗ ОСОБЕННОСТЕЙ КРИМИНАЛЬНЫХ ТЕКСТОВ В ВЕБ-КОНТЕНТЕ

Аннотация

Основой интернета является веб-контент — это совокупность текстов, изображений, видео и аудиоэлементов, которые создаются и распространяются через различные онлайн-платформы. Он влияет на наше мышление, принятие решений и формирование мнений. Однако наряду с быстрым распространением веб-контента и его пользой, интернет также создает определенные риски и социальные проблемы. Одной из таких проблем является наличие в веб-контенте криминальных текстов. Эти тексты содержат информацию о преступлениях, незаконных действиях, насилии, нелегальной торговле и других негативных явлениях.

В 2020 году, в результате массового перехода в онлайн, число киберпреступлений увеличилось на 358% по сравнению с 2019 годом, а в 2021 году количество онлайн-преступлений возросло на 125% по сравнению с предыдущим годом. В мире каждый час 97 человек становились жертвами утечки данных. В первой половине 2022 года было зафиксировано 236,1 миллиона атак с использованием вредоносных программ, и каждый пятый пользователь интернета открывал электронные письма со ссылками, содержащими вирусы. В том же году Meta обнаружила более 400 вредоносных приложений для iOS и Android, нацеленных на кражу данных учетных записей пользователей Facebook. Таким образом, киберпреступность, мошенничество, хакерские атаки и другие виды преступлений стали серьезной проблемой для общества.

Ключевые слова: Веб-контент, Семантика, iOS, Android, Facebook, Instagram, Twitter, машинное обучение.

D.A. Abdramanov, S.Zh. Kurakbaeva, Zh.D. Iztaev, P.A. Kozhabekova

master's student, M. Auezov SKU, Shymkent, Kazakhstan

Cand.Tech.Sci., Associate Professor, M. Auezov SKU, Shymkent, Kazakhstan

Cand.Ped.Sci., Associate Professor, M. Auezov SKU, Shymkent, Kazakhstan

Cand.Tech.Sci., Associate Professor, M. Auezov SKU, Shymkent, Kazakhstan

*Corresponding author's email: Zhalgasbek71@mail.ru

ANALYSIS OF FEATURES OF CRIMINAL TEXTS IN WEB CONTENT

Abstract

The foundation of the Internet is web content, a collection of texts, images, videos, and audio elements created and distributed through various online platforms. It influences our thinking, decision-making, and the formation of opinions. However, alongside its benefits, the rapid dissemination of web content also brings

certain risks and social issues. One of these issues is the presence of criminal content on the web. This includes information about crimes, illegal activities, violence, illicit trade, and other negative phenomena.

In 2020, as a result of the massive transition to online platforms, the number of cybercrimes increased by 358% compared to 2019, and in 2021, the number of online crimes rose by 125% compared to the previous year. Globally, every hour, 97 people became victims of data breaches. In the first half of 2022, 236.1 million payment program attacks were recorded, and every fifth internet user opened email messages containing malicious links. That same year, Meta identified over 400 malicious iOS and Android applications aimed at stealing Facebook login credentials from mobile users. Thus, cybercrime, fraud, hacking attacks, and other types of crimes have become significant challenges for society.

Keywords: Web content, Semantics, iOS, Android, Facebook, Instagram, Twitter, machine learning.