

Торайғыров университетінің хабаршысы  
ҒЫЛЫМИ ЖУРНАЛЫ

НАУЧНЫЙ ЖУРНАЛ  
Вестник Торайғыров университета

---

# Торайғыров университетінің ХАБАРШЫСЫ

Энергетикалық сериясы  
1997 жылдан бастап шығады



## ВЕСТНИК Торайғыров университета

Энергетическая серия  
Издается с 1997 года

ISSN 2710-3420

---

№ 1 (2023)

ПАВЛОДАР

**НАУЧНЫЙ ЖУРНАЛ**  
**Вестник Торайгыров университета**

**Энергетическая серия**

выходит 4 раза в год \_\_\_\_\_

**СВИДЕТЕЛЬСТВО**

о постановке на переучет периодического печатного издания,  
информационного агентства и сетевого издания

№ 14310-Ж

выдано

Министерство информации и общественного развития  
Республики Казахстан

**Тематическая направленность**

публикация материалов в области электроэнергетики,  
электротехнологии, автоматизации, автоматизированных и  
информационных систем, электромеханики и теплоэнергетики

**Подписной индекс – 76136**

---

<https://10.48081/BNAS6555>

**Бас редакторы – главный редактор**

Кислов А. П.  
*к.т.н., профессор*

Заместитель главного редактора

Талипов О. М., *доктор PhD*

Ответственный секретарь

Калтаев А.Г., *доктор PhD*

**Редакция алқасы – Редакционная коллегия**

Клецель М. Я., *д.т.н., профессор*  
Новожилов А. Н., *д.т.н., профессор*  
Никитин К. И., *д.т.н., профессор (Россия)*  
Никифоров А. С., *д.т.н., профессор*  
Алиферов А.И., *д.т.н., профессор (Россия)*  
Кошеков К.Т., *д.т.н., профессор*  
Приходько Е.В., *к.т.н., профессор*  
Оспанова Н. Н., *к.п.н., доцент*  
Нефтисов А. В., *доктор PhD*  
Омарова А.Р., *технический редактор*

---

За достоверность материалов и рекламы ответственность несут авторы и рекламодатели  
Редакция оставляет за собой право на отклонение материалов  
При использовании материалов журнала ссылка на «Вестник Торайгыров университета» обязательна

© Торайгыров университет

**\*А. Б. Тоқтарова<sup>1</sup>, Б. С. Омаров<sup>2</sup>, Ж. Ж. Ажибекова<sup>3</sup>  
С. А. Мамиков<sup>4</sup>**

<sup>1</sup>Қожа Ахмет Ясауи атындағы Халықаралық Қазақ – Түрік Университеті, ,  
Қазақстан Республикасы, Түркістан қ.;

<sup>2</sup>Әл - Фараби атындағы Қазақ Ұлттық университеті,  
Қазақстан Республикасы, Алматы қ.;

<sup>3</sup>С. Ж. Асфендияров атындағы Қазақ ұлттық медицина университеті  
Қазақстан Республикасы, Алматы қ.;

<sup>4</sup>Академик Ә.Қуатбеков атындағы Халықтар достығы университеті, ,  
Қазақстан Республикасы, Шымкент қ.

e-mail: toktar.aigerim@list.ru

## **ОНЛАЙН КОНТЕНТТЕГІ ҒАДАУАТ СӨЗДЕРДІ АНЫҚТАУДА ЖАСАНДЫ ИНТЕЛЛЕКТІҢ МАҢЫЗЫ**

*Бұл мақалада қазіргі таңда интернет желісінің күн санап дамуы адамзат тарихында жылдамдық пен сандық өмірге қол жеткізіп қана қоймай, оның екінші беті, яғни онлайн контентті қорлау немесе кесміту мақсатында қолданып жүрген желі қолданушыларының бар екендігі туралы, сонымен қатар, әлеуметтік желілердегі балағат тілде жазылатын пікірлерді жинау үшін оңай және автоматтандырылған қадамдарды машиналық оқыту арқылы жүзеге асыруға болатыны туралы айтылған. Бүгінде, онлайн контенттегі жағымсыз пікір қалдырушылар саны қаптап кетуі проблемалар қатарын көбейтуде. Онлайн контенттен келетін қауіптің бір түрі – кибербуллинг екені белгілі, бұл өз кезегінде желі қолданушыларды эмоциональды күйге түсіріп, қандай да бір психикалық зардап әкеледі.*

*Біз Қазақстанда пайдаланатын әлеуметтік желілер мен бұқаралық ақпарат құралдарына жариялаған бейәдеп пікірлерден тұратын деректер қорын жинап жатырмыз. Көптеген әлеуметтік желілерден түскен шағымдарды зерделей отырып байқағанымыз: онлайн контентте кемсітетін яғни, балағаттау пікірлерінің жариялануы күн сайын артып келе жатқанын көруге болады.*

*Біздің зерттеулер нәтижесі машиналық оқыту тәсілдерін қолдану арқылы әлеуметтік желілерде жарияланған дәрекі сөздердің түбірін зерттеп қана қоймай, сонымен қатар, гадауат пікірлердің түрлерін ажыратып және автоматтандырылған деректер жиынына қол жеткізуге көмектеседі.*

*Кілтті сөздер: онлайн контент, жасанды интеллект, бейәдеп сөздер, парсинг, табиғи тіл*

## **Кіріспе**

Интернеттің дамуы әртүрлі мәтін форматтағы ақпарат көлемінің бақыланбайтын экспоненциалды өсуіне әкеледі. Интернетті пайдалану уақытының шексіз, яғни 24 сағат қол жетімділігі, жасырын пайдаланушы ретінде және саны көп онлайн аудиториялармен жұмыс жасай алу мүмкіндігін беріп отыр. Қазіргі уақытта интернеттегі зорлық- зомбылық, жағымсыз пікірлер адамдардың психологиялық – эмоциональды денсаулығына қауіп төніп тұрғандықтан өзекті болып отыр.

Технологиялық процесстің дамуы оң мүмкіндіктерімен қатар оны теріс әсер ететін жақтары мен салдары бар екенін ескеруді талап етеді, бұл дегеніміз деструктивті іс әрекеттердің өсуіне алып келеді. Біздің жағдайымыздағы деструктивті деректер – бұл онлайн контент қолданушысына қандай да бір түрде залал келтіру, яғни бейәдеп пікірлерді қалдыру арқылы зиянын тигізу болып табылады.

Заманауи желідегі орасан зор ақпарат көлемі жеке тұлғалардың мүдделері мен құқықтарын қорғауды талап етеді. Тіпті, жаһандық деңгейде де өз салдарын тигізуге болатынын көрумізге болады, мысалы ретінде онлайн контенттегі халықаралық даулар мен қақтығыстардан көруге болады.

Мысалы, 2021 жылы жүргізілген әлем мемлекеттерінің интернетті қолдану деңгейі статистикасы бойынша Қазақстан мемлекеті жалпы халықтың 15,47 миллионы, яғни 81,9 % көрсетіп отыр. Желідегі ақпараттар аптасына 7 күн, 24 сағат бойы қолжетімді, интернеттегі «шабуылдан» қорғана алмайсың, себебі электрондық пошта, ұялы телефонға, түрлі жеке әлеуметтік желілер сайтына кез- келген уақытта келіп түсуі мүмкін. Бұл адамдарға, оның ішінде жасөспірімдерге психологиялық түрде әсер әтеді. Сонымен қатар, инетернетте жасырын қолданушы бола білу мүмкіндігі бар, яғни кімнен келген «шабуыл» екенін білмеуіңіз мүмкін. Физикалық зорлық – зомбылыққа қарағанда эмоционалдық «шабуылдың» соны адамды психологиялық ауытқуларға әкеліп соғу қаупі жоғары екенін психологтар айтуда.

Зерттеудің нысаны: онлайн контенттегі бейәдеп сөздерді машиналық оқыту арқылы анықтау

Зерттеудің пәні: машиналық оқытуды онлайн контенттегі ғадауат сөздер топтамасын анықтауда қолдану

**Мақсаты:** онлайн контент қолданушыларынан келіп түскен ғадауат сөздер тізбегінен автоматтандырылған мәліметтер қорын жасақтау және машиналық оқыту арқылы анықтау

**Зерттеу жұмысының міндеттері:**

Ғадауат сөздерді парсерлеу арқылы, яғни көрсетілген желі ресурстарындағы жағымсыз пікірлерді жинап, сұрыптап, қажетті форматта ұсыну;

Есептелген алгоритмдердің орындалуын қамтамасыз ету;

Үлкен көлемді іздеу модельін құрып, массивті мәліметтер қорына сұрыптау жасау;

**Материалдар мен әдістері**

Жасанды интеллект дегеніміз дайын дайын мәліметтер қорын пайдалана отырып, тілді оқуды қайталауға және іздеу процестерін автоматты түрде жүзеге асыруға мүмкіндік береді. Дегенмен, жасанды интеллект машиналық техникадан айырмашылығы өте жоғары. Жасанды интеллекттің мақсаты – қолмен жасалатын жұмыстарды автоматтандыру ғана емес, сонымен қатар, программаланған ауқымды жүйелермен жұмыс істеу болып табылады. Жүйені инициализациялаумен қатар адамның қолымен жасалуды талап етеді [1].

Жасанды интеллекттің ерекшеліктеріне тоқталып өтер болсақ, автоматтандырудан бастап, коммуникациялық платформалар, боттар және смарт компьютерлер үлкен көлемдегі деректермен біріктіріліп, қауіпсіздік деректерін талдаудан инвестициялық талдау құралдарына дейін үйлер мен кеңселерде қолданылатын әртүрлі технологияларды жақсарта алады.

Жасанды интеллект көптеген жасырын қабаттары бар нейрондық желілерді пайдалана отырып, деректердің үлкен көлемін тереңірек талдауды жүзеге асырады. Терең оқыту үлгілері деректердің үлкен көлемін қажет етеді, өйткені бұл жасанды интеллекттің негізі болып табылады. Сондықтан деректер неғұрлым көп болса, модель соғұрлым дәлірек болады [2].

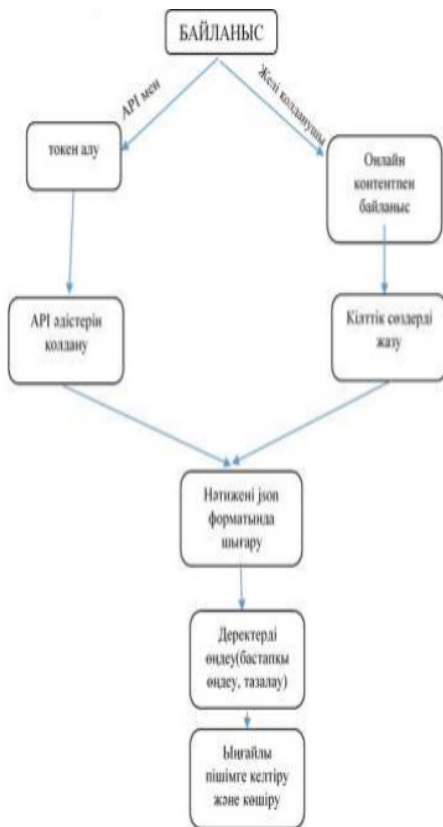
**Зерттеу нәтижесі**

Онлайн контенттегі қазақ тілді тұтынушысы бар әлеуметтік желілерден мәліметтерді жинау барысында 2 типті парсинг құру туралы шешім қабылданды: API негізінде парсинг жасау және кілттік сөздерді пайдалану арқылы өзіңнің «веб - шолғышыңды» жасау. Әр типтің өз артықшылықтары мен кемшіліктері бар.

Қолданбалы бағдарламалау интерфейсі (API) – бір бағдарлама басқа бағдарламамен әрекеттесе алатын жаңа бағдарламалық қолданбаларды құруға арналған интерфейс түріндегі класстар, процедуралар мен функциялар жиынтығы [3]. Бұл бағдарлама жасаушыларға өз өнімдерінің

функционалдығын кеңейтуге және оны бағдарламалық жасақтамалармен байланыстыруға мүмкіндік береді.

1 - суретте мәліметтер қорын жинау процессіндегі 2 типті парсинг жасаудың сызбасы көрсетілген.

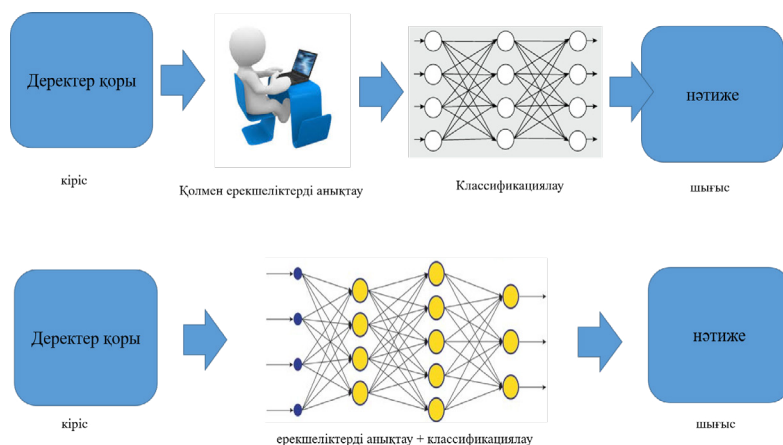


Сурет 1– Мәліметтер қорын жинау процессіндегі 2 типті парсинг жасаудың сызбасы

Біз формалды іске асыру моделінің нұсқасын ұсынамыз бұл процесс инженерлік тәсіл болып табылады, оған сәйкес компьютерлік объект (құбылыс) компьютердің көмегімен ойнатылады.

Белгілі бір тілдегі балағат сөздерге қатысты мәтіндерді табу қиын. Қазақ тіліне арналған жағымысыз сөздері бар өз корпусымызды құруды жөн көрдік.

Сызба екі схемадан тұрады: мәтіндік жазбаларды қамтитын әлеуметтік желілердегі балағат сөздер және вербалды емес мәлімдемелер. Біз деректерді жинау үшін аналитикалық технологияны қолдандық. Бірінші схема үшін API әзірлеушісі әзірлеген аналитикаға сәйкес, бұл әдіс жұмыс істеуі үшін алдымен қосылымды орнату керек, яғни, токенді алу арқылы қосылымға қол жеткізуге болады. Токенді қолданбаның әзірлеуші бөлімінен тікелей алуға болады, содан кейін дерекқорға қол жеткізу үшін пайдаланылады. Кез келген формат үшін деректерді түрлендіру қызметтерін ұсына алады. Содан кейін, деректер қоры «тазаланады», содан кейін бұл деректер сақталады және оны машиналық оқытуда пайдалануға болады. Сызбаның екінші тармағында басқаша конфигурацияланған, мұнда сіз веб-беттерге кіру үшін өзіңіздің пайдаланушы интерфейсін жасай аласыз және деректер қорына қол жеткізуге мүмкіндік бар [4]. Яғни, біз онлайн контент сайттарынан, әлеуметтік желілерден «эмоционалды» деректерді жинаймыз. Бұл әдіс бастапқыда жалықтыратын болса да, деректерді жинауды аяқтаған кезде операция логикасы біріктірілгенін байқайсыз.



Сурет 2 – Мәліметтер қорын жинаудағы 2 типті парсингтің жұмыс істеу схемасы

Онлайн контенттегі мәліметтер қорын жинап болған соң, электрондық зорлық-зомбылықтың болуын анық көрсететін ауызша маркерлердің сегіз класы анықталды: [5]

1 Тыйым салу және әдепсіз сөздер (бір құрамды, екі компонентті, үш құрамды және көп компонентті) – ол жәбірленушінің беделін түсіру, оны

манипуляциялау және оның кемшілігін көрсету үшін ашық ауызша агрессия құралы ретінде қолданылады;

2 Адамның интимдік өміріне қатысты сөздер – көбінесе бұзақылар қолданушының сезімін жаралау үшін қолданылады. Көптеген тілдерде жыныс мүшелерінің атаулары мен жыныстық әрекет түрлері тыйым салынған сөздік болып табылады және адресаттың агрессиясын білдіру үшін жиі қолданылады;

3 Жыныстық қатынас ұғымдарына қатысты сөздер жыныстық бағдары және сексизм - жоғарыда көрсетілген мақсаттарға жету үшін құрбандар көбінесе гей, лесбиян, педофил/педофил, сексист, транс және т.б. және адресат сияқты дәстүрлі емес жыныстық бағдарын атайды және/немесе атап көрсетеді;

4 Зұлымдықты және өлімді қалауды білдіретін сөздер – мұндай лексикалық бірліктерді қолдану қылмыскерді қатты ұнатпаудан, тіпті жәбірленушіні жек көруден туындайды;

5 Адамды қорлауды білдіретін сөздер – көбінесе «бірнеше бұзақы – бір құрбан» схемасы бойынша электронды қудалау орын алған кезде қолданылады. Лексикалық сөздер (намыссыз, мылқау, лас, надан, ақымақ, т.б.) үнемі сынға алып, олардың қабілетін төмендетіп, жәбірленушінің өзін-өзі бағалауы мен өзін-өзі бағалауын түсіруге бағытталған;

6 Ұлтқа және нәсілшілдікке қатысты ұғымдарды атайтын сөздер, мысалы, қарабет, хакер, өзбек, нәсілшіл, социопат және т.б. – агрессорға жәбірленушіге моральдық ғана емес, сонымен бірге әлеуметтік зиян келтіруге мүмкіндік беруі мүмкін;

7 Жануарлар атымен аталатын сөздер, жануарлармен салыстыру арқылы қолдану яғни, сиыр, есек, шошқа, мал, қой, сұржылан, т.б. жәбірленушінің мінез-құлқын жануарлардың қасиеттерімен теңестіреді, бұл оның қоғамдағы мәртебесін төмендетеді және адамның өзін-өзі бағалауын төмендетеді;

8 Физикалық және ақыл-ой кемістігі арқылы адамдарды балағаттаушы сөздер;

### **Талқылаулар**

Мәтінді талдау үш негізгі форматта жүзеге асырылады: классификация, мазмұнды рефлексия және көңіл-күйді талдау. Барлық мәтінді жіктеу есептерін екі түрге бөлуге болады:

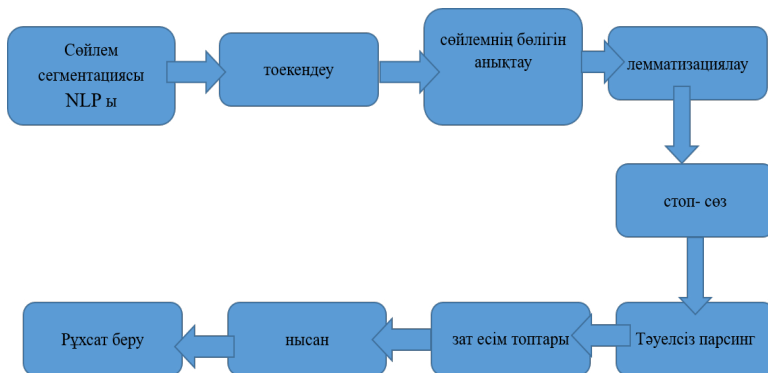
– екілік классификация пайдаланушыға ұсынылатын құжаттың өзектілігін анықтауға мүмкіндік береді;

– көп класты классификация құжаттың тақырыбын анықтауға және оны жүздеген пәндік сыныптардың біріне тағайындауға мүмкіндік береді [6].

бастады және NLP мәселелерін шешудің жалпы тәсілін қалыптастырды.

NLP конвейері





Сурет 3- NLP конвейері сызбасы

Бұл тәсілдің мәні мәселені бірнеше ретгі ішкі тапсырмаларға бөліп, олардың әрқайсысын жеке шешу болып табылады. Мәселені шешу кезінде екі бөлікті шартты түрде ажыратуға болады: кіріс деректерді алдын ала өңдеу (әдетте бұл ең көп уақытты алады) және модельді құру [7].

#### Жеті негізгі кезеңі бар:

1) Кез келген NLP тапсырмаларын шешу үшін орындалатын конвейердің алғашқы екі қадамы сегменттеу (мәтінді сөйлемдерге бөлу) және лексемалау (сөйлемдерді лексемаларға бөлу, яғни бөлек сөздер) болып табылады.

2) Әрбір жетонның ерекшеліктерін есептеу. Токеннің контекстен тәуелсіз атрибуттары есептеледі.

Күрделі морфологиясы бар тілдер үшін (қазақ тілі) морфологиялық белгілер де маңызды: мысалы, қай жағдайда зат есім, сын есімнің қай түріне жататындығы. Морфология сөздерді бастапқы пішініне келтіру үшін де қажет, оның көмегімен біз ерекшелік кеңістігінің көлемін азайта аламыз. Мысалы ретінде: «Саған бүгін келу қажет», «Мен келдім». Екі сөйлемде де «кел» етістігі бар, бірақ әр түрлі жалғаулармен аяқталған. Егер мәтіндер компьютерде өңделетін болса, біз «кел» сөзін пайдаланып бірдей ұғым туралы айтып жатқанымызды түсіну үшін ол әрбір сөздің бастапқы түбірін білуі керек. Әйтпесе, «келу» және «келдім» таңбалауыштары (токен) мүлдем басқаша қабылданады. NLP-де бұл процесс лемматизация деп аталады. [8]

3) Тоқтау (стоп-слова) сөздердің маңыздылығын анықтау және сүзгілеу. Қазақ тілінде көптеген көмекші сөздер бар, мысалы, «және», «сонымен», «немесе». Мәтінді статистикалық талдау кезінде бұл таңбалауыштар басқаларға қарағанда жиі кездесетіндіктен, қиындық тудырады. Сондықтан олар тоқтау сөздері ретінде белгіленіп, сөйлем «арамшөптерден» тазартылады.

4 Сөзді алмастыру. Қазақ тілінде ол, оның, бұның, бұл жерде деген сияқты есімдіктер бар. Мысалы, «Қазір мен Алматы қаласындамын. Бұл жерде тұратын адамдар саны басқа қалалармен салыстырғанда көп.» Яғни көріп отырғанымыздай «Алматы» сөзін «бұл жер» деген сөзімен алмастырдық. Адам контекст негізінде бұл сөздердің сөйлемнен сөйлемге қатынасын бақылай алады. Бірақ NLP моделі есімдіктердің нені білдіретінін білмейді, өйткені ол бір уақытта тек бір сөйлемді қарастырады.

5 Тәуелділіктерді талдау. Бұл қадамның түпкі мақсаты – әрбір таңбалауыш жалғыз түбірі бар «ағаш» құру. Түбір негізгі етістік бола алады. Сондай-ақ екі сөздің арасындағы байланыс түрін орнату керек.

6 Өңделген мәтінді векторлық түрге аудару. Бұл қадам сөздердің векторлық көрінісін құруға мүмкіндік береді. Осылайша, бір контексте қолданылатын сөздердің ұқсас векторлары болады [9].

7 Мақсатқа байланысты модель құру. Мысалы, жаңа мәтіндерді жіктеуге немесе генерациялауға арналған модель.

Белгілі бір мәселені шешу үшін кейбір қадамдарды алып тастауға немесе жаңа қадам қосуға болады. Дегенмен, бұл конвейер NLP-тен практикалық мән алуға мүмкіндік беретін барлық типтік қадамдар мен тәсілдерді қамтиды [10].

### **Қорытынды**

Жоғарыда айтылғандарды түйіндей келе, жағымсыз пікірлерді табу ақпараттық технологияның жұмысы ретінде қарастыруға болады. Қойылып отырған проблеманы шешуде екі шешімді көруге болады: адам өз қолымен жасайды немесе автоматтандырылған жүйеге жүгінеді. Машиналық оқыту ғылымда үнемі өсіп, жаңа қарқынға ие болып отыратын система ретінде қарастыруға болады қорқыту – үркітудің түрлері және оларды анықтаудың белгілері осы ғылыми зерттеу жұмысында келтірілді. Сонымен қатар, деструктивті мәліметтер қорын іздеуде автоматтандырылған жүйе қажет екендігі көрсетілген. Осы мақсатта қолданылған әдістер жиынтығын көрсетіп отырымыз. Іздеу жолдарын және олардың бір – бірінен ерекшелігін байқауға болады.

Жинақталған мәліметтер қорын алдыңғы уақытта табиғи тілдерді өңдеу және машиналық оқыту әдістерін пайдалана отырып онлайн контентте анықтауға қолдануға болатындығын, ғадауат сөздер тізбегінің қай түріне жататындығын анықтауға болатындығын байқап отырымыз. Сонымен қатар, онлайн контенттегі ғадауат сөздерді анықтаудың моделін құруға болатындығын байқап, болашақтағы ғылыми зерттеуіміздің басты мәселесі ретінде алуға болатыны талқыланды.

### **ПАЙДАЛАНЫЛҒАН ДЕРЕКТЕР ТІЗІМІ**

1 **Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K.** (2020, April). Machine learning and deep learning techniques for cybersecurity: a review. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 50-57). Springer, Cham.

2 **Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C.** (2019). Deep learning for real-time social media text classification for situation awareness—using Hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth*, 12(11), 1230–1247.

3 **Mohan, A., Singh, A. K., Kumar, B., & Dwivedi, R.** Review on remote sensing methods for landslide detection using machine and deep learning. *Transactions on Emerging Telecommunications Technologies*, (2021). 32(7), e3998.

4 **Mussiraliyeva, S., Omarov, B., Yoo, P., & Bolatbek, M.** . Applying machine learning techniques for religious extremism detection on online user contents. *Computers, Materials and Continua*, (2021) 70(1), 915–934.

5 **Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K.** Attention-based LSTM network for rumor veracity estimation of tweets. *Information Systems Frontiers*, 1–16. (2020).

6 **Canhoto, A. I., & Clear, F.** Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, (2020). 63(2), 183-193

7 **Yuan, S., & Wu, X.** Deep learning for insider threat detection: Review, challenges and opportunities. *Computers & Security*, (2021). 104, 102221..

8 **Mühlhoff, R.** Human-aided artificial intelligence: Or, how to run large computations in human brains? *Toward a media sociology of machine learning. new media & society*, (2020). 22(10), 1868–1884.

9 **Neupane, D., & Seok, J.** (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8, 93155-93178.

10 **Kumar, A., Singh, J. P., Dwivedi, Y. K., & Rana, N. P.** A deep multi-modal neural network for informative Twitter content classification during emergencies. *Annals of Operations Research*, (2020). 1–32.

## REFERENCES

1 **Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K.** (2020, April). Machine learning and deep learning techniques for cybersecurity: a review. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 50-57). Springer, Cham.

2 **Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C.** (2019). Deep learning for real-time social media text classification for situation awareness—

using Hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth*, 12(11), 1230-1247.

3 **Mohan, A., Singh, A. K., Kumar, B., & Dwivedi, R.** (2021). Review on remote sensing methods for landslide detection using machine and deep learning. *Transactions on Emerging Telecommunications Technologies*, 32(7), e3998.

4 **Mussiraliyeva, S., Omarov, B., Yoo, P., & Bolatbek, M.** (2021). Applying machine learning techniques for religious extremism detection on online user contents. *Computers, Materials and Continua*, 70(1), 915-934.

5 **Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K.** (2020). Attention-based LSTM network for rumor veracity estimation of tweets. *Information Systems Frontiers*, 1-16.

6 **Canhoto, A. I., & Clear, F.** (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, 63(2), 183-193

7 **Yuan, S., & Wu, X.** (2021). Deep learning for insider threat detection: Review, challenges and opportunities. *Computers & Security*, 104, 102221..

8 **Mühlhoff, R.** (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? *Toward a media sociology of machine learning. new media & society*, 22(10), 1868–1884.

9 **Neupane, D., & Seok, J.** (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8, 93155–93178.

10 **Kumar, A., Singh, J. P., Dwivedi, Y. K., & Rana, N. P.** (2020). A deep multi-modal neural network for informative Twitter content classification during emergencies. *Annals of Operations Research*, 1–32.

Материал баспаға 13.03.23 түсті.

\**А. Б. Токтарова*<sup>1</sup>, *Б. С. Омаров*<sup>2</sup>, *Ж. Ж. Ажибекова*<sup>3</sup>, *С. А. Мамиков*<sup>4</sup>

<sup>1</sup>Международный Казахско–Турецкий Университет имени Ходжа Ахмет Ясауи, Республика Казахстан, г. Туркестан;

<sup>2</sup>Казахский Национальный университет им. аль-Фараби, Республика Казахстан, г. Алматы;

<sup>3</sup>Казахский Национальный медицинский университет имени С. Д. Асфендиярова, Республика Казахстан, Алматы;

<sup>4</sup>Университет дружбы народов имени академика А. Куатбекова Республика Казахстан г. Шымкент.

Материал поступил в редакцию 13.03.23.

## **ВАЖНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ИДЕНТИФИКАЦИИ ОСКОРБИТЕЛЬНЫХ СЛОВ В ОНЛАЙН-КОНТЕНТЕ**

*В настоящее время известно, что простые и автоматизированные шаги по сбору ненормативной лексики в социальных сетях можно реализовать с помощью машинного обучения. Сегодня огромное количество негативных комментариев к онлайн-контенту усугубляет проблему.*

*Мы собираем базу нецензурных комментариев, размещенных в социальных сетях и СМИ, используемых в Казахстане. Что мы заметили, проанализировав жалобы, полученные из многих социальных сетей, так это то, что публикация уничижительных, то есть оскорбительных комментариев в онлайн-контенте увеличивается с каждым днем.*

*Результаты нашего исследования с использованием методов машинного обучения помогут нам не только исследовать корень грубых слов, размещенных в социальных сетях, но и различать типы грубых комментариев и получать доступ к автоматизированным наборам данных.*

*Ключевые слова: онлайн контент, искусственный интеллект, нецензурная лексика, парсинг, NLP.*

\*A. B. Toktarova<sup>1</sup>, B. S. Omarov<sup>2</sup>, Zh. Zh. Azhibekova<sup>3</sup>, S. A. Mamikov<sup>4</sup>

<sup>1</sup>Khoja Akhmet Yassawi International Kazakh–Turkish University, Republic of Kazakhstan, Turkistan;

<sup>2</sup>AL-Farabi Kazakh National University, Republic of Kazakhstan, Almaty;

<sup>3</sup>Asfendiyarov Kazakh National Medical University, Republic of Kazakhstan, Almaty;

<sup>4</sup>University of friendship of people's academician A. Kuatbekov, Republic of Kazakhstan, Shymkent.

Material received on 13.03.23

## **THE IMPORTANCE OF ARTIFICIAL INTELLIGENCE IN IDENTIFYING OFFENSIVE WORDS IN ONLINE CONTENT**

*It is now known that simple and automated steps to collect profanity on social networks can be implemented using machine learning. Today, the sheer volume of negative comments on online content exacerbates the problem.*

*We collect a database of obscene comments posted on social networks and media used in Kazakhstan. What we noticed after analyzing the complaints received from many social networks is that the posting of derogatory, i.e. offensive comments in online content is increasing every day.*

*The results of our research using machine learning methods will help us not only explore the root of rude words posted on social networks, but also distinguish between types of rude comments and access automated datasets.*

*Keywords: online content, artificial intelligence, obscene language, parsing, NLP*

Теруге 13.03.2023 ж. жіберілді. Басуға 31.03.2023 ж. кол қойылды.

Электронды баспа

3,44 Мб RAM

Шартты баспа табағы 23.59. Таралымы 300 дана. Бағасы келісім бойынша.

Компьютерде беттеген: А. К. Мыржикова

Корректор: А. Р. Омарова

Тапсырыс № 4039

Сдано в набор 13.03.2023 г. Подписано в печать 31.03.2023 г.

Электронное издание

3,44 Мб RAM

Усл. печ. л. 23.59. Тираж 300 экз. Цена договорная.

Компьютерная верстка: А. К. Мыржикова

Корректор: А. Р. Омарова, Д. А. Кожас

Заказ № 4039

«Toraighyrov University» баспасынан басылып шығарылған

Торайғыров университеті

140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы

Торайғыров университеті

140008, Павлодар қ., Ломов к., 64, 137 каб.

67-36-69

E-mail: [kereku@tou.edu.kz](mailto:kereku@tou.edu.kz)

[www.vestnik-energy.tou.edu.kz](http://www.vestnik-energy.tou.edu.kz)