

ISSN 2518-1726 (Online),  
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

# Х А Б А Р Л А Р Ы

---

---

**ИЗВЕСТИЯ**

НАЦИОНАЛЬНОЙ АКАДЕМИИ  
НАУК РЕСПУБЛИКИ КАЗАХСТАН  
Қазақстан Республикасының  
Ғылым Академиясының  
Әл-Фараби атындағы  
Қазақ ұлттық университетінің

**N E W S**

OF THE ACADEMY OF SCIENCES  
OF THE REPUBLIC OF  
KAZAKHSTAN  
al-Farabi Kazakh National University

**PHYSICO-MATHEMATICAL SERIES**

**1 (345)**

**JANUARY – MARCH 2023**

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

## БАС РЕДАКТОР:

**МУТАНОВ Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас директорының м.а. (Алматы, Қазақстан), **Н=5**

## РЕДАКЦИЯ АЛҚАСЫ:

**КАЛИМОЛДАЕВ Мақсат Нұрәліұлы** (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), **Н=7**

**МАМЫРБАЕВ Өркен Жұмажанұлы** (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), **Н=5**

**БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), **Н=3**

**ВОЙЧИК Вальдемар**, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

**СМОЛАРЖ Анджей**, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), **Н=17**

**ӘМІРҒАЛИЕВ Еділхан Несіпханұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Жасанды интеллект және робототехника зертханасының меңгерушісі (Алматы, Қазақстан), **Н=12**

**КИЛАН Әлімхан**, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония), ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=6**

**ХАЙРОВА Нина**, техника ғылымдарының докторы, профессор, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=4**

**ОТМАН Мохаммед**, PhD, Информатика, коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті (Селангор, Малайзия), **Н=23**

**НЫСАНБАЕВА Сауле Еркебұланқызы**, техника ғылымдарының докторы, доцент, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының аға ғылыми қызметкері (Алматы, Қазақстан), **Н=3**

**БИЯШЕВ Рустам Гакашевич**, техника ғылымдарының докторы, профессор, Информатика және басқару мәселелері институты директорының орынбасары, Ақпараттық қауіпсіздік зертханасының меңгерушісі (Қазақстан), **Н=3**

**КАПАЛОВА Нұрсұлтан Алдажарқызы**, техника ғылымдарының кандидаты, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), **Н=3**

**КОВАЛЕВ Александр Михайлович**, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

**МИХАЛЕВИЧ Александр Александрович**, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), **Н=2**

**ТИГИНЯНУ Ион Михайлович**, физика-математика ғылымдарының докторы, академик, Молдова Ғылым академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

## «ҚР ҰҒА Хабарлары. Информатика сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика-математикалық сериясы*.

Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБҚ ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 218 бөл., тел.: 272-64-39*

*<http://www.physico-mathematical.kz/index.php/en/>*

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2023  
Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

## Главный редактор:

**МУТАНОВ Галимжаир Мутанович**, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=5**

## Редакционная коллегия:

**КАЛИМОЛДАЕВ Максат Нурадилович**, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), **Н=7**

**МАМЫРБАЕВ Оркен Жумажанович**, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), **Н=5**

**БАЙГУНЧЕКОВ Жумадил Жанабаевич**, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Саптаева (Алматы, Казахстан), **Н=3**

**ВОЙЧИК Вальдемар**, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **Н=23**

**СМОЛАРЖ Анджей**, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), **Н=17**

**АМИРГАЛИЕВ Едилхан Несипханович**, доктор технических наук, профессор, академик Национальной инженерной академии РК, заведующий лабораторией «Искусственного интеллекта и робототехники» (Алматы, Казахстан), **Н=12**

**КЕЙЛАН Алимхан**, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=6**

**ХАЙРОВА Нина**, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=4**

**ОТМАН Мохамед**, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), **Н=23**

**НЫСАНБАЕВА Сауле Еркебулановна**, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

**БИЯШЕВ Рустам Гакашевич**, доктор технических наук, профессор, заместитель директора Института проблем информатики и управления, заведующий лабораторией информационной безопасности (Казахстан), **Н=3**

**КАПАЛОВА Нурсулу Алдажаровна**, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

**КОВАЛЕВ Александр Михайлович**, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **Н=5**

**МИХАЛЕВИЧ Александр Александрович**, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), **Н=2**

**ТИГИНЯНУ Ион Михайлович**, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **Н=42**

«Известия НАН РК. Серия информатики».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика-математическая.*

В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 218, тел.: 272-64-39*

*<http://www.physico-mathematical.kz/index.php/en/>*

© Национальная академия наук Республики Казахстан, 2023  
Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

### Chief Editor:

**MUTANOV Galimkair Mutanovich**, doctor of technical sciences, professor, academician of NAS RK, acting General Director of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=5**

### EDITORIAL BOARD:

**KALIMOLDAYEV Maksat Nuradilovich**, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), **H = 7**

**Mamyrbayev Orken Zhumazhanovich**, (Academic Secretary, PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H = 5**

**BAIGUNCHEKOV Zhumadil Zhanabaevich**, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

**WOICIK Waldemar**, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), **H=23**

**SMOLARJ Andrej**, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), **H= 17**

**AMIRGALIEV Edilkhan Nesipkhanovich**, Doctor of Technical Sciences, Professor, Academician of NAS RK, Head of the Laboratory of Artificial Intelligence and Robotics (Almaty, Kazakhstan), **H= 12**

**KEILAN Alimkhan**, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 6**

**KHAIROVA Nina**, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 4**

**OTMAN Mohamed**, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), **H= 23**

**NYSANBAYEVA Saule Yerkebulanovna**, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H= 3**

**BIYASHEV Rustam Gakashevich**, doctor of technical sciences, professor, Deputy Director of the Institute for Informatics and Management Problems, Head of the Information Security Laboratory (Kazakhstan), **H= 3**

**KAPALOVA Nursulu Aldazarhovna**, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=3**

**KOVALYOV Alexander Mikhailovich**, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), **H=5**

**MIKHALEVICH Alexander Alexandrovich**, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), **H=2**

**TIGHINEANU Ion Mihailovich**, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

### News of the National Academy of Sciences of the Republic of Kazakhstan.

#### Series of informatics.

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan No. 16906-Ж, issued 14.02.2018

Thematic scope: *series physical-mathematical series.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 218, Almaty, 050010, tel. 272-64-39*

*<http://www.physico-mathematical.kz/index.php/en/>*

© National Academy of Sciences of the Republic of Kazakhstan, 2023

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES  
OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X

Volume 1, Number 345 (2023), 191-203  
<https://doi.org/10.32014/2023.2518-1726.177>

IRSTI 28.23.37  
UDC 004.032.26

© **A.B. Toktarova**<sup>1\*</sup>, **B.S. Omarov**<sup>2</sup>, **G.N. Kazbekova**<sup>1</sup>, **S.A. Mamikov**<sup>3</sup>,  
**F.E. Temirbekova**<sup>4</sup>, 2023

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

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

<sup>3</sup>University of friendship of people's academician A. Kuatbekov,  
Shymkent, Kazakhstan;

<sup>4</sup>M. Auezov South Kazakhstan University, Shymkent, Kazakhstan.  
E-mail: [toktar.aigerim@list.ru](mailto:toktar.aigerim@list.ru)

## COLLECTING HATE SPEECH DATABASE ON SOCIAL NETWORK IN KAZAKH LANGUAGE BY USING MACHINE LEARNING

**Toktarova Aigerim** — doctoral student, department of Computer engineering, faculty of Engineering, Khoja Akhmet Yassawi International Kazakh – Turkish University. 161200. Turkistan, Kazakhstan.  
E-mail: [toktar.aigerim@list.ru](mailto:toktar.aigerim@list.ru), <https://orcid.org/0000-0002-6265-9236>;

**Omarov Batyrkhan** — PhD, department of Information System, faculty of Information technologies AL-Farabi Kazakh National University, 050040. Almaty, Kazakhstan,  
E-mail: [batyahan@gmail.com](mailto:batyahan@gmail.com), <https://orcid.org/0000-0002-8341-7113>;

**Kazbekova Gulnur** — Cand. Sci (Technical), head of department of Computer engineering, faculty of Engineering, Khoja Akhmet Yassawi International Kazakh – Turkish University. 161200. Turkistan, Kazakhstan,  
E-mail: [kazbekova.gulnur@auy.edu.kz](mailto:kazbekova.gulnur@auy.edu.kz), <https://orcid.org/0000-0002-2756-7926>;

**Mamikov Satmyrza** — senior lector, department of computer science and mathematics, faculty of Faculty of Science and Engineering, University of friendship of people's academician A. Kuatbekov. 160000. Shymkent, Kazakhstan.  
E-mail: [Satmyrza85@mail.ru](mailto:Satmyrza85@mail.ru), <https://orcid.org/0009-0000-85054-2157>;

**Temirbekova Fariza** — lector, department of Information communication technologies, faculty of Information communication technologies, M. Auezov South Kazakhstan University. 160012. Shymkent, Kazakhstan.  
E-mail: [temirbekova.fariza@inbox.ru](mailto:temirbekova.fariza@inbox.ru), <https://orcid.org/0009-0007-6834-4968>.

**Abstract.** Today, it is possible to use machine learning to develop automated methods that make it simple to gather negative comments on social networks. The situation is getting worse nowadays due to the overwhelmingly unfavorable comments on online content. We are compiling a database of vulgar remarks made

on Kazakh social networks and news outlets. By examining the complaints received from numerous social networks, we discovered that the publication of disparaging, or insulting, comments in online content is growing daily. The outcomes of our research, which used machine learning techniques, will aid in not only analyzing the origins of abusive phrases posted on social media but also in categorizing different forms of offensive remarks and gaining access to automated data sources. Measurement of the accuracy of vocabulary addition by using different indicators. We are creating a dataset of Internet users in Kazakhstan, who use social networks and media to share their opinions. The description of the proposed research work describes the collection of data to the classification and identification of sets of words with destructive content.

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

© **А.Б. Тоқтарова<sup>1\*</sup>, Б.С. Омаров<sup>2</sup>, Г.Н. Казбекова<sup>1</sup>, С.А. Мамиков<sup>3</sup>,  
Ф.Е. Темірбекова<sup>4</sup>, 2023**

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

<sup>2</sup>Әл –Фараби атындағы Қазақ Ұлттық университеті Алматы, Қазақстан;

<sup>3</sup>Академик Ә. Қуатбеков атындағы Халықтар достығы университеті,  
Шымкент, Қазақстан;

<sup>4</sup>М. Әуезов атындағы Оңтүстік Қазақстан университеті,  
Шымкент, Қазақстан.

E-mail: toktar.aigerim@list.ru

## **ӘЛЕУМЕТТІК ЖЕЛДЕГІ ҚАЗАҚ ТІЛДІ БЕЙӘДЕП СӨЗДЕР ҚОРЫН МАШИНАЛЫҚ ОҚЫТУДА ЖИНАҚТАУ**

**Тоқтарова Айгерим Бастарбекқызы** — Докторант, кафедра Компьютерная инженерия, Қожа Ахмет Ясауи атындағы Халықаралық Қазақ – Түрік Университеті. 161200. Түркістан, Қазақстан.

E-mail: toktar.aigerim@list.ru, <https://orcid.org/0000-0002-6265-9236>;

**Омаров Батырхан Султанович** — Phd, кафедра Ақпараттық жүйелер, факультет Ақпараттық технологиялар, Әл –Фараби атындағы Қазақ Ұлттық университеті. 050040. Алматы, Қазақстан Алматы, Казахстан.

E-mail: batyahan@gmail.com, <https://orcid.org/0000-0002-8341-7113>;

**Казбекова Гулнур Нагиметовна** — т.ғ.к., доцент, “Компьютерлік инженерия” кафедрасы жетекшісі, факультет Инженерия, Қожа Ахмет Ясауи атындағы Халықаралық Қазақ – Түрік Университеті. 161200. Түркістан, Қазақстан.

E-mail: kazbekova.gulnur@auy.edu.kz, <https://orcid.org/0000-0002-2756-7926>;

**Мамиков Сатмурза Абаевич** — аға оқытушы, кафедра Информатика және Математика, факультет Жаратылыстану ғылымдары және инженерия, Академик Ә.Қуатбеков атындағы Халықтар достығы университеті. 160000. Шымкент, Қазақстан.

E-mail: Satmyrza85@mail.ru, <https://orcid.org/0009-0000-85054-2157>;

**Темірбекова Фариза Ерланқызы** — оқытушы, кафедра Ақпараттық коммуникациялық технологиялар, факультет Ақпараттық коммуникациялық технологиялар, М. Әуезов атындағы Оңтүстік Қазақстан университеті. 160012. Шымкент, Қазақстан.

E-mail: temirbekova.fariza@inbox.ru, <https://orcid.org/0009-0007-6834-4968>.

**Аннотация.** Бүгінде әлеуметтік желілерде жағымсыз пікірлерді жинауды жеңілдететін автоматтандырылған әдістерді әзірлеу үшін машиналық оқытуды қолдануға болады. Қазіргі уақытта желідегі мазмұндағы жағымсыз пікірлердің салдарынан әлеуметтік желі қолданушылардың психо-эмоционалдық жағдайы қиындап бара жатқанын психологтар айтып дабыл қағуда. Қазақ тілді әлеуметтік желілер мен ақпарат құралдарында жазылған дәрекі сөздердің базасын жинап, оны класстар бойынша топтап жатырмыз. Көптеген әлеуметтік желілерден келіп түскен арыз-шағымдарды зерделей келе, біз желідегі контентте кемситетін немесе қорлайтын пікірлердің жариялануы күн сайын артып отырғанын анықтадық. Машиналық оқыту әдістерін пайдаланған зерттеуіміздің нәтижелері әлеуметтік желілерде жарияланған балағат сөздердің шығу тегін талдауға ғана емес, сондай-ақ қорлайтын пікірлердің әртүрлі түрлерін санаттауға және автоматтандырылған деректер көздеріне қол жеткізуге көмектеседі. Әртүрлі көрсеткіштер арқылы сөздік толықтыру дәлдігін өлшеу. Біз әлеуметтік желілермен БАҚ арқылы өз пікірлерімен бөлісетін Қазақстандағы интернет пайдаланушылардан келіп түскен пікірлерден деректер жинағын құрадамыз. Ұсынылған зерттеу жұмысының сипаттамасы деструктивті мазмұндағы сөздердің топтамасын жіктеуге, анықтауға және деректер жинауды сипаттайды.

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

© А.Б. Токтарова<sup>1\*</sup>, Б.С. Омаров<sup>2</sup>, Г.Н. Казбекова<sup>1</sup>, С.А. Мамиков<sup>3</sup>,  
Ф.Е. Темирбекова<sup>4</sup>, 2023

<sup>1</sup>Международный Казахско–Турецкий университет им. Ходжа Ахмет Ясауи,  
Туркестан, Қазақстан;

<sup>2</sup>Казахский национальный университет им. аль-Фараби, Алматы, Қазақстан;

<sup>3</sup>Университет дружбы народов имени академика А.Куатбекова,  
Шымкент, Қазақстан;

<sup>4</sup>Южно–Казахстанский университет имени М. Ауезова, Шымкент, Қазақстан.  
E-mail:toktar.aigerim@list.ru

## СБОР БАЗЫ ДАННЫХ О ЯЗЫКЕ НЕНАВИСТИ В СОЦИАЛЬНОЙ СЕТИ НА КАЗАХСКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

**Токтарова Айгерім Бастарбекқызы** — Докторант, кафедра Компьютерная инженерия, факультет Инженерия, Международный Казахско–Турецкий университет им. Ходжа Ахмет Ясауи. 161200. Туркестан, Қазақстан.

E-mail: toktar.aigerim@list.ru, <https://orcid.org/0000-0002-6265-9236>;

**Омаров Батырхан Султанович** — Phd, кафедра Информационных систем, факультет Информационных технологии, Казахский Национальный университет им. аль-Фараби. 050040. Алматы, Қазақстан.

E-mail: batyahan@gmail.com, <https://orcid.org/0000-0002-8341-7113>;

**Казбекова Гулнур Нагиметовна** — К.т.н., доцент, зав кафедрой “Компьютерная инженерия”, факультет Инженерия, Международный Казахско–Турецкий университет им. Ходжа Ахмет Ясауи. 161200. Туркестан, Казахстан.

E-mail: kazbekova.gulnur@auy.edu.kz, <https://orcid.org/0000-0002-2756-7926>;

**Мамиков Сатмурза Абаевич** — старший преподаватель, кафедра Информатика и Математика, Факультет Естественных наук и инженерии, Университет дружбы народов имени академика А. Куатбекова. 160000. Шымкент, Казахстан.

E-mail: Satmyrza85@mail.ru, <https://orcid.org/0009-0000-85054-2157>;

**Темирбекова Фариза Ерланкызы** — преподаватель, кафедра Информационная коммуникационная технология, факультет Информационная коммуникационная технология, Южно – Казахстанский университет имени М. Ауезова. 160012. Шымкент, Казахстан.

E-mail: temirbekova.fariza@inbox.ru, <https://orcid.org/0009-0007-6834-4968>.

**Аннотация.** Сегодня можно использовать машинное обучение для разработки автоматизированных методов, упрощающих сбор негативных комментариев в социальных сетях. В настоящее время ситуация ухудшается из-за крайне неблагоприятных комментариев к онлайн-контенту. Мы формируем базу пошлых высказываний в казахстанских социальных сетях и новостных агентствах. Изучив жалобы, полученные из многочисленных социальных сетей, мы обнаружили, что количество публикаций пренебрежительных или оскорбительных комментариев в онлайн-контенте растет с каждым днем. Результаты нашего исследования, в котором использовались методы машинного обучения, помогут не только проанализировать происхождение оскорбительных фраз, размещенных в социальных сетях, но и классифицировать различные формы оскорбительных замечаний и получить доступ к автоматизированным источникам данных. Измерение точности пополнения словарного запаса с помощью различных показателей. Мы создаем набор пользователей Интернета в Казахстане, которые используют социальные сети и СМИ, чтобы поделиться своим мнением. Описание предлагаемой исследовательской работы описывает сбор данных для классификации и выявления наборов слов деструктивного содержания.

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

## **Introduction**

The volume of information in different text formats is growing exponentially and out of control as a result of Internet development. You can interact with a variety of online audiences and operate anonymously with unlimited Internet access, which entails 24-hour access. Violence and offensive words are still significant today because they put people’s mental and emotional well-being in danger.

Along with its positive aspects, technological advancement necessitates consideration of its drawbacks and effects, which means that it encourages more destructive behavior users. In this instance, destructive data is hurting the online content user by posting offensive comments (Huang et al., 2019).



The vast amount of information on today's networks necessitates the protection of people's interests and rights. Even on a worldwide scale, we can observe the effects of it, as evidenced, for instance, by international disagreements and conflicts involving online information (Salloum et al., 2020).

For instance, the state of Kazakhstan has 15.47 million people, or 81.9% of the total population, as of 2021, according to statistics on the percentage of countries around the world that utilize the Internet. You cannot protect yourself against «attacks» on the Internet since emails, mobile phones, and numerous personal social networking sites can arrive at any moment. Information on the Internet is accessible twenty-four hours a day, seven days a week. People, even teenagers, mentally impacted by it.

Additionally, it is possible to access the Internet anonymously, thus you might not be able to identify the source of the «attack.» According to psychologists, psychological illnesses are more likely to result from the conclusion of an emotional «assault» than physical violence.

### **Materials and methods. Research analysis and problem statement**

It is suggested that a variant of the formal model of this process' implementation inside the framework of the engineering method, in which a linguistic object (phenomenon) can be replicated using a computer (Mohan et al., 2021).

To examine what showed, felt, feeling mood in texts, we should collect database. Unfortunately, when it comes to non-normative vocabulary in a particular language, it might be challenging to discover texts that are appropriate for this purpose. Our recommendation is for you to develop your own corpus of non-normative Kazakh terminology.

The corpus divided into two schemes: open social network sites without any non-normative vocabulary messages and text entries with non-normative vocabulary. Syntax analysis technology uses to collect data.

The first parsing strategy devised by the API developer states that in order to use this technique, a connection must first be made, which entails obtaining a token and gaining access to the connection. The token c directly obtained in the program's developer part, and used to access the database, may provide any kind of data conversion services. Following the requisite database's «cleaning,» the information saved and made available for machine learning (Mussiraliyeva et al., 2021).

On the second point, the connection made differently here; you can build a database and your own user interface to access web pages. In other words, we gather «emotional» data from social networks and internet content sites. Although this procedure is laborious at first, once you finished gathering data, you might realize that the logic of operation unified (Singh et al., 2020) and (Canhoto et al., 2020).



Figure 1- Data collection

Eight classes of verbal markers identified that clearly indicate the presence of electronic bullying:

The taboo and obscene language (one-component, two-component, three-component and multi-component) - is used as a means of pronounced verbal aggression with the aim of socially discrediting the victim, manipulating, emphasizing one's own superiority.

The word that related to the intimate life of a person — often used by hooligans in order to hurt the feelings of the victim, humiliate and insult her. In many languages, the names of genitals and types of sexual acts are taboo vocabulary and often used to express a high degree of aggression in the addressee.

The words related to concepts related to sexual orientation and sexism — often, in order to achieve the above goals, victims are attributed and / or emphasize their non-traditional sexual orientation, for example, gay, lesbian, pedophile / sexist, trans, etc., and the addressee is not necessarily a representative of a sexual minority;

The words expressing a wish for evil and death - the use of such lexical units is due to the strong personal dislike of the buller and even hatred of the victim. The use of lexical items such as committing suicide, dying in hell, cutting one's wrists, stopping breathing, etc. shows that the buller wants to emphasize his superiority over the victim, to prove his greatness;

The words expressing humiliation and insult to a person — most often used when electronic harassment occurs according to the «several bullies - one victim» scheme. Lexical items brash, dumb, dirty, ignorant, stupid, etc. aimed at undermining the self-esteem and self-respect of the victim through their constant criticism and downplaying of their abilities. Thus, Buller seeks to demonstrate his power over him or her;

The words that name concepts related to nationality and racism, such as black-faced, hack, Uzbek, racist, sociopath, etc. — allow the aggressor to cause the victim not only moral, but also social harm;

Animal names are insulting terms that are used. By equating the attributes of the victim with those of animals, such as cows, donkeys, pigs, mongrel sheep, and

worms, bullies are able to diminish the victim’s place in society and boost their own self-esteem.

The word which call people with physical and mental disabilities - the main purpose of using this vocabulary is to humiliate honor, reduce its social attractiveness, destroy it as a person. Buller often uses expressions to indicate the victim’s limited mental capacity when the victim is smarter than him or her, and Buller wants to prove otherwise (Yuan et al., 2021).

In terms of AI, automation, communication platforms, bots, and smart computers can combine with big data to improve a variety of technologies used in homes and offices, from security data analytics to investment analysis tools.

Artificial intelligence performs in-depth analysis of large amounts of data using neural networks with many hidden layers. Deep learning models require large amounts of data, as this is the basis of artificial intelligence. Therefore, the more data, the more accurate the model (Mühlhoff et al., 2020).

In the process of collecting data from social networks with Kazakh-speaking customers on online content, it was decided to create 2 types of parsing: API-based parsing and creating your own «web browser» using keywords. Each type has its advantages and disadvantages.

An application programming interface (API) is a collection of classes, procedures, and functions in the form of an interface for creating new software applications through which one program can interact with another program. This allows developers to extend the functionality of their products and connect them with software.

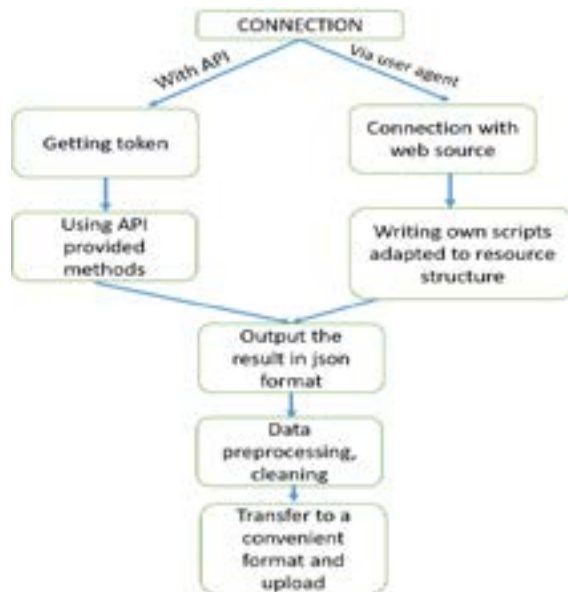


Figure 2 - Data Collection Method Using Two Different Parsers

A parser that utilizes the API readily accessible by the developers themselves makes up the first branch of the schema. You must first establish a connection in order to begin using these methods. In this instance, a token obtained to accomplish this. In the section for app developers, you can directly obtain the token, which you can then use to gain access to the services. The service offers the option to convert the data in any format for ease of usage after obtaining the data. The required data then cleansed and put through the primary processing stage. Following that, this information kept and can be used to train and test models (Neupane et al., 2020).

The connection built slightly differently in the second branch because the parser created entirely from scratch. You must gain access to the browser in order to perform this imitation of web page browsing behavior. Since browsers not give third-party programs and websites access to their APIs, the de-tailed description that follows explains how this parser fragment made. Although this procedure is quite laborious, once we have finished and extracted the required data, we can once more begin primary processing, at which point the logic of work crosses with the logic of using the API.

The initial parser do not cause any issues because it adheres to particular algorithms, and because of an established system, the API operates flawlessly and consistently provides accurate information. An E-token supplied to each user when connecting to the API utilized, for instance, while using the Vk API for Vkontatkte.

Then the Instagram parser underwent the similar treatment. To do this, lists of communities (accounts) with a sizable subscriber base compiled; personal and entertainment accounts disregarded because they were unnecessary. Then, Instagram-specific methods substituted using the same reasoning as for the Vk API. Vkontakte and Instagram are both social networks, however their working theories are very different from one another. Actually, one thing hasn't changed: we still need to obtain a token before we can begin using the given methods.

The app needs to be whitelisted in order to use a specially configured Instagram URL scheme. For this, modify the Info. plist file of the app by adding `instagram:/` to the `LSApplicationQueriesSchemes` key.

After adding the schema to your whitelist, you can use it with the following parameters to carry out specific activities within the Instagram app.

However this is only the first step; since we can not upload the data we need right away, we must first clear the data we received and then choose the pertinent data that contains the text we need from it (in our case, keywords for cyberbullying). After that, you may save this processed data and use it with our models' training and testing procedures (Kumar et al., 2020).

The Instagram API exactly the same across all of them, even though the methods may alter slightly in appearance.

*Getting a token -> getting access to services ->Using methods and getting relevant data -> Clearing them of unnecessary data.*

The drawback of such a parser is that you have to create your own filter to extract the required data after the data received and even after preprocessing.

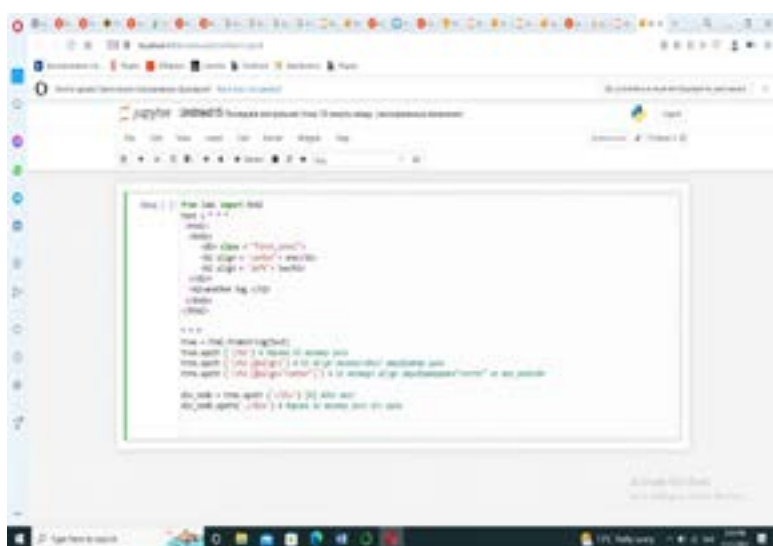
This architecture has the benefit of requiring less time and money to construct a system of this kind uploading data does not require us to develop our own functions.

The second kind of parser entirely created by a person. To put it another way, the scripts were customized for the architecture of a given web resource (Sarker et al., 2020).

The servers will kick us out if they «understand» that this is code, rather than a person, in this situation, when the token's function is to mimic human behavior. The user-agent used to establish the initial connection with the browser. User agent is a character string that enables servers and network nodes to recognize the requesting user agent's application, operating system, vendor, and version gives us the option to connect to the server, eradicating the potential.

```
Listing 1:  
  
import requests  
user_id = 12345  
url = 'http://www.kinopoisk.ru/user/%d/votes/list/ord/date/page/2/#list' %  
(user_id) #  
headers = { 'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9;  
rv:45.0) Gecko/20100101 Firefox/45.0' }  
r = requests.get(url, headers = headers)
```

A script that extracts all of the headers from this page provided by the code block above. The data we require is stored in the html structure, so we will need to put in a little effort next. As a result, once we have identified the blocks in which our records are stored, we can cycle through each article or record and choose a particular block, then extract the relevant data from there. For instance:



The information we require in this block is found in blocks with «h2,» which is found inside the «first level» block.

Once we get the required information, we may instantly filter it to remove any records that are not pertinent to our inquiry. The data then simply written into our files or given to the training algorithms.

```
Listing 2:  
  
f = open("demofile2.txt", "a") # opening the file for editing  
f.write("---") # writing data  
f.close() # close the file
```

After we have the necessary data, we may immediately filter it to get rid of any records that are not relevant to our inquiry. The information was then either submitted to the training algorithms or just written into our files.

### Result and discussion

Data Preprocessing and Filtering which we obtained a dataset of social media texts using the built parsers scarping comments from youtube which showed on Fig.3 . We therefore have unstructured data that is irrelevant. We now faced with the issue. Using uncleaned data prevents you from using it properly. We want pertinent data, together with a sufficient number of training and testing sets. All of this aids in obtaining the most precise findings.

Depending on the objective, you can use one or a combination of several data cleaning or data preparation approaches. It depends on your preferences and the task's requirements.

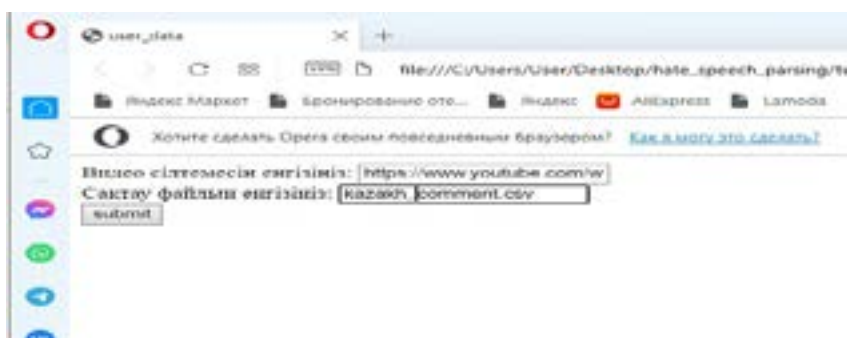


Figure 3- Youtube scarper comments (interface)

### Avoiding duplicates

The probability of parsing the same text high because we are using multiple sources of data. For instance, we have posts from the news website https://

tengrinews.kz, Instagram, and VK. So, occasionally we must delete. In this instance, we made VK a privileged source because all of the data on <https://tengrinews.kz/> semantically reviewed before posting and may change for a variety of reasons.

Although we have not encountered those circumstances yet, we must keep an eye out for their emergence and set up functionality to remove copies.

**Avoid typos**

Typos are grammatical errors made by people most frequently, and they can happen everywhere. Grammar errors corrected using a variety of calculations and techniques. Since models address distinct features in unexpected ways, mistakes are simple to remedy. Spelling and case are very important for strings.

For example, although having the same spelling, «қзақ» and «казак» are distinct words. This kind of mistakes will need to check for and properly fixed.

**Data Filtering**

Not all of the information we gathered used, because we analyze all the news posts' data. This indicates that the parser we created covers all the information, even news about the coronavirus, which has no bearing on our issue. Therefore, it is not a problem. We just created a code that determines whether certain terms, such as «piggy,» «I hate you,» «fatty,» «nigga,» and others, are present in each cell. Additionally, we created a bag of hateful words in Kazakh, which we simply added to the filtering mechanism. Following that, we have a suitable data collection of our own for identifying and preventing possible bullying (Jain et al., 2020).

	A	B	C	E
B704	Малдарды ату жазасына беру керек!дадан емес бунар МАЛДАР БУЛАР		1	
B705	Тауында алып алтуну ерес.		1	
B706	Жалмауыздардын мойундарына релушак тастарын байлап койбу керек устың кунің астына шығарып байлап тастап қоралп өлтіру керек мурна неборіс.		1	
B707	Талаңдай салмайт емес, аяқтарыңа мадақтаран болса, жұмысшыларды қорлағануылардаға нәлімді бермей құялестерін немел алып домогайтеңе, бір ныз ақсамады ала алаңай шылап өлтім дид, төн күніңгі әрэн сұрап алдым дид, нәлімлетпелерды мүлмеңа алып домогайтеңе, аяқтарыңа падофил!!		1	
B708	ешқатарды малан әнеліңдері менде тышартап өлтірейім аяптап өлтірменің		1	
B709	Бұрынғы барыптысы, банды жанды қаңгелерлер қазіргі хренерал, жаулап, қанж, қреністрлар.. Сірлеріңнен тайда болғандары "ақуаларын" әрі қарай жалмастырда "Жаңа Қазақстанда"... Қлы қалық ұста балтасын кіналайт дименші "жазрын" Қазақстаным осы		1	
B710	Тәі қлаңбастың алып айырып өлтір бойы бас бостандығыңнан айырып жіберу керек!д, сарла өсіріп		1	
B711	ИСЛАМГА кір кілтіріп сұралар		1	
B712	Бұл айттыңдар сенасың екен, қанша айтсада құлақтарыңа ермейді.		1	
B713	Болмасада сенің болмаңды шетінен қан ішер анша ұйле		1	
B714	Фуу қалпа тірне наравмай мастаның аттарын білемеймін деп мастаның қолды қой найс. Ділапа сен		1	
B715	не қызым		1	
B716	Ужас осы айла бесіт салындаған қалай нақар журналист көр бет		1	
B717	Баланы тастап көткен адам ешқашан онбайды, онбайтың. Мүретпай ай. Екінші рет күйеуге шығатын кезде ойлансаңа әнелер.		1	
B718	Мына Гузель деген мадам үйіңді тана деп ұалмай қалай әркені беріңді, маңқара. Байла тіле		1	
B719	бермей, бір балаңды асырап, жұмысқа алыңдайсың ба, мармас ат.		1	
	Умангас Селарымбетовна			
	3 месіңа непад			
B720	Мына күдіргі қараң өнеріңің тұрағын айтып тұр Мына Еркен қауым өнеріңің молмаңдар		1	
B721	Фууу.....ты жүрсе не педерь, меніңнен адамдар өнеріңің, «Собоймаған Қо Яғи ны сі		1	
B722	Қалы сорлы өнер өнер айтып деген		1	
B800	Алла сардасын мынаңдай аяқдан ұят өнер өрсі ай фууу маған кілді ай өрмен көткім		1	

Figure 4- hate speech comments in Kazakh language

We have created a database that consists of Kazakh obscene speeches Fig. 4. The database consists of three columns, one of them is the number of sentences which have hate speech comments and attributes that are divided into classes if «1» means there is obscene speech in the sentence if «0» is neutral words.

### **Conclusion**

Summarizing the above, finding negative reviews show as the job of information technology. Two solutions describes in solving the given problem: a person does it with his own hands or turns to an automated system. Machine learning considered as a system that is constantly growing and gaining new momentum in science. Intimidation - the types of intimidation and the signs of their detection presented in this research paper. In addition, it shows that an automated system needed in the search of destructive databases. We present a set of methods used for this purpose. You can see the search paths and their differences from each other.

We have noticed that the collected database used to identify online content using natural language processing and machine learning methods, and to determine the type of a string of offensive words. In addition, it observed that it is possible to create a model for identifying offensive words in online content, and it discussed that it takes as the main issue of our future scientific research.

### **REFERENCES**

- Canhoto A.I., & Clear F. (2020). — *Canhoto A.I., & Clear F.* Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, 63(2), 183-193  
<https://vk.com/dev/methods> - VK API methods  
Instagram.com - Instagram Social Network
- Jain A. & Mandowara J. (2016). — *Jain A. & Mandowara J.* Text classification by combining text classifiers to improve the efficiency of classification. *International Journal of Computer Application* (2250-1797), 6(2), 2250-1797.
- Kumar A., Singh J.P., Dwivedi Y.K. & Rana N.P. (2020). — *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*, 1-32.
- Mohan A., Singh A.K., Kumar B. & Dwivedi R. (2021). — *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*, 32(7), e3998.
- Mühlhoff R. (2020). — *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*, 22(10), 1868-1884.
- Mussiraliyeva S., Omarov B., Yoo P. & Bolatbek M. (2021). — *Mussiraliyeva S., Omarov B., Yoo P. & Bolatbek M.* Applying machine learning techniques for religious extremism detection on online user contents. *Computers, Materials and Continua*, 70(1), 915-934.
- Neupane D. & Seok J. (2020). — *Neupane D. & Seok J.* Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8, 93155-93178.
- Salloum S.A., Alshurideh M., Elnagar A. & Shaalan K. (2020, April). — *Salloum S.A., Alshurideh M., Elnagar A. & Shaalan K.* 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.
- Sarker I.H., Kayes A.S.M., Badsha S., Alqahtani H., Watters P. & Ng A. (2020). — *Sarker I.H.,*



Kayes A.S.M., Badsha S., Alqahtani H., Watters P. & Ng A. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7(1), 1-29.

Singh J.P., Kumar A., Rana N.P. & Dwivedi Y.K. (2020). — 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.

VK.com - V Kontakte Social Network

Yu M., Huang Q., Qin H., Scheele C. & Yang C. (2019). — Yu M., Huang Q., Qin H., Scheele C. & Yang C. 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.

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

NEWS OF THE NATIONAL ACADEMY OF SCIENCES  
OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X

Volume 1, Number 345 (2023), 204-227  
<https://doi.org/10.32014/2022.2518-1726.178>

УДК 004.931

© **A. Shekerbek** , **G. Abdikerimova** \*, **Zh. Lamasheva** , **M. Baibulova** ,  
**A. Tokkuliyeva** , 2023

Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.  
E-mail: [gulzira1981@mail.ru](mailto:gulzira1981@mail.ru)

## CLASSIFICATION OF X-RAY IMAGES USING THE DEEP LEARNING ALGORITHM

**Shekerbek Ainur** — Doctoral student of the Department of Information Systems of the Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.

E-mail: [shekerbek80@mail.ru](mailto:shekerbek80@mail.ru), <https://orcid.org/0000-0002-1088-4239>;

**Abdikerimova Gulzira** — PhD, Associate Professor, Department of Information Systems, Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.

E-mail: [gulzira1981@mail.ru](mailto:gulzira1981@mail.ru), <http://orcid.org/0000-0002-4953-0737>;

**Lamasheva Zhanar** — PhD, Senior Lecturer at the Department of Information Systems of the Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.

E-mail: [zhanarlb@mail.ru](mailto:zhanarlb@mail.ru), <https://orcid.org/0000-0002-9535-2636>;

**Baibulova Makbal** — Senior Lecturer, Department of Information Systems, Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.

E-mail: [m.gabbasovnaa@gmail.com](mailto:m.gabbasovnaa@gmail.com);

**Tokkulieva Aizhan** — Senior Lecturer, Department of Information Security, Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.

E-mail: [aizhan\\_tk@mail.ru](mailto:aizhan_tk@mail.ru).

**Abstract.** The relevance of the topic in me in particular, if we take one of the information flows, whether it is a human factor action or a specific object, it is true that special machine learning language processing and automatic output of information significantly optimize human life. With the help of neural networks and their chest radiography is one of the most accessible radiological studies for screening and diagnosis of many lung diseases a special deep learning language is to study the flow of information about him and the same object in real time using neural networks. The article describes the terminology of the problem of X-ray recognition using deep learning algorithms, examines the relevance of the problem, and analyzes the current state of the problem in the field of X-ray recognition. The aspects of the problem being solved, identified during the analysis, in the form of solved problems, approaches, methods, information technologies used,

## МАЗМҰНЫ

<b>Ж.К. Абдугулова, Г.А. Ускенбаева, М.Н. Тлеген, А.К. Шукирова</b> ҚҰБЫР ЖАБДЫҒЫНДА МАЙДЫ ҚЫЗДЫРУДЫҢ ТЕХНОЛОГИЯЛЫҚ ПРОЦЕСІН АВТОМАТТАНДЫРУ.....	5
<b>Ж.С. Авкурова, С. Гнатюк, Л.М. Кыдыралина, Н.К. Курмангалиева</b> АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ЖҮЙЕЛЕРДЕ ҚҰҚЫҚ БҰЗУШЫНЫ ЕРТЕ АНЫҚТАУ ЖӘНЕ СӘЙКЕСТЕНДІРУДІҢ ИНТЕЛЛЕКТУАЛДЫ ӘДІСІ.....	22
<b>А. Бекарыстанкызы, Ө. Ж. Мамырбаев</b> АГГЛЮТИНАТИВТІ ТІЛДЕРГЕ АРНАЛҒАН СӨЙЛЕУДІ АВТОМАТТЫ ТҮРДЕ ТАҢУ ЖҮЙЕСІ.....	37
<b>А.С. Еримбетова, Э.Н. Дайырбаева, Л. Черикбаева</b> БИКУБТЫҚ ИНТЕРПОЛЯЦИЯҒА НЕГІЗІНДЕ СУРЕТТЕРГЕ ЖАСЫРЫН АҚПАРАТТЫ ЕНГІЗУ.....	50
<b>М.Б. Есенова, Г.Б. Абдикеримова, А. Толстой, Ж.Б. Ламашева, А.А. Некесова</b> БИДАЙДАҒЫ АРАМШӨПТЕР ОШАҒЫН АНЫҚТАУ ҮШІН ТЕКСТУРАЛЫҚ БЕЛГІЛЕР ӘДІСТЕРІН ҚОЛДАНУ.....	64
<b>Л.З. Жолшиева, Т.К. Жукабаева, Ш. Тураев, М.А. Бердиева, Р.К. Сенгирбаева</b> НАҚТЫ УАҚЫТ РЕЖИМІНДЕ МЕДИАРИПЕ ЖӘНЕ SVM АРҚЫЛЫ ҚАЗАҚ ҰМ ТІЛІН ТАҢУ.....	82
<b>Ж.С. Иксебаева, К. Жетписов, А.Б. Медешова, И.М. Бапиев, Ж.Ж. Багисов</b> ҒАЛЫМДАРДЫҢ ҒЫЛЫМИ ЖОБАЛАР БОЙЫНША ГРАНТТЫҚ ҚАРЖЫЛАНДЫРУҒА ҚАТЫСУҒА ӨТІНІМДЕРІН ДАЙЫНДАУДЫҢ АҚПАРАТТЫҚ ЖҮЙЕСІ.....	94
<b>А.А. Иманберді, Р.Н. Молдашева</b> ӘЛЕУМЕТТІК МЕДИА ТАРАТУ ҮЛГІЛЕРІНЕ ШОЛУ.....	107
<b>Г. Қалман, М.Ғ. Есмағанбет, М.М. Жаманкарин, А.И. Габдулина, Д.В. Плескачев</b> КЛАСТЕРЛЕУ ӘДІСІН ҚОЛДАНЫП КОРЕФЕРЕНЦИЯН ШЕШУ.....	121

<b>Қ.Т. Қырғызбай, Е.Х. Какимжанов</b> ГАЗ ТЕХНОЛОГИЯЛАРЫ НЕГІЗІНДЕ АЛМАТЫ ОБЛЫСЫНЫҢ ГЕОДЕРЕКТЕР БАЗАСЫН ҚҰРУ ВІТСОІН ЖЕЛІСІНДЕГІ КҮДІКТІ ТРАНЗАКЦИЯЛАРДЫ АНЫҚТАУ.....	136
<b>Ш.Ж. Мусиралиева, М.Ж. Шайзат, А.К. Бекетова, Е. Абайұлы, А.Б. Манасова</b> ВІТСОІН ЖЕЛІСІНДЕГІ КҮДІКТІ ТРАНЗАКЦИЯЛАРДЫ АНЫҚТАУ.....	154
<b>А.Ұ. Мұхиядин, Ұ.Т. Махажанова, М.У. Мукашева, А.А. Муханова</b> АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР ТӨТЕНШЕ ЖАҒДАЙДА ҚАШЫҚТАН ОҚЫТУДА ЭКСПЕРИМЕНТТЫҚ ДЕРЕКТЕРДІ ТАЛДАУ ҚҰРАЛЫ РЕТІНДЕ.....	170
<b>А.Б. Тоқтарова, Б.С. Омаров, Г.Н. Казбекова, С.А. Мамиков, Ф.Е. Темірбекова</b> ӘЛЕУМЕТТІК ЖЕЛІДЕГІ ҚАЗАҚ ТІЛДІ БЕЙӘДЕП СӨЗДЕР ҚОРЫН МАШИНАЛЫҚ ОҚЫТУДА ЖИНАҚТАУ.....	191
<b>А.Ә. Шекербек, Г.Б. Абдикеримова, Ж.Б. Ламашева, М.Г. Байбулова, А.К. Токкулиева</b> ТЕРЕҢ ОҚЫТУ АЛГОРИТМІМЕН РЕНТГЕНДІК КЕСКІННІҢ КЛАССИФИКАЦИЯСЫ.....	204
<b>Э.Э. Эльдарова</b> JPEG2000 ҚЫСУЫНАН KEЙІН ЦИФРЛІК БЕЙНЕЛЕРДІҢ ВИЗУАЛДЫ САПАСЫН ЖАҚСАРТУ.....	228

## СОДЕРЖАНИЕ

<b>Ж.К. Абдугулова, Г.А. Ускенбаева, М.Н. Глеген, А.К. Шукирова</b> АВТОМАТИЗАЦИЯ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА ПОДОГРЕВА НЕФТИ НА ТРУБОПРОВОДНОМ ОБОРУДОВАНИИ.....	5
<b>Ж.С. Авкурова, С.А. Гнатюк, Л.М. Кыдыралина, Н.К. Курмангалиева</b> ИНТЕЛЛЕКТУАЛИЗИРОВАННЫЙ МЕТОД РАННЕГО ВЫЯВЛЕНИЯ ИДЕНТИФИКАЦИИ НАРУШИТЕЛЯ В ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ СИСТЕМАХ.....	22
<b>А. Бекарыстанқызы, О. Ж. Мамырбаев</b> ИНТЕГРАЛЬНАЯ СИСТЕМА АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ СЛИТНОЙ РЕЧИ ДЛЯ АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ.....	37
<b>А.С. Еримбетова, Э.Н. Дайырбаева, Л. Черикбаева</b> ВНЕДРЕНИЕ СКРЫТОЙ ИНФОРМАЦИИ В ИЗОБРАЖЕНИИ НА ОСНОВЕ БИКУБИЧЕСКОЙ ИНТЕРПОЛЯЦИИ.....	50
<b>М.Б. Есенова, Г.Б. Абдикеримова, А. Толстой, Ж.Б. Ламашева, А.А. Некесова</b> ПРИМЕНИМОСТЬ МЕТОДОВ АНАЛИЗА ТЕКСТУРНЫХ ИЗОБРАЖЕНИЙ ДЛЯ ВЫЯВЛЕНИЯ ОЧАГОВ СОРНЫХ ТРАВ ПШЕНИЦЫ.....	64
<b>Л.З. Жолшиева, Т.К. Жукабаева, Ш. Тураев, М.А. Бердиева, Р.К. Сенгирбаева</b> РАСПОЗНАВАНИЕ КАЗАХСКОГО ЖЕСТОВОГО ЯЗЫКА В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ MEDIAPIPE и SVM.....	82
<b>Ж.С. Иксебаева, К. Жетписов, А.Б. Медешова, И.М. Бапиев, Ж.Ж. Багисов</b> ИНФОРМАЦИОННАЯ СИСТЕМА ПОДГОТОВКИ ЗАЯВОК ДЛЯ УЧАСТИЯ В ГРАНТОВОМ ФИНАНСИРОВАНИИ УЧЕНЫХ ПО НАУЧНЫМ ПРОЕКТАМ.....	94
<b>А.А. Иманберді, Р.Н. Молдашева</b> ОБЗОР МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ.....	107

<b>Г. Қалман, М.Ғ. Есмағанбет, М.М. Жаманқарин, А.Г. Габдулина, Д.В. Плескачев</b> РЕШЕНИЕ КОРЕФЕРЕНЦИИ С ПОМОЩЬЮ МЕТОДА КЛАСТЕРИЗАЦИИ.....	121
<b>Қ.Т. Қырғызбай, Е.Х. Какимжанов</b> СОЗДАНИЕ БАЗЫ ГЕОДАНЫХ АЛМАТИНСКОЙ ОБЛАСТИ НА ОСНОВЕ ГИС-ТЕХНОЛОГИЙ О МЕТОДЕ ИДЕНТИФИКАЦИИ ПОДОЗРИТЕЛЬНЫХ ТРАНЗАКЦИЙ В БИТКОИН СЕТИ.....	136
<b>Ш.Ж. Мусиралиева, М.Ж. Шайзат, А.К. Бекетова, Е. Абайұл, А.Б. Манасова</b> О МЕТОДЕ ИДЕНТИФИКАЦИИ ПОДОЗРИТЕЛЬНЫХ ТРАНЗАКЦИЙ В БИТКОИН СЕТИ.....	154
<b>А.Ұ. Мұхиядин, У.Т. Махажанова, М.У. Мукашева, А.А. Муханова</b> ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ КАК СРЕДСТВО АНАЛИЗА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ПРИ ЭКСТРЕННОМ ДИСТАНЦИОННОМ ОБУЧЕНИИ.....	170
<b>А.Б. Токтарова, Б.С. Омаров, Г.Н. Казбекова, С.А. Мамиков, Ф.Е. Темирбекова</b> СБОР БАЗЫ ДАННЫХ О ЯЗЫКЕ НЕНАВИСТИ В СОЦИАЛЬНОЙ СЕТИ НА КАЗАХСКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ.....	191
<b>А.А. Шекербек, Г.Б. Абдикеримова, Ж.Б. Ламашева, М.Г. Байбулова, А.К. Токкулиева</b> КЛАССИФИКАЦИЯ РЕНТГЕНОВСКИХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ АЛГОРИТМА ГЛУБОКОГО ОБУЧЕНИЯ.....	204
<b>Э.Э. Эльдарова</b> УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ ПОСЛЕ СЖАТИЕ JPEG2000.....	228

## CONTENTS

<b>J.K. Abdugulova, G.A. Uskenbayeva, M.N. Tlegen, A.K. Shukirova</b> AUTOMATION OF THE TECHNOLOGICAL PROCESS OF HEATING OIL PIPELINE EQUIPMENT.....	5
<b>Z. Avkurova, S. Gnatyuk, L. Kydyralina, N. Kurmangaliev</b> THE INTELLECTUALIZED METHOD OF EARLY DETECTION AND IDENTIFICATION OF THE VIOLATOR IN INFORMATION AND COMMUNICATION SYSTEMS.....	22
<b>A. Bekarystankyzy, O. Zh. Mamyrbayev</b> INTEGRATED AUTOMATIC SPEECH RECOGNITION SYSTEM FOR AGGLUTINATIVE LANGUAGES.....	37
<b>A. Yerimbetova, E. Daiyrbayeva, L. Cherikbayeva</b> EMBEDDING HIDDEN INFORMATION IN IMAGES BASED ON BICUBIC INTERPOLATION.....	50
<b>M. Yessenova, G. Abdikerimova, A. Tolstoy, Zh. Lamasheva, A. Nekessova</b> APPLICABILITY OF TEXTURE IMAGE ANALYSIS METHODS FOR DETECTION OF WHEAT WEED POCKS.....	64
<b>L. Zholshiyeva, T. Zhukabayeva, Sh. Turaev, M. Berdieva, R. Sengirbayeva</b> REAL-TIME KAZAKH SIGN LANGUAGE RECOGNITION USING MEDIAPIPE AND SVM.....	82
<b>Zh.S. Ixebayeva, K. Jetpisov, A.B. Medeshova, I.M. Bapiyev , Zh.Zh. Bagisov</b> AN INFORMATION SYSTEM FOR THE PREPARATION OF APPLICATIONS FOR PARTICIPATION IN GRANT FUNDING OF SCIENTISTS IN SCIENTIFIC PROJECTS.....	94
<b>A. Imanberdi, R. Moldasheva</b> REVIEW OF MODELS OF DISSEMINATION OF INFORMATION IN SOCIAL NETWORKS.....	107
<b>G. Kalman, M.G. Esmaganbet, M.M. Zhamankarin, A.I. Gabdulina, D.V. Pleskachev</b> COREFERENCE SOLUTION USING THE CLUSTERING METHOD.....	121

<b>K. Kyrgyzbay, E. Kakimzhanov</b> CREATION OF A GEODATABASE OF ALMATY REGION BASED ON GIS TECHNOLOGIES.....	136
<b>Sh. Mussiraliyeva, M. Shaizat, A. Beketova, Y. Abayuly, A. Manassova</b> IDENTIFICATION OF SUSPICIOUS TRANSACTIONS IN THE BITCOIN NETWORK.....	154
<b>A. Mukhiyadin, U. Makhazhanova, M. Mukasheva, A. Mukhanova</b>  INFORMATION TECHNOLOGIES AS A MEANS OF EXPERIMENTAL DATA ANALYSIS IN EMERGENCY DISTANCE LEARNING.....	170
<b>A.B. Toktarova, B.S. Omarov, G.N. Kazbekova, S.A. Mamikov, F.E. Temirbekova</b> COLLECTING HATE SPEECH DATABASE ON SOCIAL NETWORK IN KAZAKH LANGUAGE BY USING MACHINE LEARNING.....	191
<b>A. Shekerbek, G. Abdikerimova, Zh. Lamasheva, M. Baibulova, A. Tokkulyeva</b> CLASSIFICATION OF X-RAY IMAGES USING THE DEEP LEARNING ALGORITHM.....	204
<b>E.E. Eldarova</b> IMPROVING THE VISUAL QUALITY OF DIGITAL IMAGES AFTER JPEG2000 COMPRESSION.....	228



**Publication Ethics and Publication Malpractice  
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ([http://publicationethics.org/files/u2/New\\_Code.pdf](http://publicationethics.org/files/u2/New_Code.pdf)). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

**[www.nauka-nanrk.kz](http://www.nauka-nanrk.kz)**

**<http://physics-mathematics.kz/index.php/en/archive>**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Заместитель директор отдела издания научных журналов НАН РК *Р. Жалиқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 30.03.2023.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

15,5 п.л. Тираж 300. Заказ 1.