

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



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ
АКАДЕМИИ НАУК РЕСПУБЛИКИ
КАЗАХСТАН»

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN

PHYSICO-MATHEMATICAL SERIES

4 (352)

OCTOBER – DECEMBER 2024

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 ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н=5**

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

МАМЫРБАЕВ Өркен Жұмажанұлы, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

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

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

БОШКАЕВ Қуантай Авғазыұлы, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

QUEVEDO Nemando, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

ЖҮСІПОВ Марат Абжанұлы, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

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

РАМАЗАНОВ Тілекқабұл Сәбитұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

ТАКИБАЕВ Нұрғали Жабағаұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

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

ХАРИН Станислав Николаевич, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

ДАВЛЕТОВ Асқар Ербуланович, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

КАЛАНДРА Пьетро, Ph.D (физика), Нанокұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

«ҚР ҰҒА Хабарлары. Физика және 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 және ақпараттық коммуникациялық технологиялар сериясы*. Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

ГЛАВНЫЙ РЕДАКТОР:

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

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

МАМЫРБАЕВ Оркен Жумажанович, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **H=5**

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

КАЛИМОЛДАЕВ Максат Нурадилович, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **H=7**

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

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

БОШКАЕВ Куантай Авгазыевич, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=10**

QUEVEDO Hemando, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **H=28**

ЖУСУПОВ Марат Абжанович, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=7**

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

РАМАЗАНОВ Тлексабул Сабитович, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=26**

ТАКИБАЕВ Нурғали Жабағевич, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=5**

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

ХАРИН Станислав Николаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **H=10**

ДАВЛЕТОВ Аскар Ербуланович, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=12**

КАЛАНДРА Пьетро, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **H=26**

«Известия НАН РК. Серия физика и 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 и информационные коммуникационные технологии.* 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. 219, тел.: 272-13-19*

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

EDITOR IN CHIEF:

MUTANOV Galimkair Mutanovich, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

DEPUTY EDITOR-IN-CHIEF

MAMYRBAYEV Orken Zhumazhanovich, Ph.D. in the specialty "Information systems, executive secretary of the RSE "Institute of Information and Computational Technologies", Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

BAYGUNCHEKOV Zhumadil Zhanabayevich, 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 Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

BOSHKAYEV Kuantai Avgazievich, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

QUEVEDO Hemando, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

ZHUSSUPOV Marat Abzhanovich, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

KOVALEV Alexander Mikhailovich, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

RAMAZANOV Tlekkabul Sabitovich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

TAKIBAYEV Nurgali Zhabagaevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

TIGHINEANU Ion Mikhailovich, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

KHARIN Stanislav Nikolayevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

DAVLETOV Askar Erbulanovich, Doctor in Physics and Mathematics, Professor, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

CALANDRA Pietro, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

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

Series of physics and 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-ЖК**, issued 14.02.2018
Thematic scope: *series physics and information technology.*

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. 219, Almaty, 050010, tel. 272-13-19*

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

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 4. Number 352 (2024). 5–16

<https://doi.org/10.32014/2024.2518-1726.303>

УДК 004.931

©**M. Aitimov**¹, **R. U Almenayeva**^{1*}, **K.K. Makulov**², **A. B. Ostayeva**¹,
R. Muratkhan³, 2024.

¹Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan;

²Caspian University of Technologies and Engineering named after. Sh. Yessenov,
Aktau, Kazakhstan;

³Karaganda Buketov University, Karaganda, Kazakhstan.

E-mail: a_raihan@mail.ru

APPLICATION OF MACHINE LEARNING METHOD TO ANALYZE AND EXTRACT SEMANTIC STRUCTURES FROM SCIENTIFIC TEXTS

Aitimov Murat – PhD, senior lecturer, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan,
E-mail: aitimovmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>;

Almenayeva Raikhan - PhD, senior lecturer, Korkyt Ata Kyzylorda University, Kyzylorda,
Kazakhstan, E-mail: a_raihan@mail.ru, <https://orcid.org/0000-0001-7468-8088>;

Makulov Kaiyrbek - PhD, Associate Professor of the Department of Computer Science of the
Caspian University of Technologies and Engineering named after. Sh. Yessenov, Aktau, Kazakhstan,
E-mail: kaiyrbek.makulov@yu.edu.kz, <https://orcid.org/0000-0002-0826-0371>;

Ostayeva Aiymkhan - candidate of Pedagogical Sciences, senior lecturer, Korkyt Ata Kyzylorda
University, E-mail: aimak73@mail.ru, <https://orcid.org/0000-0003-3361-2022>;

Muratkhan Raikhan – Karaganda Buketov University, Associate Professor, PhD, Karaganda,
Kazakhstan, e-mail: raikhan.muratkhan@mail.ksu.kz, <https://orcid.org/0000-0002-2030-8948>.

Abstract. This paper discusses a method for extracting text summaries from scientific documents using the DistilBART model, which is an improved and reduced version of the Bidirectional and Auto-Regressive Transformers (BART) model. The DistilBART model, trained on large volumes of text data, allows one to effectively solve natural language processing (NLP) tasks such as text summarization, machine translation, and text generation. This paper focuses on the application of DistilBART to analyzing and extracting text summaries from scientific documents. The goal of this work is to develop a universal tool based on the DistilBART model that will be effective in extracting and structuring information from scientific documents in various fields. Traditional text processing methods are often not powerful enough and require significant computational resources, which makes them inapplicable to analyzing large volumes of data. The use of advanced machine learning models such as DistilBART is a significant step forward. The relevance of this work is also due to the growing need for effective research support systems. Extracting text summaries using DistilBART can significantly improve

the quality of analytical reviews, simplify the search for relevant literature, and facilitate a deeper understanding of research questions. Ultimately, this helps accelerate scientific progress and improve the efficiency of work in various fields of science and technology.

Key words: Automated document analysis, Machine learning, DistilBART, Natural language processing, Text summarization

Conflict of interest: The authors declare that there is no conflict of interest.

**©М. Айтимов¹, Р.У Альменаева^{1*}, К.К. Мақұлов², А.Б. Остаева¹,
Р. Муратхан³, 2024.**

¹Қорқыт Ата атындағы Қызылорда университеті, Қызылорда, Қазақстан;

²Ш. Есенов атындағы Каспий технологиялар және инжиниринг университеті, Ақтау, Қазақстан;

³Е.А. Бөкетов атындағы Қарағанды университеті, Қарағанды, Қазақстан.
E-mail: a_raihan@mail.ru

ҒЫЛЫМИ МӘТІНДЕРДЕН СЕМАНТИКАЛЫҚ ҚҰРЫЛЫМДАРДЫ ТАЛДАУ ЖӘНЕ АЛУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСІН ҚОЛДАНУ

Айтимов Мурат – PhD, аға оқытушы, Қорқыт Ата атындағы Қызылорда университеті Қызылорда, Қазақстан, E-mail: aitimovmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>;

Альменаева Райхан Умирзаковна – PhD, аға оқытушы, Қорқыт Ата атындағы Қызылорда университеті Қызылорда, Қазақстан, E-mail: a_raihan@mail.ru, <https://orcid.org/0000-0001-7468-8088>;

Мақұлов Кайырбек Калданбекович – э.ғ.к., Ш. Есенов атындағы Каспий технологиялар және инжиниринг университетінің Компьютерлік ғылымдар кафедрасының қауымдастырылған профессор м.а., Ақтау, Қазақстан, E-mail: kaiyrbek.makulov@yu.edu.kz, <https://orcid.org/0000-0002-0826-0371>;

Остаева Айымхан Батырхановна – педагогика ғылымдарының кандидаты, аға оқытушы, Қорқыт Ата атындағы Қызылорда университеті, E-mail: aimak73@mail.ru, <https://orcid.org/0000-0003-3361-2022>;

Муратхан Райхан – Е.А.Бөкетов атындағы Қарағанды университеті, қауымдастырылған профессор, PhD, Қарағанды қаласы, Қазақстан, e-mail: raikhan.muratkhan@mail.ksu.kz, <https://orcid.org/0000-0002-2030-8948>.

Аннотация. Бұл мақалада екі бағытты және авторегрессивті трансформаторлар (BART) үлгісінің жақсартылған және қысқартылған нұсқасы болып табылатын DistilBART үлгісін пайдаланып, ғылыми құжаттардан қысқа мәтіндік сипаттамаларды алу әдісі талқыланады. Мәтіндік деректердің үлкен көлеміне үйретілген DistilBART моделі мәтінді қорытындылау, машиналық аударма және мәтінді құру сияқты табиғи тілді өңдеу (NLP) тапсырмаларын тиімді шеше алады. Бұл жұмыс ғылыми құжаттардан мәтіндік қорытындыларды талдау және шығару үшін DistilBART қолданбасын қолдануға бағытталған. Бұл жұмыстың мақсаты әртүрлі

салалардағы ғылыми құжаттардан ақпаратты алу және құрылымдау үшін тиімді болатын DistilBART моделіне негізделген әмбебап құралды әзірлеу болып табылады. Дәстүрлі мәтінді өңдеу әдістері жиі қуатсыз және есептеуді қажет етеді, сондықтан оларды деректердің үлкен көлемін талдау үшін жарамсыз етеді. DistilBART сияқты машиналық оқытудың озық үлгілерін пайдалану алға жасалған маңызды қадам болып табылады. Бұл жұмыстың өзектілігі ғылыми зерттеулерді қолдаудың тиімді жүйелеріне деген қажеттіліктің артуына да байланысты. DistilBART көмегімен мәтіндік қорытындыларды шығару аналитикалық шолулардың сапасын айтарлықтай жақсартады, сәйкес әдебиеттерді іздеуді жеңілдетеді және зерттеу сұрақтарын тереңірек түсінуге ықпал етеді. Нәтижесінде бұл ғылыми прогресті жеделдетуге және ғылым мен техниканың әртүрлі салаларындағы жұмыстың тиімділігін арттыруға көмектеседі.

Түйін сөздер: Мәтіндерді автоматты талдау, Машиналық оқыту, DistilBART, Табиғи тілдерді өңдеу, Мәтіндерді қысқаша сипаттау

©М. Айтимов¹, Р.У Альменаева^{1*}, К.К. Макулов², А.Б. Остаева¹,
Р. Муратхан³, 2024.

¹Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан;

²Каспийский университет технологий и инжиниринга имени Ш. Есенова,
Ақтау, Казахстан;

³Карагандинский университет имени Е.А. Букетова, Караганда, Казахстан.

E-mail: a_raihan@mail.ru

ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА И ИЗВЛЕЧЕНИЯ СЕМАНТИЧЕСКИХ СТРУКТУР ИЗ НАУЧНЫХ ТЕКСТОВ

Айтимов Мурат – PhD, старший преподаватель, Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан, E-mail: aitimvmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>;

Альменаева Райхан Умирзаковна – PhD, старший преподаватель, Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан, E-mail: a_raihan@mail.ru, <https://orcid.org/0000-0001-7468-8088>;

Макулов Кайырбек Калданбекович – к.э.н., и.о. ассоциированного профессора кафедры Компьютерные науки Каспийского Университета технологий и инжиниринга имени Ш. Есенова, Ақтау, Казахстан, E-mail: kaiyrbek.makulov@yu.edu.kz, <https://orcid.org/0000-0002-0826-0371>;

Остаева Айымхан Батырхановна – кандидат педагогических наук, старший преподаватель Кызылординского университета имени Коркыт Ата, Кызылорда, Казахстан, E-mail: aimak73@mail.ru, <https://orcid.org/0000-0003-3361-2022>;

Муратхан Райхан – Карагандинский университет им. Е.А. Букетова, ассоциированный профессор, PhD, Караганда, Казахстан, e-mail: raikhan.muratkhan@mail.ksu.kz, <https://orcid.org/0000-0002-2030-8948>.

Аннотация. В данной работе рассматривается метод извлечения краткого описания текста из научных документов с использованием модели DistilBART, представляющей собой усовершенствованную и сокращенную версию модели Bidirectional and Auto-Regressive Transformers (BART). Модель DistilBART, обученная на больших объемах текстовых данных, позволяет эффективно решать задачи обработки естественного языка (NLP), такие как обобщение текста, машинный перевод и генерация текста. В данной работе основное внимание уделяется применению DistilBART для анализа и извлечения краткого описания текста из научных документов. Цель данной работы заключается в разработке универсального инструмента на основе модели DistilBART, который будет эффективным в извлечении и структурировании информации из научных документов различных областей. Традиционные методы обработки текста часто оказываются недостаточно мощными и требуют значительных вычислительных ресурсов, что делает их неприменимыми для анализа больших объемов данных. Использование передовых моделей машинного обучения, таких как DistilBART, представляет собой значительный шаг вперед. Актуальность данной работы также обусловлена растущей потребностью в эффективных системах поддержки научных исследований. Извлечение краткого описания текста с помощью DistilBART может существенно повысить качество аналитических обзоров, упростить поиск релевантной литературы и способствовать более глубокому пониманию исследовательских вопросов. В итоге, это способствует ускорению научного прогресса и повышению эффективности работы в различных областях науки и техники.

Ключевые слова: автоматический анализ документов, машинное обучение, DistilBART, обработка естественного языка, краткое описание текста

Кіріспе

Жыл сайын ғылыми әдебиеттердің көлемі ұлғайып келеді, бұл ғылыми құжаттардан ақпаратты (Шай, et al, 2023) тиімді алу және құрылымдау міндетін өзекті етеді. Зерттеушілер мен сарапшылар соңғы жетістіктерден хабардар болу және өз зерттеулері үшін сәйкес деректерді табу үшін үлкен көлемдегі мәтінді (Дагделен, et al, 2024) жылдам және дәл талдауы керек. Ғылыми құжаттардан қысқаша мәтіндік сипаттамаларды (Хартманн, et al, 2023) (Тревисо, et al, 2023) алу қазіргі заманғы табиғи тілді өңдеу (NLP) әдістерін қолдануды талап ететін маңызды міндетке айналуда (Йвги, et al, 2023). Бұл мәселені шешудің перспективті тәсілдерінің бірі екі бағытты және авторегрессивті трансформаторлар (BART) сияқты трансформатор архитектурасына негізделген модельдерді пайдалану болып табылады. BART (Қайбасова, et al, 2022), Facebook AI әзірлеген, екі бағытты кодтауыш пен авторегрессивті декодерді біріктіретін гибриді модель (Сантана, et al, 2023). Бұл оны мәтінді құру, қорытындылау және мәтіннің мағынасын түсіну тапсырмаларының қуатты құралына айналдырады. Дегенмен, жоғары

өнімділігіне қарамастан, бастапқы BART моделі (Ландолси, et al, 2023) айтарлықтай есептеу ресурстарын қажет етеді (Вэй, et al, 2023). Сонымен қатар, DistilBART қолдану ғылыми және білім беру мекемелерінде шешім қабылдау процесін жақсартуға көмектесетін әртүрлі ғылыми және білім беру құжаттарынан ақпаратты жинақтау және құрылымдау, талдау есептерін жасауға көмектеседі (Полак, et al, 2024) (Садирмекова, et al, 2023). Осылайша, бұл жұмыстың ғылыми мақсаты әртүрлі салалардағы ғылыми құжаттардан семантикалық ақпаратты алу және құрылымдау үшін тиімді болатын DistilBART моделіне негізделген әмбебап құралды жасау болып табылады.

Тиімділікті арттыру және есептеу шығындарын азайту үшін DistilBART, BART моделінің тазартылған нұсқасы жасалды. DistilBART BART-тың негізгі артықшылықтарын кішірейтілген өлшемде және жоғары жұмыс жылдамдығында сақтайды. Бұған білімді айдау техникасы арқылы қол жеткізіледі, мұнда жинақы модель (студент) үлкенірек модельдің (мұғалім) мінез-құлқын еліктеу үшін оқытылады. Бұл жұмыста ғылыми құжаттардан семантикалық құрылымды алу үшін DistilBART моделін пайдалану талқыланады. BART үлгісі бастапқыда жаңалықтар мақалалары мен олардың қысқаша мазмұнын қамтитын CNN/Daily Mail сияқты қысқа мәтіндік деректер жиынында оқытылды. Модельді ғылыми мәтіндерге бейімдеу үшін біз arXiv сияқты ашық көздерден ғылыми мақалалардың үлкен деректер жинағын жинадық және мәтіндерді алдын ала өңдедік, оның ішінде кіші әріптерді азайту, қажетсіз таңбалар мен бос орындарды жою, таңбалау және тоқтату сөздерді алып тастау. Әрі қарай, алдын ала дайындалған DistilBART моделі біздің ғылыми мақалаларымыздың деректер жиынтығында қосымша оқытылды. Жаттығу кезінде біз үлгінің жақсаруын бақылау және артық орнатуды болдырмау үшін партия өлшемі, үйрену жылдамдығы және дәуірлер саны сияқты гиперпараметрлерді реттедік және жоғалту және ROUGE мәндері сияқты көрсеткіштерді бақыладық. Эксперименттік нәтижелер қайта оқытылған DistilBART моделі бастапқы үлгімен салыстырғанда ғылыми құжаттардан ақпаратты алу және құрылымдау сапасын айтарлықтай жақсартқанын көрсетті. Осылайша, ғылыми құжаттардан семантикалық құрылымды алу үшін DistilBART пайдалану ғылыми мәтіндерді өңдеуді автоматтандыру және тиімділігін арттыру үшін жаңа мүмкіндіктер ашады, бұл ғылыми зерттеулерді жылдамдатуға және деректерді талдау сапасын жақсартуға көмектеседі.

Әдістер мен материалдар.

DistilBART моделін дамыту және одан әрі оқыту үшін arXiv сияқты ашық көздерден ғылыми мақалалардың кең деректер жинағы жиналды. Деректер жинағы геологиялық деректерді, клиникалық есептер мен білім беру бағдарламаларын қоса алғанда, әртүрлі ғылыми салалардағы мақалаларды қамтиды (Ли, 2023) (Лю, 2023). Деректерді жинаудың негізгі қадамдары arXiv API арқылы мақалаларды іздеу және жүктеп алу, содан кейін мәтіндерді

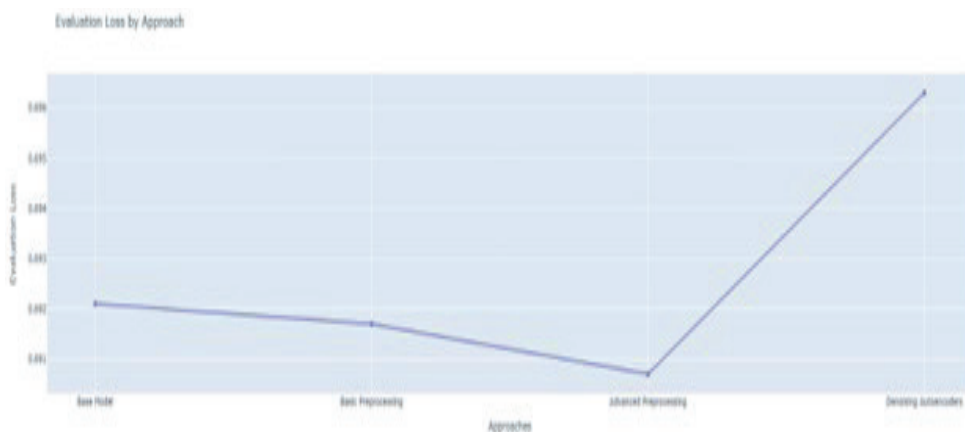
кейінгі өңдеу үшін бір корпусқа біріктіру болды. Мәтінді алдын ала өңдеу деректерді машиналық оқытуға дайындаудағы маңызды қадам болып табылады. Негізгі алдын ала өңдеу қадамдары регистр сезімталдығын жою үшін кіші әріптегі мәтінді, мәтінді тазалау үшін қажет емес танбалар мен бос орындарды жоюды, мәтінді таңбалауыштарға (сөздерге немесе сөйлемдерге) бөлу үшін таңбалауды және жоғары жиілікті, бірақ ақпаратсыз сөздерді алып тастау үшін тоқтату сөздерді жоюды қамтиды. Бұл қадамдар бізге үлгіні әрі қарай оқытуға дайын таза және құрылымды мәтінді алуға мүмкіндік берді.

Енгізу сапасын одан әрі жақсарту үшін деноизация автокодерлері әдісі де қолданылды. Бұл әдіс деректерден шуды жою және оның сапасын қалпына келтіру арқылы модельге мәнді мүмкіндіктерді шығаруға мүмкіндік береді. Содан кейін тазартылған деректер DistilBART үлгісін пайдалана отырып, мәтінді талдау сапасын одан әрі жақсарту үшін мәтінменге сезімтал ендірулерді жасау үшін пайдаланылды. Эксперимент барысында бұл тәсілдерді DistilBART-пен үйлестіре қолдану ғылыми құжаттардағы ақпаратты жинақтау және құрылымдау тапсырмаларында модельдің тиімділігін айтарлықтай арттырып, оны қолданудың әртүрлі салаларына арналған әмбебап құралға айналдыратыны анықталды. DistilBART моделін ғылыми мәтіндерге бейімдеу үшін партия өлшемі, оқу жылдамдығы және дәуірлер саны сияқты белгілі бір гиперпараметрлер бапталды және пайдаланылды. Алдын ала дайындық процесі Hugging Face Transformers кітапханасы арқылы алдын ала дайындалған DistilBART үлгісін жүктеуді, жаттығу параметрлерін оңтайландыру үшін гиперпараметрлерді баптауды, дайындалған ғылыми мақала деректер жинағында алдын ала дайындық процесін іске қосуды және модельдің жақсаруын бақылау және шамадан тыс орнатудың алдын алу үшін жоғалту және ROUGE сияқты көрсеткіштерді бақылауды қамтыды.

Нәтижелер және оларды талқылау

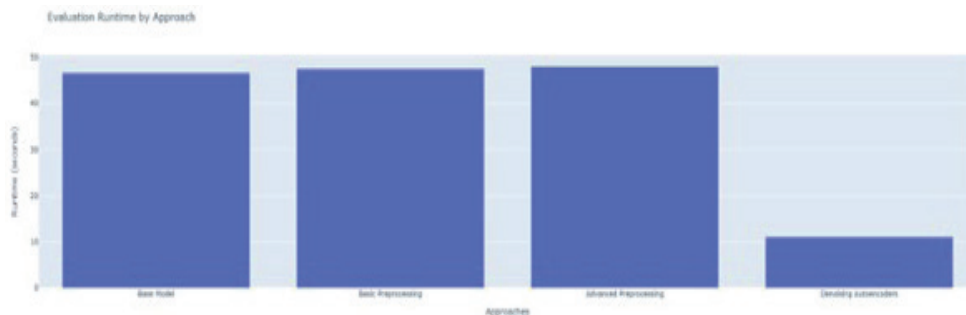
Қосымша оқыту нәтижелері бастапқы үлгімен салыстырғанда ғылыми құжаттардан ақпаратты алу және құрылымдау сапасының айтарлықтай жақсарғанын көрсетті. Модельдің сапасы ROUGE метрикасының көмегімен бағаланды, бұл мәтіндерді қорытындылау және құрылымдау тапсырмаларында үлгі өнімділігінің жақсаруын анықтауға мүмкіндік берді. Нәтижелерді көрнекі түрде көрсету үшін графиктер мен кестелер оқу процесі кезінде метрикадағы өзгерістерді көрсету және қосымша оқытуға дейін және одан кейінгі үлгі өнімділігін салыстыру үшін пайдаланылды. Мысалы, жаттығу жоғалту сюжеті үлгінің біртіндеп жетілдірілуін көрсететін дәуірлер санының өсуімен жоғалту мәндерінің тұрақты төмендеуін көрсетті. DistilBART моделінің негізгі гиперпараметрлері, мысалы, партия өлшемі (16), оқу жылдамдығы (3e-5) және дәуірлер саны (10), оқыту сапасы мен тиімділігі арасындағы оңтайлы теңгерімге қол жеткізу үшін таңдалды. DistilBART моделінің құрылымын визуализациялау және қосымша оқыту кезінде ROUGE метрикасындағы өзгерістер графигі ғылыми құжаттардан семантикалық

құрылымды алу үшін осы модельді пайдаланудың артықшылықтарын анық көрсетті. Ғылыми құжаттардан семантикалық құрылымды алу үшін DistilBART моделін пайдалану тиімді екенін көрсетті. Қосымша оқытылған модель ақпаратты жинақтау және құрылымдау мәселелерінде жоғары нәтиже көрсетті, бұл автоматтандыру және ғылыми мәтіндерді талдау тиімділігін арттыру үшін жаңа мүмкіндіктер ашады. Төменде әртүрлі деректерді алдын ала өңдеу және оқыту тәсілдерін пайдалану кезінде DistilBART үлгісінің өнімділігінің әртүрлі аспектілері көрсетілген. 3-суретте модельдің өнімділігін бағалауға мүмкіндік беретін әрбір тәсіл үшін шығынды бағалауды қадағалау көрсетілген. Бақылаулар жетілдірілген алдын ала өңдеуде үлгі өнімділігінің жақсырақ екенін көрсететін ең төменгі жоғалту ұпайы бар екенін көрсетеді. Негізгі алдын ала өңдеу және бастапқы үлгі ұқсас, бірақ сәл жоғары шығынды бағалауды көрсетеді. Сонымен қатар, деноизациясы бар автокодерлер жоғары жоғалтуларды көрсетеді, бұл нашар өнімділікті көрсетеді. Жетілдірілген алдын ала өңдеумен төмен жоғалту ұпайы жақсырақ жалпылау мүмкіндігін және үлгі өнімділігін көрсетеді (1-сурет)



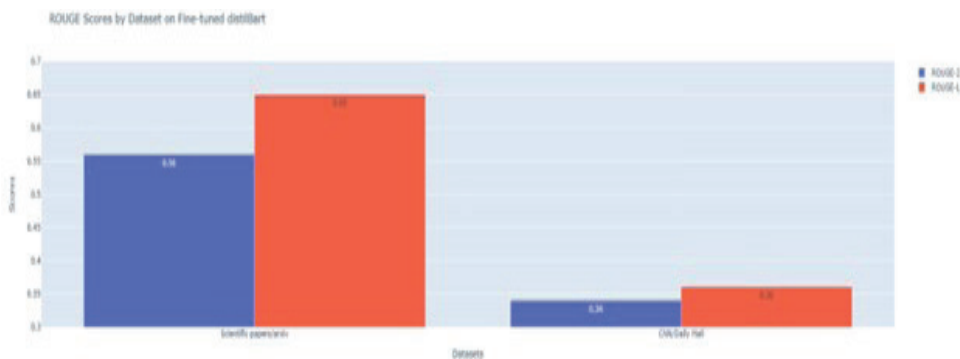
Сур. 1. Тәсіл бойынша шығынды бағалау
(Fig. 1. Evaluation Loss by Approach)

2-сурет әрбір тәсілді бағалау және есептеу тиімділігін бөлектеу үшін орындау уақытының салыстыруын көрсетеді. Бақылаулар көрсеткендей, деноизациялау автокодерлерінің басқа тәсілдермен салыстырғанда орындау уақыты қысқа болады, ал негізгі үлгі, негізгі алдын ала өңдеу және кеңейтілген алдын ала өңдеу ұзағырақ және ұқсас орындау уақыттарын көрсетеді. Автокодерлердің жоғары есептеу тиімділігіне қарамастан, олардың жоғары жоғалту ұпайлары мен төмен ROUGE ұпайларында көрінетін төмен өнімділігі жылдамдық пен сапа арасындағы ымырасыздықты көрсетеді..



Сур. 2. Тәсіл бойынша бағалаудың орындалу уақыты
(Fig. 2. Evaluation Runtime by Approach)

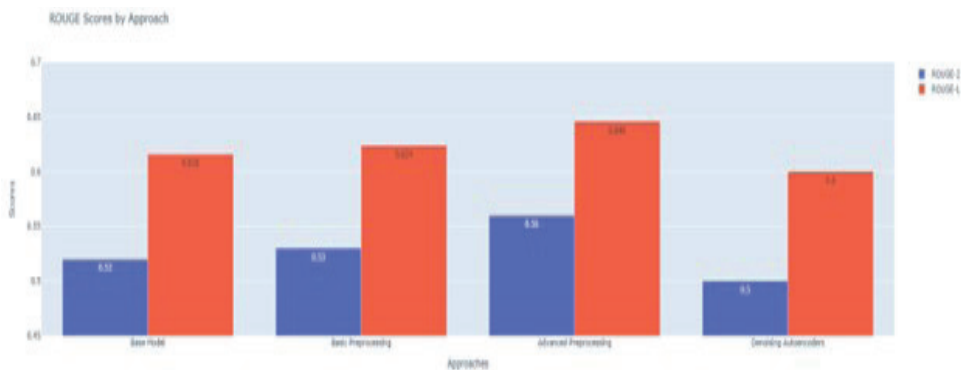
3-сурет әртүрлі деректер жиынындағы үлгі өнімділігіндегі айырмашылықтарды анық көрсетеді, бұл нақты тапсырма үшін үлгіні таңдау кезінде ескеру маңызды. Біз модельді DistilBART бағдарламасының алдын ала дайындалған нұсқасында бағаладық және ол CNN/Daily Mail-пен салыстырғанда Scientific papers/архив деректер жинағында айтарлықтай жақсы нәтиже көрсетті. “Scientific papers/архив” үшін ROUGE-2 және ROUGE-L көрсеткіштері айтарлықтай жоғары, бұл ғылыми мақалаларды өңдеу кезінде үлгінің жақсырақ өнімділігін көрсетеді. Сонымен қатар, CNN/Daily Mail деректер жинағындағы нәтижелер төмен болды, бұл модельдің жаңалықтар мақалаларының мәтінін өңдеу қиынырақ екенін немесе деректердің осы түріне үлгіні қосымша баптау қажет екенін көрсетуі мүмкін.



Сур. 3. Тәсіл бойынша бағалаудың орындалу уақыты
(Fig. 3. Evaluation Runtime by Approach)

4-суретте әртүрлі алдын ала өңдеу және оқыту тәсілдері мәтінді қорытындылау үшін пайдаланылатын үлгілердің сапасына қалай әсер ететіні анық көрсетілген. ROUGE-2 және ROUGE-L метрикасының нәтижелері деректерді алдын ала өңдеу мен үлгіні оқытудың төрт тәсілі үшін ұсынылған. Жетілдірілген алдын ала өңдеуі бар модель күрделі деректерді алдын ала

өңдеудің тиімділігін, соның ішінде таңбалауды, сөзді жоюды тоқтатуды және тіректерді бөлуді көрсететін екі көрсеткіш бойынша да ең жақсы нәтижелерді көрсетеді. Бұған қоса, негізгі алдын ала өңдеу үлгісі деректерді алдын ала өңдеудің маңыздылығын растайтын негізгі үлгімен салыстырғанда жақсартылған нәтижелерді көрсетеді. Деректерді тазалау үшін автокодерлерді пайдаланатын модель (Denoising Autoencoders) ROUGE-2 метрикасындағы өнімділікті аздап төмендетеді. Дегенмен, ол ROUGE-L метрикасында әлі де жақсы нәтижелерді көрсетеді, бұл шуды жоюдың тиімділігін көрсетеді, бірақ ол әрі қарай реттеуді қажет етеді. Осылайша, әртүрлі алдын ала өңдеу және оқыту тәсілдері үшін ROUGE-2 және ROUGE-L метрикасының нәтижелерін талдау оқыту үлгілерінен бұрын деректерді мұқият өңдеудің маңыздылығын көрсетеді. Жетілдірілген деректерді алдын ала өңдеу ең тиімді болып көрінеді, ал негізгі алдын ала өңдеу және автокодерлеу әдісі де негізгі үлгіге қарағанда жақсартуларды көрсетеді, бірақ жақсы нәтижелерге қол жеткізу үшін одан әрі жетілдіруді қажет етеді.



Сур. 4. Тәсіл бойынша бағалаудың орындалу уақыты
(Fig. 4. Evaluation Runtime by Approach)

Осылайша, салыстырмалы талдау ғылыми мәтіндерден семантикалық құрылымды алу тапсырмаларында DistilBART моделінің сапасын жақсартудың ең тиімді тәсілі деректерді кеңейтілген алдын ала өңдеу екенін көрсетеді. Негізгі алдын ала өңдеу және негізгі үлгі әдістері де жақсы нәтижелер көрсетеді, ал деноизациялық автокодерлер әдісі оның жұмысын жақсарту үшін одан әрі оңтайландыруды қажет етеді.

Қорытынды

Бұл мақалада BART моделінің жақсартылған және қысқартылған нұсқасы болып табылатын DistilBART үлгісін пайдаланып ғылыми құжаттардан семантикалық құрылымды алу әдісі зерттелді. Мәтіндік деректердің үлкен көлеміне үйретілген DistilBART моделі мәтінді қорытындылау, машиналық аударма және мәтінді құру сияқты табиғи тілді өңдеу мәселелерін шешуде

жоғары өнімділікті көрсетті. Бұл жұмыстың негізгі мақсаты әртүрлі салалардағы ғылыми құжаттардан семантикалық ақпаратты алуда және құрылымдауда тиімді болатын DistilBART моделі негізінде әмбебап құрал жасау болды. Тәжірибелер arXiv сияқты ашық дереккөздерден ғылыми мақалалардың кең деректер жинағын жинады. Модельдерді оқыту сапасын арттыру үшін деректерді алдын ала өңдеудің әртүрлі әдістері, соның ішінде негізгі және кеңейтілген алдын ала өңдеу, сонымен қатар деноизациялау автокодерлері әдісі қолданылды. Нәтижелер кеңейтілген деректерді алдын ала өңдеуі бар DistilBART үлгісі ROUGE-2 және ROUGE-L өлшемдерінде ең жақсы нәтиже көрсеткенін көрсетті, бұл күрделі деректерді алдын ала өңдеудің, соның ішінде токенизацияны, сөзді жоюды және тіректерді жоюдың тиімділігін көрсетеді. Модельдің «Scientific papers/archiv» және «CNN/Daily Mail» сияқты әртүрлі деректер жиынындағы өнімділігінің салыстырмалы талдауы DistilBART моделінің жаңалықтар мәтіндеріне қарағанда ғылыми мақалаларды өңдеуде айтарлықтай жақсы жұмыс істейтінін көрсетті. Бұл әртүрлі деректер түрлері үшін үлгіні қосымша баптау қажеттілігін көрсетеді. Модельдің өнімділігін бағалауға қажетті уақыт, сонымен қатар қолданылатын деректерді алдын ала өңдеу тәсіліне байланысты өзгерді, деноизациялаушы автокодерлер басқа әдістерге қарағанда орындау уақытын қысқартады, бірақ дәлдігі төмен. Осылайша, ғылыми құжаттардан семантикалық құрылымды алу үшін DistilBART моделін қолдану мәтінді талдауды автоматтандыру және тиімділігін арттыру үшін жаңа мүмкіндіктер ашты. Жетілдірілген деректерді алдын ала өңдеу үлгі сапасын жақсартуда ең тиімді болды, ал негізгі алдын ала өңдеу және деноизациялау автокодерлері де жақсы нәтиже көрсетті, бірақ одан әрі оңтайландыруды қажет етеді. Бұл зерттеу табиғи тілді өңдеу тапсырмаларында жақсы нәтижелерге қол жеткізу үшін деректерді мұқият өңдеудің және гиперпараметрлерді реттеудің маңыздылығын растайды.

Әдебиеттер

Вэй, Х., Цуй, Х., Чен, Н., Ван, Х., Чжан, Х., Хуан, С.,... & Хан, В. (2023). Chatgpt-пен сөйлесу арқылы нөлдік ақпаратты алу. arXiv алдын ала басып шығару arXiv:2302.10205.

Дагделен, Дж., Данн, А., Ли, С., Уокер, Н., Розен, А. С., Седер, Г.,... & Джейн, А. (2024). Үлкен тілдік үлгілері бар ғылыми мәтіннен құрылымдық ақпаратты алу. Табиғат Коммуникациялары, 15 (1), 1418.

Ивги, М., Шахам, У. Және Берант, Дж. (2023). Қысқа мәтінді модельдермен ұзақ мәтінді тиімді түсіну. Есептеу Лингвистикасы Қауымдастығының операциялары, 11, 284-299.

Қайбасова, Д., & Нұртай, М. (2022, Сәуір). Мәтіндік Оқу Жұмыстарының Сапасын Бағалау Үшін Машиналық Оқыту Модельдерін Салыстырмалы Талдау. 2022 Жылы Ақылды Ақпараттық Жүйелер мен Технологиялар Бойынша Халықаралық Конференция (SIST) (1-4 беттер). ИӘ.

Ландолси, М.Ю., Хлауа, Л. Және Бен Ромдхейн, Л. (2023). Электрондық медициналық құжаттардан ақпарат алу: қазіргі жағдайы және болашақ зерттеу бағыттары. Білім Және Ақпараттық Жүйелер, 65 (2), 463-516.

Ли, Х., Ай, К., Чен, Дж., Донг, К., Ву, Ю., Лю, Ю.,... & Тянь, Q. (2023, Шілде). SAILER: заңды істерді іздеу үшін құрылымды білетін алдын-ала дайындалған тілдік модель. АҚПАРАТТЫ Іздеу Саласындағы Зерттеулер мен Өзірлемелер бойынша ACM SIGIR 46-Шы Халықаралық Конференциясының Материалдарында (1035-1044 беттер).

Лю, Х., Чжоу, Г., Конг, М., Инь, З., Ли, Х., Инь, Л. Және Чжэн, В. (2023). Твиттердегі қысқа мәтіндердің көп таңбаланған корпусын жасау: жартылай автоматты әдіс. *Жүйелер*, 11 (8), 390.

Полак, М. П., Моды, С., Латосинска, А., Чжан, Дж., Ван, К. В., Ван, С.,... & Морган, Д. (2024). Жалпы мақсаттағы тілдік модельдерді қолдана отырып, мәтіннен материалдар деректерін алудың икемді, модельдік-агностикалық әдісі. *Сандық Жаңалық*, 3 (6), 1221-1235.

Полак, М.П. Және Морган, Д. (2024). Ауызекі сөйлеу тілінің модельдері мен жедел инженериясы бар ғылыми-зерттеу жұмыстарынан нақты материалдар туралы мәліметтер алу. *Табиғат Коммуникациялары*, 15 (1), 1569.

Садирмекова, З., Түсіпов, Ж., Мурзахметов, А., Жидекулова, Г., Тунгатарова, А., Төленбаев, М.,... & Боранқұлова, Г. (2023). Мәтінді өңдеудің автоматты әдістерінің онтологиялық инженериясы. *Халықаралық Электротехника Және Есептеу Техникасы Журналы(IJESCE)*, 13 (6), 6620-6628.

Сантана, Б., Кампос, Р., Аморим, Э., Хорхе, А., Сильвано, П., Және Нуньес, С. (2023). Мәтіндік деректерден баяндауды алу бойынша сауалнама. *Жасанды Интеллектке Шолу*, 56 (8), 8393-8435.

Тревисо, М., Ли, Дж. У., Джи, Т., Акен, Б. В., Цао, К., Чосичи, М. Р.,... & Шварц, Р. (2023). Табиғи тілді өңдеудің тиімді әдістері: сауалнама. *Есептеу Лингвистикасы Қауымдастығының операциялары*, 11, 826-860.

Умер, М., Имтиаз, З., Ахмад, М., Наппи, М., Медалия, К., Чой, Г. С. Және Мехмуд, А. (2023). Конволюциялық нейрондық желінің және Жылдам Мәтінді ендірудің мәтінді жіктеуге әсері. *Мультимедиялық Құралдар мен Қосымшалар*, 82(4), 5569-5585.

Хартманн, Дж., Хейтманн, М., Сиберт, С., Шамп, С. (2023). Сезімнен гөрі: Сезімді талдаудың Дәлдігі және қолданылуы. *Маркетингтегі Халықаралық Зерттеулер Журналы*, 40 (1), 75-87.

Шай, С. Р. (2023). Мәтінді алдын-ала өңдеу әдістерін салыстыру. *Табиғи Тіл Инженериясы*, 29 (3), 509-553.

References

Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509-553.

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., ... & Jain, A. (2024). Structured information extraction from scientific text with large language models. *Nature Communications*, 15(1), 1418.

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, 40(1), 75-87.

Ivgi, M., Shaham, U., & Berant, J. (2023). Efficient long-text understanding with short-text models. *Transactions of the Association for Computational Linguistics*, 11, 284-299.

Kaibassova, D., & Nurtay, M. (2022, April). The Comparative Analysis of Machine Learning Models for Quality Assessment of Textual Academic Works. In *2022 International Conference on Smart Information Systems and Technologies (SIST)* (pp. 1-4). IEEE.

Landolsi, M. Y., Hlaoua, L., & Ben Romdhane, L. (2023). Information extraction from electronic medical documents: state of the art and future research directions. *Knowledge and Information Systems*, 65(2), 463-516.

Li, H., Ai, Q., Chen, J., Dong, Q., Wu, Y., Liu, Y., ... & Tian, Q. (2023, July). SAILER: structure-aware pre-trained language model for legal case retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1035-1044).

Liu, X., Zhou, G., Kong, M., Yin, Z., Li, X., Yin, L., & Zheng, W. (2023). Developing multi-labelled corpus of twitter short texts: a semi-automatic method. *Systems*, 11(8), 390.

Polak, M. P., & Morgan, D. (2024). Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nature Communications*, 15(1), 1569.

Polak, M. P., Modi, S., Latosinska, A., Zhang, J., Wang, C. W., Wang, S., ... & Morgan, D. (2024). Flexible, model-agnostic method for materials data extraction from text using general purpose language models. *Digital Discovery*, 3(6), 1221-1235.

Sadirmekova, Z., Tussupov, J., Murzakhmetov, A., Zhidekulova, G., Tungatarova, A., Tulenbayev, M., ... & Borankulova, G. (2023). Ontology engineering of automatic text processing methods. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(6), 6620-6628.

Santana, B., Campos, R., Amorim, E., Jorge, A., Silvano, P., & Nunes, S. (2023). A survey on narrative extraction from textual data. *Artificial Intelligence Review*, 56(8), 8393-8435.

Treviso, M., Lee, J. U., Ji, T., Aken, B. V., Cao, Q., Ciosici, M. R., ... & Schwartz, R. (2023). Efficient methods for natural language processing: A survey. *Transactions of the Association for Computational Linguistics*, 11, 826-860.

Umer, M., Imtiaz, Z., Ahmad, M., Nappi, M., Medaglia, C., Choi, G. S., & Mehmood, A. (2023). Impact of convolutional neural network and FastText embedding on text classification. *Multimedia Tools and Applications*, 82(4), 5569-5585.

Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S., ... & Han, W. (2023). Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

CONTENTS

INFORMATION AND COMMUNICATION TECHNOLOGIES

M. Aitimov, R.U Almenayeva, K.K. Makulov, A.B. Ostayeva, R. Muratkhan APPLICATION OF MACHINE LEARNING METHOD TO ANALYZE AND EXTRACT SEMANTIC STRUCTURES FROM SCIENTIFIC TEXTS.....	5
A.K. Aitim, G.K. Sembina MODELING OF HUMAN BEHAVIOR FOR SMARTPHONE WITH USING MACHINE LEARNING ALGORITHM.....	17
G. Aksholak, A. Bedelbayev, R. Magazov ANALYSIS AND COMPARISON OF MACHINE LEARNING METHODS FOR MALWARE DETECTION.....	29
A.L. Alexeyeva SUBSONIC VIBROTRANSPORT SOLUTIONS OF THE WAVE EQUATION IN SPACES OF DIMENSION $N=1,2,3$	42
K. Bagitova, Sh. Mussiraliyeva, K. Azanbai ANALYSIS OF SYSTEMS FOR RECOGNIZING POLITICAL EXTREMISM IN ONLINE SOCIAL NETWORKS.....	60
A.S. Baegizova, G.I. Mukhamedrakhimova, I. Bapiyev, M.Zh. Bazarova, U.M. Smailova EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING METHODS FOR KEYWORD COVERAGE.....	73
G. Bekmanova, B. Yergesh, G. Yelibayeva, A. Omarbekova, M. Strecker MODELING THE RULES AND CONDITIONS FOR CONDUCTING PRE-ELECTION DEBATES.....	89
M. Bolatbek, M. Sagynay, Sh. Mussiraliyeva USING MACHINE LEARNING METHODS FOR DETECTING DESTRUCTIVE WEB CONTENT IN KAZAKH LANGUAGE.....	99
Y. Golenko, A. Ismailova, K. Kadirkulov, R. Kalendar DEVELOPMENT OF AN ONLINE PLATFORM FOR SEARCHING FOR TANDEM REPEATS USING WHOLE GENOME SEQUENCING.....	112

T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, Sh. Akhmetzhanova A BIBLIOMETRIC ANALYSIS OF EDGE COMPUTING IN INDUSTRIAL INTERNET OF THINGS (IIoT) CYBER-PHYSICAL SYSTEMS.....	123
S.S. Koishybay, N. Meirambekuly, A.E. Kulakaeva, B.A. Kozhakhmetova, A.A. Bulin DEVELOPMENT OF THE DESIGN OF A MULTI-BAND DISCONE ANTENNA.....	138
A. Kydyrbekova, D. Oralbekova SPEAKER IDENTIFICATION USING DISTRIBUTION-PRESERVING X-VECTOR GENERATION.....	152
B. Medetov, A. Nurlankyzy, A. Akhmediyarova, A. Zhetpisbayeva, D. Zhexebay COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF NEURAL NETWORKS WITHIN THE LOW SNR.....	163
A.A Myrzatay, L.G. Rzaeva, B. Zhumadilla, A.A. Mukhanova, G.A. Uskenbayeva DOUBLE EXPONENTIAL SMOOTHING AND TIME WINDOW METHODS FOR PREDICTIVE LAN MONITORING: ANALYSIS, COMPARISON AND APPLICATION.....	174
L. Naizabayeva, M.N. Satymbekov PREDICTING URBAN SOIL POLLUTION USING MACHINE LEARNING ALGORITHMS.....	194
A.U. Mukhiyadin, U.T. Makhazhanova, A.Z. Alimagambetova, A.A. Mukhanova, A.I. Akmoldina PREDICTING STUDENT LEARNING ENGAGEMENT USING MACHINE LEARNING TECHNIQUES: ANALYSIS OF EDUCATION DATA IN KAZAKHSTAN.....	204
Zh. Tashenova, Zh. Abdugulova, Sh. Amanzholova, E. Nurlybaeva PENETRATION TESTING APPROACHES EMPLOYING THE OPENVAS VULNERABILITY MANAGEMENT UTILITY.....	218
D.B. Tyulemissova, A.K. Shaikhanova, V. Martsenyuk, G.A. Uskenbayeva MODERN APPROACHES TO STUDYING THE DYNAMICS OF INFORMATION FLOW IN SOCIAL MEDIA BASED ON MACHINE LEARNING METHODS.....	231

МАЗМҰНЫ

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР

М. Айтимов, Р.У Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан
ҒЫЛЫМИ МӘТІНДЕРДЕН СЕМАНТИКАЛЫҚ ҚҰРЫЛЫМДАРДЫ
ТАЛДАУ ЖӘНЕ АЛУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСІН
ҚОЛДАНУ.....5

Ә.Қ. Әйтiм, Г.К. Сембина
МАШИНАЛЫҚ ОҚУ АЛГОРИТМІН ПАЙДАЛАНЫП СМАРТФОН
ҮШІН АДАМ МІНЕЗІН МОДЕЛДЕУ.....17

Г.И. Ақшолақ, А.А. Бедельбаев, Р.С. Мағазов
ЗИЯНДЫ БАҒДАРЛАМАЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ТАЛДАУ ЖӘНЕ САЛЫСТЫРУ.....29

А.Л. Алексеева
N=1,2,3 ӨЛШЕМДІ КЕҢІСТІГІНДЕГІ ТОЛҚЫНДЫҚ ТЕҢДЕУДІҢ
ДЫБЫСҚА ДЕЙІНГІ ДІРІЛКӨЛІКТІК ШЕШІМДЕРІ.....42

Қ.Б. Бағитова, Ш.Ж. Мусиралиева, Қ. Азанбай
ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕГІ САЯСИ ЭКСТРЕМИЗМДІ ОНЛАЙН ТАҢУ
ЖҮЙЕЛЕРІН ТАЛДАУ.....60

**А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова,
У.М. Смайлова**
ТҮЙІН СӨЗДЕРДІ ҚАМТУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ
ТИІМДІЛІГІН БАҒАЛАУ.....73

**Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова,
М. Strecker**
САЙЛАУ АЛДЫНДАҒЫ ПІКІРТАЛАСТАРДЫ ӨТКІЗУ ЕРЕЖЕЛЕРІ
МЕН ШАРТТАРЫН МОДЕЛЬДЕУ.....89

М.А. Болатбек, М.Сағынай, Ш.Ж. Мусиралиева
ҚАЗАҚ ТІЛІНДЕГІ ДЕСТРУКТИВТІ ВЕБ-КОНТЕНТТІ АНЫҚТАУ ҮШІН
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ.....99

Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь
ТОЛЫҚ ГЕНОМДЫҚ СЕКВЕНИРЛЕУДЕ ТАНДЕМДІК
ҚАЙТАЛАНУЛАРДЫ ІЗДЕУ ҮШІН ОНЛАЙН ПЛАТФОРМАСЫН
ӘЗІРЛЕУ.....112

- Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова**
ӨНДІРІСТІК ЗАТТАР ИНТЕРНЕТІ (IoT) КИБЕРФИЗИКАЛЫҚ
ЖҮЙЕЛЕРІНДЕ ШЕТКІ ЕСЕПТЕУЛЕРДІ ҚОЛДАНУҒА
БИБЛИОМЕТРИЯЛЫҚ ТАЛДАУ.....123
- С.С. Қойшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова,
А.А. Булин**
КӨПДИАПАЗОНДЫДИСКОНУСТЫҚАНТЕННАКОНСТРУКЦИЯСЫН
ӘЗІРЛЕУ.....138
- А.С. Кыдырбекова, Д.О. Оралбекова**
ТАРАТУДЫ САҚТАЙТЫН Х-ВЕКТОРЛАР ГЕНЕРАЦИЯСЫН
ПАЙДАЛАНЫП ДАУЫСТЫ ИДЕНТИФИКАЦИЯЛАУ.....152
- Б. Медетов, А. Нурланқызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай**
СИГНАЛШУЫЛ ҚАТЫНАСЫ ТӨМЕН ЖАҒДАЙДА НЕЙРОНДЫҚ
ЖЕЛЛЕРДІҢ ТИІМДІЛІГІНЕ САЛЫСТЫРМАЛЫ ТАЛДАУ ЖАСАУ.....163
- А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова,
Г.А. Ускенбаева**
ЖЕРГІЛІКТІ ЖЕЛІНІ БОЛЖАМДЫ БАҚЫЛАУҒА АРНАЛҒАН ҚОС
ЭКСПОНЕНЦИАЛДЫ ТЕГІСТЕУ ЖӘНЕ УАҚЫТ ТЕРЕЗЕЛЕРІНІҢ
ӘДІСТЕРІ: ТАЛДАУ, САЛЫСТЫРУ ЖӘНЕ ҚОЛДАНУ.....174
- Л. Найзабаева, М.Н. Сатымбеков**
МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ПАЙДАЛАНУ АРҚЫЛЫ
ҚАЛА ТОПЫРАҒЫНЫҢ ЛАСТАНУЫН БОЛЖАУ.....194
- А.Ұ. Мұхиядин, У.Т. Махажанова, А.З. Алимагамбетова, А.А.Муханова,
А.И. Акмолдина**
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНА ОТЫРЫП,
ОҚУШЫЛАРДЫҢ БІЛІМ АЛУҒА ЫНТАСЫН БОЛЖАУ:
ҚАЗАҚСТАНДАҒЫ БІЛІМ БЕРУ ДЕРЕКТЕРІН ТАЛДАУ.....204
- Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева**
OPENVAS ОСАЛДЫҒЫН БАСҚАРУ УТИЛИТАСЫН ҚОЛДАНА
ОТЫРЫП, ЕНУДІ ТЕСТІЛЕУ ТӘСІЛДЕРІ.....218
- Д.Б. Тюлемисова, А.К. Шайханова, В.П. Мартценюк, Г.А. Ускенбаева,
Г.В. Бекешева**
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН ӘЛЕУМЕТТІК
ЖЕЛЛЕРДЕГІ АҚПАРАТ АҒЫНЫНЫҢ ДИНАМИКАСЫН ЗЕРТТЕУДІҢ
ЗАМАНАУИ ТӘСІЛДЕРІ.....231

СОДЕРЖАНИЕ

ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

М. Айтимов, Р.У. Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА И ИЗВЛЕЧЕНИЯ СЕМАНТИЧЕСКИХ СТРУКТУР ИЗ НАУЧНЫХ ТЕКСТОВ.....	5
А.К. Айтим, Г.К. Сембина МОДЕЛИРОВАНИЕ ЧЕЛОВЕЧЕСКОГО ПОВЕДЕНИЯ ДЛЯ СМАРТФОНА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ.....	17
Г.И. Акшолок, А.А. Бедельбаев, Р.С. Магазов АНАЛИЗ И СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ВРЕДОНОСНОГО ПО.....	29
Л.А. Алексеева ДОЗВУКОВЫЕ ВИБРОТРАНСПОРТНЫЕ РЕШЕНИЯ ВОЛНОВОГО УРАВНЕНИЯ В ПРОСТРАНСТВАХ РАЗМЕРНОСТИ $N=1,2,3$	42
К.Б. Багитова, Ш.Ж. Мусиралиева, К. Азанбай АНАЛИЗ СИСТЕМ РАСПОЗНАВАНИЯ ПОЛИТИЧЕСКОГО ЭКСТРЕМИЗМА В СОЦИАЛЬНЫХ СЕТЯХ ОНЛАЙН.....	60
А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова, У.М. Смайлова ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОХВАТА КЛЮЧЕВЫХ СЛОВ.....	73
Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова, М. Strecker МОДЕЛИРОВАНИЕ ПРАВИЛ И УСЛОВИЙ ПРОВЕДЕНИЯ ПРЕДВЫБОРНЫХ ДЕБАТОВ.....	89
М.А. Болатбек, М. Сагынай, Ш.Ж. Мусиралиева ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ДЕСТРУКТИВНОГО ВЕБ-КОНТЕНТА НА КАЗАХСКОМ ЯЗЫКЕ.....	99
Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь РАЗРАБОТКА ОНЛАЙН-ПЛАТФОРМЫ ДЛЯ ПОИСКА ТАНДЕМНЫХ ПОВТОРОВ ПРИ ПОЛНОГЕНОМНОМ СЕКВЕНИРОВАНИИ.....	112

- Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова**
БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИМЕНЕНИЯ ГРАНИЧНЫХ
ВЫЧИСЛЕНИЙ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ
ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ (IIoT).....123
- С.С. Койшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова,
А.А. Булин**
РАЗРАБОТКА КОНСТРУКЦИИ МНОГОДИАПАЗОННОЙ
ДИСКОНУСНОЙ АНТЕННЫ.....138
- А.С. Кыдырбекова, Д.О. Оралбекова**
ИДЕНТИФИКАЦИЯ ГОВОРЯЩЕГО С ИСПОЛЬЗОВАНИЕМ
ГЕНЕРАЦИИ X-ВЕКТОРОВ С СОХРАНЕНИЕМ РАСПРЕДЕЛЕНИЯ...152
- Б. Медетов, А. Нурланкызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай**
СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ НЕЙРОННЫХ
СЕТЕЙ ПРИ НИЗКОМ ЗНАЧЕНИИ ОТНОШЕНИЯ С/Ш.....163
- А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова, Г.А. Ускенбаева**
МЕТОДЫ ДВОЙНОГО ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ
И ВРЕМЕННЫХ ОКОН ДЛЯ ПРЕДИКТИВНОГО МОНИТОРИНГА
ЛВС: АНАЛИЗ, СРАВНЕНИЕ И ПРИМЕНЕНИЕ.....174
- Л. Найзабаева, М.Н. Сатымбеков**
ПРОГНОЗИРОВАНИЕ ЗАГРЯЗНЕНИЯ ГОРОДСКОЙ ПОЧВЫ С
ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ.....194
- А.У. Мухиядин, У.Т. Махажанов, А.З. Алимагамбетова, А.А. Муханова,
А.И. Акмолдина**
ПРОГНОЗИРОВАНИЕ МОТИВАЦИИ УЧАЩИХСЯ К ОБУЧЕНИЮ С
ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ:
АНАЛИЗ ДАННЫХ ОБ ОБРАЗОВАНИИ В КАЗАХСТАНЕ.....204
- Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева**
ПОДХОДЫ К ТЕСТИРОВАНИЮ НА ПРОНИКНОВЕНИЕ С
ИСПОЛЬЗОВАНИЕМ УТИЛИТЫ УПРАВЛЕНИЯ УЯЗВИМОСТЯМИ
OPENVAS.....218
- Д.Б. Тюлемисова, А.К. Шайханова, В. Мартценюк, Г.А. Ускенбаева,
Г.В. Бекешева**
СОВРЕМЕННЫЕ ПОДХОДЫ К ИЗУЧЕНИЮ ДИНАМИКИ
ИНФОРМАЦИОННОГО ПОТОКА В СОЦИАЛЬНЫХ
МЕДИА НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....231

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

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Редакторы: *Д.С. Аленов, Ж.Ш. Әден*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 2.12.2024.

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

16,0 п.л. Тираж 300. Заказ 4.