

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
Казахский национальный
университет имени аль-Фараби

N E W S

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

SERIES
PHYSICO-MATHEMATICAL

3 (343)

JULY – SEPTEMBER 2022

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 Нұрәділұлы (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 Нұрсұлу 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. Физика-математикалық 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: *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/>

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022
Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

Главный редактор:

МУТАНОВ Галимкаир 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 физика-математическая».

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 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/>

© Национальная академия наук Республики Казахстан, 2022
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 Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cyber-security, 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. Physical-mathematical series.

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 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/>

© National Academy of Sciences of the Republic of Kazakhstan, 2022

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 3, Number 343 (2022), 260-274
<https://doi.org/10.32014/2022.2518-1726.151>
УДК 004.4, 004.8, 004.6, 004.89:004.4

**К. Якунин^{1,2,*}, Р.И. Мухамедиев^{1,2}, М. Елис^{1,2}, Я. Кучин^{1,2},
А. Сымагулов^{1,2}, Н. Юничева¹, Е. Мухамедиева¹**

¹Институт информационных и вычислительных технологий МОН РК,
Казахстан, Алматы;

²Satbayev University, Казахстан, Алматы;
E-mail: ykuchin@mail.ru

АНАЛИЗ ТЕМАТИЧЕСКИХ КЛАСТЕРОВ ПУБЛИКАЦИЙ СМИ РЕСПУБЛИКИ КАЗАХСТАН ПО ТЕМЕ ПАНДЕМИИ COVID-19

Аннотация. В настоящей работе сформирован корпус документов по данным русскоязычных СМИ Казахстана с помощью автоматического скрапинга. Корпус состоит из 761831 документа, которые относятся к ведущим новостным изданиям страны. Одним из основных инструментов, применяемых для анализа крупных корпусов текстов, является тематическое моделирование. Наиболее часто для формирования тематической модели исследователи используют так называемое латентное размещение Дирихле (LDA).

Мы использовали ARTM – расширение LDA, отличие которого заключается в применении конфигурируемых регуляризаторов, которые позволяют тонко настроить желаемый результат модели: в том числе уменьшить/увеличить склонность модели к включению слова и/или документа сразу в несколько топиков, изменить склонность модели к большему/меньшему количеству ненулевых весов в итоговой матрице. Анализ результатов показывает, как меняется отношение общества к проблемам COVID-19 в 2021-2022 годах. Во-первых, результаты отражают устойчивую тенденцию снижения интереса электронных СМИ к теме пандемии, хотя и в неравной степени для разных тематических групп. Во-вторых, выявилась тенденция к переносу

фокуса внимания на более прагматичные вопросы, такие как вопросы удалённого обучения, удалённой работы, влияния карантинных ограничений на экономику.

Ключевые слова: машинное обучение (machine learning), геопространственный искусственный интеллект (geoAI), здравоохранение (health care), мультикритериальный анализ решений (multi criteria decision analysis – MCDA), объяснимое машинное обучение (explainable machine learning), обработка естественного языка (natural language processing).

**К. Якунин^{1,2*}, Р.И. Мухамедиев^{1,2}, М. Елис^{1,2}, Я. Кучин^{1,2},
Н. Юничева¹, А. Сымагулов^{1,2}, Е. Мухамедиева¹**

¹Ақпараттық және есептеуіш технологиялар институты,
Қазақстан, Алматы;

²Сатбаев университеті, Қазақстан, Алматы.

E-mail: ykuchin@mail.ru

КОВИД-19 ПАНДЕМИЯСЫ ТАҚЫРЫБЫ БОЙЫНША ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БАҚ БАСЫЛЫМДАРЫНЫҢ ТАҚЫРЫПТЫҚ КЛАСТЕРЛЕРІН ТАЛДАУ

Аннотация. Осы жұмыста Қазақстандағы орыстілді БАҚ деректері бойынша автоматты қырғышты қолдану арқылы құжаттар корпусы қалыптастырылды. Корпус еліміздің жетекші ақпарат агенттіктеріне тиесілі 761 831 құжаттан тұрады. Үлкен мәтіндік корпусы талдау үшін қолданылатын негізгі құралдардың бірі тақырыптық модельдеу болып табылады. Көбінесе зерттеушілер тақырыптық модельді қалыптастыру үшін жасырын Дирихле орналастыру (LDA) деп аталады.

Бұл жұмыста ARTM қолданылды - LDA кеңейтімі, оның айырмашылығы үлгінің қалаған нәтижесін дәл келтіруге мүмкіндік беретін конфигурацияланатын регуляризаторларды пайдалануда жатыр: соның ішінде модельдің сөзді және/немесе қосуға бейімділігін азайту / арттыру. Соңғы матрицадағы нөлдік емес салмақтар санының көп/кеміне модель үрдісін өзгерту арқылы бірден бірнеше тақырыпта құжат. Нәтижелерді талдау 2021-2022 жылдары қоғамның COVID-19 проблемаларына деген көзқарасының қалай өзгеретінін көрсетеді. Біріншіден, нәтижелер әртүрлі тақырыптық топтар үшін бірдей емес

дәрежеде болса да, электронды БАҚ-тың пандемия тақырыбына қызығушылығының тұрақты төмендеу үрдісін көрсетеді. Екіншіден, назарды қашықтықтан оқыту, қашықтан жұмыс істеу және карантиндік шектеулердің экономикаға әсері сияқты прагматикалық мәселелерге аудару үрдісі байқалды.

Түйін сөздер: Машиналық оқыту (machine learning), геокеңістіктік жасанды зият (geoAI), Денсаулық сақтау (health care), мультикритериалдық шешімдерді талдау (multi criteria decision analysis-MCDA), түсінікті машиналық оқыту (explainable machine learning), табиғи тілді өңдеу (natural language processing).

**K. Yakunin^{1,2*}, R.I. Mukhamediev^{1,2}, M. Elis^{1,2}, Ya. Kuchin^{1,2},
N. Yunicheva¹, A. Symagulov^{1,2}, E. Mukhamedieva¹**

¹Institute of Information and Computational Technologies,
Kazakhstan, Almaty;

²Satbayev University, Kazakhstan, Almaty.

E-mail: ykuchin@mail.ru

ANALYSIS OF THEMATIC CLUSTERS OF KAZAKHSTAN MEDIA PUBLICATIONS ON THE TOPIC OF THE COVID-19 PANDEMIC

Abstract. In this paper, a corpus of documents based on the Russian-language media of Kazakhstan was formed using automatic scraping. The corpus consists of 761831 documents, which belong to the leading news publications of the country. One of the main tools used to analyze large corpora of texts is thematic modeling. Most often researchers use the so-called latent Dirichlet placement (LDA) to form a thematic model.

We used ARTM, an extension of LDA, the difference of which lies in the use of configurable regularizers that allow us to fine-tune the desired model result: including reducing/increasing the propensity of the model to include a word and/or document in several topics at once, changing the model propensity to have more/less non-zero weights in the resulting matrix. Analysis of the results shows how public attitudes toward COVID-19 issues change between 2021 and 2022. First, the results reflect a steady downward trend in electronic media interest in the pandemic topic, albeit to an unequal degree for different thematic groups. Second, there is a tendency to shift the

focus to more pragmatic issues, such as distance learning, telecommuting, and the impact of quarantine restrictions on the economy.

Key words: Machine Learning, geospatial artificial intelligence (geoAI), health care, multi criteria decision analysis – MCDA, explainable machine learning, natural language processing.

Введение. Период пандемии ясно показал, что системы здравоохранения почти во всем мире сталкиваются с многочисленными проблемами, связанными с повышенным спросом на медицинские услуги, высокими ожиданиями населения и увеличивающимися расходами (Атун Р., 2015:2). При этом обнаруживается, что в системе здравоохранения важны не только экономическая, но и социальная, и медицинская эффективность, поскольку, как отмечается в (Орлов и др., 2010:6), «медицинские мероприятия лечебно-профилактического характера могут быть экономически невыгодными, но медико-социальный эффект требует их проведения». И это особенно справедливо в условиях пандемии. Однако пандемия COVID-19 является хорошим примером того, как слухи и неполнота знаний влияют на общество. Массмедиа для людей становится основным источником информации, и они ощущают неопределенность, когда в окружающей среде возникают угрозы (Болл-Рокич и др., 1976:18). В соответствии (Казахстан и COVID-19: «Делюкс Типография», 2021) пандемия спровоцировала всплеск слухов и дезинформации, которые препятствовали рациональному поведению населения и в некоторой степени способствовали ускорению распространения вируса. Во время пандемии COVID-19 сообщения СМИ существенно повлияли на эмоции людей и их психологическую устойчивость (Гири и др., 2021). В этот период более 51% заголовков новостей в англоязычных СМИ были негативными и только около 30% – позитивными (Аслам и др., 2020:8). Подаваемая в таком ключе информация может вызывать отрицательные эмоции у большого числа людей и может представлять угрозу для психики человека (Хамидин и др., 2020).

Материалы. В результате сопутствующего стресса страдает иммунитет и человек становится более восприимчивым к инфекционным заболеваниям. Значительную часть информации население получает с помощью электронных СМИ, поскольку большая часть населения планеты является пользователями Интернет, например, в Казахстане – практически все взрослое население (14,73 млн.) (Кемп, 2020). Поэтому оценка объективности и качества подачи материалов электрон-

ных СМИ в период пандемии позволяет понять, как СМИ реагируют на текущую ситуацию и на необходимые мероприятия в системе здравоохранения. Эта оценка может отражать качество подачи материалов, «эмоциональную перегретость» информации и может использоваться для коррекции публикуемых материалов с целью увеличения эмоционально-психологической устойчивости читателей в период серьезных социальных потрясений. Такую оценку можно провести с использованием одного из подразделов искусственного интеллекта (AI) - методов обработки естественного языка (natural language processing – NLP) (Мухамедиев и др., 2021:26) (рисунок 1):

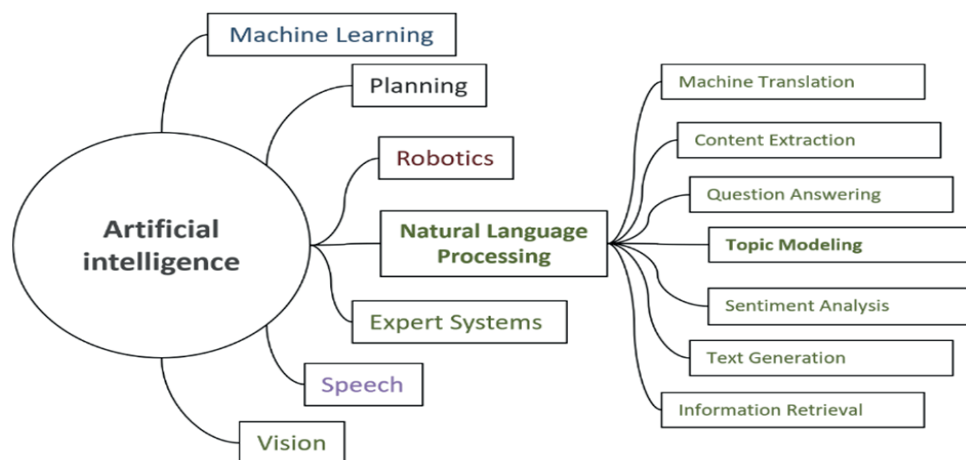


Рисунок 1. Области ИИ и НЛП

Одним из методов, продуктивно применяемых в области NLP, является тематический анализ или тематическое моделирование. Тематическое моделирование – метод, основанный на статистических характеристиках коллекций документов, который применяется в задачах автоматического реферирования, извлечения информации, информационного поиска и классификации (Машечкин и др., 2013:12). Смысл данного подхода заключается в интуитивном понимании того, что документы в коллекции образуют группы, в которых частота встречаемости слов или сочетаний слов различается.

Основой современных тематических моделей является статистическая модель естественного языка. Вероятностные тематические модели

описывают документы дискретным распределением на множестве тем, а темы – дискретным распределением на множестве терминов (Воронцов и др., 2012::13). Другими словами, тематическая модель определяет, к каким темам относится каждый документ и какие слова образуют каждую тему. Кластеры терминов и фраз, формируемые в процессе тематического моделирования, в частности, позволяют решать задачи синонимии и полисемии терминов (Пархоменко и др., 2017:139).

Недавние исследования предлагают использовать для тематического моделирования векторное представление слов (word embeddings), что позволяет учесть контекст, в котором используется тот или иной термин (Диенг и др., 2020:14).

Для построения тематической модели корпуса документов применяют: Вероятностный латентно-семантический анализ (PLSA), весьма популярное, латентное размещение Дирихле (LDA) (Блей и др., 2003:29), ARTM (Additive regularization of topic models) (Воронцов и др., 2015:11).

Иными словами, методы NLP можно использовать и в контуре обратной связи здравоохранения для оценки того, как СМИ и общество реагируют на предпринимаемые меры в чрезвычайных обстоятельствах.

При этом мы исходим из следующих предположений. Во-первых, СМИ влияет на общественное мнение, а, следовательно, те вопросы, которые освещаются в СМИ, начинают сильнее волновать общество. Во-вторых, СМИ, как и всякий бизнес, пытается предоставлять продукт (публикации) в соответствии с общественным спросом, и освещают те вопросы, которые волнует общество в большей степени. Поэтому в данной работе проведен анализ тематических кластеров публикаций СМИ Республики Казахстан по теме пандемии COVID-19 на основе автоматического скрапинга и использовании конфигурируемых регуляризаторов.

Метод анализа СМИ. Для решения поставленной задачи исследован основной аспект, связанный с представлением динамики публикационной активности СМИ в числовой форме: автоматически формируемые темы. Метод включает следующие основные шаги, продемонстрированные на рисунке 2:

- 1) Формирование корпуса документов с помощью автоматических систем сбора данных.

- 2) Создание иерархической тематической модели с применением методов, описанных в (Мухамедиев и др., 2020).

$$R(\varphi, \theta) = \sum_{i=1} \tau_i R_i(\varphi, \theta) \tag{2}$$

где n_{dw} количество вхождений слова w в документе d , φ_{wt} распределение слова в теме t , θ_{td} распределение темы t по документам d . Это слагаемое, $\sum_{i=1} \tau_i R_i(\varphi, \theta)$, является взвешенной линейной комбинацией регуляризаторов с неотрицательными весами τ_i . ARTM предлагает набор регуляризаторов, реализованных на основе дивергенции Кульбака-Лейблера, в данном случае демонстрирующих энтропийные различия между распределениями исходной матрицы $p'(w/d)$ и модели $p'(w/d)$: Сглаживающий регуляризатор, Уменьшающий регуляризатор, декоррелирующий регуляризатор.

Результаты и обсуждение. На основе библиотеки BigARTM (Воронцов и др., 2015:11), сформирована тематическая модель, состоящая из 200 кластеров документов, из которых экспертами было выбрано 12 тех, которые относятся к медицине. Эти 12 кластеров образовали субкорпус из 119956 документов, на котором было вновь выполнено тематическое моделирование с формированием 150 тематических групп из которых затем вновь отобраны 47 наиболее релевантных с точки зрения матрицы принадлежности (порог выше 0.05). Это позволило сформировать финальную модель на субкорпусе из 100481 документа. В каждый получившийся кластер помещены тексты с принадлежностью более 0.1. Итоговая тематическая модель с наиболее релевантными каждому кластеру словами показана на рисунке 3.

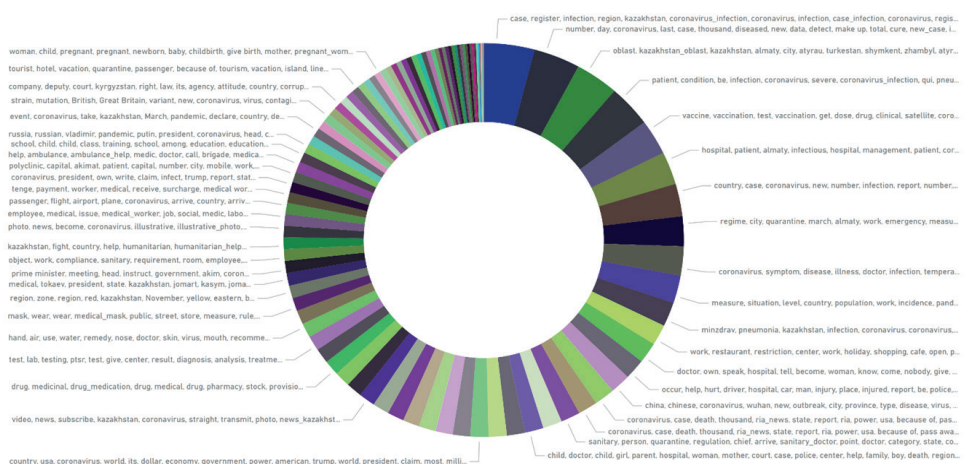


Рисунок 3. Тематические группы текстов, ранжированные по количеству документов

Количество тематических кластеров (топиков) и другие параметры тематической модели задаются вручную, обычно исходя из объективных метрик качества (Perplexity, Sparsity Score (Воронцов и др., 2015:11) и так называемого принципа плеча (Маруто и др., 2018:6). Соответственно, ввиду ограниченности количества топиков, некоторые субъективно важные темы не обязательно будут сформированы автоматически.

По этой причине, предлагается дополнительно использовать набор сформированных вручную поисковых запросов, которые позволяют проверить отдельные гипотезы о связи публикационной активности определенной тематики с эпидемиологической ситуацией.

В (Якунин и др., 2021:23) были предложены запросы, представлены ниже:

- Фальсификация, дезинформация, антивакцинация;
- Безработица, бедность;
- Кризис, экономический упадок;
- Голод, голод, бездомные, нищета;
- Удаленное образование;
- Фриланс, удаленная работа, утечка мозгов;
- Криминал, грабежи, кражи, убийства;
- Кризис, кредитование, долги, микрокредиты;
- Здоровоохранение, больницы, проблемы, скандалы в здравоохранении;
- Вакцинация, вакцины COVID-19.

В настоящей работе динамика публикационной активности анализируется по этому же списку запросов. Поиск по этим запросам (на русском языке) выполнялся с помощью ElasticSearch с использованием метода полнотекстового поиска, который возвращал список совпадающих документов с относительными весами релевантности.

Например, проанализирован тематический кластер «Заболееваемость, Школа, Ребенок, Рост, Эпидемиологический». Определено, что в 2020 году тема удалённого обучения для школьников активно поднималась только с началом учебного года (в сентябре).

Аналогичная ситуация с еще большей амплитудой наблюдалась в 2021 году. Можно предположить, что население в целом ожидало ослабления карантинных мер и переход на очное обучение, однако ухудшение эпидемиологической ситуации вызвало рост интереса к данной теме.

Рассмотрен пример анализа динамики по поисковому запросу «Вакцинация, Вакцины, Прививка, COVID». В то же время резкий

рост заболеваемости начала 2022 года, связанный с распространением омикрон-штамма коронавируса, напротив, не вызвал отклика в медиапространстве, а динамика публикационной активности продолжала снижение, в соответствии с общим трендом уменьшения интереса к теме COVID-19.

Далее проанализирована динамика публикационной активности по топику, отражающему доступность и цены лекарственных средств. Анализ показал, что, например, к весне 2021 постепенно снижается публикационная активность по этой теме, что может свидетельствовать как о стабилизации ситуации с поставками лекарственных средств, так о и снижении напряжённости в обществе по поводу поставок лекарственных препаратов во время пандемии.

При рассмотрении тематического кластера «Фейк, ложная информация, дезинформация» определено, что тема фейков в области здравоохранения продолжает оставаться актуальной, и в отличие от темы вакцинации, лекарственных средств и общей эпидемиологической обстановки, публикационная активность не затухает.

Для анализа динамики интереса СМИ к теме пандемии в целом был выбран самый крупный топик, прямо относящийся к COVID-19, и рассмотрена динамика его изменения с начала 2020 года по 23 февраля 2022 года. На рисунке 4 показана эта динамика.



Рисунок 4. График нормированной публикационной активности по топику «Случай, Инфекция, Коронавирусный, Коронавирусный_Инфекция, Коронавирус, Зарегистрировать».

По рисунку 4 видно, что интерес к COVID-19 стабильно падает. Так, если сравнить январь 2021 и январь 2022 года, интерес к данной теме упал примерно в два раза. Значение на оси ординат представляет

собой отношение суммы весов документов в этом топике к сумме всех весов по всем топикам за данный период. То есть, значение может интерпретироваться как доля данной темы в потоке информации. Видно, что в пике интереса в начале 2020 года единственный из топиков имеющий отношение к COVID-19 занимал около 8% всей информации в масс-медиа, а на начало 2022 года этот показатель упал до 1%. Цифра в один процент является значимой, но вполне сравнимой с другими темами. Например, тема искусственного интеллекта, оцененная подобным образом, составляет от 1 до 5 процентов. Если учесть все топика, связанные с COVID-19, то общая доля информации СМИ, касающаяся пандемии достигала в пике 10-15%.

Заключение. В работе на основе сформированного с помощью автоматического скрапинга корпуса документов по данным русскоязычных СМИ Казахстана относящихся к ведущим новостным изданиям страны выполнен тематический анализ публикаций относящихся к COVID-19. Мы использовали ARTM – модификацию латентного размещения Дирихле LDA, отличие которого от оригинала заключается в применении конфигурируемых регуляризаторов, которые позволяют настроить желаемый результат модели. Например, с его помощью можно уменьшить либо увеличить склонность модели к включению слова или документа сразу в несколько топиков; изменения склонности модели к большему либо меньшему количеству ненулевых весов в итоговой матрице и т.п. Проведенный анализ результатов показал изменение отношения общества к проблемам COVID-19 в период последних двух лет: появилась общая устойчивая тенденция снижения интереса электронных СМИ к теме пандемии на фоне постепенного увеличения интереса к другим аспектам общественной жизни.

Задачей ближайших исследований является разработка методики корреляционного анализа между информационными трендами в электронных СМИ Казахстана и показателями эпидемиологической ситуации по COVID-19 по данным Всемирной организации здравоохранения (ВОЗ).

Благодарности. Данная публикация профинансирована Комитетом по науке Министерства образования и науки Республики Казахстан, грант № AP09259587 "Разработка методов и алгоритмов интеллектуальной ГИС для многокритериального анализа данных здравоохранения".

Information about authors:

Yakunin Kirill – PhD, senior researcher of the Institute of Information and Computational Technologies of the Ministry of Education and Science of the Republic of Kazakhstan, yakunin.k@mail.ru, <https://orcid.org/0000-0002-7378-9212>;

Mukhamediev Ravil – PhD, professor, K.I. Satpayev KazNITU, professor, ravil.muhamedyev@gmail.com, <https://orcid.org/0000-0002-3727-043X>;

Symagulov Adilkhan – PhD student, software engineer, ISRT KN MES RK; asmogulove00@gmail.com; ORCID ID: <https://orcid.org/0000-0001-9974-3215>;

Kuchin Yan – PhD student, senior researcher of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan; ykuchin@mail.ru; ORCID ID: <https://orcid.org/0000-0002-5271-9071>;

Yunicheva Nadiya – PhD, leading researcher of the Institute of High Technologies of the National Academy of Sciences of the Republic of Kazakhstan; naduni@mail.ru; ORCID ID: <https://orcid.org/0000-0001-6351-3450>;

Yelis Marina – PhD student, junior researcher of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan; k.marina92@gmail.com; ORCID ID: <https://orcid.org/0000-0003-4203-800X>;

Mukhamedieva Elena – is a researcher at the Institute of Information and Computational Technologies of the Ministry of Education and Science of the Republic of Kazakhstan, muhamediyeva@gmail.com, ORCID ID: <https://orcid.org/0000-0001-9596-4432>.

ЛИТЕРАТУРА:

Агун Р. (2015) Переход систем здравоохранения на мультиморбидность. Ланцет 386: 721-722.

Орлов Е.М. (2010) Категория эффективности в системе здравоохранения. Фундаментальные исследования 4:70-75. (На русском).

Болл-Рокич С.Дж., Дефлер М. (1976) Модель зависимости эффектов средств массовой информации. Коммуникационные исследования, 3(1): 3–21. <https://doi.org/10.1177/009365027600300101>.

«ДЕЛЮКС Типография». (2021) КАЗАХСТАН И COVID-19: СМИ, КУЛЬТУРА, ПОЛИТИКА, <http://library.fes.de/pdf-files/bueros/kasachstan/18218.pdf>.

Гири С.П., Маурья А.К. (2021) Забытая реальность средств массовой информации во время COVID-19: влияние новостей о пандемии на положительные и отрицательные

эмоции и психологическую устойчивость человека. Личность и индивидуальные различия, 180:110962.

Аслам Ф., Аван Т.М., Сайед Дж.Х., Кашиф А., Парвин М. (2020) Чувства и эмоции, вызванные заголовками новостей о вспышке коронавирусной болезни (COVID-19). Коммуникации гуманитарных и социальных наук, 7: 1-9.

Хамидин З., Хатам Дж., Резапур Т. (2020) Как люди эмоционально реагируют на новости о COVID-19: онлайн-опрос. Базовая и клиническая неврология, 11:171.

Кемп С. (2020) Digital 2020: Казахстан. www.datareportal.com. Дата обращения:16.10.2020.

Мухамедиев Р.И., Сымагулов А., Кучин Ю., Якунин К., Елис М. (2021) От классического машинного обучения к глубоким нейронным сетям: упрощенный наукометрический обзор. Прикладные науки, 11(12):5541, <https://doi.org/10.3390/app1125541>.

Машечкин И.В., Петровский М.И., Царев Д.В. (2013) Методы расчета релевантности текстовых фрагментов на основе тематических моделей в задаче автоматического аннотирования. Вычислительные методы и программирование, 14(1):91-102.

Воронцов К.В., Потапенко А.А. (2012) Регуляризация, надежность и разреженность вероятностных тематических моделей. Компьютерные исследования и моделирование, 4(4):693-706.

Пархоменко П.А., Григорьев А.А., Астраханцев Н.А. (2017) Обзор и экспериментальное сравнение методов кластеризации текста. Труды ИСП РАН, 29(2):161-200. DOI: 10.15514/ИСПР РАН-2017-29(2)-6.

Диенг А.Б., Руис Ф.Дж., Блей Д.М. (2020) Тематическое моделирование в пространстве вложений. Труды Ассоциации компьютерной лингвистики, 8:439-453.

Блей Д., Нг Э.Й., Джордан М. (2003) Скрытое распределение Дирихле. Журнал исследований машинного обучения, 3(1):993-1022.

Воронцов К. и соавт. (2015) Bigartm: библиотека с открытым исходным кодом для регуляризованного мультимодального тематического моделирования больших коллекций. Международная конференция по анализу изображений, социальных сетей и текстов, Springer, Cham. С. 370-381.

Мухамедиев Р.И., Якунин К., Мусабаев Р., Булдыбаев Т., Кучин Ю., Мурзахметов С., Елис М. (2020) Классификация негативной информации на общественно значимые темы в СМИ. Симметрия, 12(12):1945.

Якунин К., Калимолдаев М., Мухамедиев Р.И., Мусабаев Р., Баракнин В., Кучин Ю., Мурзахметов С., Булдыбаев Т., Оспанова Ю., Елис М. (2021) KazNewsDataset: Единый национальный корпус цифровых СМИ, 6:31. <https://doi.org/10.3390/data6030031>.

Джелодар Х. (2018) Скрытое распределение Дирихле (LDA) и тематическое моделирование: модели, приложения, обзор. Мультимедийные инструменты и приложения. 78(5): 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>.

Воронцов К., Фрей О., Апишев М., Ромов П., Дударенко М. (2015) BigARTM: библиотека с открытым исходным кодом для регуляризованного мультимодального тематического моделирования больших коллекций. На Международной конференции по анализу изображений, Soc. Сети и тексты, Springer: Cham, Швейцария. стр. 370-381. https://doi.org/10.1007/978-3-319-26123-2_36.

Маруто Д., Хандака С.Х., Виджая Э. (2018) Определение числа кластеров по

к-среднему с использованием метода локтя и оценка чистоты в заголовках новостей. Международный семинар по применению информационных и коммуникационных технологий 2018 г., стр. 533–538, <https://doi.org/10.1109/ISEMANTIC.2018.8549751>.

Якунин К., Мухамедиев Р.И., Зайцева Е., Левашенко В., Елис М., Сымагулов А., Кучин Ю., Мухамедиева Е., Аубакиров М., Гопеженко В. (2021) СМИ как зеркало пандемии COVID-19. Вычисления. 9(12):140. <https://doi.org/10.3390/computation9120140>.

REFERENCES:

Atun R. (2015) Transitioning health systems for multimorbidity. *The Lancet* 386:721-722.

Orlov E.M. (2010) The category of effectiveness in the health care system. *Basic research* 4:70-75. (In Russian).

Ball-Rokeach S.J., DeFleur M.L. (1976) A dependency model of mass-media effects. *Communication Research*, 3(1): 3–21. <https://doi.org/10.1177/009365027600300101>.

«DELUXE Printery». (2021) KAZAKHSTAN AND COVID-19: MEDIA, CULTURE, POLITICS, <http://library.fes.de/pdf-files/bueros/kasachstan/18218.pdf>.

Giri S.P., Maurya A.K. (2021) A neglected reality of mass media during COVID-19: Effect of pandemic news on individual's positive and negative emotion and psychological resilience. *Personality and Individual Differences*, 180:110962.

Aslam F., Awan T.M., Syed J.H., Kashif A., Parveen M. (2020) Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, 7:1-9.

Hamidein Z., Hatam J., Rezapour T. (2020) How people emotionally respond to the news on COVID-19: An online survey. *Basic and Clinical Neuroscience*, 11:171.

Kemp S. (2020) Digital 2020: Kazakhstan. www.datareportal.com. Date of access:16.10.2020.

Mukhamediev R.I., Symagulov A., Kuchin Y., Yakunin K., Yelis M. (2021) From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review. *Applied Sciences*, 11(12):5541, <https://doi.org/10.3390/app11125541>.

Mashechkin I.V., Petrovsky M.I., Tsarev D.V. (2013) Methods for calculating the relevance of text fragments based on thematic models in the problem of automatic annotation. *Computational methods and programming*, 14(1):91-102.

Vorontsov K.V., Potapenko A.A. (2012) Regularization, robustness and sparseness of probabilistic thematic models. *Computer research and modeling*, 4(4):693-706.

Parkhomenko P.A., Grigoriev A.A., Astrakhantsev N.A. (2017) Review and experimental comparison of text clustering methods. *Proceedings of ISP RAS*, 29(2):161-200. DOI: 10.15514/ISPRAS-2017-29(2)-6.

Dieng A.B., Ruiz F.J., Blei D.M. (2020) Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439-453.

Blei D.M., Ng A.Y., Jordan M.I. (2003) Latent dirichlet allocation. *Journal of machine Learning research*, 3(1):993-1022.

Vorontsov K. et al. (2015) Bigartm: Open source library for regularized multimodal topic modeling of large collections. *International Conference on Analysis of Images, Social Networks and Texts*, Springer, Cham. P. 370-381.

Mukhamediev R.I., Yakunin K., Mussabayev R., Buldybayev T., Kuchin Y., Murzakhmetov S., Yelis M. (2020) Classification of Negative Information on Socially Significant Topics in Mass Media. *Symmetry*, 12(12):1945.

Yakunin K., Kalimoldayev M., Mukhamediev R.I., Mussabayev R., Barakhnin V., Kuchin Y., Murzakhmetov S., Buldybayev T., Ospanova U., Yelis M. (2021) KazNewsDataset: Single Country Overall Digital Mass Media Publication Corpus, 6:31. <https://doi.org/10.3390/data6030031>.

Jelodar H. (2018) Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications*. 78(5): 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>.

Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. (2015) BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. In *International Conference on Analysis of Images, Soc. Networks and Texts*, Springer: Cham, Switzerland. pp.370-381. https://doi.org/10.1007/978-3-319-26123-2_36.

Marutho D., Handaka S.H., Wijaya E. (2018) The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. "2018 International Seminar on Application for Technology of Information and Communication. pp. 533-538, <https://doi.org/10.1109/ISEMANTIC.2018.8549751>.

Yakunin K., Mukhamediev R.I., Zaitseva E., Levashenko V., Yelis M., Symagulov A., Kuchin Y., Muhamedijeve E., Aubakirov M., Gopejenko V. (2021) Mass Media as a Mirror of the COVID-19 Pandemic. *Computation*. 9(12):140. <https://doi.org/10.3390/computation9120140>.

МАЗМҰНЫ

А.С.Ақанова, А.А.Макашев, С.А. Наурызбаева, Н.Н.Оспанова ИНТЕРНЕТТЕН ТАҚЫРЫП БОЙЫНША ДЕРЕКТЕРДІ АЛУЫН МОДЕЛДЕУ.....	5
Ж.С. Авкурова, С.А. Гнатюк, Б.К. Абдураимова, Л.М. Кыдыралина КИБЕРКЕҢІСТІКТЕГІ АРТ-ШАБУЫЛДАРДЫ ЕРТЕ АНЫҚТАУ ЖӘНЕ БҰЗУШЫЛАРДЫ СӘЙКЕСТЕНДІРУ ҮШІН ЭТАЛОН МОДЕЛЬДЕРІ АНЫҚТАУШЫ ЕРЕЖЕЛЕР.....	19
М.А. Болатбек, К.Б. Багитова, Ш.Ж. Мусиралиева КИБЕРҚАУІПСІЗДІК МӘСЕЛЕЛЕРІН ТАБИҒИ ТІЛДІ ӨНДЕУ ӘДІСТЕРІ АРҚЫЛЫ ШЕШУ ТАҚЫРЫБЫНА ЖҮЙЕЛІК ШОЛУ.....	52
А.К. Жумадиллаева, М.Д. Кабибуллин, Б.Б. Оразбаев, К.Н. Оразбаева, Ж.Н. Тулеуов КАТАЛИТИКАЛЫҚ РИФОРМИНГ ҚОНДЫРҒЫСЫ РИФОРМИНГТЕУ РЕАКТОРЛАРЫ ЖҰМЫС РЕЖИМДЕРІН КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ НЕГІЗІНДЕ ОПТИМИЗАЦИЯЛАУ.....	71
Ж.Д. Изтаев, Г.Т. Джусупбекова, Г.К. Ордабаева УНИВЕРСИТЕТ ҮШІН АҚПАРАТТЫҚ ҚАУІПСІЗДІК ҚАТЕРЛЕРІНІҢ ЖЕКЕ МОДЕЛІН ӨЗІРЛЕУ.....	91
Ж.С. Каженова, Ж.Е. Кенжебаева, А.М. Прудник MQTT (ТЕЛЕМЕТРИЯ ХАБАРЛАМАЛАРЫ КЕЗЕГІН ТАСЫМАЛДАУ) ХАТТАМАСЫНЫҢ ҚАУІПСІЗДІК МЕХАНИЗМДЕРІ.....	117
А.Ж. Картбаев, Г.С. Ыбытаева, О.Ж. Мамырбаев, К.Ж. Мухсина, Б.Ж. Жумажанов АВТОМАТТЫ ҚЫЛМЫС ОНТОЛОГИЯСЫН ҚҰРУ ҮШІН ҚЫЛМЫС ЖАҒАЛЫҚТАРЫНДА СУБЪЕКТИЛЕРДІ ФОРМАЛЬДЫ КӨРСЕТУ ӘДІСТЕРІ.....	136
А.Т. Мазақова, Қ.Б. Бегалиева, Т.Ж. Мазаков, Ш.А. Жомартова, Г.З. Зиятбекова КВАДРАТ ҚИМАСЫ БАР ӨЗЕКШЕНІҢ ЖЫЛУ ӨТКІЗГІШТІК ТЕҢДЕУІН ҚАРАПАЙЫМ ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕР ЖҮЙЕСІНЕ ҚОЮ АРҚЫЛЫ ШЕШУ.....	153

Ж.Ж. Молдашева, Б.Б. Оразбаев, Б.У. Асанова, С.Ш. Исакова, К.Н. Оразбаева МҰНАЙ ҚҰБЫРЫ АГРЕГАТТАРЫНЫҢ ЖҰМЫС РЕЖИМДЕРІН БАСҚАРУ ҮШІН ЭВРИСТИКАЛЫҚ ТӘСІЛ ҚҰРУ.....	164
А.Б. Мименбаева, А.С. Аканова СОЛТҮСТІК ҚАЗАҚСТАН ОБЛЫСЫНЫҢ АУЫЛШАРУАШЫЛЫҒЫ ДАҚЫЛДАРЫНЫҢ КҮЙІН NDVI СЫЗЫҚТЫҚ ТРЕНДТЕРІ АРҚЫЛЫ ЗЕРТТЕУ.....	185
М.О. Ногайбаева, Б. Ахметов, Дж.Дж. Расулзаде, Е.А. Максум, С. Рустамов U-NET КОНВОЛЮЦИЯЛЫҚ НЕЙРОНДЫҚ ЖЕЛІ НЕГІЗІНДЕ ТОПОЛОГИЯЛЫҚ ОҢТАЙЛАНДЫРУДЫҢ ЕСЕПТЕУ ПРОЦЕСІН ЖЕДЕЛДЕТУ.....	198
Г.Б. Туребаева, А.К. Сыздықов, А.Р. Тенчурина, Ж.Б. Дошакова ҚОЛДАНБАЛЫ БАҒДАРЛАМАЛАРДЫ ҚОЛДАНА ОТЫРЫП ДИФФЕРЕНЦИАЛДЫҚ ТЕНДЕУЛЕРДІ ШЕШУДІҢ САҢДЫҚ ӘДІСТЕРІ.....	214
К.С. Чезимбаева, А.Н. Хайруллина LORA ҚАБЫЛДАҒЫШ/ТАРАТҰЫШЫНЫҢ ӨНІМДІЛІГІН БАҒАЛАУ.....	228
А.Г. Шаушенова, А.А. Нурпейсова, Ж.С. Муталова, Д.Б. Досалянов, М.Б. Онгарбаева ҚАШЫҚТЫҚТАН ОҚЫТУДА БІЛІМ АЛУШЫНЫ ИДЕНТИФИКАЦИЯЛАУ ЖӘНЕ БЕЙНЕМОНИТОРИНГТЕУ ШЕТЕЛДІК ЖҮЙЕЛЕРІНІҢ ЕРЕКШЕЛІКТЕРІ.....	247
К. Якунин, Р.И. Мухамедиев, М. Елис, Я. Кучин, Н. Юничева, А. Сымагулов, Е. Мухамедиева КОВИД-19 ПАНДЕМИЯСЫ ТАҚЫРЫП БОЙЫНША ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БАҚ БАСЫЛЫМДАРЫНЫҢ ТАҚЫРЫПТЫҚ КЛАСТЕРЛЕРІН ТАЛДАУ.....	260

СОДЕРЖАНИЕ

А.С. Аканова, А.А. Макашев, С.А. Наурызбаева, Н.Н. Оспанова МОДЕЛИРОВАНИЕ ТЕМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ДАННЫХ ИЗ ИНТЕРНЕТА.....	5
Ж.С. Авкурова, С.А. Гнатюк, Б.К. Абдураимова, Л.М. Кыдыралина МОДЕЛИ ЭТАЛОНОВ И ОПРЕДЕЛЯЮЩИЕ ПРАВИЛА ДЛЯ СИСТЕМРАННЕГО ВЫЯВЛЕНИЯ АРТ-АТАКИ ИДЕНТИФИКАЦИИ НАРУШИТЕЛЕЙ В КИБЕРПРОСТРАНСТВЕ.....	19
М.А. Болатбек, К.Б. Багитова, Ш.Ж. Мусиралиева СИСТЕМАТИЧЕСКИЙ ОБЗОР ТЕМЫ РЕШЕНИЯ ЗАДАЧ КИБЕРБЕЗОПАСНОСТИ С ПОМОЩЬЮ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА.....	52
А.К. Жумадиллаева, М.Д. Кабибуллин, Б.Б. Оразбаев, К.Н. Оразбаева, Ж.Н. Тулеуов ОПТИМИЗАЦИЯ РЕЖИМОВ РАБОТЫ РЕАКТОРОВ РИФОРМИНГА УСТАНОВКИ КАТАЛИТИЧЕСКОГО РИФОРМИНГА НА ОСНОВЕ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ.....	71
Ж.Д. Изтаев, Г.Т. Джусупбекова, Г.К. Ордабаева РАЗРАБОТКА ЧАСТНОЙ МОДЕЛИ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ДЛЯ УНИВЕРСИТЕТА.....	91
Ж.С. Каженова, Ж.Е. Кенжебаева, А.М. Прудник МЕХАНИЗМЫ БЕЗОПАСНОСТИ ПРОТОКОЛА MQTT (ТРАНСПОРТ ТЕЛЕМЕТРИИ ОЧЕРЕДИ СООБЩЕНИЙ).....	117
А.Ж. Картбаев, Г.С. Ыбыгаева, О.Ж. Мамырбаев, К.Ж. Мухсина, Б.Ж. Жумажанов МЕТОДЫ ФОРМАЛЬНОГО ПРЕДСТАВЛЕНИЯ СУЩНОСТЕЙ В КРИМИНАЛЬНЫХ НОВОСТЯХ ДЛЯ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ОНТОЛОГИИ ПРЕСТУПЛЕНИЙ.....	136
А.Т. Мазакова, К.Б. Бегалиева, Т.Ж. Мазаков, Ш.А. Жомартова, Г.З. Зиятбекова РЕШЕНИЕ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ СТЕРЖНЯ С КВАДРАТНЫМ СЕЧЕНИЕМ ПРИВИДЕНИЕМ К СИСТЕМЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ.....	153

Ж.Ж. Молдашева, Б.Б. Оразбаев, Б.У. Асанова, С.Ш. Искакова, К.Н. Оразбаева РАЗРАБОТКА ЭВРИСТИЧЕСКОГО МЕТОДА ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ УПРАВЛЕНИЯ РЕЖИМАМИ РАБОТЫ АГРЕГАТОВ НЕФТЕПРОВОДА.....	164
А.Б. Мименбаева, А.С. Аканова ИССЛЕДОВАНИЕ СОСТОЯНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР СЕВЕРО-КАЗАХСТАНСКОЙ ОБЛАСТИ ПО ЛИНЕЙНЫМ ТРЕНДАМ NDVI.....	185
М.О. Ногайбаева, Б. Ахметов, Дж.Дж. Расулзаде, Е.А. Максум, С. Рустамов УСКОРЕНИЕ ВЫЧИСЛИТЕЛЬНОГО ПРОЦЕССА ТОПОЛОГИЧЕСКОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ U-NET.....	198
Г.Б. Туребаева, А.К. Сыздыков, А.Р. Тенчурина, Ж.Б. Дошаков ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ИСПОЛЬЗОВАНИЕМ ПРИКЛАДНЫХ ПРОГРАММ.....	214
К.С. Чежимбаева, А.Н. Хайруллина ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРИЕМОПЕРЕДАТЧИКА LORA.....	228
А.Г. Шаушенова, А.А. Нурпейсова, Ж.С. Муталова, Д.Б. Досалянов, М.Б. Онгарбаева ОСОБЕННОСТИ ЗАРУБЕЖНЫХ СИСТЕМ ВИДЕОМОНИТОРИНГА И ИДЕНТИФИКАЦИИ ОБУЧАЮЩЕГОСЯ В ДИСТАНЦИОННОМ ОБУЧЕНИИ.....	247
К. Якунин, Р.И. Мухамедиев, М. Елис, Я. Кучин, А. Сымагулов, Н. Юничева, Е. Мухамедиева АНАЛИЗ ТЕМАТИЧЕСКИХ КЛАСТЕРОВ ПУБЛИКАЦИЙ СМИ РЕСПУБЛИКИ КАЗАХСТАН ПО ТЕМЕ ПАНДЕМИИ COVID-19.....	260

CONTENTS

A.S. Akanova, A.A. Makashev, C.A. Наурызбаева, N.N. Ospanova MODELING OF THEMATIC DATA EXTRACTION FROM THE INTERNET.....	5
Zh. Avkurova, S. Gnatyuk, B. Abduraimova, L. Kydyralina MODELS OF STANDARDS AND GOVERNING RULES FOR THE SYSTEMS OF EARLY DETECTION OF APT-ATTACKS AND IDENTIFICATION OF VIOLATORS IN CYBERSPACE.....	19
M. Bolatbek, K. Bagitova, Sh. Musiralieva A SYSTEMATIC REVIEW ON CYBERSECURITY ISSUES USING NATURAL LANGUAGE PROCESSING TECHNIQUES.....	52
A. Zhumadillayeva, M. Kabibullin, B. Orazbayev, K. Orazbayeva, Zh. Tuleuov OPTIMIZATION OF THE OPERATING MODES OF THE REFORMING REACTORS OF THE CATALYTIC REFORMING UNIT BASED ON COMPUTER MODELING.....	71
Zh.D. Iztayev, G.T. Dzhusupbekova, G.K. Ordabaeva DEVELOPMENT OF A PRIVATE MODEL OF INFORMATION SECURITY THREATS FOR THE UNIVERSITY.....	91
Zh.S. Kazhenova, Zh.E. Kenzhebayeva, A.M. Prudnik SECURITY MECHANISMS OF PROTOCOL MQTT (MESSAGE QUEUEING TELEMETRY TRANSPORT).....	117
A.Zh. Kartbayev, G.S. Ybytayeva, O.Zh. Mamyrbayev, K.Zh. Mukhsina, B.Zh. Zhumazhanov METHODS FOR FORMAL REPRESENTATION OF ENTITIES IN CRIME NEWS FOR AUTOMATIC CRIME ONTOLOGY CONSTRUCTION.....	136
A.T. Mazakova, K.B. Begaliyeva, T.Zh. Mazakov, Sh.A. Jomartova, G.Z. Ziyatbekova SOLUTION OF THE THERMAL CONDUCTIVITY EQUATION OF A ROD WITH A SQUARE SECTION BY CASTING TO A SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS.....	153

Zh. Moldasheva, B. Orazbayev, B. Assanova, Sh. Iskakova, K. Orazbayeva OPTIMIZATION OF OPERATION MODES OF REFORMING REACTORS OF A CATALYTIC REFORMING UNIT ON THE BASIS OF COMPUTER MODELING.....	164
A.B. Mimenbayeva, A.C. Akanova RESEARCH OF THE STATE OF AGRICULTURAL CROPS NORTH KAZAKHSTAN REGION ACCORDING TO LINEAR NDVI TRENDS.....	185
M. Nogaibayeva, B. Akhmetov, J. Rasulzade, Y. Maksim, S. Rustamov ACCELERATION OF THE COMPUTATIONAL PROCESS OF TOPOLOGICAL OPTIMIZATION BASED ON THE CONVOLUTIONAL NEURAL NETWORK U-NET.....	198
G. Turebaeva, A. Syzdykov, A. Tenchurina, J. Doshakov NUMERICAL METHODS FOR SOLVING DIFFERENTIAL EQUATIONS USING APPLICATION PROGRAMS.....	214
K.S. Chezimbayeva, A.N. Khairullina EVALUATION OF LORA TRANSCEIVER PERFORMANCE.....	228
A.G. Shaushenova, A.A. Nurpeisova, Z.S. Mutalova, D.B. Dosalyanov, M.B. Ongarbaeva FEATURES OF FOREIGN SYSTEMS OF VIDEO MONITORING AND IDENTIFICATION OF STUDENTS IN DISTANCE LEARNING.....	247
K. Yakunin, R.I. Mukhamediev, M. Elis, Ya. Kuchin, N. Yunicheva, A. Symagulov, E. Mukhamedieva ANALYSIS OF THEMATIC CLUSTERS OF KAZAKHSTAN MEDIA PUBLICATIONS ON THE TOPIC OF THE COVID-19 PANDEMIC.....	260

**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*

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

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

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 15.09.2022.

Формат 60x88/8. Бумага офсетная. Печать – ризограф.

17,5 п.л. Тираж 300. Заказ 3.