

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

2 (342)

APRIL – JUNE 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 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:

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

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

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

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

«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: *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/>

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

ГЛАВНЫЙ РЕДАКТОР:

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

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

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

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**

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

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

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

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

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

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

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

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

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

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

«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 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/>

© Национальная академия наук Республики Казахстан, 2022
Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

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.

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 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/>

© 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 2, Number 342 (2022),5–18
<https://doi.org/10.32014/2022.2518-1726.126>

UDC 338(575.2); 340.13

T.I. Ganieva¹, N.S. Semenov^{2*}, S.R. Semenov²

¹Kyrgyz National University named after J. Balasagyn,
Bishkek, Kyrgyzstan;

²International University of Kyrgyz Republic, Bishkek, Kyrgyzstan.
E-mail: frindland@mail.ru

CYBERSECURITY OF INFORMATION RELATIONS IN THE FIELD OF INFORMATION INFRASTRUCTURE OF A GLOBAL SOCIETY

Abstract. The article deals with cybersecurity issues in the field of information relations, with an analysis of the legislation of the countries of the European Union (EU), Japan and the states of the Eurasian Economic Union (EAEU). Information relations in the world are becoming one of the important components of the interaction of various subjects of law, where, in the practical implementation of relationships, legal issues arise in the field of data security. The information environment, expressed in the form of the Internet space, is a key factor that is based on the information infrastructure. The rapid development of information infrastructure gives impetus to the process of the emergence of new legal norms, institutions, branches in the field of law, which makes it possible to ensure legal regulation of cybersecurity both at the national and international levels, taking into account the legislation of certain countries included in the integration associations.

The purpose of the work is to analyze and formulate proposals for the further development of new directions in the field of cybersecurity based on the development of a common information infrastructure based on the development of legal, economic relations aimed at ensuring data security.

Cybersecurity, as an area of information technology protection, is intended to regulate information and communication technologies in the

field of protection and security of the global network and its users; this area is gaining relevance and becomes a concentrated reflection of the ongoing globalization.

Research on the development of information relations contributes to the formation of information legislation, including the EAEU member states, which create a common structure for the exchange of legal and economic information. Establishing a mechanism for international control over the safety of digital processes is becoming a priority area for the development of information and communication technologies, Internet networks, ensuring the integration interaction of states and a number of countries in building a global infrastructure.

Key words: Cybersecurity, information relations, information space, information infrastructure, EU, EAEU, strategy, concept, legislation, economics.

Т.И. Ганиева¹, Н.С. Семенов^{2*}, С.Р. Семенов²

¹Ж. Баласағын атындағы Қырғыз ұлттық университеті,
Бішкек, Қырғызстан;
²Қырғыз Республикасының Халықаралық университеті,
Бішкек, Қырғызстан.
E-mail: frindland@mail.ru

ЖАҒАНДЫҚ ҚОҒАМНЫҢ АҚПАРАТТЫҚ ИНФРАҚҰРЫЛЫМЫ САЛАСЫНДАҒЫ АҚПАРАТТЫҚ ҚАТЫНАСТАРДЫҢ КИБЕРҚАУІПСІЗДІГІ

Аннотация. Мақалада Еуропалық Одақ (ЕО) елдерінің, Жапонияның және Еуразиялық экономикалық одақ (ЕАЭО) мемлекеттерінің заңнамасын талдай отырып, ақпараттық қатынастар саласындағы киберқауіпсіздік мәселелері қарастырылады. Әлемдегі ақпараттық қатынастар әртүрлі құқық субъектілерінің өзара әрекеттесуінің маңызды құрамдас бөліктерінің біріне айналады, онда қатынастарды іс жүзінде жүзеге асыру кезінде мәліметтер қауіпсіздігі саласында құқықтық мәселелер туындайды. Интернет кеңістігі түрінде көрсетілген ақпараттық орта ақпараттық инфрақұрылымға негізделген негізгі фактор болып табылады. Ақпараттық инфрақұрылымның қарқынды дамуы құқық саласындағы жаңа құқықтық нормалардың, институттардың, салалардың пайда болу процесіне серпін береді, бұл интеграциялық бірлестіктерге кіретін жекелеген елдердің заңна-

масын ескере отырып, ұлттық және халықаралық деңгейлерде киберқауіпсіздікті құқықтық реттеуді қамтамасыз етуге мүмкіндік береді.

Жұмыстың мақсаты – деректер қауіпсіздігін қамтамасыз етуге бағытталған құқықтық, экономикалық қатынастарды дамыту негізінде, жалпы ақпараттық инфрақұрылымды дамыту негізінде киберқауіпсіздік саласындағы жаңа бағыттарды одан әрі дамыту бойынша ұсыныстарды талдау және тұжырымдау. Киберқауіпсіздік, Ақпараттық технологияларды қорғау саласы ретінде, ғаламдық желіні және оны пайдаланушылардың қауіпсіздігі мен ақпараттық-коммуникациялық технологияларды реттеуге арналған; бұл сала өзекті болып, жалғасып жатқан жаһандандудың шоғырланған көрінісіне айналады.

Ақпараттық қатынастарды дамыту жөніндегі зерттеулер құқықтық және экономикалық ақпарат алмасу үшін ортақ құрылым құратын ақпараттық заңнаманы, оның ішінде ЕАЭО-ға мүше мемлекеттерді қалыптастыруға ықпал етеді. Цифрлық процестердің қауіпсіздігін халықаралық бақылау тетігін құру ақпараттық-коммуникациялық технологияларды, интернет желілерін дамытудың, жаһандық инфрақұрылымды құруда мемлекеттер мен бірқатар елдердің интеграциялық өзара іс-қимылын қамтамасыз етудің басым бағыты болып отыр.

Түйін сөздер: киберқауіпсіздік, ақпараттық қатынастар, ақпараттық кеңістік, ақпараттық инфрақұрылым, ЕО, ЕАЭО, стратегия, тұжырымдама, заңнама, экономика.

Т.И. Ганиева¹, Н.С. Семенов^{2*}, С.Р. Семенов²,

¹Кыргызский Национальный Университет им. Ж. Баласагына,
Бишкек, Кыргызстан;

²Международный Университет Кыргызской Республики,
Бишкек, Кыргызстан.

E-mail: frindland@mail.ru

КИБЕРБЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ОТНОШЕНИЙ В СФЕРЕ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ ГЛОБАЛЬНОГО ОБЩЕСТВА

Аннотация. В статье рассматриваются вопросы кибербезопасности в области информационных отношений, с проведением анализа законодательства стран Европейского союза (ЕС), Японии и государств Евразийского экономического союза (ЕАЭС). Информационные отношения в мире становятся одним из важных компонентов

взаимодействия различных субъектов права, где при практической реализации взаимоотношений возникают правовые моменты в области обеспечения безопасности данных. Информационная среда, выраженная в виде интернет-пространств является ключевым фактором, который основан на информационной инфраструктуре. Быстрое развитие информационной инфраструктуры дает толчок процессу появления новых правовых норм, институтов, отраслей в области права, что дает возможность обеспечить правовое регулирование кибербезопасности, как на национальном, так и на международном уровнях, с учетом законодательства тех или иных стран входящих в интеграционные объединения.

Цель работы - на основе развития правовых, экономических отношений направленных на обеспечения безопасности данных дать анализ и сформировать предложения для дальнейшего развития новых направлений в области кибербезопасности, основанных на развитии общей информационной инфраструктуры.

Кибербезопасность, как область защиты информационных технологий предназначена для регулирования информационных и коммуникационных технологий в области защиты и безопасности глобальной сети и ее пользователей, это направление приобретает актуальность и становится концентрированным отражением происходящей глобализации.

Исследование по развитию информационных отношений способствует формированию информационного законодательства, в том числе стран-участниц ЕАЭС, которые создают общую структуру обмена правовой и экономической информации. Установления механизма международного контроля за безопасностью цифровых процессов, становится приоритетным направлением развития информационно-коммуникационных технологий, интернет сетей, обеспечивая интеграционное взаимодействие государств и ряда стран в вопросах построения глобальной инфраструктуры.

Ключевые слова: Кибербезопасность, информационные отношения, информационное пространство, информационная инфраструктура, ЕС, ЕАЭС, стратегия, концепция, законодательство, экономика.

Introduction. The development of the information sphere requires the solution of general issues of information security, since the development of the Internet based on technologies of open standards of networks has turned into a developed international commercial infrastructure with extensive information exchange. The basic technology of the Internet structure is

becoming the main factor in the dissemination of information in the global information space. Cyberspace involves more and more new users and new network management capabilities, including the use of the “Internet of Things” systems, that is, the involvement of technically managed objects. Since the Internet web servers contain a variety of information and distribute information resources for the documents sought, the protection of content resources from malicious software becomes the main direction of the fight against cybercrime and ensures the cybersecurity of information relations in information storage and exchange systems.

The development of integration processes in the field of digital transformation of the EAEU member states sets certain tasks in the fight against cybercrime, threats in the information infrastructure are currently becoming relevant for the national and economic security of states. The overall coordination of states in the direction of the development and adoption of appropriate legal documents protecting state structures in the field of information relations, ensuring the security of economic, financial, social, political information resources and the consequences of cybercrime actions should be aimed at improving and interacting with common state security structures. Studying the experience of countries in the field of cybersecurity will help avoid the risks of losing its own information infrastructure and develop legal methods to prevent cybercrime.

Research materials and methods. The research materials were legislation and economic strategies in the field of information relations of a number of countries, including the European Union, Japan and the EAEU member countries in the field of information relations development.

The purpose of the work is a legal and economic analysis of interaction in the study of cybersecurity in information relations in some advanced countries, including the EAEU. An opportunity to assess the regulatory sphere and the current situation, identify gaps and develop recommendations for filling them. Assess the principles of data protection and their proportionality in matters of cybersecurity, taking into account the global crisis and the spread of COVID-19, to develop a number of proposals that can improve the work of the EAEU.

Information relations in the global information space depend on information resources, which are the basis for the exchange, storage and creation of information products. The protection of information products, including information databases, taking into account the global crisis phenomena, is becoming the main criterion for cyber security.

The methodological basis of the research was the use of general and general scientific methods: - the method of the empirical level, where facts

are collected, judgments are developed and theoretical generalizations are made; - the method of the theoretical level, where a systematic, integrated approach to statistical, economic, legal information is applied and a comparative analysis of legal and economic data is proposed.

Results. Cybersecurity of information relations in the 21st century has become a major component of supporting and ensuring the safety of data. The global community is currently concerned about the growth of cybercrime, and the lack of legal regulation of entire sectors of the economy, as well as national information security at the level of states and individual countries, may lead to the risk of losing information / digital sovereignty. Today information technologies are aimed at meeting the information needs of all users, including users of leading businesses using various information databases, electronic directories, financial, scientific and technical information and other electronic resources related to international, national, regional information systems (Semenov S.R., 2018: 55). Information / digital sovereignty is one of the modern foundations of any state, which should be based on various information content and have a degree of national influence and support for the implementation of an independent information policy. A number of states are already actively promoting the ideas of cybersecurity and its institution, in particular:

European Union (EU). In 2013, at the level of the European Commission, the EU Cybersecurity Strategy is adopted, where in Part 1 it is recognized that Cybersecurity is necessary to integrate the social and political environment together, to support and exercise freedom through democratic institutions, with the possibility of promoting European values in the field of human rights and freedoms and citizen, but the values themselves must be protected from illegal actions (Cybersecurity Strategy of the European Union, 2013: 2). At the same time, clause 1.2 establishes the basic principles of Cybersecurity, such as - The core values of the EU should be applied both in physical and digital space (the same force of EU legal sources), protection of fundamental rights and freedoms (recognition and support of natural rights enshrined in legislation), open access for all subjects of law, democratic governance (participation of many stakeholders represented by the private sector), the presence of shared responsibility (formation of a legal model of responsibility). Accordingly, these principles create priority areas consisting of cyber resilience, cybercrime reduction, the formation of a common cybersecurity policy (including defense policy), the development of cybersecurity production capacities, and the development of a common EU vision on international cyberspace. When developing this strategy, the EU took the 2001 Budapest Declaration as a basis,

which approved a common vision of legal measures (computer-assisted privacy crimes, copyright infringements, etc.), with further harmonization of law, both nationally and at the international level (Convention on Cybercrime, 2021: 5-7). Separately, the scope of procedural law was designated under Article 14, with the possibility of conducting criminal investigations, judicial proceedings. Moments in the field of search and seizure of computer data, data collection in real time, jurisdiction are identified. In 2020, a new EU Cybersecurity Strategy is adopted, where part 2 outlines the key trends in the development of the industry, in particular: 1) technological sovereignty, sustainable politics, leadership; 2) strengthening technological capacity to prevent risks; 3) promotion of global cyberspace (The EU's Cybersecurity strategy for the digital decade, 2020: 1). Thus, the EU intentionally supports and promotes the idea of an open and secure cyber network space with the ability to attract international investment and promote European values (openness of data, freedom of information, freedom of communication, data security, etc.). The EU is a unified data security system, which is supported by the EU Directive 2016/1148, as a major element in the formation and implementation of cyber law in the national legislation of the participating countries in various areas of the economy (implementation and support of online trading with the conclusion of online contracts; banking infrastructure; development of a system of online search engines with multilingual support), building a CSIRT (computer security incident response team to detect risks, warnings, mutual assistance, cross-border data exchange, establish fines in this area, etc.) (Directive (EU) 2016/1148, 2016: 1). Accordingly, support for cybersecurity has been established in the national legislation of the EU member states. For example: the Republic of Poland has introduced its plan for the development of cybersecurity for 2017-2022, where part 4 indicates that secure cyberspace will provide the necessary potential for building the country's digital economy, while ensuring the necessary level of provision of digital services, both from private and public sectors (National framework of cybersecurity policy of the Republic of Poland for 2017-2022, 2017 : 4-5). In addition to general trends, the Czech Republic, in its Cybersecurity Strategy for 2015-2020, has taken a course towards research and scientific work in this area, strengthening educational policy from schools to universities, training government officials, especially paying attention to the training of prosecutors and judges (National cyber security strategy of the Czech Republic for the period from 2015 to 2020, 2015:5-10). An important component of EU cybersecurity is a supranational body - the EU Cybersecurity Agency (ENISA), which aims to promote

and develop the cybersecurity of the EU member states (starting with the construction of a single network, an integration gateway). At the same time, by the end of 2019, all EU countries had implemented cybersecurity at the national level in the format of strategies, concepts, programs or plans.

Japan: In 2000, the Law on the Formation of a Modern Information and Telecommunication Network Society was adopted, which establishes that Japan's society becomes networked, where legal entities are in a single data stream (Basic Act on the formation of an advanced information and telecommunications network society, 2000: 1-5). At the same time, the network society, in accordance with Article 3, gives its possibilities of open access to the Internet, to various technologies, which leads to the emergence of the creative development of the individual and the recognition of his rights of digital rights. In 2013, the first draft of the Cybersecurity Strategy (based on the results of the first conference in Japan on cybersecurity) is being developed, which leads to a rethinking of the process of information flows and information protection (private and public) (Cyberdefense report, 2020: 14). In 2014, the Law on Cybersecurity was adopted, indicating the importance of cyber threats to the information space, which, on the basis of Article 12, leads to the formation of a cybersecurity strategy plan for the country, and, on the basis of Article 24, to form a Strategic Cybersecurity Headquarters under the Cabinet of Ministers, which will allow developing standards and measures for state bodies, adopt functions and plans for intergovernmental cooperation. The Chief of Staff is the Chief Secretary of the Cabinet of Ministers. In 2015, as a product of previous decisions, the Cybersecurity Strategy was adopted, which, according to 4. established the basic security principles, such as the free flow of information (without censorship, political pressure, but with the protection of users' personal data), the rule of law (the ratio of national law to international law) , openness of sources, autonomy (formation of mechanisms that allow legal entities to independently develop legal relations within the information environment), cooperation (in this area) (Cybersecurity strategy, 2015: 1-12). The main goal of this Strategy is to build a safe society where every subject of law can feel safe. Consequently, the planning process (development, analysis and adoption of a three-year plan (2015-2018)) with a display of financial costs, the degree of implementation in the field of state bodies, etc., serves as a tool for implementing the Strategy. In 2018, a new Cybersecurity Strategy is adopted, where from the innovations it can be noted that there is no longer a distinction between cyberspace and real space (Cyberdefense report, 2020:15). Now there is only their synthesis - "mutually interacting subjects". The emphasis was also placed on new

threats in the field of cryptocurrency, the Internet of Things (data exchange between different devices), 5G, etc.

Eurasian Economic Union (EAEU). Information security, and in the future also cybersecurity, should become key trends in the development of information relations among the EAEU member states. It is necessary to distinguish between normative legal acts (NLA) at the level of the EAEU integration law and the national law of the participating countries. Integration law is a number of legal acts approved by the EAEU bodies. First of all, it is possible to single out Article 367 of the Customs Code of the EAEU, which indicates the legal protection of information data based on the legislation of the participating countries, including the authorized customs authorities (Treaty on the Customs Code of the Eurasian Economic Union, 2017: 983). But this article does not indicate the application of cybersecurity and the identification of possible risks in this area. Suggestion: to supplement this article - with the wording "including in the field of cyberspace", thus, the legal level of legal relations will expand. In 2016, a project was laid for the cryptographic protection of information resources and integration gateways, which can become practical elements of cybersecurity (On the implementation of a project for the joint development of specialized means of cryptographic protection of information of the Eurasian Economic Union, 2016: 1-20). Another NLA is the 2017 Decision of the Supreme Eurasian Economic Council on the development of the digital agenda until 2025, where Part 1 states that the EAEU member states independently develop policies in the field of information space (On the main directions for the implementation of the digital agenda of the Eurasian Economic Union until 2025, 2017: 3-5). Thus, there is only the establishment of trends and directions, without a specific program for the information or cybersecurity system. At the same time, the digital agenda itself should not contradict national legislation. National law is expressed in the following legal acts:

-Armenia. In 2017, the Concept of Information Security is adopted, which confirms the main directions of the country's information policy (Concept of Information Security and Information Policy of the Republic of Armenia, 2017: 1-5), and in 2020, the National Security Strategy of the country is adopted, where Part 7 establishes that information security is a key area of the future, while cybersecurity will serve as protection from possible cyberattacks and crimes in the digital field (National Security Strategy of the Republic of Armenia, 2020: 1-10). And the main challenges in this area are: the lack of specialized legislation, issues of critical infrastructure, a unified state policy, but in the future these challenges will be overcome.

-Belarus. In 2019, the Information Security Concept is adopted, which confirms the importance of cyber threats, including the emergence of cybercrimes in the Internet space (Information security concept of the Republic of Belarus, 2019: 1-5). Clause 61 recognizes under the cyber threat the failure of technical equipment, software failures, which may be caused by the illegal actions of certain groups of people. Accordingly, in order to ensure cyber security, it is necessary to create a national security system that would allow responding to modern challenges. The important areas of clause 65.66 confirm the provision of the national segment of the Internet as an important space for interaction between the public and private sectors.

-Kazakhstan. In 2017, the Cybersecurity Concept was approved, which forms a model of approaches to this area, including the development of a data monitoring system, a unified policy in the field of information resources, the display of legal principles (observance of human rights and freedoms, ensuring the legitimate interests of legal entities, personal security and etc.) (Kazakhstan Cyber Shield, 2017: 1-10). An important point of this concept is the idea of promoting “Cyberhygiene” (security rules in the electronic network, when working with information flows). In addition, in 2018, to strengthen the country’s financial sector, a separate Cybersecurity Strategy for 2018-2022 was adopted, which should strengthen the banking sector with the formation of a national security system in the field of banking operations (On the approval of the Cybersecurity Strategy of the Financial Sector of the Republic of Kazakhstan for 2018-2022, 2018: 1-10).

-Kyrgyzstan. In 2019, the Cybersecurity Strategy for 2019-2023 was adopted, with the approval in part 5 of the profile directions for the development of state policy in the field of cybersecurity, in particular, the formation of a unified system of measures (strengthening the interaction of state bodies in the field of information; teaching computer / digital literacy among state employees; uniform state policy), organization of information infrastructure security (identification of problematic aspects of infrastructure, development of criteria for assessing critical infrastructure), creation of a threat prevention system, development and implementation of the institution of cybercrime into national legislation, creation of a protection system, etc. (Cybersecurity Strategy of the Kyrgyz Republic for 2019-2023, 2019: 1-10). The issue of security of the country’s critical infrastructure is being worked out, with the introduction of the rights and obligations of users into the legislation, the introduction of monitoring and security assessment, which will strengthen the current Cybersecurity Strategy (Semenov N.S., 2020: 190-192).

-Russia. In 2013, a draft Concept of the country's cybersecurity strategy is being developed, which identifies the main problems, such as causing damage to the rights of individuals, government agencies and legal entities; the presence of cyberattacks against information resources; cyber warfare (Concept of the Cybersecurity Strategy of the Russian Federation, 2013:2-6). At the same time, part 2 of this draft mentions that "Cybersecurity" is not yet separated from the concept of "information security", which is an existing gap in the legislation. Part 5 denotes the basic principle of cyber security: guaranteed constitutional human and civil rights in the field of working with information, maximum personal protection, constructive cooperation, balance of responsibility, etc. In the future, the Cybersecurity Strategy will be adopted, and the presented draft of 2013 can serve as a basis for it.

Discussion. Any document, regulatory legal acts approving the idea of cybersecurity should be based on national legislation, with the implementation of the information space, including interaction with information, forms of implementation and information infrastructure and the possibility of economic application. This direction is confirmed in international legislation, since cybersecurity has, first of all, cross-border coverage (two or more countries), expressed in the form of a regional segment (Asia, Europe, etc.), and secondly, in the form of a transcontinental segment (North and South America, etc.). Accordingly, each country, integration association has its own policy and vision. For example: the EU through building information relations and ensuring their protection in the cyber space, where it promotes its European values and rights in the global information space. There is a specialized supranational agency (ENISA) that deals with the cybersecurity of the entire EU, monitoring relevant actions in the field of legislation, information infrastructure, and technical means. Japan is giving impetus to the formation of a number of areas in the field of legislation, from the Laws on the formation of a modern information and telecommunication network society to Cybersecurity, and ending with three profile strategies, which indicates a high level of preparation of the society. The main feature was the creation of a ready-made model of an information and network society and state with specific legal principles, institutions and a security structure, where the differences between cyberspace and real space are becoming a thing of the past. There is a recognition of equal opportunities both in the physical and in the virtual worlds, with different legal relationships. Within the framework of integration law, the EAEU is still inferior to the EU, but at the level of national legislation, the experience of the EU and Japan is taken in

the field of informatization and building information infrastructure. From the general analysis of the development of cybersecurity, a number of proposals are relevant that can improve the work of the EAEU, including:

1. To form a separate area of cybersecurity on the basis of the Information Technology Department of the EAEU, which will harmonize the legislation of the participating countries, develop general recommendations at the national levels.

2. To develop a common strategy and plan for the cybersecurity of the EAEU, which will lead to the formation of a common vision on the problems and prospects.

3. Finalize national strategies and plans for cybersecurity of the EAEU member states.

4. Expand the capabilities of the EAEU integration gateway, as there is the experience of the Customs Union and its information infrastructure.

5. To create conditions for the promotion of national integration projects in the field of cybersecurity, data processing, electronic services, etc.

Conclusion. The unifying structure for both the EU, Japan and the EAEU is the provision of information / digital sovereignty, which plays an important role in the security of the country or any territorial association. The more successfully this sovereignty is protected, the more opportunities the states will have to advance their priority areas in the field of data security, both in the national information field and in the global space. The basis for this will be a strong information infrastructure, with its own scientific and resource base.

Information about authors:

Ganieva Tamara Imangalievna – doctor of law sciences, professor, department of theory and history and law, Kyrgyz National University named after J.Balasagyn; *ganieva1000@gmail.com* – <https://orcid.org/0000-0003-2930-8722>;

Semenov Nikolai Sergeevich – candidate of law sciences, acting associate professor of Department of jurisprudence and international law, International University of Kyrgyzstan; *frindland@mail.ru* – <https://orcid.org/0000-0001-5183-7482>;

Semenov Sergei Rudolfovich – candidate of economic sciences; acting associate professor of Department of international business, International University of Kyrgyzstan; *ssr2002@list.ru* – <https://orcid.org/0000-0001-7871-6541>.

REFERENCES:

Basic Act on the formation of an advanced information and telecommunications network society. Act № 144 of December 6, 2000. URL: https://japan.kantei.go.jp/it/it_basiclaw/it_basiclaw.html#:~:text=Article%205.-,The%20formation%20of%20an%20advanced%20information%20and%20telecommunications%20network%20society,range%20of%20a%20high%2Dquality (last visit 14.02.2021) (in Eng.).

Concept of Information Security and Information Policy of the Republic of Armenia. Approved by the Order of the President of the Republic of Armenia dated October 23, 2017 NK-146-A. URL: http://igf.am/wp-content/uploads/2019/12/%D5%80%D5%80-%D5%8F%D5%A5%D5%B2%D5%A5%D5%AF%D5%A1%D5%BF%D5%BE%D5%A1%D5%AF%D5%A1%D5%B6-%D4%B1%D5%B6%D5%BE%D5%BF%D5%A1%D5%B6%D5%A3%D5%B8%D6%82%D5%A9%D5%B5%D5%A1%D5%B6-%D5%80%D5%A1%D5%B5%D5%A5%D6%81%D5%A1%D5%AF%D5%A1%D6%80%D5%A3_20171023.pdf (last visit 14.02.2021) (in Russ.).

Concept of the Cybersecurity Strategy of the Russian Federation. Federation Council of the Federal Assembly of the Russian Federation. URL: <http://council.gov.ru/services/discussions/themes/38324/> (last visit 14.02.2021) (in Russ.).

Convention on Cybercrime (Budapest Convention). Adopted on 23 November 2001 in Budapest, Hungary. URL: <https://rm.coe.int/1680081580> (last visit 14.02.2021) (in Russ.).

Cyberdefense report. Japans National cyber security and defense posture. Policy and organizations. Zurich, September 2020. P.14,15. URL: <https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2020-08-Japans-national-cybersecurity-defense-posture.pdf> (last visit 14.02.2021) (in Eng.).

Cybersecurity Concept (Kazakhstan Cyber Shield). Approved by the Decree of the Government of the Republic of Kazakhstan dated June 30, 2017 No. 407. URL: <https://tengrinews.kz/zakon/pravitelstvo-respubliki-kazahstan-premer-ministr-rk/hozyaystvennaya-deyatelnost/id-P1700000407/> (last visit 14.02.2021) (in Russ.).

Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace. European commission. Brussels, 7.2.2013. URL:

Cybersecurity Strategy of the Kyrgyz Republic for 2019-2023 Adopted by the Resolution of the Government of the Kyrgyz Republic dated July 24, 2019 No. 369. URL: <http://cbd.minjust.gov.kg/act/view/ru-ru/15479> (last visit 14.02.2021) (in Russ.).

Cybersecurity strategy. The Government of Japan. September 4, 2015. URL: https://www.itu.int/en/ITUDE/Cybersecurity/Documents/National_Strategies_Repository/Japan_2015_cs-strategy-en.pdf (last visit 14.02.2021) (in Eng.).

Directive (EU) 2016/1148 of the European Parliament and the Council of 6 July 2016 concerning measures for a high common level security of network and information systems across the Union. Eur-lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.194.01.0001.01.ENG&toc=OJ:L:2016:194:TOC (last visit 14.02.2021) (in Eng.).

https://eeas.europa.eu/archives/docs/policies/eu-cyber-security/cybsec_comm_en.pdf (last visit 14.02.2021) (in Eng.).

Information security concept of the Republic of Belarus. Approved by the Resolution of the Security Council of the Republic of Belarus dated March 18, 2019 No. 1. <https://www.sb.by/articles/kontseptsiya-informatsionnoy-bezopasnosti-respubliki-belarus.html> (last visit 14.02.2021) (in Russ.).

National cyber security strategy of the Czech Republic for the period from 2015 to 2020. URL: <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national-cyber-security-strategies-interactive-map?selected=Czech%20Republic> (last visit 14.02.2021) (in Eng.).

National framework of cybersecurity policy of the Republic of Poland for 2017-2022. EU Agency for cybersecurity. URL: https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Cybersecuritystrategy_PL.pdf/view (last visit 14.02.2021) (in Eng.).

National Security Strategy of the Republic of Armenia 2020. URL: <https://mil.am/files/LIBRARY/%D0%A1%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D0%B8%D1%8F%20%D0%BD%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D0%B9%20%D0%B1%D0%B5%D0%B7%D0%BE%D0%BF%D0%B0%D1%81%D0%BD%D0%BE%D1%81%D1%82%D0%B8%20%D0%A0%D0%B5%D1%81%D0%BF%D1%83%D0%B1%D0%BB%D0%B8%D0%BA%D0%B8%20%D0%90%D1%80%D0%BC%D0%B5%D0%BD%D0%B8%D1%8F%202020.pdf> (last visit 14.02.2021) (in Russ.).

On the approval of the Cybersecurity Strategy of the Financial Sector of the Republic of Kazakhstan for 2018-2022. Adopted by the Resolution of the Board of the National Bank of the Republic of Kazakhstan dated October 29, 2018 No. 281. URL: <https://cdb.kz/sistema/pravovaya-baza/ob-utverzhenii-strategii-kiberbezopasnosti-finansovogo-sektora-respubliki-kazakhstan-na-2018-2022-gody/> (last visit 14.02.2021) (in Russ.).

On the implementation of a project for the joint development of specialized means of cryptographic protection of information of the Eurasian Economic Union. Approved by the Order of the Supreme Eurasian Economic Council dated December 26, 2016 No. 7. URL: https://docs.eaeunion.org/docs/ru-ru/01413600/sco_11042017_7 (last visit 14.02.2021) (in Russ.).

On the main directions for the implementation of the digital agenda of the Eurasian Economic Union until 2025. Approved by the Decision of the Supreme Eurasian Economic Council dated October 11, 2017 No. 12. URL: https://docs.eaeunion.org/docs/ru-ru/01415258/scd_10112017_12.

Semenov N.S. Legal features of information security of the Kyrgyz Republic. In the collection: Russian entrepreneurs - philanthropists and patrons. Collection of materials of the III All-Russian Morozov readings. Orekhovo-Zuevo, 2020. Pp.187-192. (in Russ.).

Semenov S.R. Information infrastructure development and digital transformation. / Scientific and practical journal of the Almaty Academy of Economics and Statistics - "Statistics, accounting and audit", No. 3 (70) 2018, pp. 50-56 (in Russ.).

The EU's Cybersecurity strategy for the digital decade. Brussels, 16.12.2020. URL: <https://ec.europa.eu/digital-single-market/en/news/eus-cybersecurity-strategy-digital-decade> (last visit 14.02.2021) (in Eng.).

Treaty on the Customs Code of the Eurasian Economic Union. Adopted on April 11, 2017. URL: https://docs.eaeunion.org/docs/ru-ru/01413569/itia_12042017 (last visit 14.02.2021) (in Russ.).

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

ISSN 1991-346X

Volume 2, Number 342 (2022), 19–38

<https://doi.org/10.32014/2022.2518-1726.127>

УДК 519.68.02

МРНТИ 53.31.23

Е.С. Голенко^{1*}, А.А. Исмаилова¹, А.С. Жумаханова²

¹Казахский агротехнический университет им. С. Сейфуллина,
Нур-Султан, Казахстан;

²Евразийский национальный университет им. Л.Н. Гумилёва,
Нур-Султан, Казахстан.

E-mail: golenko.katerina@gmail.com

ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКОВ ПРИ ПОМОЩИ БАЗЫ ДАНЫХ «GENE ONTOLOGY» И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Аннотация. Прогнозирование функций белков является важной частью аннотации генома. В последнее время методы прогнозирования быстро развиваются благодаря появлению высокопроизводительных технологий секвенирования. Среди доступных баз данных для определения терминов функций белка важным ресурсом, описывающим функциональные свойства белков, является Gene Ontology (GO). Gene Ontology предлагает классификацию функций, которая базируется на некотором списке или словаре четко сформулированных терминов, каждый из которых принадлежит одной из категорий – молекулярным функциям, биологическим процессам и клеточным компонентам. Из этой базы данных можно по названию белка или его идентификационному номеру найти присвоенные ему термины Gene Ontology или аннотации, которые были сделаны на основе расчётных или экспериментальных данных. Каждому термину в Gene Ontology присваивается ряд атрибутов: уникальный цифровой идентификатор, название, словарь, к которому термин принадлежит, и определение. Термины могут иметь синонимы, которые делятся на точно соответствующие значению термина, более широкие, более узкие и имеющие

некоторое отношение к термину. Исследователи используют различные подходы для эффективного прогнозирования терминов GO. Между тем, глубинное обучение, быстро развивающаяся дисциплина в подходе, основанном на данных, демонстрирует впечатляющий потенциал в отношении присвоения терминов GO аминокислотным последовательностям. В данной статье авторами рассмотрены доступные сегодня вычислительные методы аннотации GO для белков, начиная от традиционного и заканчивая методом глубинного обучения. Также вынесены на обсуждение основные проблемы в этой области и подчеркнуты будущие направления предсказания функций белка с помощью GO.

Ключевые слова: Gene Ontology, предсказание функций белка, машинное обучение, глубинное обучение, аннотация белков.

Е.С. Голенко^{1*}, А.А. Исмаилова¹, А.С. Жумаханова²

¹С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Нұр-Сұлтан, Қазақстан;

²Л.Н. Гумилёв атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан.

E-mail: golenko.katerina@gmail.com

«GENE ONTOLOGY» БАЗАСЫН ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ ҮЛГІЛЕРІН ПАЙДАЛАНА ОТЫРЫП АҚУЫЗ ФУНКЦИЯЛАРЫН БОЛЖАУ

Аннотация. Ақуыз функциясын болжау геномдық аннотацияның маңызды бөлігі болып табылады. Соңғы уақытта жоғары өнімді секвенирлеу технологияларының пайда болуына байланысты болжау әдістері қарқынды дамып келеді. Ақуыз функциясының терминдерін анықтауға арналған дерекқорлардың арасында Gene Ontology (GO) белоктардың функционалдық қасиеттерін сипаттайтын маңызды ресурс болып табылады. Гендік онтология әрқайсысы категориялардың біріне жататын - молекулалық функцияларға, биологиялық процестерге және жасушалық компоненттерге жататын нақты анықталған терминдердің тізбесіне немесе сөздік қорына негізделген функциялардың жіктелуін ұсынады. Бұл дерекқордан ақуыздың атын немесе оның сәйкестендіру нөмірін оған тағайындалған Гендік онтология терминдерін немесе есептелген немесе эксперименттік деректер негізінде жасалған аннотацияларды табу үшін пайдалануға болады. Гендік онтологиядағы

әрбір терминге бірнеше атрибуттар тағайындалады: бірегей сандық идентификатор, атау, термин жататын сөздік және анықтама. Терминдердің синонимдері болуы мүмкін, олар терминнің мағынасына сәйкес келетін, кеңірек, тар және терминге қандай да бір қатысы бар болып бөлінеді. Зерттеушілер GO терминдерін тиімді болжау үшін әртүрлі тәсілдерді пайдаланады. Сонымен қатар, терең оқыту, деректерге негізделген тәсілдегі жылдам дамып келе жатқан пән, аминқышқылдарының тізбегіне GO терминдерін тағайындаудың әсерлі әлеуетін көрсетеді. Бұл мақалада авторлар дәстүрліден терең оқыту әдісіне дейінгі протеиндерге арналған GO аннотациясының қазіргі уақытта қол жетімді есептеу әдістерін қарастырады. Ол сондай-ақ осы саладағы негізгі проблемаларды талқылайды және GO көмегімен ақуыз функцияларын болжаудың болашақ бағыттарын көрсетеді.

Түйін сөздер: Gene Ontology, ақуыз функциясын болжау, машиналық оқыту, терең оқыту, ақуыз аннотациясы.

Y.S. Golenko^{1*}, A.A. Ismailova¹, A.S. Zhumakhanova²

¹S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan;

²L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan.

E-mail: golenko.katerina@gmail.com

PREDICTING PROTEIN FUNCTIONS USING THE «GENE ONTOLOGY» DATABASE AND MACHINE LEARNING MODELS

Abstract. Protein function prediction is an important part of genome annotation. Recently, predictive methods have advanced rapidly due to the advent of high throughput sequencing technologies. Among the available databases for defining protein function terms, Gene Ontology (GO) is an important resource describing the functional properties of proteins. Gene Ontology proposes a classification of functions that is based on a list or vocabulary of well-defined terms, each of which belongs to one of the categories - molecular functions, biological processes, and cellular components. From this database, you can use the name of the protein or its identification number to find the Gene Ontology terms assigned to it or annotations that were made on the basis of calculated or experimental data. Each term in Gene Ontology is assigned a number of attributes: a unique numeric identifier, a name, the dictionary to which the term belongs, and a definition. Terms can have synonyms, which are divided into exactly

corresponding to the meaning of the term, broader, narrower, and having some relation to the term. Researchers use a variety of approaches to efficiently predict GO terms. Meanwhile, deep learning, a rapidly evolving discipline in a data-driven approach, is showing impressive potential for assigning GO terms to amino acid sequences. In this article, the authors consider the currently available computational methods for GO annotation for proteins, ranging from the traditional one to the deep learning method. It also discusses the main problems in this area and highlights future directions for predicting protein functions using GO.

Key words: Gene Ontology, protein function prediction, machine learning, deep learning, protein annotation.

Введение. Белки представляют собой органические макромолекулы, которые являются фундаментальными детерминантами структуры и функций живых организмов. Они играют роль во многих процессах, включая биохимические реакции, передачу сигналов, транспорт питательных веществ и т.д. Поэтому понимание свойств белков важно не только с биологической и эволюционной точек зрения, но и с точки зрения использования их потенциала в биомедицинских и фармацевтических сферах и других областях.

Как правило, идентификация функций белка осуществляется с помощью ручной или компьютерной аннотации. Первый подход является золотым стандартом для функциональных аннотаций, потому что он реализуется экспертами и дает высококачественные результаты. Тем не менее, этот подход является дорогостоящим и трудоемким, и поэтому его трудно масштабировать. Кроме того, из-за развития высокопроизводительных технологий секвенирования, таких как секвенирование нового поколения (Next-Generation Sequencing или NGS), количество последовательностей, подлежащих аннотированию, резко возросло. Таким образом, методы вычислительной аннотации были разработаны как необходимость для автоматической обработки большого объема вновь сгенерированных последовательностей, а также для повышения точности аннотированных данных.

Из-за изменчивости словаря, используемого для определения функций белка, были предложены различные базы данных для обеспечения стандартизированной схемы, такие как Enzyme Commission (EC), Функциональный каталог (FunCat) и Киотская энциклопедия генов и геномов (KEGG). В настоящее время Gene Ontology (GO) (Consortium GO, 2015) является наиболее полным ресурсом, поскольку он обладает всеми нужными свойствами системы функциональной классификации.

Консорциум GO создал базу данных для контролируемого словаря, описывающего функциональные свойства геномных продуктов (например, генов, белков и РНК). Каждая онтология (словарь) принадлежит к одной из трех категорий: молекулярная функция (МФ, MF, Molecular Function), биологический процесс (БП, BP, Biological Process) и клеточный компонент (СС, Cellular Component). С точки зрения структуры, GO следует иерархической организации в виде направленного ациклического графа (DAG), в котором каждый термин является узлом, а каждое ребро, соединенное с двумя узлами, представляет отношение родитель-потомок. Его можно использовать для вывода многих типов информации, таких как «является» («is-a») или «является частью» («part-of»). Кроме того, «является частью» подразумевает, что дочерний узел обязательно является частью родителя. Это позволяет гибко аннотировать белки в отношении различных уровней функции — от общих терминов до конкретных — в зависимости от доступных данных.

Автоматизированное прогнозирование функций (Automated Function Prediction - AFP) на основе системы GO представляет собой сложную задачу биоинформатики. Во многих исследованиях обсуждалась функциональная аннотация белка с разных точек зрения. Предыдущие обзоры были посвящены AFP (Rost, 2003) с точки зрения используемого типа данных (Shehu, 2016), недостатков и соответствующих решений, сети взаимодействия белков, типов классифицированных функций и назначения GO на основе информации о последовательности. В (Bonetta, 2020) демонстрируется предсказание функции белка в рабочем процессе машинного обучения. В (Zhao, 2022) рассматривают прогнозирование функции генов с точки зрения моделирования GO. Все эти исследования представили независимые взгляды на проблему, однако не представлено подробного обзора глубинного обучения, которое является новым подходом к прогнозированию функций белков с помощью терминов GO. Таким образом, в статье предлагается возможность прогнозирования функций белка с использованием как обычного, так и глубинного обучения, что также указывает на то, что можно ожидать более высокой прогностической эффективности при сравнении нескольких методов друг с другом.

Основываясь на предыдущих исследованиях и вышеупомянутых обзорах, наша статья разделена на две основные части. Первая часть охватывает традиционный подход и включает решения, не использующие глубинное обучение, а вторая часть описывает методы, кото-

рые основываются на глубинном обучении для решения проблемы функциональной аннотации белка.

Ниже кратко рассмотрен традиционный подход, так как по этому вопросу уже доступно большое количество работ и обзоров литературы. В этой части суммируются три основные подкатегории традиционного подхода и упоминаются известные и самые последние соответствующие исследования.

Основная часть. Материалы и методы. Стандартный подход для прогнозирования GO терминов белкам.

В целом, было предложено несколько методов для присвоения терминов GO белкам с использованием различных типов данных и методов. Ниже рассмотрены репрезентативные решения, внимание сосредоточено на трех категориях: методы на основе подобия, вероятностные методы и методы машинного обучения.

Методы, основанные на сходстве. Первоначально функциональные аннотации присваивались неохарактеризованным белкам на основе простого принципа: последовательность белка искалась в базах данных экспериментально выбранных белков и, если были извлечены какие-либо похожие белки («похожие» специально определенным образом), термины GO, связанные с извлеченными последовательностями, были присвоены запрашиваемому белку. Самое первое решение методов, основанных на сходстве, было основано на гомологии. Эти инструменты основывались на инструментах поиска локальных выравниваний, таких как Базовый инструмент поиска локальных выравниваний (BLAST). В этих методах неизвестная последовательность ищется в базе данных, в которой хранятся хорошо аннотированные белки. Затем идентифицируется извлеченная последовательность с наивысшей оценкой выравнивания в соответствии с заданным порогом, и ее аннотация передается запрашивающему белку. OntoBlast, GOFigure, GOblet и GOtcha являются типичными системами аннотаций, использующими сходство последовательностей, определяемое поисковым механизмом BLAST. Более подробная информация об этих инструментах представлена в Таблице 1.

Таблица 1 Предикторы аннотаций GO, использующие традиционный подход

Метод	Название программы	Описание	Год исследования
Основанный на сходстве	OntoBlast	Веб-сервер, который является частью инструмента «Ontologies TO GenomeMatrix». Он предсказывает функцию белка по взвешенному списку терминов GO, связанных с последовательностями попаданий BLAST в девяти базах данных генома.	2003
	GOFigure	Производит вывод в виде кликабельного графа в четыре этапа, включая поиск гомологичной последовательности, построение графа с минимальным покрытием и назначение онтологий после их оценки.	2003
	Goblet	Программный пакет, позволяющий пользователю определять чувствительность к E-значению, и базы данных, используемые для поиска BLAST. Все онтологии, идентифицированные после сопоставления белков, строятся в сводной структуре DAG, где количество последовательностей, имеющих общий термин GO, является кумулятивным, чтобы представить значимость термина.	2004
	Gotcha	Переносит ассоциации GO, полученные из BLAST, в генные продукты с помощью новой схемы ранжирования. E-значения производных терминов распространяются на их родительские термины в DAG, обеспечивая нормализованные оценки достоверности для отдельной онтологии GO.	2009
	PFP	Функциональная система аннотаций, основанная на совпадениях PSI-BLAST, использующая совпадающие последовательности до необычно высокого порога и извлекающая очень важные термины GO из UniProt для оценки функций для неизвестных белков.	2015
	INGA	Функциональный сервер аннотаций выводит прогнозы GO из множества источников данных.	2016
	GoFDR	Метод на основе множественного выравнивания с использованием FDR и PSSM для ранжирования терминов GO для аминокислотных последовательностей.	2003

Вероятностный	-	Вероятностная структура, основанная на PPI, прогнозирует назначение GO для 10% неаннотированных белков в дрожжах.	2007
	-	Использует разнородные данные для улучшения вероятностной модели функционального прогнозирования <i>Saccharomyces cerevisiae</i> .	2010
	BMRF	Вариант подхода на основе MRF к PPI, который позволяет делать новые прогнозы для неаннотированных белков.	2006
Машинное обучение	GOPET	Инструмент прогнозирования и оценки терминов GO, предоставляющий термины GO для молекулярных функций и биологических процессов для любого организма.	2010
	PoGO	Модель ансамбля со сходством последовательностей InterPro, биохимическими свойствами и третичной структурой белка для прогнозирования GO белков грибов.	2018
	FFPred3	Инструмент на основе SVM, обеспечивающий независимое от гомологии назначение терминов GO для эукариотических белков.	2018
	PANNZER2	Взвешенный классификатор k-ближайших соседей, обеспечивающий функциональную аннотацию для неохарактеризованных белков с терминами GO и описаниями в произвольном тексте.	2018
	Deep Text2GO	Сочетание текстового метода и метода, основанного на последовательностях, для улучшения крупномасштабного предсказания функции белка.	2018
	NetGo	Эффективный веб-сервер на основе LTR, объединяющий как последовательность, так и массивную сетевую информацию о белках для аннотирования генных продуктов.	2019

PFP (Piovesan, 2015) это еще один метод, который использует функциональную информацию, связанную с удаленными гомологами, путем применения позиционно-специфического итерированного BLAST (PSIBLAST). Инструмент является улучшением с точки зрения охвата и точности, что определяется путем анализа эталонного набора данных по сравнению с GOtcha, PSI-BLAST и InterProScan. Другим вариантом решения, основанным на выравнивании последовательностей, является INGA (Piovesan, 2015), в котором данные сети белок-белкового взаимодействия (Protein-Protein Interaction - PPI) объединяются с назначением домена и сходством последовательностей

из BLAST для достижения консенсуса в прогнозировании функций GO с использованием анализа обогащения. Кроме того, GoFDR (Gong, 2016) берет соответствующие термины GO из запросов множественного выравнивания последовательностей (MSA) с помощью поиска BLAST или PSI-BLAST. Вероятность присвоения термина последовательности запроса определяется функционально различающимися остатками (Functionally Discriminating Residues - FDR), оценочной матрицей для конкретных позиций (PSSM) для FDR и таблицей оценки с максимальной вероятностью, подготовленной с использованием обучающих последовательностей.

Хотя методы, основанные на поиске локального выравнивания, все еще актуальны, просты и в некоторой степени работают хорошо, они также имеют некоторые недостатки, включая ошибки аннотаций базы данных или чрезмерную передачу функций, относительность порога и низкую чувствительность или специфичность. В дополнение к методам, основанным на выравнивании, доступно несколько других предикторов, которые аннотируют функции переноса на основе сходства на уровне структуры белка, семейства белков или филогеномики (Shehu, 2016).

Вероятностные методы. Для определения функции белков был разработан ряд вероятностных моделей. В (Narai, 2007) использовали граф функциональных связей, построенный на основе сети PPI дрожжей *Saccharomyces cerevisiae*. Предположение заключалось в том, что вероятность совместного использования функций между (узлами) белками, находящимися близко друг к другу на графе, выше, чем вероятность для узлов, которые не находятся в непосредственной близости. В этом методе вероятности терминов GO, присвоенных белковой последовательности, выводятся из биномиальной модели, включающей алгоритм случайного поля Маркова (MRF). Позже, в (Narai, 2007) объединили несколько источников данных (PPI, данные об экспрессии генов, информацию о белковых мотивах, данные о мутантном фенотипе и данные о локализации белка), используя байесовскую структуру, чтобы повысить эффективность прогнозирования по сравнению с моделью, которая использует только PPI.

Работая с тем же модельным организмом, *S. cerevisiae*, который использовался в описанных выше методах, в (Kourmpetis, 2010) было предложено применить байесовский подход к модели MRF. Это предложение считалось улучшением для оценки параметров модели и предоставления прогнозов на основе сетевых данных. Кроме того, Pinoli, Chicco и Masseroli (Pinoli, 2015) сравнили различные

схемы взвешивания в сочетании с тремя алгоритмами, а именно: разложение усеченного сингулярного значения (truncated Singular Value Decomposition - tSVD), семантически улучшенное tSVD (SIM) и вероятностный латентно-семантический анализ с нормализацией (probabilistic Latent Semantic Analysis with normalization - pLSAnorm). Первые два метода основаны на линейной алгебре, а последний представляет собой модифицированную вероятностную модель, основанную на байесовском выводе. Эти методы были успешно использованы для создания новых аннотаций GO для трех модельных организмов: *Bos taurus*, *Danio rerio* и *Drosophila melanogaster*.

Методы машинного обучения. Инструменты на основе машинного обучения были разработаны для выявления скрытых взаимосвязей между различными характеристиками белка (последовательностью, структурой или другими соответствующими эволюционными доказательствами) и функциональными метками на основе обучающего набора (группы полностью охарактеризованных макромолекул) и использования этой информации для создания аннотаций для новых белков. Были предложены различные методы, основанные на машинном обучении, включающие различные линии доказательств в качестве функций для обучения классификаторов и прогнозирования терминов GO.

Использование классификаторов множественной машины опорных векторов (Support Vector Machine - SVM) является преобладающим выбором в нескольких исследованиях. Например, GOPET передает несколько функций, связанных с терминами GO (меры сходства последовательностей, основанные на поиске BLAST, частоте, качестве аннотаций гомологов и уровне аннотаций в иерархии GO) для 99 классификаторов SVM. Конкретный термин будет «правильным» или «неправильным» ярлыком для неизвестной последовательности, с вычисленными показателями достоверности голосования. В качестве другого примера, FFPred первоначально был установлен для неаннотированного протеома человека, но продемонстрировал обобщение на другие протеомы. Последняя версия FFPred – FFPred3, которая по-прежнему основана на SVM, расширена для исследования корреляций между характеристиками признаков, извлеченными из последовательностей и структур в рамках трех подонтологий (MF, BP и CC). Другой инструмент, Prediction of Gene Ontology terms (PoGO), был разработан на основе автоматической аннотации функционального класса белка (Automatic Annotation of Protein Functional Class - AAPFC). Вместо того, чтобы использовать только

термины InterPro в качестве характеристик, PoGO объединяет еще три источника (сходство последовательностей, биохимические свойства и третичная структура белка). Впоследствии SVM и линейный классификатор используются в качестве классификаторов базового уровня перед этапом метаобучения.

Используя алгоритм k-ближайшего соседа (k-Nearest Neighbor - k-NN), PANNZER2 предоставляет быстрые функциональные системы аннотаций, основанные на гомологии последовательностей и других предикторах аннотаций. Между тем, MS-k NN объединяет разнородные данные, чтобы предложить конкурентоспособную модель для предсказания функции белка. DeepText2GO – это согласованный подход, объединяющий глубокое семантическое представление текста из цитат MEDLINE (NCB, 2018) в виде текстовой информации и информации о последовательности, полученной с помощью BLAST и InterProScan. Эти функции передаются в модели k-NN и логистической регрессии для создания функций белков без знаний в большом масштабе. NetGO является расширением GoLabeler, в котором используется модель обучения для ранжирования (Learning-To-Rank - LTR) для интеграции доказательств на основе последовательностей. NetGO повышает производительность крупномасштабного AFP за счет доступа к огромной белок-белок сети из более чем 2000 видов в базе данных STRING.

Подходы машинного обучения считаются будущим направлением для AFP. Некоторые новые белки либо не имеют идентифицируемых гомологичных последовательностей, либо их обнаруживаемым гомологам не были присвоены какие-либо GO-метки. Следовательно, определение функции белка с нуля с использованием машинного обучения, то есть прямого вывода аннотации из аминокислотной последовательности без доступа к каким-либо дополнительным ссылкам или базам данных, является актуальной задачей.

Аннотации терминов белков с использованием подхода глубинного обучения

Потенциал глубинного обучения был продемонстрирован в нескольких областях применения, включая биоинформатику (Li, 2019). Основной характеристикой, которая отличает этот подход от других методов, является процесс обучения. Они автоматически извлекают функции высокого уровня из необработанных данных и обеспечивают комплексные прогнозы. В отличие от обычных моделей машинного обучения, точная классификация достигается с помощью созданных вручную функций. В настоящее время количество генерируемых

геномных данных, а также количество сложных алгоритмов и вычислительных ресурсов быстро растут. Эти ресурсы поддерживают глубинное обучение для решения проблемы функциональных аннотаций. Ниже рассмотрены предлагаемые методы, основанные на двух критериях: модели и используемых входных данных. Сводная таблица рассмотренных методов представлена в Таблице 2.

Таблица 2 Методы, основанные на глубинном обучении, для назначения терминов GO белкам

Используемая функция	Название программы	Описание	Используемая модель DL
Основанный на последовательности		Метод глубокого АЕ для прогнозирования термина GO	АЕ
		Выводит отсутствующие аннотации GO для белков <i>Homo sapiens</i> , <i>S. cerevisiae</i> , <i>Mus musculus</i> и <i>Drosophila</i> с использованием машин Больцмана с глубоким ограничением	DRBM
	ProLanGO	Модель на основе LSTM, использующая идею NMT для функциональной аннотации аминокислотных последовательностей.	LSTM
	SECLEF	Генерирует функцию GO и семейство белков UniProt из данных последовательности. Основные модели были разработаны Szalkai & Grolmusz (2018a). Веб-интерфейс доступен по адресу https://pitgroup.org/seclaf/ .	CNN
	DEEPred	Прогнозирование функции белка на основе GO с набором связанных MTDNN. Исходный код и данные доступны по адресу https://github.com/cansyl/DEEPred .	DNN
	DeepSeq	Предсказывает функцию белка <i>H. sapiens</i> только на основе данных о последовательности, используя архитектуру глубокого обучения на основе CNN. Исходный код и данные доступны на https://github.com/recluze/deepseq .	CNN
	DeepGOPlus	Расширенный метод DeepGO, основанный на структуре нескольких слоев CNN. Доступно на http://deepgoplus.bio2vec.net/ .	CNN

Интеграция на основе данных/ на основе структуры		Глубинная модель CNN, классифицирующая функциональность на основе третичной структуры белков человека.	CNN
	DeepGO	Фреймворк глубинного обучения, использующий один сверточный слой. Аннотации Protein GO выводятся на основе иерархически структурированного классификатора. Онлайн-инструмент доступен по адресу https://deepgo.cbrc.kaust.edu.sa/deepgo/ .	CNN
	deepNF	Глубинное слияние сетей, фиксирует высокоуровневые функции нескольких сетевых данных для AFP. Исходный код и данные доступны на https://github.com/VGligorijevic/deepNF .	AE
		Реализация двух сложенных многослойных структур для прогнозирования терминов GO на основе свойств последовательности и структуры. Исходный код и данные доступны по адресу http://bioinf.cs.ucl.ac.uk/downloads/mtdnn .	DNN
	DeepFunc	Выводит аннотации GO путем объединения функций сети на основе InterPro и взаимодействия белков.	FCDN
	DeepGOA	Интегрирует исчерпывающую информацию о последовательности и PPI для AFP.	Bi-LSTM, CNN
	SDN2GO	Интегрирует архитектуры с тремя подмоделями и классификатором взвешивания для прогнозирования терминов GO. Исходный код и данные доступны на https://github.com/Charrick/SDN2GO .	CNN
	DeepAdd	Структура на основе CNN, предсказывающая функцию белка на основе последовательности и дополнительной информации (PPI или SSP).	CNN
	FFPred-GAN	FFPred, в котором используется обучающая выборка, дополненная архитектурой глубокого обучения.	GAN
	GONET	Глубинная модель для прогнозирования функции белка с использованием обучения представлению для встраивания белковых последовательностей и сетей.	CNN, RNN

Модельный подход. Обучение с учителем. Обучение с учителем является важным подходом для прогнозирования функции белков *in silico*. Модели под наблюдением учатся аннотировать функции, руководствуясь обучающим набором данных, который пред-

ставляет собой группу охарактеризованных белков с надежными, экспериментально подтвержденными аннотациями. После этапа обучения создается модель, которая фиксирует взаимосвязь между признаком и функцией и используется для прогнозирования терминов GO для новых аминокислотных последовательностей.

Глубокая нейронная сеть (Deep Neural Network - DNN) представляет собой архитектуру с прямой связью. Обычно она состоит из входного слоя, нескольких скрытых слоев и выходного слоя. Входные данные обрабатываются однонаправленно по слоям, от первого до конечного этапа. Что касается функциональной аннотации на основе GO, DNN использовались в качестве многозадачных DNN (MTDNN) (Fa, 2018) или серии MTDNN (Rifaioglu, 2019).

Сверточная нейронная сеть (Convolutional Neural Network - CNN) изначально была разработана для обработки двумерных изображений для распознавания рукописных цифр. Однако с тех пор CNN стала эффективной архитектурой не только для многомерных данных, но и для одномерных входных данных, таких как предложения или геномные последовательности. Модель CNN начинается со сверточных слоев, единицами которых являются карты объектов. Каждая единица получается путем вычисления операций свертки между локальными участками единицы в предыдущем слое и фильтром (набором весов). Слои нелинейного пула объединяются в сверточные слои, что позволяет интегрировать функции разного масштаба в эту архитектуру. CNN могут применяться в AFP либо отдельно (Du, 2020), либо в сочетании с другими архитектурами (Spalević, 2020).

Рекуррентная нейронная сеть (Recurrent Neural Network - RNN) – это архитектура глубокого обучения, разработанная специально для последовательных данных. RNN имеют магистраль DNN, блоки скрытого уровня связаны между собой. Таким образом, скрытая единица получает информацию от входного слоя на текущем временном шаге, а также от скрытых единиц на предыдущем этапе. Долгосрочная кратковременная память (Long Short-Term Memory - LSTM) – это вариант модели RNN, разработанный для фиксации долгосрочных зависимостей между входными последовательностями. Двухнаправленный LSTM (Bi-LSTM) – это модель LSTM, которая обрабатывает данные в двух направлениях, вперед и назад. Эти нейронные сети используются для обработки белковых последовательностей на естественном языке или в сочетании с моделью CNN для предоставления аннотаций GO.

Другой тип нейронной сети, полносвязная глубокая сеть (Fully

Connected Deep Network - FCDN), представляет собой серию полностью связанных слоев. Эта модель использовалась для преобразования входных векторов из InterPro и прогнозирования терминов GO.

Обучение без учителя. В отличие от обучения с учителем, обучение без учителя независимо выявляет скрытые закономерности в распределении входных данных. Это имеет решающее значение для изучения немаркированных данных и предоставляет важную информацию для контролируемых архитектур. Как правило, модели обучения без учителя используются для кластеризации, уменьшения размеров и преобразования данных.

Autoencoder (AE) – это модель обучения без учителя, разработанная для присвоения терминов GO аминокислотным последовательностям. AE – это нейронная сеть для преобразования данных. После кодирования и декодирования целевой размер вывода совпадает с размером ввода. Используя обученную модель AE, прогнозы GO для белков можно вывести непосредственно из сгенерированной матрицы. В качестве альтернативы низкоразмерные представления признаков в скрытом слое извлекаются и передаются в классификатор SVM для окончательной классификации или в модель CNN для окончательной классификации.

Ограниченная машина Больцмана (Restricted Boltzmann Machine - RBM) имеет один скрытый слой для представления скрытых признаков и входной слой, кодирующий наблюдаемые данные. Для аннотации GO единицами входного слоя являются доступные термины GO. Zou, Wang и Yu (Zou, 2017) представили глубокий RBM (DRBM), в котором многослойные RBM обучаются и разворачиваются, что приводит к прогнозированию функции белка.

Наконец, порождающая состязательная сеть (Generative Adversarial Network - GAN) состоит из двух сетей: порождающей модели, которая создает синтетические реалистичные данные, и дискриминационной модели, которая оценивает, являются ли данные реальными или нет. В AFP модель GAN улучшает прогноз, создавая синтетические функции для контролируемых классификаторов или используя онтологические корреляции (Seyyedsalehi, 2021).

Подход на основе данных. Что касается входных данных, методы, основанные на глубинном обучении, в основном используют два подхода для присвоения терминов GO неизвестным генным продуктам. Одним из них является подход только к последовательности, который полезен для предсказания функций новых белков в отсутствие гомологичной информации или других ссылок. Второй основан на струк-

туре или иным образом использует большие данные из нескольких доступных ресурсов.

Модели на основе последовательностей (результаты и обсуждение). Одним из первых методов прогнозирования аннотаций GO, основанных на глубинном обучении, является метод, предложенный в (Chicco, 2014). Авторы сравнили два решения для аннотаций, tSVD и нейронную сеть AE. Форма вывода двух методов была одинаковой; однако последний показал лучшие результаты на шести разных наборах данных. Позже, вместо использования архитектуры AE, Zou, Wang и Yu (Zou, 2017) предложили DRBM для аннотирования продуктов генов четырех модельных видов: *Homo sapiens*, *S.cerevisiae*, *Musmusculus* и *Drosophila*.

ProLanGo – это первый инструмент, который применяет нейронный машинный перевод (Neural Machine Translation - NMT), разработанный Google в AFP. Аминокислотные последовательности и термины GO были преобразованы в языки «ProLan» и «GOLan» соответственно. Затем после трансляции с использованием модели с тремя слоями RNN была сгенерирована аннотация GO нового белка. DeepSeq — еще один метод присвоения терминов GO аминокислотным последовательностям, основанный на сверточных слоях. Однако авторы предсказали только пять наиболее частых онтологий MF для белков *H. sapiens*.

Используя архитектуру DNN, DEEPred (Rifaioglu, 2019) реализует стек многозадачных DNN с прямой связью. Каждая отдельная сеть построена для определенного уровня терминов GO в DAG, что позволяет выполнять иерархическую постобработку прогнозов. Белковые последовательности экспериментально представлены тремя типами дескрипторов (карта профиля подпоследовательности, состав псевдоаминокислот и функция объединенной триады), при этом карта профиля подпоследовательности лучше всего подходит для анализа.

DeepGOPlus был разработан для преодоления существующих ограничений DeepGO, таких как длина последовательности, недоступные функции PPI и количество меток GO. Функциональная аннотация предсказывается многослойной структурой CNN в сочетании со сходством последовательностей. TALE был разработан для создания прогнозов GO путем интеграции шаблонов последовательности на основе кодировщика преобразователя и совместного сходства термина последовательности. PFP-WGAN (Seyyedsalehi, 2021) – одна из последних идей, использующих GAN для определения функциональных возможностей белков. В то время как сеть генератора

обрабатывает необработанные последовательности, дискриминатор получает два входа; один состоит из аннотированных белков из базы данных SwissProt, а другой – из необработанных последовательностей с соответствующими аннотациями, синтезированными из генератора.

Интегрированные модели на основе данных/структуры. Функциональное назначение аминокислотных последовательностей трехмерных структур было предложено в (Tavanaei, 2016). В основном анализе использовалась модель CNN, которая была обучена и протестирована на пяти наборах данных белков человека, каждому из которых было присвоено два термина GO. Однако прогноз не распространялся на дерево DAG для унаследованных терминов GO. DeepGO – это функциональный сервер аннотаций, основанный на архитектуре CNN, который был разработан для изучения особенностей аминокислотных последовательностей и сети PPI. Метки GO назначаются входным белкам посредством иерархической классификации, структурированной в виде дерева DAG. DeepAdd был вдохновлен сервером DeepGO и предоставляет решение для AFP, используя структуру CNN для изучения векторных представлений из последовательностей и дополнительной информации. Кроме того, в DeepAdd добавляется профиль белковой последовательности (SSP), если информация о сети PPI недоступна. На основе аналогичной концепции объединения первичной структуры белка и PPI была построена новая модель GONET с использованием CNN, RNN и Уровня внимания (Attention layer) для последовательностей человека и мыши.

Работая с набором данных, идентичным тому, который был собран FFPred3, Fa et al. (Fa, 2018) протестировали MTDNN, в которой предсказание функции белка рассматривалось как проблема классификации с несколькими метками. Решение состояло из слоев, общих для всех задач (меток GO), которые сложены параллельно слоям, специфичным для задачи. DeepFunc – это новый предиктор, который превосходит DeepGO, FFPred3 и BLAST. Во-первых, информация о белковом домене, семействе и мотиве запрашивается у InterPro и кодируется перед прохождением через полностью связанные слои. Затем алгоритмом Deepwalk получают топологические характеристики PPI. Наконец, этот метод объединяет два типа признаков (последовательный и сетевой ввод), чтобы соответствовать FCDN. Архитектура DeepGOA более сложная, чем у DeepFunc. В дополнение к информации, генерируемой InterPro и PPI, глобальные и локальные семантические признаки аминокислотных последовательностей извлекаются с помощью Bi-LSTM и сверточного слоя соответственно.

Используя те же типы функций, SDN2GO использует три подмодели для каждого источника информации, при этом все выходные данные интегрированы в окончательную взвешенную модель.

В другом исследовании deepNF был создан с помощью мульти-модального глубокого АЕ для захвата скрытой информации в белках из различных типов сетей взаимодействия. DeepMNE-CNN имеет более высокую производительность, чем deepNF в человеческих данных, за счет использования слоев CNN вместо SVM для модели классификации. С другой стороны, FFPred-GAN (Wan, 2020) использует GAN для расширения возможностей традиционных моделей машинного обучения, особенно SVM. Реальные особенности — это биофизическая информация, извлеченная из необработанных аминокислотных последовательностей с помощью FFPred, а генератор использует скрытые переменные для увеличения синтетических образцов.

Выводы. Вычислительная GO-аннотация белков была активно решаемой и сложной задачей в биоинформатике примерно с 2000-х годов; это ответ на необходимость преодолеть разрыв между известными и неизвестными, недавно открытыми аминокислотными последовательностями.

С одной стороны, понимание функции белка необходимо для расшифровки биологической эволюции и для многочисленных задач, таких как разработка лекарств и лечение болезней. База данных GO облегчила всеобъемлющий словарь для функциональной аннотации, поскольку она представляет структурированные функции GO в трех доменах (MF, BP и CC), тем самым эффективно поддерживая назначение функций белков *in silico*. Здесь мы представили текущее состояние области и сравнили решения AFP на основе GO, которые классифицируются как традиционные подходы и подходы глубинного обучения.

Одной из основных трудностей в этой области является использование входных функций для достижения эффективной работы. Хотя гетерогенные входные данные полезны для предсказания GO, для большинства неаннотированных белков доступна только информация о последовательности. Что касается конкретных подходов, модели машинного обучения могут быть ограничены неоднородностью геномных данных при анализе различных источников информации, в то время как настройка параметров и гиперпараметров является сложным этапом подхода глубинного обучения. Что касается выходных данных, база данных GO обновляется, потому что аннотации GO все еще несбалансированы и не полны для всех видов. Его сложная структура идеально подходит для описания функциональных ролей

белков, но также делает задачу предсказания сложной задачей с несколькими метками.

Тем не менее, несмотря на существующие проблемы, поиск успешного аннотирования будет продолжаться во многих направлениях благодаря усилиям научного сообщества. Наряду с постоянным расширением баз данных -omics методы, основанные на данных, продемонстрировали превосходное применение в различных областях с многообещающей тенденцией в области функциональной аннотации. В зависимости от доступных источников они могут быть основаны только на интегрированных данных или информации о последовательности, учитывая гибридный подход. Были предложены и разработаны некоторые подходы для работы с несбалансированными данными; они включают работу над подгруппами, в которых классы более сбалансированы, увеличение данных с использованием GAN и рассмотрение показателей оценки, указанных для несбалансированных данных, таких как AUPR. Проблема с несколькими метками может быть решена за счет продвижения вычислительных ресурсов и четко определенных решений, например, путем объединения множества отдельных решений. Наконец, успешные решения для прогнозирования срока ГО могут быть расширены за счет включения других функциональных ресурсов (ЕС, путей и т. д.), чтобы понять биологическую роль и потенциал белков в науке о жизни.

Information about authors:

Golenko Y.S. – Doctoral Student, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; golenko.katerina@gmail.com; <https://orcid.org/0000-0002-4643-4571>;

Ismailova A.A. – Ph.D., Associate Professor, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; a.ismailova@mail.ru; <https://orcid.org/0000-0002-8958-1846>;

Zhumakhanova A.S – Master's degree, L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan; zhumanar@gmail.com; <https://orcid.org/0000-0003-4128-2107>;

REFERENCES

Bonetta R., Valentino G. (2020) Machine learning techniques for protein function prediction, *Proteins: Structure, Function, and Bioinformatics*, 88(3):397–413. DOI: 10.1002/prot.25832.

Chicco D., Sadowski P., Baldi P. (2014) Deep autoencoder neural networks for gene ontology annotation predictions, In: *Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics*, 533–540.

Consortium GO. (2015) Gene ontology consortium: going forward, *Nucleic Acids Research*, 43(D1):D1049–D1056. DOI: 10.1093/nar/gku1179.

Du Z., He Y., Li J., Uversky V.N. (2020) Deepadd: protein function prediction from k-merembedding and additional features, *Computational Biology and Chemistry*, 89:107379. DOI: 10.1016/j.compbiolchem.2020.107379.

Fa R., Cozzetto D., Wan C., Jones D.T. (2018) Predicting human protein function with multi-task deep neural networks, *PLOS ONE* 13(6):e0198216. DOI: 10.1371/journal.pone.0198216.

Gong Q., Ning W., Tian W. (2016) Gofdr: a sequence alignment-based method for predicting protein functions, *Methods*, 93:3–14. DOI: 10.1016/j.ymeth.2015.08.009.

Kourmpetis Y.A., Van Dijk A.D., Bink M.C., Van Ham R.C., ter Braak C.J. (2010) Bayesian markov random field analysis for protein function prediction based on network data, *PLOS ONE*, 5(2):e9293. DOI: 10.1371/journal.pone.0009293.

Li Y., Huang C., Ding L., Li Z., Pan Y., Gao X. (2019) Deep learning in bioinformatics: introduction, application, and perspective in the big data era, *Methods*, 166:4–21. DOI: 10.1016/j.ymeth.2019.04.008.

Nariai N., Kolaczyk E.D., Kasif S. (2007) Probabilistic protein function prediction from heterogeneous genome-wide data, *PLOS ONE*, 2(3):e337. DOI: 10.1371/journal.pone.0000337.

Pinoli P., Chicco D., Masseroli M. (2015) Computational algorithms to predict gene ontology annotations, *BMC Bioinformatics*, 16(6):1–15.

Piovesan D., Giollo M., Leonardi E., Ferrari C., Tosatto S.C. (2015) Inga: protein function prediction combining interaction networks, domain assignments and sequence similarity, *Nucleic Acids Research*, 43(W1):W134–W140. DOI: 10.1093/nar/gkv523.

Rifaioğlu A.S., Doğan T., Martin M.J., Cetin-Atalay R., Atalay V. (2019) Deepred: automated protein function prediction with multi-task feed-forward deep neural networks, *Scientific Reports*, 9(1):1–16.

Rost B., Liu J., Nair R., Wrzeszczynski K.O., Ofra Y. (2003) Automatic prediction of protein function, *Cellular and Molecular Life Sciences*, 60(12):2637–2650. DOI: 10.1007/s00018-003-3114-8.

Seyyedsalehi S.F., Soleymani M., Rabiee H.R., Mofrad M.R. (2021) Pfp-wgan: protein function prediction by discovering gene ontology term correlations with generative adversarial networks, *PLOS ONE*, 16(2):e0244430. DOI:10.1371/journal.pone.0244430.

Shehu A., Barbará D., Molloy K. (2016) A survey of computational methods for protein function prediction, Cham, Switzerland: Springer International Publishing, 225–298.

Spalević S., Veličković P., Kovačević J, Nikolić M. (2020) Hierarchical protein function prediction with tails-gnns, *ArXiv preprint. arXiv:2007.12804*.

Tavanaei A., Maida A.S., Kaniyattam A., Loganantharaj R. (2016) Towards recognition of protein function based on its structure using deep convolutional networks, In: 2016 IEEE international conference on bioinformatics and biomedicine (BIBM), Piscataway: IEEE, 145–149.

Wan C., Jones D.T. (2020) Protein function prediction is improved by creating synthetic feature samples with generative adversarial networks, *Nature Machine Intelligence*, 2(9):540–550. DOI: 10.1038/s42256-020-0222-1.

Zhao Y., Wang J., Chen J., Zhang X., Guo M., Yu G. (2020) A literature review of gene function prediction by modeling gene ontology, *Frontiers in Genetics*, 11:400. DOI: 10.3389/fgene.2020.00400.

Zou X., Wang G., Yu G. (2017) Protein function prediction using deep restricted Boltzmann machines, *BioMed Research International*, 2017:1729301.

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

Volume 2, Number 342 (2022), 39–53

<https://doi.org/10.32014/2022.2518-1726.128>

UDC 004.4: 551.481.1 (574)

**R.M. Moldasheva^{1*}, A.A. Ismailova¹, A.K. Zhamangara²,
A.M. Zadagali³**

¹S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; ²Astana botanical garden, a branch RSE REM «Institute of botany and phytointroduction» CFW MEGNR RK, Nur-Sultan, Kazakhstan;

³L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan.
E-mail: raushan85_07@mail.ru

ABOUT DEVELOPMENT OF AN INFORMATION ANALYTICAL SYSTEM FOR THE STUDY OF AQUATIC ECOSYSTEMS

Abstract. The water bodies under study are habitats for many species of living organisms that provide the most important property of aquatic ecosystems - the ability to self-clean. If the anthropogenic load does not exceed certain limits, the water body copes with it. If this limit is exceeded, the aquatic ecosystem degrades. At the same time, the ability to self-clean is also reduced. Taking care of the stability of the aquatic ecosystem, humans also satisfy their needs for a high-quality water resource.

Based on the quantitative characteristics of phytoplankton and phytobenthos, the bioproductivity of the aquatic ecosystem is calculated. Data on the quantity of algae and their biomass and other quantitative indicators (pigments, proteins, fats, carbohydrates, the number of nucleic acids, ash elements, respiratory rate, photosynthesis, etc.) are used for recalculation.

The development of an information-analytical system for studying aquatic ecosystems with a set of mathematical models makes it possible to carry out quantitative studies of fundamental environmental problems, including changes in the ecological state of aquatic ecosystems (by

example of lakes in Kazakhstan). The results of model calculations allow estimating the impact of anthropogenic factors on the natural condition of aquatic ecosystems.

The development of mathematical methods of environmental forecasting is necessary both for the optimization of environmental management and for the serious scientific substantiation of programs in the field of environmental quality management and nature protection. Mathematical models of aquatic ecosystems allow us to describe non-equilibrium dynamic processes in hydrobiocenoses under various external influences, such as changes in water temperature, surface illumination, water or biogenic load, meteorological conditions. The construction of such models requires a large amount of information about the parameters of geophysical, geochemical, biological and other natural processes, the comprehension of which is impossible without the involvement of modern information technologies.

Key words: mathematical model, information-analytical system, aquatic ecosystem, database, phytoplankton, biomass.

**Р.Н. Молдашева^{1*}, А.А. Исмаилова¹, А.К. Жамангара²,
А.М. Задағали³**

¹С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Нұр-Сұлтан, Қазақстан;

²Астана ботаникалық бағы - ҚР ЭГТРМ ОШЖДК «Ботаника және
фитоинтродукция институты» ШЖҚ РМК филиалы,
Нұр-Сұлтан, Қазақстан;

³Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан.

E-mail: raushan85_07@mail.ru

СУ ЭКОЖҮЙЕЛЕРІН ЗЕРТТЕУДІҢ АҚПАРАТТЫҚ ТАЛДАУ ЖҮЙЕСІН ӘЗІРЛЕУ

Аннотация. Зерттеу объектісі болып табылатын су объектілері су экожүйелерінің маңызды қасиеті – өзін-өзі тазарту қабілетін қамтамасыз ететін тірі организмдердің көптеген түрлерінің мекендейтін орны болып табылады. Егер антропогендік жүктеме белгілі бір шектерден аспаса, су объектісі онымен күресе алады. Егер бұл шектен асып кетсе, су экожүйесінің тозуы орын алады. Бір мезгілде

өздігінен тазару қабілеті төмендейді. Су экожүйесінің тұрақтылығына қамқорлық жасай отырып, адам сапалы су ресурсына деген қажеттіліктерін қанағаттандыруды қамтамасыз етеді.

Фитопланктон мен фитобентостың сандық сипаттамалары негізінде су экожүйесінің биологиялық өнімділігі есептеледі. Балдырлар саны және олардың биомассасы туралы деректер және басқа да сандық көрсеткіштер (пигменттер, акуыздар, майлар, көмірсулар, нуклеин қышқылдарының саны, күл элементтері, тыныс алу жылдамдығы, фотосинтез және т.б.) қайта есептеу үшін пайдаланылады.

Математикалық модельдер кешенімен су экожүйелерін зерттеудің ақпараттық талдау жүйесін әзірлеу іргелі экологиялық проблемаларға, оның ішінде су экожүйелерінің экологиялық жай-күйінің өзгеруіне (Қазақстан көлдері мысалында) сандық зерттеулер жүргізуге мүмкіндік береді. Модельдер бойынша есептеу нәтижелері антропогендік факторлардың су экожүйелерінің табиғи жағдайына әсерін бағалауға мүмкіндік береді.

Экологиялық болжаудың математикалық әдістерін дамыту табиғатты пайдалануды оңтайландыру үшін де, қоршаған ортаның сапасын реттеу және табиғатты қорғау саласындағы бағдарламаларды маңызды ғылыми негіздеу үшін де қажет. Су экожүйелерінің математикалық модельдері әртүрлі сыртқы әсерлердегі гидробиоценоздардағы әркелкілігі динамикалық процестерді сипаттауға мүмкіндік береді, мысалы, су температурасының өзгеруі, оның бетінің жарықтандырылуы, су немесе биогендік жүктеме, метеорологиялық жағдайлар. Мұндай модельдерді құру геофизикалық, геохимиялық, биологиялық және басқа да табиғи процестердің параметрлері туралы ақпараттың үлкен көлемін қажет етеді, оларды қазіргі заманғы ақпараттық технологияларды тартпай түсіну мүмкін емес.

Түйін сөздер: математикалық модель, ақпараттық-аналитикалық жүйе, су экожүйесі, деректер қоры, фитопланктон, биомасса.

**Р.Н. Молдашева^{1*}, А.А. Исмаилова¹, А.К. Жамангара²,
А.М. Задағали³**

¹Казахский агротехнический университет им. С. Сейфуллина,
Нур-Султан, Казахстан;

²Астанинский ботанический сад – филиал РГП на ПХВ «Институт
ботаники и фитоинтродукции» КЛХЖМ МЭГПР РК,
Нур-Султан, Казахстан;

³Евразийский национальный университет им. Л.Н. Гумилева,
Нур-Султан, Казахстан.

E-mail: raushan85_07@mail.ru

К РАЗРАБОТКЕ ИНФОРМАЦИОННОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ ИССЛЕДОВАНИЯ ВОДНЫХ ЭКОСИСТЕМ

Аннотация. Водные объекты, являющиеся объектом исследования, являются местообитанием множества видов живых организмов, которые обеспечивают важнейшее свойство водных экосистем - способность к самоочищению. Если антропогенная нагрузка не превышает определенных пределов, водный объект справляется с ней. Если этот предел превзойден, происходит деградация водной экосистемы. Одновременно снижается и способность к самоочищению. Заботясь об устойчивости водной экосистемы, человек обеспечивает и удовлетворение своих потребностей в качественном водном ресурсе.

На основе количественных характеристик фитопланктона и фитобентоса рассчитывается биопродуктивность водной экосистемы. Данные о количестве водорослей и их биомассе и другие количественные показатели (пигменты, белки, жиры, углеводы, количество нуклеиновых кислот, элементы золы, скорость дыхания, фотосинтез и др.) используются для пересчета.

Разработка информационно-аналитической системы исследования водных экосистем с комплексом математических моделей позволяет проводить количественные исследования фундаментальных экологических проблем, в том числе изменения экологического состояния водных экосистем (на примере озер Казахстана). Результаты вычислений по моделям позволяют оценивать влияние антропогенных факторов на естественное состояние водных экосистем.

Развитие математических методов экологического прогнозирования необходимо как для оптимизации природопользования, так и для

серьезного научного обоснования программ в области регулирования качества окружающей среды и охраны природы. Математические модели водных экосистем позволяют описывать неравновесные динамические процессы в гидробиоценозах при различных внешних воздействиях, например, таких как изменение температуры воды, освещенности ее поверхности, водной или биогенной нагрузки, метеорологических условий. Построение такого рода моделей требует большого объема информации о параметрах геофизических, геохимических, биологических и других природных процессов, осмысливание которых невозможно без привлечения современных информационных технологий.

Ключевые слова: математическая модель, информационно-аналитическая система, водная экосистема, база данных, фитопланктон, биомасса.

Introduction. Kazakhstan has the opportunity to overcome the shortage of water resources by 2040 through the upgrade and development of infrastructure at a reasonable cost of \$18-20 billion, self-financing measures for the effective use of water resources, as well as the development of municipal infrastructure worth \$35 billion. By rigorously implementing deficit-reduction measures, the country can prevent adverse effects and become one of the pioneers of responsible water management among the neighboring States of Central Asia and other countries of the world with limited water resources. Industrial wastewater quality standards were developed in the Soviet era without taking into account economic factors; in addition, they are difficult to comply with and are much more complicated than EU standards. In some cases, the standards contradict each other (Integrated Environmental Permit for Activities and MPC). The norms that set the standards for wastewater quality suggest to go from the local conditions, which is a good practice, but in fact it is difficult to implement, as it requires environmental, physical and chemical data for all discharge locations (river, lake, water reservoir, etc.) and a detailed monitoring scheme tailored to local conditions, as well as a clear methodology for determining the required quality of wastewater depending on local conditions. In addition, the system used in world practice for online monitoring of the quality of discharges from the largest enterprises in Kazakhstan is virtually absent (Naydenko et al., 2003: 186).

Research materials and methods. It is necessary to create a system for monitoring water bodies – a system for monitoring the state of water bodies

and the sources of harmful effects on them. The task of environmental monitoring of reservoirs includes assessment of their current state, tracking of evolution, and forecasting as a result of anthropogenic impact (Fyodorov, 1975). It is important to preserve not just water, but the aquatic environment, with its flora and fauna, that is, aquatic ecosystems.

It was proposed to refer environmental monitoring to the system of repeated observations of one or more elements of the natural environment in space and time for certain purposes in accordance with a prearranged program.

The current system of environmental monitoring, carried out by both scientific institutions and republican regulatory bodies, is of low efficiency not only because of the poor technical resources but also, to a large extent, due to disregard of modern methods of data management and comprehensive mathematical processing of the results of multivariate observations. The wealthiest information in the hydrochemistry of natural water systems, the information on hydrobiological observations remains cast off and is lost every year. It is obvious that in addition to the traditional poorly informative reports on the part of indicators exceeding the MPC, these data could be successfully used to build both local models of seasonal and long-term dynamics of reservoirs, and generalized models of rational environmental and economic development of territorial complexes (Shitikov et al., 2003: 463).

According to the Water Code of the Republic of Kazakhstan (Art.60, paragraph 2): «State monitoring of water bodies is a system of regular observations of hydrological, hydrogeological, hydrogeochemical, sanitary-chemical, microbiological, parasitological, radiological and toxicological indicators of their state, collection, processing, and transmission of the obtained information in order to timely identify negative processes, assess and predict their development, develop recommendations for the prevention of harmful effects and determine the degree of effectiveness of water management measures».

Results. The specific nature of the tasks of monitoring complex objects imposes certain requirements, which, on the basis of the above considerations, can be expressed as follows:

- in the process of monitoring, knowledge of the monitored object should be ensured;
- the monitoring task should be solved on the basis of all available a priori information about the monitored object and its operating environment and information obtained during the measurement experiment;

- all the obtained intermediate and final monitoring results should have a metrological substantiation in the form of their quality indicators;
- to effectively solve the tasks of monitoring the monitoring objects (MO) and operational environment (OE) model, as well as methods, algorithms, and monitoring tools should reflect the main properties of complex objects, in particular, the properties of stability, self-development, and self-organization;
- monitoring results should be presented in the forms of parameter values, analytical expressions, conclusions and solutions in semantic form.

Environmental monitoring scheme is shown in Figure 1 (General, 1975: 471).

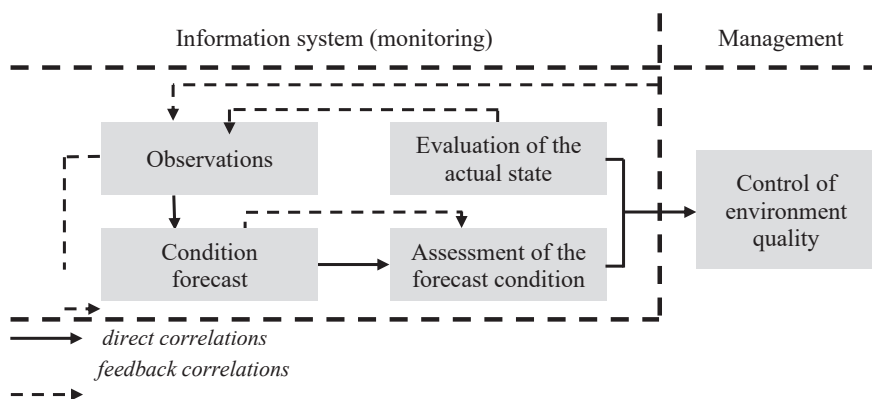


Figure 1 – Environmental monitoring flowchart

In monitoring aquatic ecosystems, the following activities should be identified:

- distinguishing of the observation object;
- survey of the distinguished observation object;
- building of an information model for the observation object;
- forecasting changes in the state of the observation object;
- presentation of information in a form convenient for the user to perceive.

In order to ensure the main objectives of environmental monitoring of aquatic ecosystems, within this work an information and analytical system for studying aquatic systems is developed. Information and analytical systems are a special class of information systems designed for analytical processing of data, their dynamic presentation, analysis of historical and current data, analysis of trends, modeling and forecasting of performance results. It is now generally accepted that the management of data from

multidisciplinary environmental studies in large regions requires new approaches and the creation of new data flow management structures (Martin, 1980: 662). The planned research is aimed at solving these problems.

The main functions of the information-analytical system are:

- extracting data from various sources, converting them and uploading them to the repository;
- data storage;
- data analysis, including operational and intellectual ones;
- preparation of the results of operational and intellectual analysis for their effective perception by the consumers.

Currently, IAS is one of the most popular classes of software in the custom software market, this is due to the global process of automating the information activities of entire industries. At the same time, in addition to the general problems of organization and management of their arrangement, they have peculiarities, requirements and problems of a more specific nature, arising from the specific nature of the functional purpose of the IAS, as well as the conditions in which the IAS is created and used.

The most important task of unified environmental monitoring is not only the collection of information, but also its rational storage, processing and presentation. IAS for Aquatic Ecosystem Studies should process a large number of indicators that need to be monitored. First of all, it is the data of hydrochemical observations. The number of components for which MPC standards are established is about 2000 substances. Hydrobiological monitoring data should reflect information on aquatic organisms living in water bodies and characterizing their environmental state – more than a hundred species. Hydrological and hydrophysical monitoring data reflect water level information and a number of important parameters (temperature, water turbidity, pH, etc.) that directly affect the state of aquatic ecosystems.

The relevance of the IAS creation to study aquatic ecosystems is related to the need to regularly receive, integrate and use significant amounts of information on the state of aquatic ecosystems of Kazakhstan, which is distributed both in time (historical and current) and in space.

From a technical point of view, the IAS is a set of procedures, methods, and regulations that lead to the regular planned collection, storage, analysis, and provision of information used for making management decisions (Boyko et al, 1989: 351, Yakubaitis, 1996: 368, Singer et al, 1987: 208).

There are two fundamentally different approaches to the design of

information and analytical systems (IAS) for environmental management and protection. In the first of them, the system is created on the basis of known types of territorial objects (rivers, lakes, air basin, anthropogenic environment) and the relations existing between them. The second approach is to solve specific problems of information management (Belov, 2005:111, Voronoi et al, 2007:194, Galliani, et al, 1982:737).

The general work plan for designing the Aquatic Ecosystem Study IAS is shown in Figure 2:

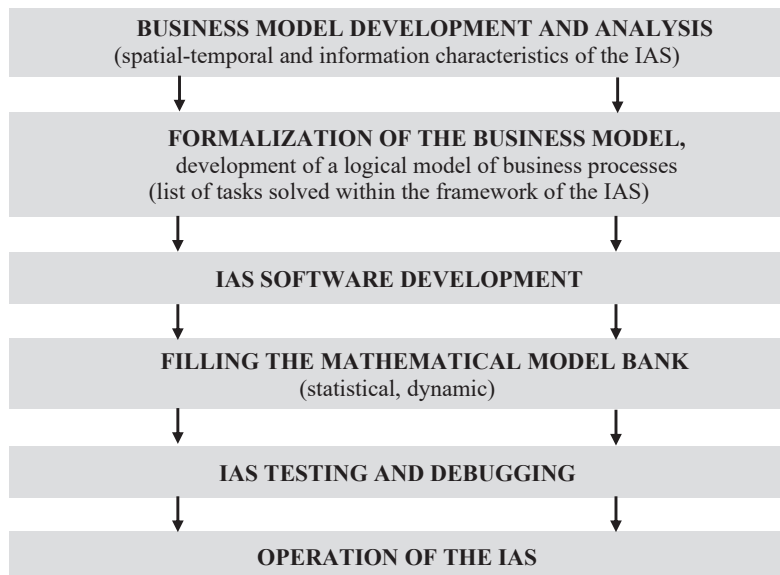


Figure 2 – IAS Designing Steps Sequence

Designing of the IAS should be aimed at solving the tasks of information management and research, for example, monitoring of aquatic ecosystems, all data should be oriented to the «problem» in focus.

The use of information systems in the field of environment (Environmental Informatics, geoinformatics) has been intensively developing in Europe, the USA, Japan in recent years, as evidenced by the large amounts of funding allocated to these studies, the huge flow of publications, regularly held scientific conferences, as well as the inclusion of environmental informatics in the curricula of various departments of universities. NSF (USA) included the creation of geographic information platforms and the development of information infrastructure for them in the list of priority funded R&D. To illustrate the magnitude of the costs, back in 1981, the United States spent about \$500 million, or about 2-2.5

% of total environmental costs, on monitoring and monitoring-based water quality management (Matsyashek, 2002: 432).

The main purpose of creating an information-analytical system for studying aquatic ecosystems (IAS) (through the example of lakes in Kazakhstan) is to implement at a qualitatively new level a procedure for analytical processing and interpretation of the results of observations to forecast supposed changes in the aquatic environment, as well as the organization of centralized storage of heterogeneous (hydrochemical and hydrobiological, etc.) data.

The purpose of such a system is not only the accumulation and visualization of monitoring data, but also the creation of a single information space and the provision of extensive system analysis of information for effective environmental quality management and ensuring the safety of the population (State Program, 2014, Matveyev et al., 2005: 96). To do this, it is necessary to establish a relationship between different types of indicators, not only purely correlative, but also causal, formalized in the form of mathematical models, which, in turn, are used to form control scenarios.

Modeling of microbial communities of plant organisms. The functioning of aquatic ecosystems is largely determined by the lower trophic levels. The biological productivity of systems is based on the productivity of phytoplankton (Alekseyev et al., 1992: 366, Abrosov et al., 1988: 333). The study of phytoplankton is an important and interesting task. Communities of microorganisms and, in particular, phytoplankton, are studied quite intensively, including by means of mathematical modeling (Riznichenko et al., 1993: 301, Jorgensen et al., 2005: 20, Adamovich et al., 2005: 5).

Essential for phytoplankton is its spatial distribution, which has a high degree of heterogeneity (Riznichenko et al., 1993: 168, Shushkina et al., 1997: 278). This heterogeneity is determined both within and between species relations in phytoplankton and environmental conditions. Among the influencing factors, experts distinguish mineral nutrition, that is, those mineral substances used by the plant organism to create organic matter in the process of photosynthesis (Riznichenko et al., 1993: 186, Silkin et al., 1988: 230,). The list of such substances is extensive, but the main role is played by compounds based on carbon, nitrogen, phosphorus and silicon (Abrosov et al., 1988: 322, Adamovich et al., 2005: 8). Mathematical models of the dynamics of microorganism biomasses based on the study of the process of consumption of mineral substances have different nature (Riznichenko et al., 1993: 201, Alekseyev et al., 1992: 185, Abrosov et al.,

1988: 142, Abakumov, 1994: 107). We address models of mineral nutrition of phytoplankton based on the concept of «cell quota» of M. Drupa (Droop, 1974: 198).

Discussion. When studying such large reservoirs as Lake Balkhash, satellite remote sensing data of the surface of the lake can be used (Shushkina et al, 1997: 102). A large part of the problems considered with the use of this material is related to the living component of the aquatic environment. Satellite information contains data on chlorophyll concentration and fluorescence, illumination and temperature, water turbidity, and a number of other characteristics of the surface layer of the reservoir (Silkin, 2011: 24, Barbini, 2004: 2095). These data are refined and supplemented by other remote and contact methods (Silkin, 2011: 122).

In mathematical models of the dynamics of biomass of microbiological communities, the properties of solutions are investigated. The model of a system with the internal flow is used to study the phytoplankton community in the aquatic ecosystem. The properties of solutions coincide with the dynamics of natural communities, demonstrating the dominance of individual species in the community.

Assessment of the biological productivity of ecosystems is important for studying the state of the natural environment and the possibilities of environmental management. For aquatic ecosystems, biological productivity can be evaluated based on phytoplankton productivity (Jorgensen, 1985: 160). The productivity of phytoplankton is largely determined by the process of consumption of mineral substances in the construction of a plant organism during photosynthesis (Riznichenko et al, 1993: 217). Remote sensing data from the surface of the seas and oceans are currently playing an important role in the study of the state and functioning of phytoplankton. In particular, artificial Earth satellites provide data on the content of minerals and chlorophyll in the surface layer. Data on chlorophyll (primarily chlorophyll «a») make it possible to assess the content of phytoplankton and give a rough estimate of primary production (Shushkina et al, 1997: 137). Data on mineral substances (based on nitrogen, phosphorus, silicon and other chemical elements), constituting the material basis for the construction of plant organisms in the process of photosynthesis, make it possible to assess the characteristics of the production processes of phytoplankton (Droop, 1974: 13). At this stage, mathematical models of the dynamics in the populations (biomass) of the main species of the phytoplankton community are useful (Riznichenko et al, 1993: 220).

Similar mathematical models are also used in describing the dynamics of microbial cultures in laboratory experiments. The paper presents groups of models of dynamics of biomass communities of microorganisms. The models are based on systems of differential equations. The qualitative properties of solutions at the structural level were investigated.

Models of functioning of phytoplankton communities describe the dynamics of the transformation of substances in the photosynthesis and construction of the plant organism. The models focus on biological species of phytoplankton and groups of mineral nutrients. Phytoplankton is represented by m species, their content in the medium is indicated by y_i for i type. Mineral nutrition of plant organisms is divided into n groups of similar substances (based on nitrogen, phosphorus, silicon, etc.). In the considered models, nutrients are assumed to be not interchangeable. The content of substances in the group j in the medium is denoted as z_j .

For a living organism, a strategy of activity is determined not only by the environment, but also by its state. The internal state of the body can be characterized in different ways. In our case, it is proposed to use the intracellular content of nutrients based on mineral compounds in the external environment as an indicator.

The content of nutrients in the group j in a cell of the i type is denoted as q_{ij} . This value is called the cellular quota. The growth rate of a separate species is determined on the basis of the Libich principle (Jorgensen, 1985: 206): it is limited by the growth rate of the least productive mineral substance. The consumption of nutrients by microorganisms is carried out at a specific rate v_{ij} , and the growth of plant biomass occurs at a specific rate $\mu_{ij}(q_{ij})$ depending on the vector $z = (z_j)_j^n = 1$, the content of minerals in the external environment and the matrix $q = (q_{ij})_j^n = 1$ a content of nutrients in plant cells. The mass dynamics model of the system has the following shape (1) (Riznichenko et al, 1993: 207):

$$\begin{aligned} \frac{\partial y_i}{\partial t} &= (\mu(q_i) - D)y_i \\ \frac{\partial z_j}{\partial t} &= (D(z_{j0} - z_j) - \sum_{i=1}^m v_{ij}(z_j, q_{ij})y_i) \text{ для } i \\ &= 1, \dots, n. \\ \frac{\partial q_{ij}}{\partial t} &= v_{ij}(z_j, \mu_i(q_i) \cdot q_{ij}) \end{aligned} \quad (1)$$

q_i denotes the vector $q = (q_{ij})_j^n = 1 \mu_i(q_i)$ the function is calculated by the formula $\mu_i(q_i) = \min \mu_{ij}(q_{ij})$. Parameter D indicates the flow rate of the substance in the system, by z_0 - the content of mineral nutrients in the incoming stream, by $v_{ij}(z_j)$ - the specific rates of absorption of the substance of the group j by organisms of type i .

The model is used by us for calculations using hydrochemical data of the characteristics of the phytoplankton community in the reservoir.

Conclusion. In conclusion, it should be noted that the purpose of the paper was to justify the creation of an Information and Analytical System for the Study of Aquatic Ecosystems, as well as in the systematization and analysis of dynamic models of biological systems that are applied to aquatic plant microorganism communities. Such communities are represented by continuous cultures in the laboratory and phytoplankton communities in natural aquatic ecosystems. We gave examples of the application of these models and the properties of their solutions, an assessment of the effectiveness of their use and the prospects for expanding this research area.

Information about authors:

Moldasheva Raushan Nurkozhaevna – Doctoral student of the Department of «Information Systems» S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-9940336, raushan85_07@mail.ru, 0000-0002-4570-0487;

Ismailova Aisulu Abzhapparovna – PhD, Associate Professor of the Department «Information Systems» S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-4606049, a.ismailova@mail.ru, 0000-0002-8958-1846;

Zhamangara Aizhan Kashagankyzy – Candidate of Biological Sciences, Deputy director for science of the Astana botanical garden branch of the RSE REM «Institute of botany and phytointroduction» CFW MEGNR RK, Nur-Sultan, Kazakhstan; Personal phone number: 8-707-8485060, Kashagankizi@mail.ru, 0000-0002-2348-1711;

Zadagali Aizhan Meirangalikyzy – Doctoral Student of the Department of «General Biology and Genomics» L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan; Personal phone number: 8-701-6677064, z.a.aizhan1993@gmail.com, 0000-0002-2537-3538.

REFERENCES

- Alekseyev V.V., Kryshev I.I., Sazykina T.G. (1992). Physical and Mathematical Modeling of Ecosystems. St. Petersburg: Gidrometeoizdat, 185-366 p.
- Abrosova N.S., Bogolyubov A.G. (1988). Ecological Genetic Patterns of Coexistence and Coevolution of Species. Novosibirsk: Nauka, 142-333 p.
- Adamovich V.V., Rogozin D.Yu., Degermenji A.G. (2005). Search for the Regulation Criterion in the Continuous Culture of Microorganisms // *Microbiology*. – Vol. 74, No. 1. – P. 5-16.
- Abakumov A.I. (1994). Model Studies of Aquatic Ecosystems // *Bulletin of the Department of Water Resources of the RAS. Vladivostok*, No. 5-6. – P. 107-111.
- Boyko V.V., Savinnikov V.M. (1989). Designing Databases of Information Systems. – ed. 2-e., revised and enlarged – M.: Finansy i Statistika, 351 p.
- Belov V.S. (2005). Information-Analytical Systems. Fundamentals of Design and Application: Textbook. Moscow: Moscow State University of Economics, Statistics and Informatics. 111 p.
- Barbini R., Colao F., Dominicis L., Fantoni R., Fiorani L., Palucci A., Artamonov E.S. (2004). Analysis of Simultaneous Chlorophyll Measurements by Lidar Fluorosensor, Modis and SeaWiFS// *International Journal for Remote Sensing*. Vol. 25, No. 11. – P. 2095-2110.
- Droop M.R. (1974). The Nutrient Status of Algal Cells in Continuous Culture // *J. Mar. Biol. Assoc.* – 1974. – Vol.54. – P. 13-855.
- General-Purpose Information Systems (DBMS analytical review)/ed. Ye.L. Yushchenko. (1975). M.: Statistika, 471p.
- Jorgensen S.E., Loffler H., Rast W., Straskraba M. (2005). Lake and Reservoir Management // Elsevier. No. 20 – 512 p.
- Galliani G., Lanzi E.A. (1982). Data Base for Environmental Management. In: *Env. Syst. Anal. And Manag* / ed. By S. Rinaldi // North-Holland Publ. Comp. IFIP. – P. 737-753.
- Jorgensen S.E. (1985). Management of Lake Systems. Moscow: Agropromizdat, 160-206p.
- Martin J. (1980). Organization of Databases in Computing Systems. Moscow: Mir, 662 p.
- Matsyashek L.A. (2002). Requirements Analysis and System Design. Development of Information Systems Using UML / translated from English - M.: Williams, 432 p.
- Matveyev A.V., Kotov V.P., Mushkudiani M.I. (2005). Application of Information Technologies in Habitat Management: Textbook. St. Petersburg: GUAP, 96 p.
- Naydenko V.V., Gubanov L.N. (2003). Ecological and Economic Monitoring of the Environment: Textbook. Nizhny Novgorod, 186 p.
- Riznichenko G. Yu., Rubin A.B. (1993). Mathematical Models of Biological Production Processes. Moscow: MSU Publishing House, 301 p.
- Shitikov V.K., Rosenberg G.S., Zinchenko T.D. (2003). Quantitative Hydroecology: Methods of Systemic Identification. - Togliatti: IEVB RAS, 463 p.
- Singer I.S., Kruglikov B.I., Sadovnikov V.I. (1987). Information Support in Organizational Management Systems. Moscow: Nauka, 208 p.
- State Program on Water Resources Management of the Republic of Kazakhstan for 2014-2040// (2014). <http://ecoinfo.iacoos.kz/wp-content/uploads/t8.pdf>.
- Silkin V.A., Khailov K.M. (1988). Bioecological Control Mechanisms in Aquaculture. Moscow: Nauka, 230 p.
- Silkin V.A., Abakumov A.I., Pautova L.A., Mikaelian A.S., Chasovnikov V.K., Lukasheva T.A. (2011). Coexistence of the Black Sea and Invasive Species in the Phytoplankton of the

North-Eastern Part of the Black Sea. Analysis of Settlement Hypotheses// Russian Journal of Biological Invasions. No.3. – P. 24-122.

Shushkina E.A., Vinogradov M.E., Gagarin V.I., Dyakonov V.Yu., Lebedeva L.P., Nezlin N.P. (1997). Assessment of Productivity, Metabolic Rate, Trophodynamics, as Well as Reserves of Planktonic Organisms in Diverse Areas of the Ocean on the Basis of Satellite and Expeditionary Observations //RFBR newsletter. Vol.5, No. 4. – C.102-278.

Voronoi A., Manko P. (2007). Information and Analytical Systems in Economics // Marketing and marketing research. - No. 03 (69). – P. 194-206.

Yakubaitis E.A. (1996). Information Networks and Systems: Reference Book. Moscow: Finansy i Statistika, 368 p.

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

Volume 2, Number 342 (2022), 54–63
<https://doi.org/10.32014/2022.2518-1726.129>

UDC 004.75

A.A. Myrzatay^{1*}, L.G. Rzayeva², G. Abitova³, M.A. Zhakenov⁴

¹Korkyt-ata Kyzylorda University, Kyzylorda, Kazakhstan;

²L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan.

³Astana IT University, Nur-Sultan, Kazakhstan;

⁴Humanitarian and Technical Academy, Kokshetau, Kazakhstan.

E-mail: mirzataitegiali@gmail.com

THE IMPLEMENTATION AND THE USE OF THE LAN MONITORING SYSTEMS FOR SYSTEMATISATION OF THE INPUT DATA OF THE INCIDENT FORECASTING SYSTEMS

Abstract. This article explores the work on the implementation of an automated LAN monitoring systems and the feasibility of the utilisation of the data produced by these systems. The purpose of this article is to identify a systematic approach to the problem identification, to model business-processes according to the set goals and to determine functions of the implemented high-tech solutions. In the process of writing the article, the authors reviewed of local computer networks and server infrastructures of five universities: L.N. Gumilyov Eurasian National University, Astana IT University, Seifullin KazATU, KazGUU and Korkyt Ata Kyzylorda University and found out that despite all efforts to modernize the server and network infrastructure in the compared universities, most of them do not have LAN monitoring systems. Authors of this article emphasise the necessity of monitoring and data storage systems for the optimisation of the functioning of the LAN maintenance services, as well as, subsequent introduction of other systems, such as forecasting systems. As for the implementation of systems using machine-learning algorithms, so far none of the reviewed Higher Education Institutions are planning such projects in the near future. Nevertheless, such projects have every reason for research

in various areas, including automation and prediction of LAN system failures. Authors emphasize that data collection by monitoring systems can significantly facilitate further work in the above areas.

Key words: network equipment' failure forecasts, local area network, monitoring systems, LAN, business-processes modelling.

А.А. Мырзатай^{1*}, Л.Г. Рзаева², Г. Абитова³, М.А. Жакенов⁴

¹Қорқыт ата атындағы Қызылорда университеті,
Қызылорда, Қазақстан;

²Еуразия ұлттық университеті. Л.Н. Гумилев.
Нұр-Сұлтан, Қазақстан;

³Astana IT University, Нұр-Сұлтан, Қазақстан;

⁴Гуманитарлық-техникалық академия, Көкшетау, Қазақстан.

E-mail: mirzataitegiali@gmail.com

ОҚИҒАЛАРДЫ БОЛЖАУ ЖҮЙЕЛЕРІНІҢ КІРІСТЕРІН ЖҮЙЕЛЕУ ҮШІН LAN МОНИТОРИНГ ЖҮЙЕСІН ЕНГІЗУ ЖӘНЕ ПАЙДАЛАНУ

Аннотация. Бұл мақалада жергілікті желіні бақылаудың автоматтандырылған жүйелерін енгізу және осы жүйелер арқылы алынған деректерді пайдалану мүмкіндігі қарастырылады. Бұл мақаланың мақсаты – проблемаларды анықтауға жүйелі көзқарасты айқындау, қойылған мақсаттарға сәйкес бизнес-процестерді модельдеу және енгізілген жоғары технологиялық шешімдердің функцияларын анықтап, келешекте қолданылатын жүйелерде немесе зерттелетін зерттеу жұмыстарында ыңғайлы қолданысқа ие болатын жүйеге келтіру. Мақала жазу барысында авторлар бес университеттердің жергілікті компьютерлік желілері мен серверлік инфрақұрылымдарын қарастырды: Л.Н. Гумилев атындағы Еуразия Ұлттық Университеті, Astana IT University, С.Сейфуллин атындағы ҚАТУ, М.С. Нарикбаев атындағы КазГЮУ, Қорқыт Ата атындағы Қызылорда университеті. Бұл оқу орындарының серверлік және желілік инфрақұрылымды модернизациялау бойынша атқарылып жатқан барлық жұмыстарына қарамастан, көпшілігінде автоматтандырылған жергілікті желіні бақылау жүйелері жоқ екені анықтады. Осы мақаланы жазу барысында, жергілікті желі қызметтерінің жұмысын оңтайландыру, сонымен қатар болжау жүйелері сияқты басқа жүйелерді кейіннен енгізу үшін

деректерді бақылау және сақтау жүйелерінің қажеттілігін авторлар атап өтеді. Машиналық оқыту алгоритмдерін қолдана отырып, әр түрлі жүйелерді қолданысқа енгізу жұмыстарына келетін болсақ, қарастырылған жоғары оқу орындарының ешқайсысы жақын арада мұндай жобаларды жоспарламайды. Алайда, мұндай жобалар жергілікті желі жүйелерінің істен шығуын автоматтандыру мен болжауды қоса алғанда, әртүрлі салаларда зерттеу жұмыстарын жүргізу үшін барлық негіздерге ие. Авторлар мониторинг жүйелерін іске енгізу арқылы деректерді жинау жұмысы – жоғарыда аталған салалардағы атқаратын келешек жұмыстарын одан әрі жеңілдететінін айтады.

Түйін сөздер: желілік жабдықтың істен шығуын болжау, жергілікті желі, мониторинг жүйелері, жергілікті-есептеу желісі, бизнес-процестерді модельдеу.

А.А. Мырзатай^{1*}, Л.Г. Рзаева², Г. Абитова³, М.А. Жакенов⁴

¹Кызылординский университет имени Коркыт-ата,
Кызылорда, Казахстан;

²Евразийский национальный университет им. Л.Н. Гумилева.
Нур-Султан, Казахстан;

³Astana IT University, Нур-Султан, Казахстан;

⁴Гуманитарно-техническая академия, Кокшетау, Казахстан.
E-mail: mirzataitegiali@gmail.com

ВНЕДРЕНИЕ И ИСПОЛЬЗОВАНИЕ СИСТЕМ МОНИТОРИНГА ЛВС ДЛЯ СИСТЕМАТИЗИРОВАНИЯ ВХОДНЫХ ДАННЫХ СИСТЕМ ПРОГНОЗИРОВАНИЯ ИНЦИДЕНТОВ

Аннотация. В этой статье рассматривается работа по внедрению автоматизированных систем мониторинга локальной сети и возможности использования данных, полученных с помощью этих систем. Целью данной статьи является определение системного подхода к выявлению проблем, моделирование бизнес-процессов в соответствии с поставленными целями и определение функций внедряемых высокотехнологичных решений. В процессе написания статьи авторы рассмотрели локальные компьютерные сети и серверные инфраструктуры пяти университетов: Евразийского Национального Университета им. Л.Н.Гумилева, Astana IT University, КазАТУ им. Сакена Сей-

фуллина, КазГЮУ им. М.С. Нарикбаева и Кызылординского университета им. Коркыт Ата и выяснили, что несмотря на все усилия по модернизации серверной и сетевой инфраструктуры в сравниваемых университетах, большинство из них не имеют каких либо систем мониторинга локальной сети. Авторы этой статьи подчеркивают необходимость внедрения и активное использование систем мониторинга и хранения данных для оптимизации функционирования служб обслуживания локальной сети, а также последующего внедрения других систем, таких как системы прогнозирования использующих алгоритмы машинного обучения. Что касается непосредственно внедрения систем с использованием алгоритмов машинного обучения в рассматриваемых университетах, то пока ни одно из них не планирует подобных проектов в ближайшем будущем. Тем не менее, такие проекты имеют все основания для исследований в различных областях, включая автоматизацию и прогнозирование отказов систем локальной сети. Соответственно, авторы подчеркивают, что сбор данных системами мониторинга могут существенно облегчить дальнейшие работы в вышеуказанных направлениях.

Ключевые слова: прогнозы отказов сетевого оборудования, локальная сеть, системы мониторинга, ЛВС, моделирование бизнес-процессов.

Introduction. Continuous development of the information technologies has led to an increase in the number of active equipment and the scale of the local networks of enterprises, as well as, problems associated with the appropriate organisation and monitoring of these networks. The existing automated enterprise management systems (ACS) must meet the stated requirement in the context of ever-growing complexity of the problem being solved: the constant growth of the scale of the provided service, the development of the route network, logistics management, the increase in time and event loads.

Materials and methods. In the past an organisation and LAN monitoring was the prioritised task for the large enterprises only, but in the modern times, full of digital solutions, the organisation and monitoring of data networks has become one the defining factors of the successful business. Taking into account that each minute, there is humongous amount of heterogeneous data being generated in a network, timely management of the LAN's network infrastructure has become a non-trivial task. First of all, a network infrastructure consists of the multitude of various devices

(communicators, servers, routers, computers, etc.). Secondly, each of these devices hosts a large amount of applied programs, which are sources of data as well. In the selection process of the software for the monitoring of data networks, authors took following two goals into consideration:

The first goal was to ensure collection, storage and real-time access to such a heterogeneous set of data.

The second goal was to reduce incident response time after automatically identifying deviant behaviours in the middle of multiple time series.

In the operational monitoring and incident management system (OM&M), incidents are recorded – incidents related to equipment breakdowns, delay or loss of traffic, overheating or failure due to external factors, and other important events. Those incidents are registered promptly from the devices (switches) of the company or through the communication channels of the end users of the services. Usually, those incidents are handled at the levels of dispatch services, monitoring centres or system/network administrator departments. Some categories of incidents require prompt resolution, for example, all kinds of breakdowns (failures) of active equipment. In order to improve the efficiency of the companies' operations, the fault tolerance of a number of active LAN equipment can be predicted.

Authors also reviewed the infrastructures of 5 universities: L.N. Gumilyov Eurasian National University, Astana IT University, S. Seifullin KazATU, KazGUU and Korkyt Ata Kyzylorda University. All of the above educational institutions have a well-developed server and network infrastructures that meets all the needs of the organization. For example, at the L.N. Gumilyov ENU uses 18 virtual servers in active way which deployed on 5-6 server machines and about 20-30 switches in the LAN contour. KazATU is actively using 4-5 server machines where virtual servers and information systems are deployed, as well as about 20-25 switches providing Internet access to educational and administrative personnel. Astana IT University rents most of its LAN infrastructure; however, they have 4 powerful servers where vital systems for the organization are deployed. KazGUU in the last 5 years have been thoroughly invested in their digital infrastructure to facilitate the work and automation of many business processes, respectively, they have a developed LAN network and one of the most optimized information systems. Korkyt Ata Kyzylorda University has 5 server arrays and LAN with a total of slightly more than 30 switches and hubs to organize access to the Internet for its staff and students.

All of the above universities and their information systems operate non-stop and all universities have staff to maintain the LAN circuit and server

infrastructure. However, not all universities have certain LAN monitoring systems for prompt response to LAN segment outages or other types of system failure. For example, in the same Kyzylorda University named Korkyt Ata the implementation of a LAN monitoring system is at the development stage, and the implementation of such systems is expected only by the end of 2023.

The main anticipated business effect from the use of forecasting tools is a reduction in financial losses through an increase of the technical availability of equipment. This goal should be achieved by providing means of predicting the fault tolerance of the equipment based on the researched forecasting methods.

These days, knowledge on the intellectual data analysis is widely explored. There are a lot of publications on this topic in the IT field. However, most of the times those publications focus on either a particular case or those works could be supplemented by the new variables which might in turn affect the final forecast results.

As for the implementation of systems using machine-learning algorithms, so far none of the listed Higher Education Institutions are planning such projects in the near future. Nevertheless, such projects have all the grounds for research in various areas, including automation and prediction of LAN system failures.

This article addresses the issue of the usage of LAN monitoring systems as an integral part of the subsequent introduction of the forecasting systems based on machine learning.

To address the first challenge, «PRTG network monitor» software was implemented on the basis of the existing LAN network of the private IT company, consisting of switches and servers. After configuring the necessary sensors, predominantly consisting of SNMP, ping, HTTP and IMAP protocols, above-mentioned software has commenced its operation, displaying up-to-date information in regards to the data transfer and active equipment of the network.

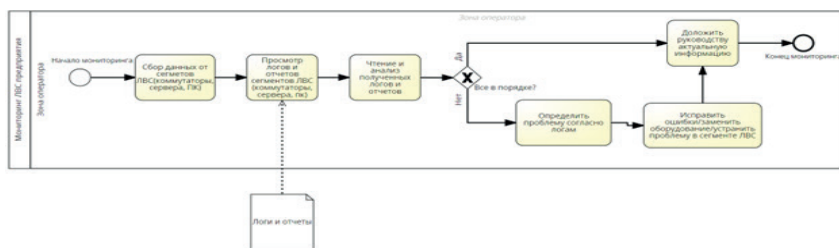


Figure 1 - «As-Is» model of the LAN monitoring process

Figure 1 outlines «As-Is» model, an example of the monitoring process of the local IT company prior to the implementation of the automated monitoring system. Process was initiated by system/network administrator, who spent 15-30 minutes on initial diagnostics of the LAN health. As one can observe from this process, at each stage of monitoring the segment of the LAN there is a room for “human factor”, which can be interpreted as the possibility of missing one message or another in the equipment logs.

After the construction of «As-Is» model, and to solve the set challenges it was decided that the LAN diagnostics time will be reduced and online monitoring of the network will be implemented through introduction of the relevant software. To get a clearer picture of how the implemented software was supposed to operate, authors constructed «To-Be» model (Figure 2).

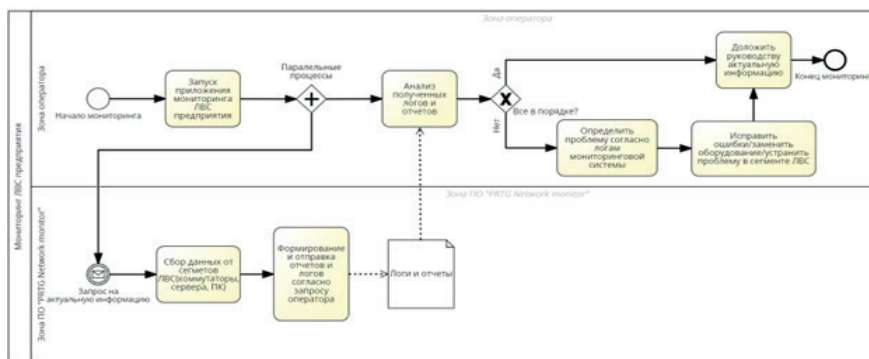


Figure 2 - «To-Be» model of the LAN monitoring process

If one were to compare two processes illustrated on Figures 1 and 2, one can observe that collection of the equipment logs of the LAN segment has changed from manual to automated way, whilst, the time required for the identification of the problem area in practice was reduced by no less than 5-10 times. In addition, the ability to see up-to-date statistical information was now not limited to engineers only, but to their higher management as well, for further decision-making on managerial level. Besides those point, the software also enables analysis of the operation of the whole system and its private sub-systems via the deployment of various sensors.

Alternatively, the authors have explored a solution on the basis of the shareware «ZABBIX» monitoring system. «ZABBIX» was implemented alongside search engine «Elasticsearch», which was responsible for the search, collection and extraction of the log messages from the LAN equipment. Various errors, exceptions and requests were used as metrics.

It is worth to mention that the above-mentioned solutions are for the most part available for free, but they do require a certain degree of proficiency in a large amount of systems, such as Apache, PHP, MySQL, VMware, KVM, Hyper-V, Virtual Box, Azure, etc., and to have at least a minimum level of knowledge of the Java, C+, Python, Jason programming languages.

Results and discussion. The authors have constructed «To-Be v.2» (Figure 3) after the applied analysis of the two LAN monitoring systems. This model could be universally applicable to a LAN of any scale. The authors presume that the implementation of the monitoring systems and the storage of the data from the sensors of the monitoring systems is a necessary minimum for the building of the incident forecasting system based on the machine learning algorithms.

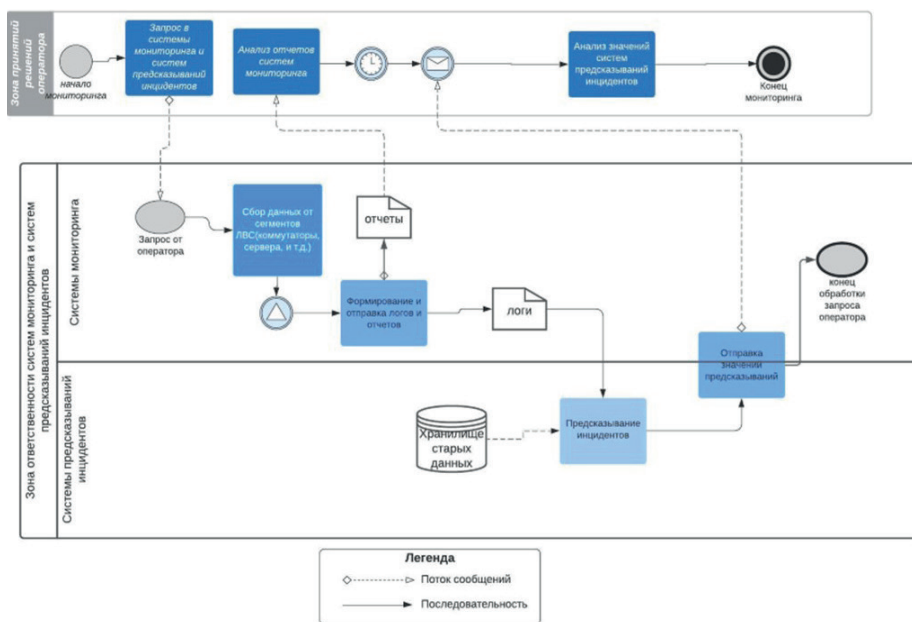


Figure 3. LAN monitoring process after the implementation of monitoring and forecasting systems.

Model «To-Be v.2» (how it should)

As an example, the authors collected a year-worth of operational data from the sensors of several switches using the system monitoring software (Figure 4). In the raw form, this data cannot be easily processed by other systems, hence authors had proceeded on structuring of it. In the future, the structuring could be automated through the use of various systems of data processing and grouping of big data.

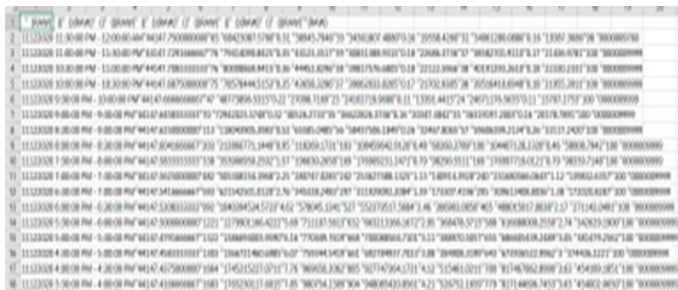


Figure 4. A sample of the unloading data from the LAN monitoring software.

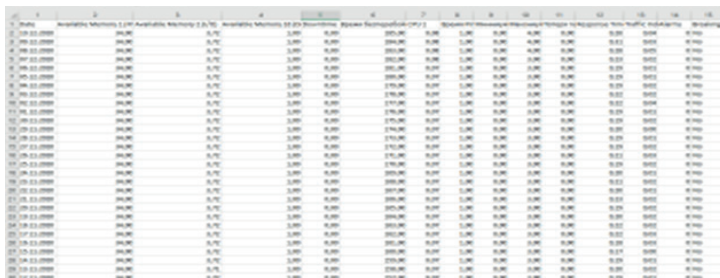


Figure 5. Structured unload of the sensor data from the switch.

Conclusion. The systematic approach to the problem identification, adequate goal-setting and modelling of business processes for the determination of the functions of the implemented high-tech solutions is the main conclusion of this article. As a result, the authors highlight the necessity of having monitoring and data storage systems for the optimisation of the functioning of the LAN maintenance services, as well as, subsequent introduction of forecasting systems. It is presumed that additional resources would be employed for the structuring of the data from databases for the further use by other implemented systems. Authors emphasize that data collection by monitoring systems can significantly facilitate further work in the LAN forecasting systems. However, economic basis for those solutions was not taken into consideration by the authors, which might provide an ample ground for discussion of the feasibility of these implementations for small enterprises, where each business-solution needs to pay off.

Information about authors:

Myrzatay Ali – teacher of “Computer science” department, Korkyt-ata Kyzylorda University, Republic of Kazakhstan; mirzataitegiali@gmail.com; ORCID ID: <https://orcid.org/0000-0002-5339-2437>;

Rzayeva L.G. – PhD, acting Associate Professor, Department of System Analysis and Management, Faculty of Information Technology, Eurasian National University. L.N. Gumilyov, Nur-Sultan, Republic of Kazakhstan.; leilarza2@gmail.com; ORCID ID: <https://orcid.org/0000-0002-3382-4685>;

Abitova G. – PhD, acting Associate Professor, Astana IT University. ORCID ID: <https://orcid.org/0000-0003-3830-6905>;

Zhakenov M.A. – candidate of Technical Sciences, lecturer of the Department of General Education and Technical Disciplines, Faculty of Economics and Finance, at the Humanitarian and Technical Academy, Kokshetau, Republic of Kazakhstan.

REFERENCES

Abomhara M., G.M. Kien. (2015) Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks, Journal of Cyber Security, Vol. 4, pgs 65–88. (in Eng).

Alestra S., Brand C., Burnaev E., Erofeev P., Papanov A., Bordry C., Silveira-Freixo C. Rare event anticipation and degradation trending for aircraft predictive maintenance // 11th World Congress on Computational Mechanics, WCCM 2014, 5th European Conference on Computational Mechanics, ECCM 2014 and 6th European Conference on Computational Fluid Dynamics, ECFD 2014 11, 2014. C. 6571-6582. - (in eng.).

Bajer M. (2017). Building an IoT Data Hub with Elasticsearch, Logstash and Kibana. 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), pgs 63–68.(in Eng).

Ishkanov S.Yu., Shelupanov A.A., Timchenko S.V. (2012). TUSUR Reports, No. 1 (25), part 2, pgs 100-103.(in Russ.).

Myrzatay A., Rzayeva L.G., Uskenbayeva G.A., Shukirova A.K., Abitova G. “The effect of the amount of data array on the results of forecasting network equipment failures” News of the National academy of sciences of the Republic of Kazakhstan, Physico-mathematical series, Volume 6, Number 340 (2021), 28-36. (in eng.).

N.I. Shahanov, I.A. Varfolomeev, E.V. Ershov, O.V. Judina Prognozirovanie otkazov oborudovaniya v usloviyah malogo kolichestva polomok. Vestnik Cherepoveckogo gosudarstvennogo universiteta 2016 №6, pgs 36-41 (in Russ).

Viktorm M. Prognoznaja analitika dlja jeffektivnogo ispol'zovaniya oborudovaniya. 2016. URL: https://filearchive.cnews.ru/files/reviews/2016_03_29/2_Maltsev.pdf - web article (in Russ).

Oliver M. Predictive maintenance & Service (PdMS) - Outline and Value Proposition. 2014. URL: <https://blogs.saphana.com/wp-content/uploads/2014/11/Predictive-Maintenance-Service-Outline-Value-Proposition.pdf> - (in eng.).

Zhilong Wang, Min Zhang, Danshi Wang, Chuang Song, Min Liu, Jin Li, Liqi Lou, and Zhuo Liu. (2017). Failure prediction using machine learning and time series in optical network. Optics Express, Vol. 25, No. 16, 18553-18565 pgs. (in Eng.).

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

Volume 2, Number 342 (2022), 64–75
<https://doi.org/10.32014/2022.2518-1726.130>

УДК 004 МРНТИ20.15.00

Ж.С. Иксебаева^{1*}, К. Жетписов¹, Ж.М. Муратова²

¹Евразийский национальный университет им. Л.Н. Гумилева,
Нур-Султан, Казахстан;

²Российский университет дружбы народов, Москва, Россия.
E-mail: iksebaeva@mail.ru

ИНФОРМАЦИОННАЯ СИСТЕМА ПОСТРОЕНИЯ ДИАГРАММЫ ГАНТА

Аннотация. Данная статья предусматривает разработку информационной системы построения диаграммы Ганта. С помощью данной диаграммы при подаче заявок на конкурс грантового финансирования пункт реализации проекта представляется графически, где четко схематически указываются задачи проекта, начало и окончание исполнения, а также его продолжительность. Авторы предлагают оригинальные способы автоматизации процесса построения диаграммы, в помощь сотрудникам и преподавателям не IT-специальностей. С помощью диаграммы Ганта можно изобразить составляющие части проекта, для удобства управления разбив его на меньшего размера задания. Диаграммы Ганта подходят практически для любых проектов, упрощая управление задачами, помогая увидеть полную картину и достичь целей. Диаграммы Ганта также популярны в сфере управления проектами по двум основным причинам. Они упрощают создание сложных планов и помогают командам планировать работы правильно распределив ресурсы. Для того чтобы получить общее представление о проекте составители используют диаграммы Ганта. Способность отображать зависимости задач, когда задача не может начаться или закончиться до тех пор, пока не начнется или не завершится другая задача является одним из наиболее важных

функций диаграммы Ганта. Основной причиной того, что проекты терпят неудачу, это плохо спланированные проекты. Поэтому четкое понимание порядка и продолжительности задач является основным ключом к успеху проекта. С помощью диаграммы руководители проектов могут определить критические точки при реализации проекта и гарантируют отсутствие задержек в выполнении ключевых задач.

Ключевые слова: информационная система, диаграмма Ганта, проект, задачи проекта, план реализации, конкурс, Python.

Ж.С. Иксебаева^{1*}, К. Жетписов¹, Ж.М. Муратова²

¹Л.Н. Гумилев атындағы Еуразиялық Ұлттық университеті,
Нұр-Сұлтан, Қазақстан;

²Ресей халықтар достығы университеті, Мәскеу, Ресей.
E-mail: iksebaeva@mail.ru

ГАНТ ДИАГРАММАСЫН ҚҰРУДЫҢ АҚПАРАТТЫҚ ЖҮЙЕСІ

Аннотация. Ұсынылып отырған мақалада Гант диаграммасын құрудың ақпараттық жүйесін әзірлеу қарастырылған. Бұл диаграмманың көмегімен гранттық қаржыландыру конкурсына өтінім берген кезде жобаны іске асыру пункті графикалық түрде ұсынылады, онда жобаның міндеттері, орындаудың басталуы мен аяқталуы, сондай-ақ оның ұзақтығы нақты сызба түрінде көрсетіледі. Авторлар IT-емес мамандықтардың қызметкерлері мен оқытушыларына көмек ретінде диаграмма құру процесін автоматтандырудың ерекше тәсілдерін ұсынады. Басқарудың ыңғайлылығы үшін Гант диаграммасын қолдана отырып жобаның құрамдас бөліктерін бейнелеуге және оны кішігірім тапсырмаға бөлуге болады. Гант диаграммалары кез-келген жобаларға жарамды, тапсырмаларды басқаруды жеңілдетеді, толық суретті көруге және мақсаттарға жетуге көмектеседі. Гант диаграммалары екі негізгі себепке байланысты жобаларды басқару саласында да танымал. Олар күрделі жоспарларды құруды жеңілдетеді және командаларға ресурстарды дұрыс бөлу арқылы жұмысты жоспарлауға көмектеседі. Жоба туралы жалпы түсінік алу үшін құрушылар Гант диаграммасын қолданады. Тапсырма басталмайынша немесе аяқталмайынша тапсырма тәуілділіктерін көрсету мүмкіндігі басқа тапсырма Гант диаграммасының ең маңызды функцияларының бірі болып табылады. Жобалардың сәтсіздікке ұшырауының басты себебі

–нашар жоспарлануы. Сондықтан тапсырмалардың тәртібі мен ұзақтығын нақты түсіну жобаның сәтті болуының негізгі кілті болып табылады. Диаграммалар көмегімен жоба жетекшілері жобаны жүзеге асырудағы сыни нүктелерді анықтай алады және негізгі тапсырмаларды орындауда кідірістердің болмауына кепілдік береді.

Түйін сөздер: ақпараттық жүйе, Гант диаграммасы, жоба, жоба міндеттері, іске асыру жоспары, конкурс, Python.

Zh.S. Ixebayeva^{1*}, K. Jetpisov¹, Zh.M. Muratova²

¹L.N. Gumilyov Eurasian national university, Nur-Sultan, Kazakhstan;

²RUDN University, Moscow, Russia.

E-mail: iksebaeva@mail.ru

INFORMATION SYSTEM FOR CONSTRUCTING GANTT CHARTS

Abstract. This article provides for the development of an information system for building a Gantt chart. With the help of this diagram, when applying for grant funding, the project implementation point is presented graphically, where the project tasks, the beginning and the end of implementation, as well as its duration are clearly schematically indicated. The authors offer original ways to automate the process of building a diagram, to help employees and teachers of non-IT specialties. You can use the Gantt chart to depict the constituent parts of the project, dividing it into smaller tasks for ease of management. Gantt charts are suitable for almost any project, simplifying task management, helping you see the big picture and achieve your goals. Gantt charts are also popular in project management for two main reasons. They simplify the creation of complex plans and help teams plan work by properly allocating resources. Gantt charts are used to get an overview of a project. The ability to show task dependencies, where a task cannot start or finish until another task starts or finishes, is one of the most important features of the Gantt chart. A major reason that projects fail is poorly planned projects. Therefore, a clear understanding of the order and duration of tasks is a major key to project success. With the help of the diagram, project managers can identify critical points in project implementation and ensure that key tasks are not delayed.

Key words: information system; Gantt chart; project; project objectives; implementation plan; competition; Python.

Введение. Организация научно-исследовательской деятельности преподавателей в образовательных учреждениях требует грамотного науч-

но обоснованного подхода и решения организационно-управленческих, учебно-методических, кадровых, организационно-методических, информационных, дидактических и психолого-педагогических задач.

Одним из факторов развития в научно-исследовательском направлении является активность сотрудников и преподавателей вузов в участии конкурсах грантового финансирования. Высокая активность исследователей в этом направлении позволит реализовать новые проекты, что в свою очередь, будет способствовать привлечению дополнительного финансирования университета, получению значимых результатов исследования.

В данной статье мы представляем и обсуждаем создание информационной системы построения диаграммы Ганта в реализации плана проекта.

Поддача заявок на участие в конкурсах грантового финансирования осуществляется согласно объявленным конкурсам на их проведение. Согласно условиям конкурсной документации на грантовое финансирование, подаваемые заявки должны соответствовать требованиям конкурса.

Материалы и методы исследования. В процессе проведения исследования для анализа и синтеза данных рассмотрены графические методы обработки информации. При подаче заявок на конкурс грантового финансирования пункт реализации проекта представляется на диаграмме Ганта, где четко схематически указываются задачи проекта, начало и окончание исполнения, а также его продолжительность.

Эта диаграмма является своеобразным стандартом в области менеджмента проектов, так как с ее помощью можно наглядно отражать структуру выполнения этапов.

В диаграмме Ганта информация содержится сжатой, для ее создания не используется запутанная таблица и большое количество текстов. Используя диаграмму, можно упростить управление проектами малого масштаба, позволяя всегда контролировать деятельность сотрудников.

Диаграммы Ганта неизменно входят в пятерку наиболее часто используемых и востребованных функций в программном обеспечении для управления проектами. Диаграмму Ганта можно создавать в табличном процессоре MS Excel, MS Visio, а также онлайн в GanttPRO, где можно использовать готовые шаблоны диаграммы, что вызывает некоторые затруднения сотрудникам и преподавателям не ИТ-специальностей. Это можно решить с помощью использования возможностей информационных технологий в организации построения

диаграммы Ганта при подаче заявок. Диаграммы Ганта подходят практически для любых проектов, упрощая управление задачами и помогая увидеть полную картину и достичь целей. Применение IT позволит организовать пользователям помощь в принятии решений.

Предметная область ИС – это та часть реального (или виртуального) мира, для автоматизации решения задач в которой предназначена ИС. Определение границ предметной области – сложная и ответственная задача. Слишком широкие границы приводят или к невозможности создания, или к чрезмерному усложнению ИС и снижают ее эффективность, а слишком узкие – не позволяют реализовать функции ИС (Саммерфилд, 2016).

Таким образом, в ходе исследований разработана информационная система построения диаграммы Ганта, используемая при планировании реализации проекта, доступного любому пользователю.

Результаты. Разработка функционала программы произведена на языке программирования Python. Универсальный язык программирования высокого уровня Python, отличный вариант для более сложных проектов по разработке приложений и программного обеспечения. Благодаря функциям анализа и визуализации данных, Python позволяет разработчикам представлять и визуализировать данные очень эффективным образом. Его упрощенная структура помогает в разработке сложных научных приложений (Вдовин и др., 2016).

Функционал построения графиков обеспечен модулем `plotly.express`.

Подключение модуля происходит строкой:

```
import plotly.express as px
```

Модуль принимает таблицы только в виде `pandas` таблиц. Модуль `pandas` подключается строкой:

```
import pandas as pd
```

Так же для функционирования программы потребуется хранить количество задач, текст задач, срок начала и окончания задач.

Количество задач хранится в целочисленной переменной `task_count`:

```
task_count = 0
```

Текст задач будет храниться в списке `tasks`:

```
tasks = []
```

Срок начала и окончания задач будет храниться в списках:

```
start_dates = []
```

```
end_dates = []
```

В `plotly.express` для создания новой строки вместо контрольного символа `'\n'` используется тег `'
'`. По этой причине нужно заме-

нить в списке все контрольные символы на тег. Делается это циклом который проходит по всему списку и заменяет каждый символ:

```
tasks2 = []
for x in range(task_count):
    txt2 = tasks[x].replace('\n', '<br />')
    tasks2.append(txt2)
```

Преобразованный список хранится в переменной tasks2. Теперь потребуется совместить все списки в одну единую таблицу:

```
final_list = []
global task_count
for x in range(task_count):
    final_list.append(dict(Задача=tasks2[x], Start=start_dates[x].
    strftime(«%Y-%m-%d»), Finish=end_dates[x].strftime(«%Y-%m-%d»)))
```

Таблица хранится в виде списка, но для рисования графика нужна таблица из модуля pandas. Список преобразуется в pandas таблицу данной строкой:

```
df = pd.DataFrame(final_list)
```

Теперь таблица готова к преобразованию в график. График будет сохранен в файл figure1.png

```
fig = px.timeline(df, x_start=»Start», x_end=»Finish», y=»Задача»,
color=»Задача»)
```

```
fig.update_yaxes(autorange=»reversed»)
```

```
fig.update_layout(autosize=False, width=1366, height=768)
```

```
fig.write_image("./figure1.png")
```

Графический интерфейс к программе разработан с использованием Qt Designer(-a). Подключение интерфейса к функциональной части программы произведен через модуль PySide2. Создание дизайна в Qt Designer (рисунок 1)

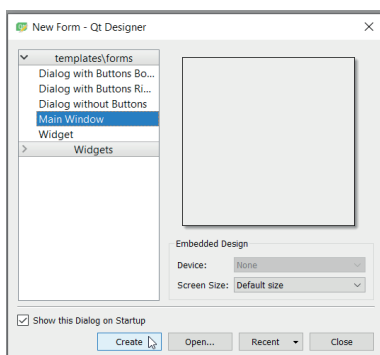


Рисунок 1. Дизайн интерфейса программы в Qt Designer

Далее в форму были добавлены меню «Файл» и 4 пункта этого меню: «Открыть», «Новый», «Сохранить», «Сохранить как...», также компонент GridLayout, который компоует элементы в виде сетки.

Добавление компонента Table Widget дает возможность отображения таблицы. Находится данный компонент во вкладке Item Widgets.

Таблица должна содержать поля: «Задача», «Начальная Дата», «Конечная Дата». Добавляются они через опцию Edit Items... (рисунок 2)

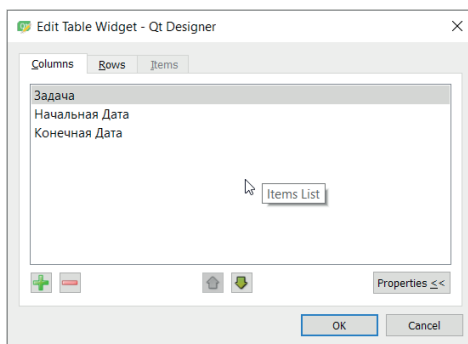


Рисунок 2. Добавление полей через опцию Edit Items...

Далее добавлены компоненты для ввода даты начального срока и конечного срока задачи. Компоненты которые нужно добавить: Label, Horizontal Spacer, DateEdit. В компонентах DateEdit свойство dateTime установлено значение 01.01.2020 (1 Января 2020 года).

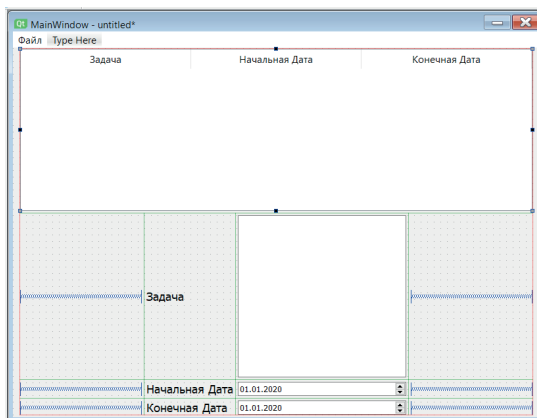


Рисунок 3. Расположение компонентов

Последние компоненты являются кнопками с помощью которых будут добавляться элементы в таблицу и запускаться алгоритм прорисовывания графика.

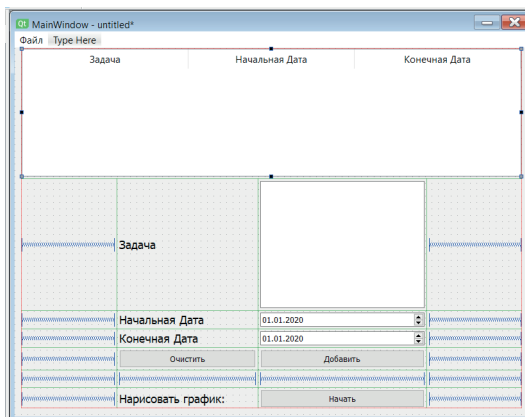


Рисунок 4. Итоговый интерфейс информационной системы диаграммы Ганта

Экспорт графического интерфейса в подключаемый модуль для Python производится командой:

```
pyrside2-uic «ganttmenu.ui» -o «ganttmenu.py»
```

Подключается модуль графического интерфейса строчкой:

```
from ganttmenu import Ui_MainWindow.
```

Обсуждение.

В подготовке научно-исследовательских проектов реализация исследования (таблица 1) представляется на диаграмме Ганта и указывается в плане работы.

Таблица 1. Задачи по реализации исследования

№ пп	Наименование задач и мероприятия по их реализации	Начало исполнения (день/месяц/год)	Продолж. в месяцах	Ожидаемые результаты от реализации проекта (по задачам и мероприятиям), форма завершения		
				1-й год	2-й год	3-й год
1	2	3	4	5	6	7
1	Анализ зарубежного и отечественного опыта part-time обучения и открытого образования	01.10.2020	4			
1.1	Анализ зарубежного опыта реализации part-time обучения при подготовке кадров в развитых странах	01.10.2020	1	Будут изучены успешные зарубежные подходы и практики применения part-time обучения		

1.2 1.3						
2	Анализ текущей ситуации внедрения технологии дистанционного обучения (ДОТ) в профессиональной образовательной системе	01.11.2020	4			
2.1	Сбор, систематизация и анализ научных, статистических, аналитических и отчетных материалов по ДОТ и оценка ее влияния на перспективы развития, трансформации и модернизации образовательных систем в Казахстане	01.11.2020	2	Будут собраны, систематизированы и проанализированы материалы для модернизации ДОТ в Казахстане		
2.2	Анализ собранного и систематизированного материала и разработка практических рекомендаций.	01.01.2021	1	Будут разработаны практические рекомендации для определения структуры открытого образовательного пространства (ООП)		
	Определение готовности будущих педагогов к работе по дистанционному обучению	01.02.2021	1		Будет проведено анкетирование обучающихся и выявлена уровень их готовности	

После добавления всех элементов в программе нужно нажать кнопку «Начать» (рисунок 5)

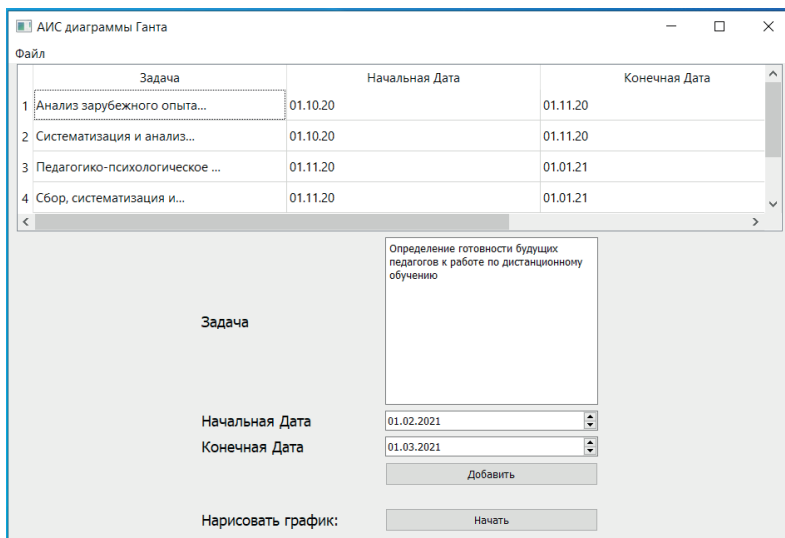


Рисунок 5. Этап заполнения программы

После внесения данных в ИС диаграммы Ганта, получаем диаграмму, где графически отображаются задачи и мероприятия по реализации проекта (рисунок 6). Итоговая диаграмма сохраняется в папке Рисунки в формате .jpg.

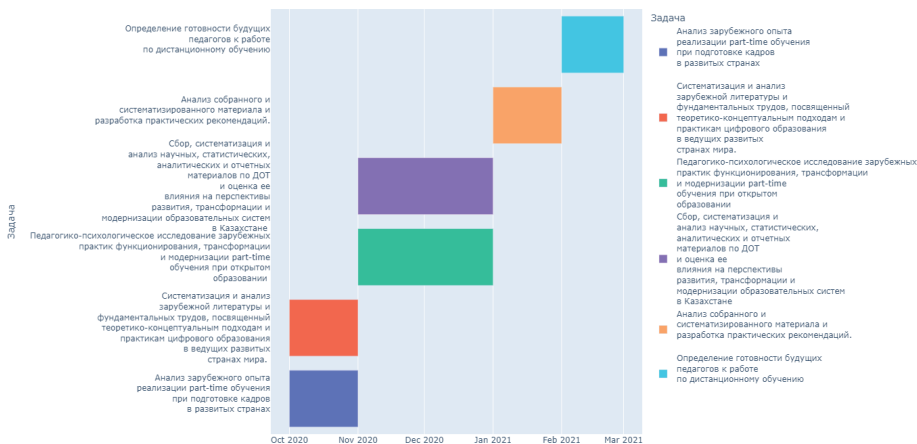


Рисунок 6. Итоговая диаграмма Ганта

Заключение. Информационная система обеспечивает человека информацией в самых различных сферах его деятельности, которая может использоваться с целями управления и выработки оптимальной стратегии и тактики всех собственных действий. Вне зависимости от вида, в информационной системе содержится информационный

фонд, в котором сохраняется информация из определённой сферы человеческой деятельности или, из соответствующей предметной области. Развитие информационных систем отражает требования к постоянному их совершенствованию.

Разработанная информационная система прошла апробацию на кафедре информатики Западно-Казахстанского университета им. Утемисова и успешно функционирует. Использование информационной системы диаграммы Ганта позволяет каждому пользователю легко создавать графики реализации задач, а также наглядно и ясно отслеживать их.

Кроме того, можно увидеть полное видение проекта, построить коммуникацию, максимально эффективно использовать имеющийся ресурс.

Таким образом, можно гибко, эффективно, успешно реализовывать стратегические цели.

Information about authors:

Ixebayeva Zhanna Sagintaevna – doctoral student L.N. Gumilyov Eurasian national university, e-mail: iksebaeva@mail.ru, <https://orcid.org/0000-0003-2619-4870>;

Kabylda Jetpisov – Candidate of Physical and Mathematical Sciences, Associate Professor L.N. Gumilyov Eurasian national university, e-mail: jetpisov_K54@mail.ru, <https://orcid.org/0000-0003-3322-2438>;

Muratova Zhansaya Muratovna – graduate student RUDN University, e-mail: jansik.86@mail.ru, <https://orcid.org/0000-0003-2841-0178>.

ЛИТЕРАТУРЫ

Баранова Е.В., Граф А.А. «Диаграмма Ганта»// Современные экономика и общество: научный взгляд молодых. Сборник статей и тезисов докладов XII международной научно-практической конференции студентов, магистрантов и аспирантов. 2016 – 26-29с.

Ваховская М.Ю. «Использование диаграммы Ганта при планировании проектной деятельности обучающихся вузов»// Анализ, моделирование, управление, развитие социально-экономических систем сборник научных трудов XII Международной школы-симпозиума АМУР-2018.// Симферополь-Судак, 14–27 сентября 2018 г. – 110-112 с.

Вдовин В.М. Предметно-ориентированные экономические информационные системы: Учебное пособие / В.М. Вдовин, Л.Е. Суркова и др. - М.: Дашков и К, 2016. – 388 с. <http://docplayer.com/34613363-Predmetno-orientirovannye-ekonomicheskie-informacionnye-sistemy.html>.

Зырянова С.А., Филимонова О.А. «Автоматизация построения диаграммы Ганта»// Прорывные научные исследования как двигатель науки сборник статей по

итогах Международной научно-практической конференции: в 3 частях. 2017 – 155-157 с.

Киямутдинова Д.Д., Баянова А.А., Трапезникова К.И., Кузьяшев А.Н. «Диаграмма Ганта и ее актуальность»/ Научный электронный журнал Меридиан, 2020 - №4(38) – 96-98 с.

Конкурсная документация на грантовое финансирование по научным и (или) научно-техническим проектам на 2020-2022 годы

Наминова А.В. «Автоматизация построения диаграммы Ганта»// Сборник научных трудов студентов. Сборник статей. Элиста, 2020 – 133-134с.

Саммерфилд, М. Программирование на Python 3. Подробное руководство / М. Саммерфилд. - М.: Символ, 2016. - 608 с.

Сухановский В.А., Короткова Т.В. «Объектные методы управления качеством, диаграмма Ганта»// Наука, техника, промышленное производство: история, современное состояние, перспективы материалы региональной научно-практической конференции студентов и аспирантов. 2020 - 397-399с.

<https://fb.ru/article/142650/diagramma-ganta---vash-pomoschnik-v-planirovanii-chto-takoe-diagramma-ganta-i-kak-ee-sostavit>.

REFERENCES

Baranova E.V., Graf A.A. “Gantt diagram”// Modern economics and society: a scientific view of the young. Collection of articles and abstracts of the XII International Scientific and Practical Conference of students, undergraduates and postgraduates. 2016 – 26-29с.

Vakhovskaya M. Yu. “The use of Gantt diagrams in the planning of project activities of university students”// Analysis, modeling, management, development of socio-economic systems collection of scientific papers of the XII International School-Symposium AMUR-2018.// Simferopol-Sudak, September 14-27, 2018 – 110-112 p.

Vdovin V.M. Subject-oriented economic information systems: A textbook / V.M. Vdovin, L.E. Surkova et al. - М.: Dashkov and K, 2016. – 388 p. <http://docplayer.com/34613363-Predmetno-orientirovannye-ekonomicheskie-informacionnye-sistemy.html>.

Zyryanova S.A., Filimonova O.A. “Automation of Gantt diagram construction”// Breakthrough scientific research as an engine of science collection of articles on the results of the International Scientific and practical Conference: in 3 parts. 2017 – 155-157 p.

Kiyamutdinova D.D., Bayanova A.A., Trapeznikova K.I., Kuzyashev A.N. “Gantt chart and its relevance”/ Scientific electronic journal Meridian, 2020 - №4(38) – 96-98 S.

Tender documentation for grant funding for scientific and (or) scientific and technical projects for 2020-2022.

Naminova A.V. “Automation of Gantt diagram construction”// Collection of scientific papers of students. Collection of articles. Elista, 2020 – 133-134с.

Summerfield, M. Programming in Python 3. A detailed guide / M. Summerfield. - М.: Symbol, 2016. - 608 p.

Sukhanovsky V.A., Korotkova T.V. “Object methods of quality management, Gantt diagram”// Science, technology, industrial production: history, current state, prospects materials of the regional scientific and practical conference of students and postgraduates. 2020 - 397-399с.

<https://fb.ru/article/142650/diagramma-ganta---vash-pomoschnik-v-planirovanii-chto-takoe-diagramma-ganta-i-kak-ee-sostavit> (in Russ)

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

Volume 2, Number 342 (2022), 76–91
<https://doi.org/10.32014/2022.2518-1726.131>

УДК 910.27

K.T. Kyrgyzbay^{1,2*}, E.Kh. Kakimzhanov^{1,2}, Jay Sagin³

¹Al Farabi Kazakh National University, Almaty, Kazakhstan;

²Kazakh Research Institute of Agriculture and Plant Growing,
Almaty, Kazakhstan;

³Western Michigan University, Kalamazoo, USA.

E-mail: kyrgyzbay.kudaibergen@gmail.com

AGRO-CLIMATIC ZONING OF ALMATY REGION USING GIS TECHNOLOGIES

Abstract. Climate change and climatic factors over the past few decades on a regional and global scale, in particular in the south-east of Kazakhstan, affect the restructuring of the agro-industrial complex. Agro-climatic zoning is one of the fundamental actions in determining the quantitative and qualitative characteristics of agro-climatic resources, as well as important factors in agriculture, thereby updating the reliability and accessibility of climatic indicators.

This study is devoted to the definition of agro-climatic zones on the territory of the Almaty region using GIS technologies. The purpose of the scientific work is to study favorable agro-climatic indicators for the growth and development of agricultural crops. To achieve this goal, the tasks of a comprehensive and geospatial study of agro-climatic components were performed, including the calculation of the sums of active temperatures and precipitation, the Selyaninov hydrothermal coefficient and the coefficient of humidification of the growing season, and their relationship is also assumed when constructing an agro-climatic model.

As a result of the study, 5 agro-climatic zones were identified with interpretation and information on the sum of active temperatures and precipitation, hydrothermal coefficient and humidification coefficient in

tabular form. Aggregates of data indicate the main characteristics of agro-climatic resources and the potential of growing certain types of crops with rational use of water resources relevant for irrigation crops.

This study is carried out under the project “BR10764908 - To develop a system of agriculture for the cultivation of agricultural crops (cereals, leguminous oilseeds and industrial crops) using elements of cultivation technology, differentiated nutrition” in Kazakh Research Institute of Agriculture and Plant Growing for further introduction in manufacture.

Key words: GIS technologies, agricultural lands, remote sensing, open climate data, climate change, agro-climatic zoning.

Қ.Т. Қырғызбай^{1,2*}, Е.Х. Какимжанов^{1,2}, Ж.М. Сагинтаев³

¹Әл-Фараби атындағы Қазақ Ұлттық университеті,
Алматы, Қазақстан;

²Қазақ егіншілік және өсімдік шаруашылығы ғылыми-зерттеу
институты, Алматы, Қазақстан;

³Батыс Мичиган Университеті, АҚШ.
E-mail: *kyrgyzbay.kudaibergen@gmail.com*

ГАЗ-ТЕХНОЛОГИЯЛАРЫ НЕГІЗІНДЕ АЛМАТЫ ОБЛЫСЫН АГРОКЛИМАТТЫҚ АУДАНДАСТЫРУ

Аннотация. Климат және климаттық факторлардың өзгеруі соңғы бірнеше онжылдықта өңірлік және жаһандық масштабта, соның ішінде Қазақстанның оңтүстік-шығысындағы агроөнеркәсіптік кешеннің құрылымына біраз өзгеріс енгізді. Агроклиматтық аудандастыру ресурстардың сандық және сапалық сипаттамаларын анықтауда климаттық көрсеткіштердің сенімділігі мен қол жетімділігін өзектендіріп, ауыл шаруашылығын жүргізудегі маңызды факторлардың бірі екендігін айқындады.

Бұл зерттеу ГАЗ-технологияларды қолдана отырып, Алматы облысының аумағындағы агроклиматтық аудандарды анықтауға бағытталған. Ғылыми жұмыстың мақсаты – дақылдардың өсуі мен дамуы үшін қолайлы агроклиматтық көрсеткіштерін зерттеу. Мақсатқа жету үшін агроклиматтық компоненттерді кешенді және геокеңістіктік зерттеу міндеттері орындалды, соның ішінде белсенді температура мен жауын-шашынның, Селяниновтың гидротермиялық коэффициенті мен вегетациялық кезеңнің ылғалдану коэффициентінің қосынды-

сын есептеу жүзеге асты. Тиісті көрсеткіштерді анықтау нәтижесінде ашық қолжетімділікті климаттық деректер пайдаланылды. CSV форматында алынған мәліметтер ГАЖ ортасына одан әрі интеграциялау үшін shaperefile-ге түрлендірілді. Растрлық форматтағы климаттық көрсеткіштер алдыңғы ковариациялармен анықталған Гаусс процесі арқылы модельденетін интерполяция - кригинг (kriging) әдісін қолдана отырып, нүктелік объектілерді кеңістіктік өзгерту арқылы құрастырылды.

Зерттеу нәтижесінде белсенді ауа температурасы мен жауын-шашынның қосындысы, гидротермиялық коэффициент және ылғалдану коэффициенті негізінде 5 агроклиматтық ауданның шекарасы анықталды. Деректер жиынтығы агроклиматтық ресурстар және ауыл шаруашылығы дақылдарының жекелеген түрлерін өсіру әлеуетін су ресурстарын ұтымды пайдалану негізінде жүзеге асыру бағытында анықталады.

Ғылыми зерттеу жұмысы Қазақ егіншілік және өсімдік шаруашылығы ғылыми-зерттеу институтында «BR10764908 - Ауыл шаруашылығы дақылдарын (дәнді дәнді-бұршақты майлы және техникалық дақылдарды) өңдеу технологиясы элементтерін қолдана отырып, егіншілік жүйесін әзірлеу» жобасы бойынша, болашақта өндіріске енгізу мақсатында жүргізілуде.

Түйін сөздер. ГАЖ-технологиялары, ауылшаруашылық алқаптары, Жерді қашықтықтан зондтау, ашық климаттық деректер, климаттың өзгеруі, агроклиматтық аудандастыру.

Қ.Т. Қырғызбай^{1,2*}, Е.Х. Какимжанов^{1,2}, Ж.М. Сагинтаев³

¹Казахский национальный университет имени аль-Фараби,
Алматы, Казахстан;

²Казахский НИИ земледелия и растениеводства, Алматы, Казахстан.

³Университет Западного Мичигана, США.

E-mail: kyrgyzbay.kudaibergen@gmail.com

АГРОКЛИМАТИЧЕСКОЕ РАЙОНИРОВАНИЕ АЛМАТИНСКОЙ ОБЛАСТИ С ПРИМЕНЕНИЕМ ГИС-ТЕХНОЛОГИЙ

Аннотация. Изменение климата и климатических факторов за последние несколько десятилетий в региональном и глобальном масштабе, в частности юго-востока Казахстана, воздействуют на перестроение

ние агропромышленного комплекса. Агроклиматическое районирование является одним из основополагающих действий при определении количественных и качественных характеристик агроклиматических ресурсов, а также немаловажным фактором при ведении сельского хозяйства, тем самым актуализируя достоверность и легкодоступность климатических показателей.

Данное исследование посвящено определению агроклиматических районов на территории Алматинской области с применением ГИС-технологий. Целью научной работы является исследование благоприятных агроклиматических показателей для роста и развития сельскохозяйственных культур. Для достижения цели выполнены задачи комплексного и геопространственного изучения агроклиматических компонентов, в том числе вычисление сумм активных температур и осадков, гидротермического коэффициента Селянинова и коэффициента увлажнения периода вегетации, а также предполагается их взаимосвязь при построении агроклиматической модели.

В результате исследования определены 5 агроклиматических зон с интерпретацией и информацией о сумме активных температур и осадков, гидротермическом коэффициенте и коэффициенте увлажнения в табличной форме. Совокупность данных указывает на главные характеристики агроклиматических ресурсов и потенциале выращивания отдельных видов сельскохозяйственных культур с рациональным использованием водных ресурсов, актуальных для оросительных сельскохозяйственных культур.

Данное исследование проводится по проекту «BR10764908-Разработать систему земледелия возделывания сельскохозяйственных культур (зерновых, зернобобовых, масличных и технических культур) с применением элементов технологии возделывания, дифференцированного питания» в Казахском НИИ земледелия и растениеводства для дальнейшего внедрения в производство.

Ключевые слова: ГИС-технологии, сельскохозяйственные угодья, дистанционное зондирование Земли, открытые климатические данные, изменение климата, агроклиматическое районирование.

Introduction. Climate change over the past decade on a global and regional scale has an impact on agricultural productivity, which is one of the fundamental factors of the economic well-being of each country. The increase in the number of industrial enterprises over the past 50 years has been the main factor in the excess concentration of CO₂ in the

atmosphere, thereby causing global warming. Climatic indicators, such as air temperature, precipitation, evaporation rate, etc. and their trends of change over time change the usual ways of cultivating crops. Global warming affects the water consumption of irrigated lands, increasing the total evaporation. This statement implies a study of the relationship between irrigation methods for each crop and changes in climatic indicators.

The main task of the agro-climatic classification is to identify zones that differ from each other in terms of agro-climatic indicators and conditions of agricultural production, to establish their geographical boundaries and to draw up a map of agro-climatic zoning (Balgabekov, 1999).

Many scientists, when allocating agro-climatic zones, take into account the peculiarities of the development of a certain type of agricultural crop with certain meteorological indicators. Taking into account the climate and weather as environmental conditions, it is necessary to evaluate the combination of agro-climatic conditions with the growth, development and formation of crop yields, based on taking into account the needs of culture to environmental factors (Bajsholanov et al, 2017a).

Araya et al (2010) investigated the agro-climatic resources of the Giba watershed in northern Ethiopia, including climatic suitability for growing teff (*Eragrostis tef*) and barley (*Hordeum vulgare*). The authors determined the beginning and duration of the growing season, evaluated the traditional method of climate classification, and also compiled complex agro-climatic zones of the studied region (Araya et al, 2010).

Nabati (2020) and others researched the relationship and complex of climatic factors, topography parameters, soil and land use in chickpea cultivation using GIS technologies. The corresponding maps of the territory were also developed. The authors proved that agro-climatic zoning and agro-land zoning play an essential role in determining the optimal areas for growing chickpeas in rain-fed and irrigated conditions (Nabati et al, 2020).

The work of Tanasijevic and others (2014) focuses on the impact of climate change on olive cultivation in the Mediterranean region, taking into account possible changes in acreage, phenological dates, crop evapotranspiration and irrigation requirements using agro-climatic zoning (Tanasijevic et al, 2014).

Zhou (2009) and other scientists conducted a spatial analysis to create maps of the climatic zoning of the irrigation zone of East Murambidji, Australia. 4 climatic indicators were selected - the average annual precipitation, the average annual evaporation value, the average annual degree of

growth per day and the average daily temperature. These indicators were used to differentiate the studied region into climatic zones and subzones based on long-term meteorological studies (Zhou et al., 2009).

In this study, methods are being determined and maps of agro-climatic components are being developed for the territory of the Almaty region using GIS technologies. The determination of the corresponding agro-climatic indicators was accompanied by the methods of calculating the sum of active temperatures and precipitation, the coefficient of humidification (K) and the HTC according to Selyaninov. The purpose of the scientific work is to create a geoinformation model based on open climate databases, which are available information for the analysis of agro-climatic indicators and their relationship.

Materials and methods. Almaty region is located in the south-east of the Republic of Kazakhstan, between 42° 20' and 47° 12' north latitude, 73° 47' and 82° 34' east longitude. The area of Almaty region is 223,911 km². It is divided into 17 districts and 3 cities of regional significance on an administrative-territorial basis.

The climate of Almaty region is mainly continental. Winter is moderately cool. The average January temperature in the northern lowland part is 10-16°C, in the southern -4-9°C. Summers are hot and dry. The average temperature in July is +25°C in the north and +27°C in the south. The average annual precipitation in the lowland regions is 110-250 mm. The climatic conditions of the foothills are mild. The average temperature in January is 5-9°C, there are often thaws. The average temperature in July in the foothills is +21-23°C, in the mountain valleys +19-22°C. Precipitation in the foothills is 400-600 mm, in mountain valleys 700-1000 mm. Precipitation falls mainly in the spring and early summer on the territory of the region. On the plains and foothills of the Northern region, the average thickness of the snow cover is 10-30 cm, on the mountain slopes-40-100 cm. There is a breeze on the shores of Balkhash and Alakol (Almaty region, 2022).

Indicators of agro-climatic zoning. Agro-climatic indicators are quantitative expressions of the relationship between the development, growth, condition and productivity of agricultural production facilities with climate factors. When studying the agro-climatic resources of territories, agro-climatic indicators obtained by averaging data in a multi-year section of the growing season are used.

This study uses such agro-climatic indicators as the sum of active temperatures (air temperature above +10°C) and precipitation, indicators

of aridity and moisture availability, as well as modes of humidification of the air and the temperature of the earth's surface for the growing season (Mishchenko, 2009).

Duration of the growing season. The duration of the growing season is defined as the period during the year when average temperatures exceed or equal +10°C ($T_{\text{mean}} \geq 10^\circ\text{C}$), and precipitation plus moisture stored in the soil exceeds half of the potential evaporation ($P > 0.5\text{PET}$). The normal growing season is defined as the period when precipitation exceeds the norm (i.e. the wet period). This period fully satisfies the needs of agricultural crops in evaporation and makes up for the lack of moisture in the soil profile. The intermediate growing season is defined as the period during which precipitation usually does not exceed the potential evaporation. The absence of a growing season occurs when the temperature does not promote crop growth or P does not exceed PET (Sauvage et al, 2022; FAO, 2014; Length of growing period, 2022).

The sum of active temperatures and precipitation during the growing season. The sum of active temperatures is an indicator characterizing the amount of heat exceeding the biological minimum temperature set for a certain period of plant development. This indicator is used to determine the heat needs of most plants, as well as to assess the thermal resources of the territory. It is defined as the sum of the average daily air temperatures for the period of time during which the vehicle was above +10°C. In many studies, the air temperature above +10°C is mentioned as a stable indicator of the growth and development of many crops. The total amount of precipitation is determined by summing up the daily precipitation indicators for the growing season.

Aridity and moisture availability of the growing season. Moisture availability is the ratio of the stock of productive moisture available in the soil to its stock at the lowest moisture capacity. Moisture availability in this study was determined using the moisture coefficient. The moisture coefficient (K) is calculated by the ratio of the average amount of precipitation to the amount of evaporation during the growing season:

$$K = R/E \quad (1)$$

where, R is the average amount of precipitation during the growing season, E is the amount of evaporation.

The assessment of moisture availability according to the coefficient of humidification (CU) is carried out according to the following gradation:less

than 0.2 – dry; 0.2 – 0.3-moisture deficiency; 0.3-0.5 – moderate moisture deficiency; 0.5-0.7 – insufficient moisture supply; 0.7-1.0 – sufficient, but unstable moisture supply.

Climatic aridity is calculated with the Hydrothermal coefficient of humidification (HTC) of Selyaninov:

$$HTC = \Sigma R \cdot 10 / \Sigma T_c \quad (2)$$

where, ΣR is the total amount of precipitation during the growing season, ΣT_c is the sum of active air temperatures for the same period.

The HTC assessment is conducted on a scale of: less than 0.3 - very dry; 0.3-0.5 - dry; 0.6-0.7 - arid; 0.8-1.0 - insufficient moisture; 1.0 - equality of the arrival and consumption of moisture; 1.0—1.5 - sufficient moisture; more than 1.5 - excess moisture; more than 2.0 - excess moisture for the tropics (Mishchenko, 2009; Bajsholanov et al, 2017 b).

Data sets and processing. The data from the National Aeronautics and Space Administration (NASA) POWER Data Access Viewer v2.0.0 includes long-term averages of climatology, estimates of meteorological quantities and solar energy fluxes on the surface. The average daily values of basic meteorological and solar data are presented in time series format. Many studies have shown that satellite and model-based products are accurate enough to provide reliable data on solar and meteorological resources in regions where surface measurements are rare or absent. The products offer two unique functions: the data is global and, as a rule, continuous in time (Kelley, 2013; POWER Data Methodology, 2022).

Spatial interpolation methods predict values for locations on a raster (grid) surface using values from sample data points. Spatial interpolation assumes that attribute data is continuous in space (Flowerdew et al, 1992), and is widely used to create continuous data when data is collected in separate locations, as well as in a Geographic Information System (GIS).

Geostatistical methods are based on statistical models that include autocorrelation, which is defined as a tendency for the observed value of a variable in one locality to depend on the values of neighboring points. In addition to creating an interpolated (predicted) surface, kriging also provides some degree of reliability and accuracy of forecasts, unlike other types of interpolation. Kriging uses the principle of weighing sample values when predicting interpolated values. Kriging is best suited for well-distributed data without discontinuities (Munyati et al, 2021).

Algorithm of work execution. The algorithm of work execution (Figure 1) includes the procedure for determining agro-climatic zones.

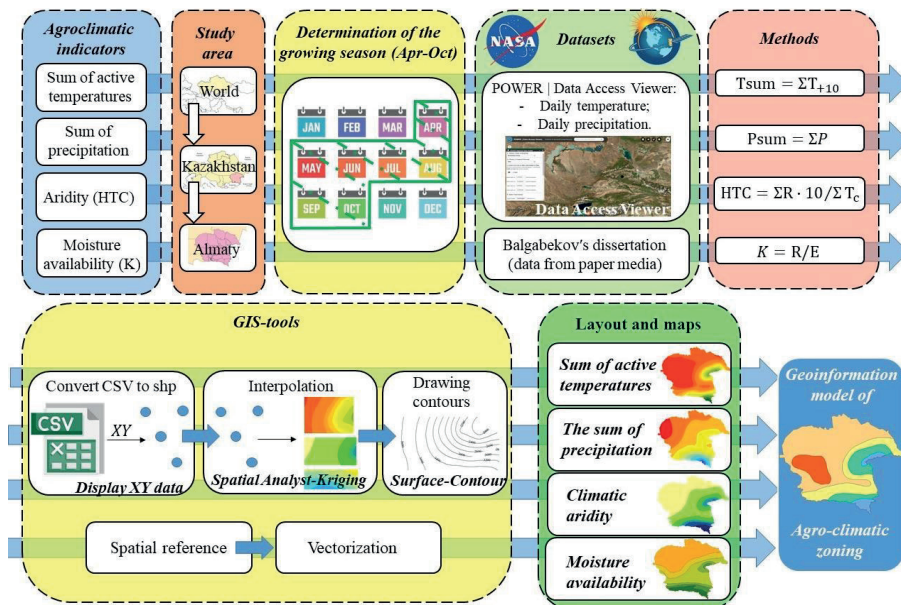


Figure 1. Sequence of work (indicators, datasets, methods, tools, result)

The agro-climatic indicators of the sum of active temperatures and precipitation, the coefficient of humidification and climatic aridity were determined within one administrative unit. The period of vegetation was chosen between April and October. The open climate maps data used in this research are downloaded from the POWER Data Access Viewer website and integrated into the GIS environment in the form of air temperature and precipitation data. The data on the moisture coefficient were obtained from the dissertation work of Balgabekov (1999). The method of calculating agro-climatic indicators was carried out using mathematical formulas (1,2). Geospatial data processing using GIS tools (display XY, kriging interpolation, data converter) was carried out in the ArcGIS program. Based on the processed data, cartographic models of agro-climatic indicators were developed, followed by the implementation of a geoinformation agro-climatic model of the Almaty region.

Results and discussion. Analysis of heat supply data and the amount of precipitation during the growing season. The determination of the amounts of active temperatures and precipitation was carried out according to the POWER Data Access Viewer (link to the Web application: <https://power.larc.nasa.gov/data-access-viewer/>). The period from April to October 2021 (the growing season) was chosen as the studied interval. Climate indicators of daily air temperature values have been uploaded in CSV

format for further integration into the ArcGIS environment. Files of this type included fields with coordinate data, as well as data on daily air temperature indicators. To allocate days with indicators above $+10^{\circ}\text{C}$, a sample was performed where the conditions were greater than or equal to 10 ($\Rightarrow 10$). The resulting point layers were interpolated using the Kriging method (Figure 2). Isolines of the sum of active temperatures were carried out using the Contour GIS tool (ArcToolbox/Spatial Analyst tools/Surface) with an interval of 200°C (McMastera et al, 1997).

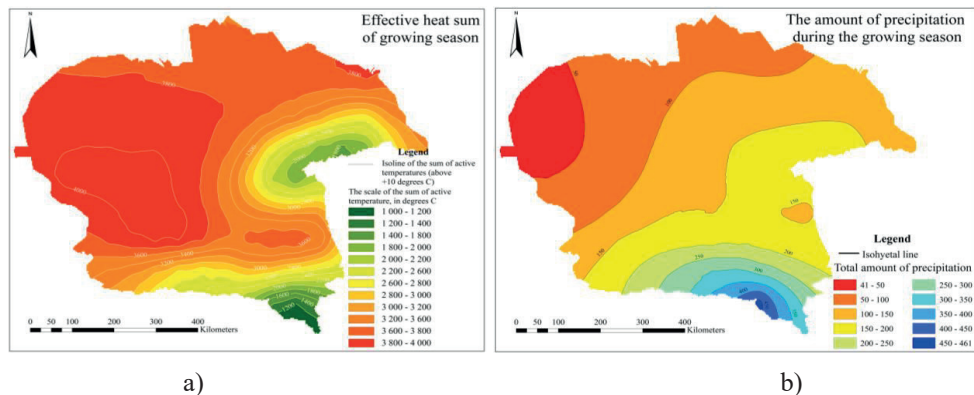


Figure 2. Results of analyses of heat supply (a) and precipitation (b) during the growing season

Figure 2a illustrates the sums of active temperatures by geospatial distribution. The temperature range includes values from 1000 to 4000. The highest indicator (4000°C) is inherent in the western part of the Almaty region. The sum of active temperatures varies gradually by natural and altitude zonality. The lowest index of 1000°C is observed in the high mountain massif of the Eastern Tien Shan (Khan Tengri Peak).

The amount of precipitation for 2021 (Figure 2b) is determined in the period from April to October. The data was also converted into a GIS environment to build an interpolated surface. The kriging method in processing the amount of precipitation was the most optimal option, due to the same distances between the point objects. Precipitation is distributed in the study area in the range from 41 mm to 461 mm. The spatial distribution of climatic indicators occurs from the smaller (north-west of the region) to the larger (south-east). Precipitation changes are correlated with changes in natural zones and altitude zonality. The highest precipitation rates (over 400 mm during the growing season) are inherent in the south-eastern part of the Almaty region, and the least amount falls in the arid north-western part.

Determination of the degree of aridity and moisture availability of the studied territory. The climatic aridity was determined using the POWER Data Access Viewer. Processing CSV files of 2021 (April-October), raster data on climatic aridity according to the Selyaninov's HTC (2) were obtained (Figure 3a).

As a result, zones with corresponding indicators from 0.1 to 3.8 were obtained. The range of indicators from 0.1 to 0.3 covers the northern and western parts of the studied region and is interpreted as a very dry indicator. Further, the range of 0.3-0.5 (dry indicator) extends from the southwest to the east of the Almaty region, covering a significant part of the central region. 0.5-0.7 covers the plain and the foothill part and is designated as an arid indicator. On the foothill part, it covers the range of 0.7-0.8 and is interpreted as an insufficiently moistened indicator. The range from 0.8 to 1.0 covers the mountainous area and is designated as an indicator of the equality of the arrival and consumption of moisture. Indicators over 1.0 of the indicator are located in mountainous areas with difficult terrain and are interpreted as indicators of sufficient and excessive moisture (Bajsholanov et al, 2017b; Młyński et al, 2021).

Moisture availability by moisture coefficient was obtained from the scientific work of Balgabekov (1999), and integrated into this study (Figure 3b). The coefficient was calculated using a special formula (1), shown in the Materials and methods section. Based on the formula, raster data of average precipitation and evaporation values in mm were needed to process the data and obtain the result. Cartographic data from paper media is scanned and digitized in ArcGIS, using ArcScan and Editor tools. The mentioned tools are focused on vectorization with pixel data information. As a result, shapefiles with attribute information about moisture availability were obtained.

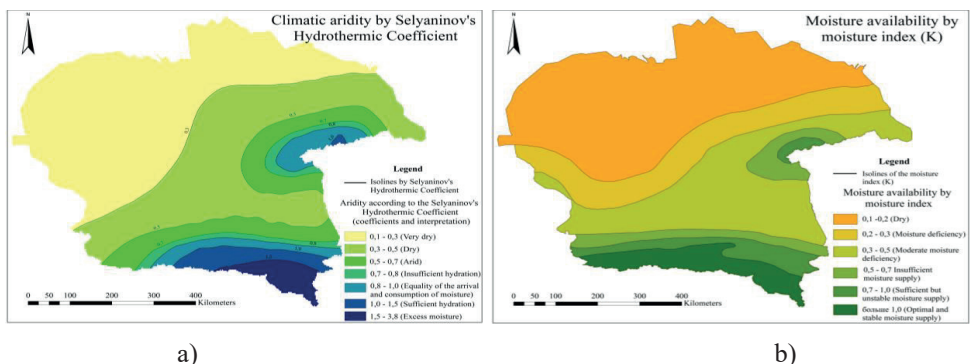


Figure 3. Indicators of climatic aridity (according to the Selyaninov HTC) (a) and moisture availability (according to the moisture coefficient - K) (b) of the territory

As a result, zones with a moisture coefficient with a step of 0.1 were obtained. It was confirmed that moisture availability decreases from north to south. The range of 0.1-0.2, which are interpreted as a dry indicator, are located in the northern and central parts of the region. 0.2-0.3 extends from west to east and is designated as an indicator of moisture deficiency. 0.3-0.5 covers the central and part of the foothill area of the study area, also interpreted as an indicator of moderate moisture deficiency. In this case, irrigation of agricultural land occurs with additional irrigation measures. Indicators 0.5-0.7 cover the foothill area of the Northern Tien Shan and the Dzungarian Alatau and are interpreted as insufficient moisture supply. Indicators above 0.7 cover the entire south of the Almaty region and are designated as areas with sufficient, but unstable, as well as further south as optimal and stable moisture supply (Balgabekov, 1999; Bajsholanov, 2017b) (Balgabekov, 1999; Bajsholanov, 2017b).

Construction of a geoinformation model of agro-climatic zoning of the Almaty region. In many studies (Araya et al, 2010; Nabati et al, 2020; Tanasijevic et al, 2014; Zhou et al, 2009) agro-climatic zoning is carried out taking into account the sum of active temperatures and humidification coefficients (Bajsholanov, 2017b). However, in this work, the development of zones of agro-climatic zoning implied a combination of thermal characteristics, data on aridity, moisture availability, as well as indicators of the amount of precipitation in the studied region (Figure 4).

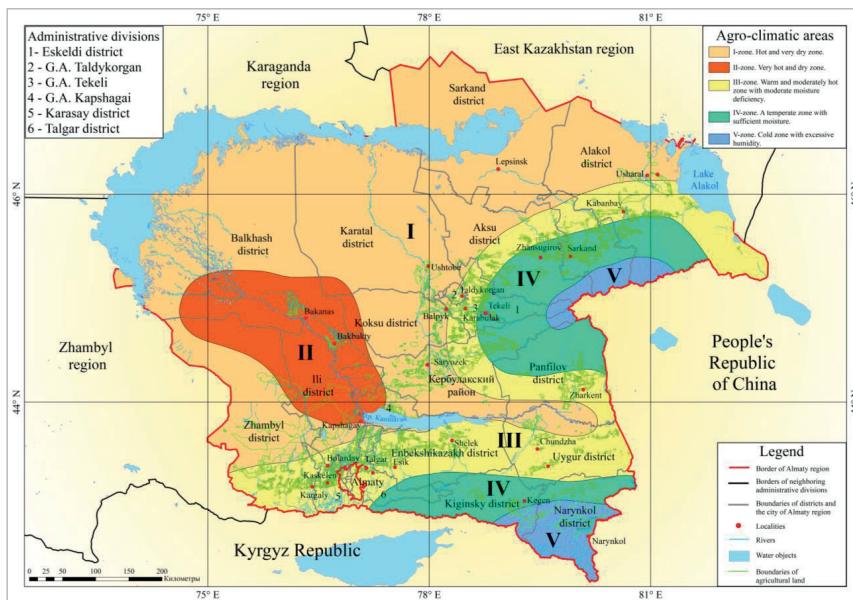


Figure 4. Agro-climatic zoning of Almaty region

As a result of the superposition of the above-mentioned indicators, agro-climatic zones on the territory of the Almaty region have been determined. Differentiation was reproduced on the basis of natural zones and altitude zonality. The results of the study identified 5 agro-climatic zones: I-hot and very dry zone; II-very hot and dry zone; III-warm and moderately hot zone with moderate moisture deficiency; IV-moderate zone with sufficient moisture; V-cold zone with excessive humidity. The complex characteristics are shown in Table-1.

The I-zone is characterized by a dry climate and covers a significant part of the studied region, including the territories of Balkhash, Zhambyl, Karatal, Koksu, Aksu, Sarkand, Alakol, Kerbulak districts and administrative units. For this zone, the sums of active temperatures in the range from 3600⁰C to 3800⁰C are inherent. The growing season lasts from 30 to 120 days. The coefficient of humidification (K) has a range of 0.1-0.2. The hydrothermal coefficient of Selyaninov varies from 0.1 to 0.5 indicators. The amount of precipitation during the growing season ranges from 41 to 200 mm.

II-zone. The second agro-climatic zone covers the territories of the southern part of Balkhash, the north-eastern part of Zhambyl and a significant part of Ili districts. The sum of active temperatures is the highest above 4000⁰C. The duration of the period is 60-150 days. The coefficient of humidification is 0.1-0.3 indicators. The HTC for Selyaninov has a range of 0.3-0.5. The amount of precipitation during the growing season is 100-250 mm. Irrigation of agricultural land requires a sufficient amount of water resources.

Table 1. Agro-climatic zones and their complex characteristics.

Agro-climatic zone	Duration of the growing season, days	The sum of active temperatures, in ⁰C	Humidification coefficient (K)	HTC by Selyaninov	Precipitation amount, in mm
I-zone. Hot and very dry zone.	30-120	3600-3800	0.1-0.2	0.1-0.5	41-200
II-zone. Very hot and dry zone.	60-150	4000 and higher	0.1-0.3	0.3-0.5	100-250
III-zone. Warm and moderately hot zone with moderate moisture deficiency.	60-150	3000-3600	0.2-0.3	0,3-0,7	150-350

IV-zone. A temperate zone with sufficient moisture.	60-120	1800-2800	0.5-0.7	1.0-1.5	150-400
V-zone. Cold zone with excessive humidity.	60-90	1000-1800	0.7-1.0	0.8 and higher	400 and higher

III-zone. The third zone covers the foothill part, including the territory of Enbekshikazakh, Karasai, Talgar, Uygur, the southern parts of Zhambyl, Panfilov and Sarkand districts, the northeastern part of Kerbulak, the central parts of Aksu and Alakol districts. The sum of active temperatures is 3000-3600°C. The duration of the growing season lasts 60-150 days. The humidification coefficient has a range of 0.2-0.3. The Selyaninov HTC covers ranges of 0.3-0.7. The total amount of precipitation in this zone is higher compared to the previous two zones and is 150-350 mm.

IV-zone. The zone covers the mountainous region, including the territories of Kegen, the northern parts of Panfilov and Narynkol districts, the eastern parts of Kerbulak and Eskeldi districts, as well as the southern parts of Aksu, Sarkand, Alakol districts. The sum of active temperatures is 1800-2800°C, the vegetation period is equal to 60-120 days. The humidification coefficient is measured in 0.5-0.7 diapason, and the HTC is 1.0-1.5 indicators. The amount of precipitation is 150-400 mm .

The V-zone covers most of the Narynkol district, and minor territories of Kegen, Kerbulak, Aksu, Sarkand and Alakol districts. The sum of active temperatures and precipitation is 1000-1800°C and 400 mm. The humidification coefficient is 0.7-1.0, and the HTC be Selyaninov has a range of 0.8 and higher indicators. The duration of the growing season is 60-90 days.

Conclusion. This article presents a study on the agro-climatic zoning of the Almaty region. Climate change has a tremendous impact on the cultivation of agricultural crops, thereby making relevant scientific work in the agro-climatic direction.

During the study, open-access climatic data were used to determine the sum of active temperatures and precipitation, the coefficient of humidity and HTC according to Selyaninov. The aggregates of the above-mentioned components were the basis for compiling a geoinformation model of agro-climatic zoning.

Agro-climatic research on the territory of the Almaty region allowed to draw the following conclusions:

1. Agro-climatic indicators change from north to south, taking into account natural zones and high-altitude zonality;

2. In most of the studied region there is insufficient humidification and high temperatures, thereby increasing the amount of water resources allocated for irrigation.

The model proposed in our article can be supplemented and modified using other climatic sources, trend analysis and error measurement. Our future research plans include a more comprehensive analysis of the relationship between climatic factors, water resources in relation to the zoning of agricultural irrigated lands, as well as the biological characteristics of individual crops.

Acknowledgments. This research is carried out under the project “To develop a system of agriculture for the cultivation of agricultural crops (cereals, leguminous oilseeds and industrial crops) using elements of cultivation technology, differentiated nutrition (no BR10764908)” in Kazakh Research Institute of Agriculture and Plant Growing.

Information about authors:

Kyrgyzbay K.T. – Ph.D. student, Al Farabi Kazakh National University, Kazakh Research Institute of Agriculture and Plant Growing, Almaty, Kazakhstan; *kyrgyzbay.kudaibergen@gmail.com*; <https://orcid.org/0000-0002-4279-6436>;

Kakimzhanov E.Kh. – Ph.D., Acting Associate Professor, Al Farabi Kazakh National University, Kazakh Research Institute of Agriculture and Plant Growing, Almaty, Kazakhstan; *erkinkakimzhanov@gmail.com*; <https://orcid.org/0000-0001-6454-681X>;

Jay Sagin – Ph.D., Adjunct Assistant Professor, Geosciences Department Western Michigan University & ESRSF Kalamzoo, MI 49008, USA; *zhanay.sagintayev@nu.edu.kz*; <https://orcid.org/0000-0002-0386-888X>.

REFERENCES

Almaty region [Almaty oblysy]. Available at: https://kk.wikipedia.org/wiki/Алматы_облысы. (accessed 2 Apr. 2022).

Araya A., Keesstra S.D., Stroosnijder L. (2010). A new agro-climatic classification for crop suitability zoning in northern semi-arid Ethiopia, *Agricultural and Forest Meteorology*, 150: 1057-1064. DOI: <https://doi.org/10.1016/j.agrformet.2010.04.003>.

Bajsholanov S.S. Agro-climatic zoning of agricultural crops in northern Kazakhstan: scientific and applied reference. Astana, Kazakhstan. ISBN: 978-601-7150-83-9.

Bajsholanov S.S., ZHakieva A.R., Gabbasova M.S., Chernov D.A. (2017) Agro-climatic zoning of agricultural crops in northern Kazakhstan [Agroklimaticheskoe rajonirovanie sel'skohozyajstvennykh kul'tur v severnom Kazahstane] 3: 17-28.

Balgabekov K.B. (1999). Hydromodule zoning and soil protection from irrigation erosion in the south-east of Kazakhstan [Gidromodul'noe rajonirovanie i zashchita pochv ot irrigacionnoj erozii na yugo-vostoke Kazahstana]. [Doctoral dissertation, Williams Kazakh Research Institute]. Williams Kazakh Research Institute.

FAO, IFAD, WFP, 2014. The State of Food Insecurity in the World 2014. Strengthening the Enabling Environment for Food Security and Nutrition. Technical Report, Rome, FAO.

Flowerdew R., Green M. (1992). Developments in areal interpolation methods and GIS, *Mathematics, Economics*, 26:67-78. DOI: 10.1007/BF01581481.

Kelley O.A. (2013). Adapting an existing visualization application for browser-based deployment: A case study from the Tropical Rainfall Measuring Mission, *Computers & Geosciences*, 51:228-237. DOI: <https://doi.org/10.1016/j.cageo.2012.10.004>.

Length of growing period. Available at: <https://www.arcgis.com/home/item.html?id=3ac478a468c245ef9bfd5533f7edbf93>. (accessed 2 Apr. 2022).

McMastera G.S., Wilhelm W.W. (1997). Growing degree-days: one equation, two interpretations, *Agricultural and Forest Meteorology*, 87: 291-300. DOI: [https://doi.org/10.1016/S0168-1923\(97\)00027-0](https://doi.org/10.1016/S0168-1923(97)00027-0).

Mishchenko Z.A. (2009). *Agroclimatology [Agroklimatologiya]*. Odessa State Ecological University, Ukraine. ISBN: 978-966-373-481-1.

Munyati C., Sinthumule N.I. (2021). Comparative suitability of ordinary kriging and Inverse Distance Weighted interpolation for indicating intactness gradients on threatened savannah woodland and forest stands, *Environmental and Sustainability Indicators*. DOI: <https://doi.org/10.1016/j.indic.2021.100151>.

Młyński D., Wałęga A., Kuriqi A. (2021). Influence of meteorological drought on environmental flows in mountain catchments, *Ecological Indicators*, DOI: <https://doi.org/10.1016/j.ecolind.2021.108460>.

Nabati J., Nezami A., Neamatollahi E., Akbari M. (2020). GIS-based agro-ecological zoning for crop suitability using fuzzy inference system in semi-arid regions, *Ecological Indicators*, 117. DOI: <https://doi.org/10.1016/j.ecolind.2020.106646>.

POWER Data Methodology. Available at: <https://power.larc.nasa.gov/docs/methodology/>. (accessed 2 Apr. 2022).

Sauvage J.S., Vitasse Y., Meier M. Delzon, S., Bigler C. (2022). Temperature rather than individual growing period length determines radial growth of sessile oak in the Pyrenees, *Agricultural and Forest Meteorology*. DOI: <https://doi.org/10.1016/j.agrformet.2022.108885>.

Tanasijevic L., Todorovic M., Pereira L.S., Pizzigalli C., Lionello P. (2014). Impacts of climate change on olive crop evapotranspiration and irrigation requirements in the Mediterranean region, *Agricultural Water Management*, 144:54-68. DOI: <https://doi.org/10.1016/j.agwat.2014.05.019>.

Zhou D., Khan S., Abbas A., Rana T., Zhang H., Chen Y. (2009). Climatic regionalization mapping of the Murrumbidgee Irrigation Area, Australia, *Progress in Natural Science*, 19:1773-1779. DOI: <https://doi.org/10.1016/j.pnsc.2009.07.007>.

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

Volume 2, Number 342 (2022), 92–111
<https://doi.org/10.32014/2022.2518-1726.132>

УДК 004.623
МРНТИ 20.53.19, 20.23.29

**А.А. Мухитова^{1,2*}, А.С. Еримбетова^{2,3}, В.Б. Барахнин^{4,5},
Э.Н. Дайырбаева^{2,6}, А. Адалбек⁷**

¹Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан;

²ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институты,
Алматы, Қазақстан;

³Сәтбаев университеті, Алматы, Қазақстан;

⁴Новосибирск мемлекеттік университеті, Новосибирск, Ресей;

⁵Есептеуіш технологиялар институты РҒА СБ, Новосибирск, Ресей;

⁶Халықаралық Білім беру Корпорациясы, Алматы, Қазақстан;

⁷Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан.

E-mail: mukhitova.aigul@gmail.com

РЕЛЯЦИЯЛЫҚ ЖӘНЕ УАҚЫТҚА ТӘУЕЛДІ XML-ДЕРЕКТЕР ҚОРЫНДАҒЫ XML-ДЕРЕКТЕРДІ ӨНДЕУДІҢ ЗАМАНАУИ ӘДІСТЕРІ

Аннотация. XML пайдаланудың соңғы екі онжылдығында әр түрлі веб-қосымшалар арасында деректер алмасудың стандартты форматына ғана емес, сонымен қатар XML дерекқорлары деп аталатын кейбір жаңа дерекқорлар немесе NoSQL дерекқорлар тобына арналған модельге айналды. NoSQL дерекқорының артықшылығы – бір уақытта бір дерекқорда құрылымдалған, ішінара құрылымдалған және құрылымданбаған деректерді сақтау және модельдеу мүмкіндігін береді. Ақпаратты басқару және деректерді бөлісу стратегиясында XML деректерін пайдалану кезінде деректерді сақтау, алу, сұрау, индекстеу және манипуляциялау мәселелері туындайды. Бастапқы ресурстардың біртектілігі мен бөлінуінің ерекшелігіне байланысты ақпараттық жүйе-

лердің жұмыс істеуінің тұжырымдамалық модельдерін жобалау және әзірлеу қажет. Осындай модельдердің бірі – деректерді басқару кешенін құруға арналған технологиялар мен бағдарламалық модульдер. Соңғы пайдаланушыға көптеген гетерогенді көздерден алынған көп мәліметтермен жұмыс істеу және сонымен бірге тиімді нәтиже алу мүмкіндігі XML форматында ұсынылған кез-келген файл құрылымын көрсетуге мүмкіндік беретін XML технологияларына негізделген графикалық веб-интерфейстерді қолдану арқылы жүзеге асырылады. Зерттеу XML деректерін енгізу және өңдеу үшін жіберілген HTML формаларын жасауға мүмкіндік беретін адаптивті графикалық веб-интерфейсті құрудың өзіндік технологиясын жасады. XML жазбаларына жауап беретін графикалық редактордың дамыған моделі кез-келген XML деректерін импорттауға және олардың құрылымын тиімді және қарапайым түрде түрлендіруге мүмкіндік береді және сол өңдеу процесі кез-келген құрылымның бастапқы деректерін бағдарламалық кодты өзгертпестен өзгертуге мүмкіндік береді. Таратылған ақпараттық жүйеде адаптивті графикалық веб-интерфейстің жасалған моделін біріктіру жоғарыда аталған технологияның қолдану саласы болып табылады.

Түйін сөздер: интеграция, xml-деректерді өңдеу, NoSQL уақытқа тәуелді деректер қоры, xslt-технологиялар.

**A.A. Mukhitova^{1,2*}, A.S. Yerimbetova^{2,3}, V.B. Barakhnin^{4,5},
E. Daiyrbayeva^{2,6}, A. Adalbek⁷**

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan;

²Institute of Information and Computational Technologies CS MES RK,
Almaty, Kazakhstan;

³Satbayev University, Almaty, Kazakhstan;

⁴Federal Research Center for Information and Computational Technologies,
Novosibirsk, Russia;

⁵Novosibirsk State University, Novosibirsk, Russia;

⁶International Educational Corporation, Almaty, Kazakhstan;

⁷L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan.

E-mail: mukhitova.aigul@gmail.com

MODERN METHODS OF PROCESSING XML DATA IN RELATIONAL AND TEMPORARY XML DATABASES

Abstract. Over the past two decades of use, XML has become both the standard format of data exchange between various web applications

as well as the model for some new databases or NoSQL databases called XML databases. The advantage of NoSQL database is in its ability to store and simulate structured, partially structured and unstructured data simultaneously in the same database. Some problems related to data storage, retrieval, querying, indexing, and manipulation arise when use of XML data in information management and data exchange strategies. Because of specific heterogeneity and distribution of the original resources, there is the necessity in designing and development of conceptual models of operation of information systems. One such model is technologies and software modules designed to create a set of data management tools. Any end user would be able to use a large amount of data from a large number of heterogeneous sources and at the same time get an effective result from the work owing to use of graphical web interfaces, built on XML-technologies, which allow to display any file structure represented in XML-format. In the course of the study, we developed our own technology for building an adaptive graphical WEB-interface, which allows for generation of HTML-forms to enter and edit XML-data. The developed model of adaptive graphical XML record editor allows importing any XML data and transforming its structure in efficient and simple way, while the same processing process allows transforming source data of any structure without any change in program code. The area of application of the above technology is the integration of the created model of adaptive graphical web-interface in a distributed information system.

Key words: integration, xml data processing, temporary NoSQL databases, xslt technologies.

**А.А. Мухитова^{1,2*}, А.С. Еримбетова^{2,3}, В.Б. Баракнин^{4,5},
Э.Н. Дайырбаева^{2,6}, А. Адалбек⁷**

¹Казахский Национальный Университет имени аль-Фараби,
Алматы, Казахстан;

²Институт информационных и вычислительных технологий КН МОН
РК, Алматы, Казахстан;

³Satbayev University, Алматы, Казахстан;

⁴ФИЦ Институт вычислительных технологий СО РАН,
Новосибирск, Россия;

⁵Новосибирский государственный университет, Новосибирск, Россия;

⁶Международная Образовательная Корпорация, Алматы, Казахстан;

⁷Евразийский национальный университет им.Л.Н.Гумилева,
Нур-Султан, Казахстан.

E-mail: mukhitova.aigul@gmail.com

СОВРЕМЕННЫЕ МЕТОДЫ ОБРАБОТКИ XML-ДАННЫХ В РЕЛЯЦИОННЫХ И ВРЕМЕННЫХ XML-БАЗАХ ДАННЫХ

Аннотация. За последние двадцать лет использование XML стало не только стандартным форматом для обмена данными между различными веб-приложениями, но и моделью для семейства некоторых новых баз данных или баз данных NoSQL, называемых базами данных XML. Преимуществом базы данных NoSQL является способность хранить и моделировать структурированные, частично структурированные и неструктурированные данные одновременно в одной базе данных. При использовании XML-данных в стратегиях управления информацией и обмена данными возникают проблемы, связанные с хранением, извлечением, запросами, индексированием и манипуляциями данными. В силу специфики разнородности и распределённости исходных ресурсов существует необходимость в проектировании и разработке концептуальных моделей функционирования информационных систем. Одной из таких моделей являются технологии и программные модули, предназначенные для создания комплекса средств управления данными. Возможность конечному пользователю работать с большим количеством данных из большего количества разнородных источников и при этом получать эффективный результат от работы осуществляется посредством использования графических веб-интерфейсов, построенных на основе XML-технологий, которые позволяют отобра-

жать любую структуру файла, представленного в XML-формате. В ходе исследования разработана собственная технология построения адаптивного графического WEB-интерфейса, позволяющая генерировать отправляемые HTML-формы для ввода и редактирования XML-данных. Разработанная модель адаптивного графического редактора XML-записей позволяет импортировать любые XML-данные и преобразовывать их структуру эффективно и просто, при этом один и тот же процесс обработки позволяет трансформировать исходные данные любой структуры без какого-либо изменения программного кода. Областью применения вышеуказанной технологии является интеграция созданной модели адаптивного графического web-интерфейса в распределенной информационной системе.

Ключевые слова: интеграция, обработка xml-данных, временные базы данных NoSQL, xslt-технологии.

Кіріспе. Қазіргі уақытта құрылымдалған деректерді ұсынудың ең тиімді және әмбебап құралы – XML тілі болып табылады. Қолданысқа енгеннен бастап XML тілі өзінің тиімділігімен тез таралып кетті. XML тілі пайдалы технология ретінде танылып және ISO – 2709-мен салыстырғанда оның бірқатар айқын артықшылықтары бар. Библиографиялық жазбаның иерархиялық құрылымы XML құжатының моделіне жақсы сәйкес келеді. XML-ді библиографиялық деректерді алмасу және сақтау пішімі ретінде пайдалану XML құжатын тексеру деңгейінде жазбалардың дұрыстығын бақылауға мүмкіндік береді. ISO – 2709 форматына қарағанда XML – бұл пайдаланушы үшін оңай оқылатын формат және оңай құжатталады, сонымен қатар құжаттың ұзындығына ешқандай шектеулер жоқ болуымен ерекшеленеді. ISO – 2709-дан және оның негізінде жасалған MARC форматтарының алуан түрлілігінен айырмашылығы: XML анағұрлым серпінді және сол тілде (XSLT түрлендірулері) тұжырымдалған және көптеген бағдарламалық жасақтама өндірушілері қолдайтын жаңа деректер сұлбалары мен олардың арасындағы ауысу ережелерін құруды жеңілдетеді. XML стандарты Unicode кодтауды қолдауды қамтиды, бұл көп тілді құжаттарды жасауды жеңілдетеді. Айта кету керек, ISO – 2709-дың XML – ге қарағанда, графикалық кескін немесе аудио немесе бейне материалдар сияқты екілік деректерді беруді қарастырмайды, ал оларсыз қазіргі уақытта ақпараттық ортаны елестету мүмкін болмайды. XML тілінің ерекше маңызды артықшылығы – оның веб-ортамен интеграциясы, сонымен қатар платформалық тәуелсіздік болып табылады. Мұның

бәрі XML – ді деректермен алмасудың стандартына айналдырды (Жижимов, 2015: 6).

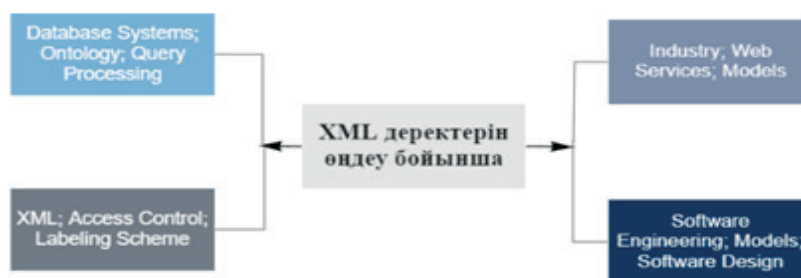
Зерттеу әдістемесі мен материалдары. Реляциялық XML деректер қорларындағы XML деректерін өңдеудің заманауи әдістері.

XML форматындағы деректерді өңдеу үшін әртүрлі технологиялар, деректерді басқару және өңдеу тілдері қолданылады. (Баазизи, т.б., 2011:11) жұмыста қарапайым және уақытқа тәуелді XML деректер қорларындағы XML деректерін өңдеудің ағымдағы жағдайына шолу жасайды, сонымен қатар осы тақырып бойынша болашақ зерттеулердің ықтимал бағыттарын ұсынады.

Соңғы онжылдықтарда шартты түрде жеке топтарға бөлуге болатын бірқатар зерттеу жұмыстары жүргізілді:

- уақытқа тәуелді және көп нұсқалы XML-құжаттар;
- XML - мәліметтер қорларында деректерді сақтау және сұраныстарды өңдеу;
- XML құжатын түрлендіру;
- XML бойынша навигация;
- XML құжаттарын басқару;
- XML үшін концептуалдық модельдеу;
- XML деректерін сұрау;
- XML деректерін кластерлеу;
- XML ағынымен тасымалдау.

XML деректерін өңдеу бойынша жоғарыда аталған барлық ғылыми зерттеулерді қолдану аясы бойынша келесі топтарға бөлуге болады (1-сурет):



1-сурет - XML деректерін өңдеу бойынша төрт тобы

XML мәліметтерімен жұмыс істеудің барлық заманауи әдістері қарапайым (стандартты) және уақытқа тәуелді деректер қорындағы XML деректерін басқаруды зерттеу нәтижелері ретінде ұсынылған. (Бидоит, т.б., 2015:10, Пардеде, т.б., 2008: 25) ғалымдардың еңбектерінде қазіргі реляциялық ДҚБЖ-де XML деректерін басқаруды зерттеуге арналған.

Baazizi және т.б. негізгі жадта XML жаңартуларын өңдеу кезінде механизмдердің жад тұтынуын азайтуға мүмкіндік беретін DTD негізінде оңтайландыру (тиімділеу) әдісін ойлап тапты. Негізгі идея – өңделетін құжаттың қай бөліктері жаңартылатындығын статикалық түрде анықтау. Бұл әдіс, әсіресе, өте үлкен XML құжаттарын жаңарту үшін пайдалы (Баазизи, 2011: 11).

Bidoit және басқалар үлкен XML құжаттарын басқарумен айналысады және өте үлкен XML құжаттарының жаңартулары мен сұраныстарын өңдеу үшін олар Andromeda жүйесін ұсынады. Жаңартулар XQuery Update Facility өрнектері ретінде көрсетілген. Andromeda жүзеге асыратын әдіс Map/Reduce кластерінің машиналарына жаңарту жүктемесін тарату мақсатында жаңартылатын үлкен құжаттың статикалық және / немесе динамикалық бөлінуінен тұрады (Бидоит, т.б., 2015:10).

Dargham және басқалар (Даргхам, т.б., 2008: 5) DTD ретінде анықталған XML сұлбаларына қатысты олардың сенімділігін сақтай отырып, XML құжаттарын жаңарту тәсілін ұсынады. Бұл тәсіл мынаған негізделген: XQuery тілін төрт жаңарту операторымен кеңейтетін XML жаңарту тілінің XUpdate-ті қолдану: insertAfter, insertBefore, delete және modify; XML құжаттарын сақтау үшін объектіге бағытталған F2 ДҚБЖ-ін пайдалану; ДҚБЖ-де орындалатын XUpdate тапсырыстарын пайдаланушылардың OQL (object Query Language) жаңарту мәлімдемелеріне түрлендіруі; ішінара әдісті қолдану (тек XML құжатының жаңартылған бөлігі тексеріледі) және жедел тексеру (құжатты тексеру жаңартуды жасамас бұрын жасалады және жаңарту операциясынан кейін ол DTD-ге сәйкес келген жағдайда ғана жаңартылады).

Kircher және т.б. XML жаңартуларын өңдеу уақытын Pre/Dist/size XML кодтауға дейін қысқартатын «құрылымдық жаппай жаңартуды» оңтайландыру әдісін ұсынады (Кирчер, т.б, 2015:9).

Klettke және т.б. (Клеттке, т.б., 2005: 12) жаңартудың төрт негізгі амалдарының жиынтығын ұсынады: туынды элементті ЖОЮ, мазмұнды КІРІСТІРУ [(Before| After) ref], туынды элементті атау үшін ҚАЙТА АТАУ және туынды элементті мазмұнмен АУЫСТЫРУ. Осы операцияларды XML құжатына қолданған кезде оның сұлбасына қатысты жарамсыз болып қалатындықтан, авторлар осы операцияларды орындау үшін төрт тәсілді (IGNORE, REJECT, REDO және EVOLVE) енгізеді. IGNORE тәсілі барлық XML жаңарту әрекеттерін оның сұлбасына қатысты жаңартылған құжаттың дұрыстығын тексерусіз қабылдайды. Оны eXcelon корпорациясының өзінің кеңейтілген XML ДҚБЖ ақпараттық сервері қолдайды. REJECT тәсілі оның сұлбасына

сәйкес келмейтін жаңартылған XML құжатын жасайтын жаңарту операцияларынан бас тартады. Оны Software AG корпорациясының Tamino компаниясы жүзеге асырады. REDO әдісі XML сұлбасын жарамсыз ететін XML жаңарту әрекеттерін келесідей өндейді: (a) біріншіден, пайдаланушы XML құжатында көрсеткен жаңарту әрекеті қабылданбайды, өйткені ол мұндай құжаттың сұлбасын бұзады; (b) содан кейін әзірлеуші XML сұлбасын қолмен бейімдейді; (c) содан кейін әзірлеуші бастапқы сұлбаға қатысты барлық басқа XML құжаттарын бейімдейді, олар бейімделген сұлбаға қатысты жарамды болуы үшін жаңартылатын құжатты қоспағанда; (d) соңында, соңғы пайдаланушы жаңарту әрекетін қайталайды, оны әдетте еш қиындықсыз орындау керек. (iv) EVOLVE тәсілі қажет болған жағдайда тізбектің эволюциясын тудыратын барлық жаңарту операцияларын қолдайды. Шынында да, жаңарту әрекеті сұлбада көрсетілген кейбір шектеулерді бұзған кезде, XML құжаты жаңартылады және оның сұлбасы пайдаланушының немесе әзірлеушінің араласуынсыз автоматты түрде бейімделеді.

(Ни, т.б., 2007: 5) еңбекте авторлар XML GLASS сұрауларының графикалық тілін кеңейтетін жаңа XML GLASSU графикалық жаңарту тілін ұсынады (жартылай құрылымдалған мәліметтерге арналған графикалық сұрау тілі). GLASSU пайдаланушыдан XML – ді жаңартудың графикалық ретін жасау кезінде XML деректерінің сұлбасы мен семантикасы туралы білімді талап етеді. XML сұлбасының моделі – ORAS, объект – қатынас – атрибут моделі жартылай құрылымдалған мәліметтер үшін қарастырылған. GLASSU жаңарту өрнегі – бұл үш бөліктен тұратын график: шарт бөлігі, көрініс бөлігі және әрекет бөлігі. Әрекеттің бір бөлігі ғана қажет; қалған бөліктер міндетті емес.

O'Connor және Roantree (O'Коннор, т.б., 2010: 9) тиімді жаңарту механизмдерін қолдау үшін XML құжаттарын динамикалық түрде белгілеу және кодтау үшін қажетті 10 қасиеттер жинағын анықтайды. Бұл қасиеттерге: құжаттың реті, кодтау көрінісі, тұрақты белгілер, XPath ұпайлары, деңгейді кодтау, толып кету мәселесі, ортогональды таңбалау сұлбасы, ықшам кодтау, бөлуді есептеу және рекурсивті таңбалау алгоритмдері жатады.

Pardede және басқалары (Пардеде, т.б., 2005: 17) XML құжаттарын сол құжаттардың концептуалдық шектеулерін бұзбай жаңартуға мүмкіндік беретін әдістемені ұсынады. Бұл әдістеме жаңарту әрекеттерін орындамас бұрын тексеретін және осы операциялар жаңартылған XML құжаттарының концептуалдық шектеулерін сақтауын қамтамасыз ететін функциялар жинағынан тұрады. Шектеулер 3 құрылымдық қатынасқа құрылған:

- қатысушылар түрлерінің саны және сілтеме тұтастығы сияқты шектеулерді қамтитын ассоциация,

- реттілік және біркелкілік сияқты шектеулерді қамтитын біріктіру,

- тек қана екі шектеуді қамтитын мұрагерлік қатынастар: ерекше дизъюнктура және жоғары тармақтарының саны.

Pardede және басқалар (Пардеде, т.б., 2008: 25) XML құжаттарының семантикасын сақтай отырып, XML қолдайтын мәліметтер базасында сақталған XML құжаттарын жаңартуға бағытталған. Авторлар XML деректері үшін көрсетілуі мүмкін XML концептуалдық шектеулерінің жинағын анықтаудан бастайды. Осыдан кейін олар бұл концептуалды шектеулерді SQL/XML стандартында көрсетілген сұлба ретінде ұсынылған логикалық модельге түрлендіреді. Содан кейін авторлар XML деректерінде орындалған кез келген жаңарту әрекетінен (яғни, кірістіру, жою немесе ауыстыру) кейін барлық анықталған XML шектеулерін сақтайтын жалпы XML жаңарту әдістерін (немесе алгоритмдерін) ұсыну үшін осы сұлбаны пайдаланады. Соңында, олар ұсынылған жаңарту әдістерін XML қосылған негізгі деректер қорларының бірінде (Oracle 10g) жүзеге асырады.

Wang және Zhang (Ванг, т.б., 2005: 4) XML – RL сұрау тілін кеңейтетін декларативті XML жаңарту тілі мен XML-RL жаңарту тілін ұсынады. Ол келесі ерекшеліктермен сипатталады: (i) ол жоғары деңгейдегі мәліметтер үлгісіне негізделген; (ii) ол құжат иерархиясындағы бірнеше деңгейлерді қамтитын күрделі жаңарту тапсырыстарын бір мәнді анықтауға мүмкіндік береді; (iii) күрделі объектілерді жаңартуды қолдайды; (iv) ол логикалық байланыстыру айнымалыларының үш түрін: атау айнымалылары, мәндер айнымалылары және объект айнымалыларын пайдаланып тег атауларын, мәндерді және объектілерді өзгертуге арналған бірыңғай синтаксисті қамтамасыз етеді.

Өз еңбегінде Wong (Вонг, т.б., 2001: 9) XML деректерін жаңартуды қолдау үшін XQL (XML Query Language) кеңейтеді. Кеңейтілген SQL бес жаңарту құрылымын - кірістіру, жаңарту, көшіру, орын ауыстыру және жоюды қамтиды. Ол SODA2(Semistructured Object DAtabase system, 2-нұсқа) жергілікті XML ДҚБЖ аясында жүзеге асырылады.

Татаринов және т.б. (Татаринов, т.б, 2001:11) XML деректерін жаңарту үшін 7 негізгі әрекеттің жиынтығын ұсынады: жою (туынды элемент), атын өзгерту (туынды элемент, атауы), кірістіру (мазмұн), InsertBefore (сілтеме, мазмұн), InsertAfter (сілтеме, мазмұн), ауыстыру (туынды элемент, мазмұн) және кірістірілген жаңарту (patternMatch, predicates, updateOp). XML жаңартуы жоғарыда аталған әрекеттерді

орындау тізбегі ретінде анықталады. Авторлар XML деректері реляциялық ДҚБЖ – де сақталатын және басқарылатын реляциялық мәліметтерде көрсетілетін ортада жаңарту операцияларын жүзеге асыру стратегиясын анықтайды.

XML құжаттарының жаңартулары XML сұлбасының эволюциясын қолдайтын орталарда да бар. Шын мәнінде, кейбір өзгерістер XML сұлбасына қолданылған кезде, бастапқы сұлба үшін рұқсат етілген XML құжаттары жаңа сұлба бойынша қайта тексерілуі керек (біртіндеп) және егер олар жарамсыз болса, XML жаңарту операцияларының реттілігі арқылы осы жаңа сұлбаға бейімделуі керек; мұндай тізбекті әзірлеуші қамтамасыз етеді немесе сұлбаның эволюция жүйесі автоматты түрде жасайды. Сұлба эволюциясы кезінде XML құжаттарының бейімделуін Guerrini тобы кеңінен зерттеді (Солимандо, т.б., 2012:9). Сонымен қатар, Guerrini сұлбаның эволюциясы процесінде бейімделетін XML құжаттарын қосымша тексеруге қарастырды. Cavalieri және т.б. (Кавальери, т.б., 2011:9) дамыған XML сұлбасына бейімделу кезінде XML құжаттарына қолданылатын XML жаңарту операцияларының реттілігін азайту әдісін ұсынады. Назар аударатын жайт: Vouchou және Duarte (Бушу, т.б., 2007:14) сияқты, Cavalieri және т.б., (Кавальери, т.б., 2011:9) XML құжатының жаңартуларына тиісті XML сұлбасына әсер етуге мүмкіндік береді. Шынында да, жаңартудан кейін XML құжаты өзінің сұлбасы үшін жарамсыз болған кезде, Cavalieri және т.б., (Кавальери, т.б., 2011:9) ұсынған тәсіл – сұлбаны жаңартылған құжаттың құрылымын сипаттайтын етіп жаңартуға негізделген.

Қарапайым XML мәліметтер базасында XML деректерін басқару үшін ғылыми қауымдастық ұсынған негізгі тәсілдерді қорытындылай келе, 2-суреттегідей 4 критерийлерді бөліп қарастыруға болады:

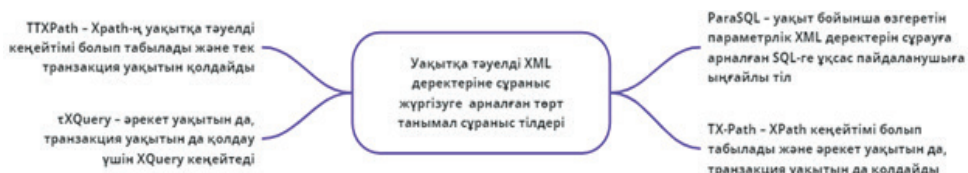


2 сурет - XML деректерін басқару үшін ұсынылған негізгі тәсілдердің 4 критерийлері

Уақытқа тәуелді XML деректер қорында XML деректерін өңдеудің заманауи әдістері

Соңғы онжылдықтарда XML икемділігі мен практикалық қолдануына байланысты уақытқа тәуелді деректерді сақтау, өңдеу және сұраныстар үшін стандартты емес мәліметтер базасының жаңа буыны – құжат бойынша NoSQL-ге бағытталған моделге айналды. XML уақытқа тәуелді ақпараттың ең нақты әрі тиімді көрінісі болып саналатын уақытқа тәуелді топтастырылған деректер модельдеріне керемет қолдау көрсетеді. XML деректерін өңдеудің көптеген әдістерінде сұлбаға бағытталған және жартылай құрылымдалған ақпараттық ресурстарды құруға байланысты бірқатар қиындықтар бар.

3 – суретте уақытқа тәуелді XML деректеріне сұраныс жүргізуге арналған төрт танымал сұраныс тілдері көрсетілген:



3 сурет - Уақытқа тәуелді XML деректеріне сұраныс жүргізуге арналған төрт танымал сұраныс тілдері

Қолданыстағы алғашқы XML-дің ДҚБЖ (eXist, xDB және Sedna сияқты) және коммерциялық XML қолдайтын ДҚБЖ (Oracle 12c және DB2 v10 сияқты) уақытқа тәуелді XML деректерін өңдеуге (енгізуге, жоюға және өзгертуге) ешқандай қолдау көрсетпейтінін ескеру қажет.

Norvag және басқалары (Norvag, т.б., 2003:10) уақытқа тәуелді деректерді қолдау үшін ішкі ДҚБЖ модульдерін өзгерту/кеңейту үшін көп деңгейлі және интеграцияланған тәсілді пайдаланады. Авторлар уақытқа тәуелді деректердің нұсқаларын сақтауға және іздеуге, белгілі бір уақытта жарамды құжаттарды алуға және құжаттарға өзгертулерді сұрауға, сондай-ақ деректерді манипуляциялау операцияларын (енгізу, жаңарту және жою) орындау үшін жеке ДҚБЖ-ң V2 – ні ұсынады.

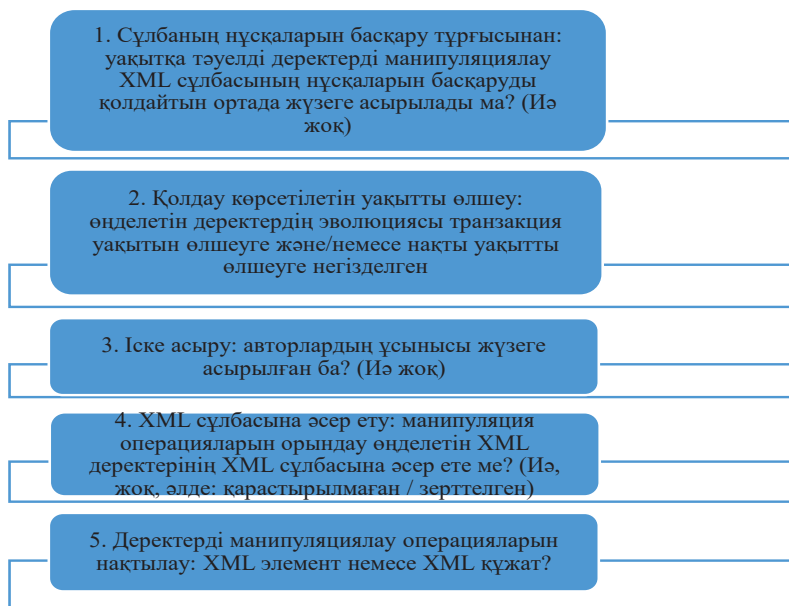
Wang және Zaniolo (Ванг, т.б., 2008:18) уақытқа тәуелді топтастырылған мәліметтер модельдерін қолдана отырып, XML құжаттары түрінде транзакция уақыты, әрекет ету уақыты және аз уақыттық реляциялық мәліметтер базасының мазмұнын ұсынатын және жариялайтын тәсілді ұсынады. Мұндай тәсіл Xquery – де жазылған реляциялық кестелер үшін SQL тілінде құруға қиын келетін күрделі уақытқа тәуелді сұрауларды және пайдаланушы анықтаған функцияларды қолдана

отырып, уақытқа тәуелді деректерді енгізу, жою және өзгертулерді қолдайды.

XML уақытқа тәуелді деректер қорлары саласында тXSchema (Temporal XML Schema) (Роддик, т.б., 2018:12) уақытқа тәуелді XML деректерін құруға және тексеруге арналған белгілі және толық құрылым (мәліметтер үлгісі, тілі және құралдар жинағы) болып табылады. Соңғылары тұрақты XML деректерінен логикалық және физикалық аңдатпалар жиынтығын қолдану арқылы жасалады.

Уақытқа тәуелді сұлба кәдімгі сұлба мен оған сәйкес логикалық және физикалық аңдатпалар жиынын біріктіру арқылы жасалады. тXSchema сонымен қатар сұлба нұсқасын қолдау үшін кеңейтілді [19]. Дегенмен, мұндай құрылым уақытқа тәуелді XML деректерін өңдеу тілін қамтамасыз етпейді.

Зерттеу қоғамдастығы тарапынан уақытқа тәуелді XML деректерін басқаруға ұсынған негізгі тәсілдерді қорытындылай келе, келесі бес критерийді ескеру керекпіз (4 – сурет):



4 сурет - XML деректерін басқаруға ұсынған негізгі тәсілдерді қорытындысы бойынша бес критерий

Зерттеу нәтижесінде қарапайым XML деректерін манипуляциялау бойынша көптеген зерттеулер жүргізілді, ал уақытқа тәуелді XML деректерін манипуляциялау шектеулі дәрежеде ғана қарастырылды. Демек, бұл салада әлі көп жұмыс істеу керек.

Зерттеу нәтижелері. XML деректерін басқарудың барлық заманауи әзірлемелерін және оларға қойылатын талаптарды ескере отырып, авторлар гетерогенді ақпараттық жүйенің нақты ақпарат көзінен XML форматындағы жазбаны алудың өзіндік моделін және оны редакциялау үшін осы жазбаны ұсынуды жасады.

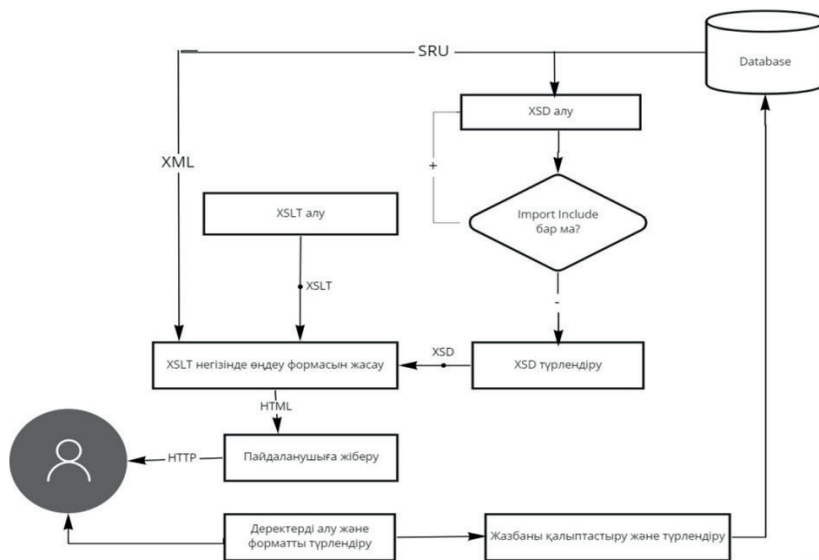
Бұл гетерогенді ақпараттық жүйенің нақты ақпарат көзінен алынған XML жазбасының мүмкін құрылымының (XSD) толық сипаттамасын алу және оны өңдеу үшін осы жазбаны ұсыну технологиясы болып табылады. XML құжатының өзінде оны браузер терезесінде қандай түрде көрсету керектігі туралы нұсқаулар жоқ. XSL стиль кестесін қолдана отырып, XSLT тілінің мүмкіндіктері XML құжатынан қажетті ақпаратты бөліп алуға және оны оқуға ыңғайлы түрде ұсынуға, оны кез-келген басқа формада, атап айтқанда HTML түрінде шығаруға мүмкіндік береді (Мухитова, т.б., 2018:4).

Біз өңделген XML құжатының құрылымы (XSD) туралы ақпаратты алудың келесі моделін ұсындық. XSD алу үшін 1- кестеде көрсетілген 3 тәсілдің біреуін орындау керек:

1 кесте - XSD алу модельдері

XSD алу қадамдары		
1-тәсіл	2-тәсіл	3-тәсіл
Пайдаланылған атаулар кеңістігін анықтаған кезде schemaLocation атрибутының мәніне қарай URL мекен-жайы ретінде XML жазбасының түпкі элементіндегі XSD мәліметтер сұлбасына сілтеме алу	Егер XML өңдеуге арналған жазбада атаулар кеңістігінің идентификаторы (URI) болса, бірақ URL түрінде пайдаланылатын XSD мәліметтер сұлбасына сілтеме болмаса, XSD атаулар кеңістігі идентификаторы арқылы XSD ұсыну туралы ақпараттық жүйеге сұраныс жүргізу. Ұқсас сұранысты, мысалы, Explain қызметі өңдей алады.	XML редакциялау үшін алынатын жазбада атаулар кеңістігінің анықтамалары болмаған жағдайда, деректерді алуға сұраныс құруды қалыптастыру кезінде сұраныс жүргізілетін сұлбаға сәйкес келетін ақпараттық ресурстың (мәліметтер базасының) атауы бойынша XSD беру немесе XSD пайдалану туралы ақпараттық жүйеге сұраныс құру.

Сұлбалық түрде XML құжатын түрлендіру процесін келесідей ұсынуға болады (5 – сурет):

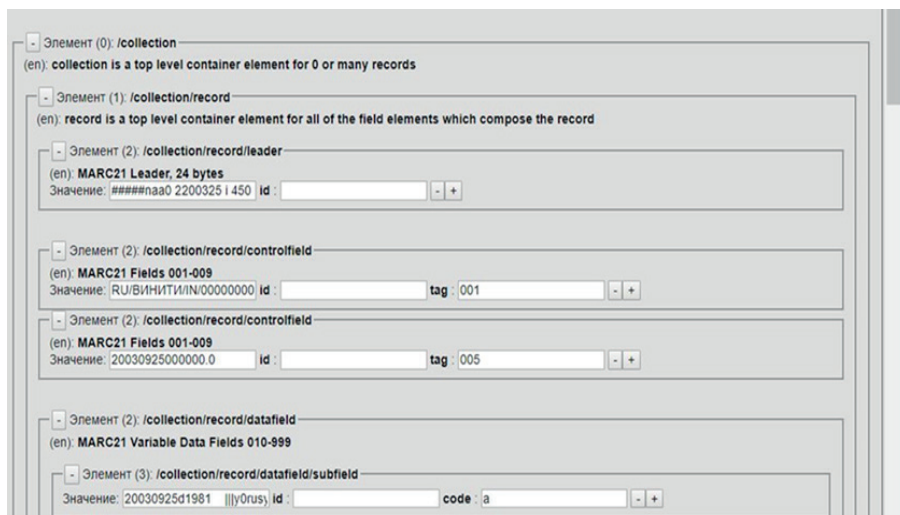


5 сурет - XML құжатын түрлендіру процесі

Модельдің жұмыс принципі келесідей: файл кірісінде*.xml деректер схемасының тақырыбында *.xsd болуын тексереді. Болмаған жағдайда оны қалыптастыруға сұраныс жасалады. Одан кейін бар немесе жаңадан құрылған *.xsd файлдың дұрыстығы тексеріліп, XSLT – түрлендірулері арқылы xml жазбалары өңделеді.

Келесі этапта, сервер XML құжатына стиль кестесін (CSS) қолданады және оны пайдаланушы интерфейсі ретінде ұсынылған басқа форматқа (әдетте HTML) түрлендіреді. Пайдаланушы ақпаратты енгізеді немесе өңдейді және ақпараттың сыртқы көрінісі өзгерген сайын жаңа сұраныс орындалады және HTML – парағы қайтадан жасалады. Жаңадан құрылған XML құжаты ДҚБЖ-ге жіберіледі.

Бағдарлама таңдалған XSD схемасына сәйкес XML құжаттарын жасау және өңдеу үшін экрандық нысандарды ұсынатын серверлік веб-қосымша болып табылады. Экран формалары XSLT технологиясын қолдана отырып, XSD анықтамалары негізінде сервер жағында қалыптасады және пайдаланушыға HTML – парақтары түрінде беріледі. Толтырылған экран формалары (6-сурет) XML құжатты құру үшін сервермен өңделеді.



6 сурет - XML жазуды өңдеудің экрандық формасы

XSD мәліметтер сұлбасының анықтамаларында ағымдағы атаулар кеңістігіндегі (xsd элементі:include) және басқа атаулар кеңістігіндегі (XSD элементі:import) анықтамаларды толықтыратын басқа XSD мәліметтер сұлбаларына сілтемелер болуы мүмкін. Сондықтан XSLT процессорымен өңдеуден бұрын XSD бастапқы құрылымы қосымша анықтамаларды есепке алу үшін модификацияны қажет етеді.

Зерттеулер көрсеткендей, әзірленген модель XSLT түрлендірулерін қолдана отырып, XSD қосымшасының деректер қорының сұлбалары негізінде гетерогенді ақпараттық ресурстарды біріктіру мәселесін тиімді шешуге мүмкіндік береді. Сипатталған әдіснаманы қолданудың жалпы тәсілдері бар және оны деректерді енгізу және өңдеу үшін жіберілген HTML формаларын құруға мүмкіндік беретін бейімделген графикалық веб-интерфейстерді құру үшін қолдануға болады.

XSD-ді XSLT көмегімен өңдеу кезінде XSD анықтамаларының әртүрлі ерекшеліктері, соның ішінде сілтемелер мен рекурсия ескеріледі. Бағдарламаны XML форматында ұсынылған деректерді құру және өңдеудің тәуелсіз модулі ретінде, сондай-ақ гетерогенді ақпараттық жүйелер үшін әртүрлі серверлік бағдарламалық жасақтамаға енгізілген модуль ретінде пайдалануға болады.

Ұсынылған модельдің XML деректері бойынша контекстік іздеуді ұйымдастырудың тиімділігін арттыру, пайдаланушының метадеректерге қол жеткізуін және импортын жеделдету, өзгертілген деректерді ыңғайлы өңдеу және сақтау процестері артықшылығы болып келеді.

Талқылау. Зерттеу көрсеткендей, XML форматындағы библиографиялық ақпаратты сипаттау стандартына көшу және оларды өңдеу үшін жауап беретін веб-интерфейстерді (клиент-серверлік қосымшалар) құру әртүрлі бағдарламалық және аппараттық платформаларды қолданатын тәуелсіз пайдаланушыларға бір-бірінен алынған ақпаратты оңай пайдалануға мүмкіндік береді.

Адаптивті графикалық веб-интерфейсті құрудың дамыған моделі және оны таратылған ақпараттық жүйеге тікелей енгізу, деректерді енгізу және өңдеу үшін ыңғайлы графикалық интерфейсін ұсыну арқылы сұраныс пен сұралған ресурсқа жауап арасындағы уақытты азайтудың жоғары тиімділігін көрсетті. Графикалық веб-интерфейсті таратылған жүйенің ішкі ресурстарына бейімдеу мүмкіндігі есебінен гетерогенді ақпараттық ресурстарды интеграциялау міндеттері тиімді шешілуде.

Соңғы уақытта интернеттегі әлеуметтік желілер, ақылды қалалар сияқты заманауи қосымшалар сұлбаның барлық нұсқаларын, тиісінше қарапайым және уақытқа тәуелді NoSQL деректерімен сақтайтын сұлбаларының нұсқасын басқаруды кеңінен қолданады. Енді NoSQL деректерді сақтау және бөлісу үшін ең танымал JSON деректер форматтарының бірін қолдануды жөн көреді.

JSON-ны жасау оңай және деректерді өңдеусіз таратуға қолайлы, ал XML құрылымдық жағынан күрделірек, бірақ деректерді таратуға ғана емес, файлдарды өңдеуге және пішімдеуге де мүмкіндік береді. JSON атаулар кеңістігін қолдай алмайды, бірақ Ajax құралдарын толығымен қолдай алады. XML атаулар кеңістігін қолдайды, бірақ Ajax құралдарын қолдай алмайды.

Метадеректерді өңдеуді талдау JSON форматындағы құжаттарды XML-ге түрлендіру моделін жасау қажеттілігін көрсетеді. Осыған байланысты келесі негізгі шешімдер бар: XML өңдеуге арналған құжаттың объектілік моделін қолдана отырып, JSON форматтарын XML-ге түрлендіру және шаблон механизмін қолдана отырып, индекстелген массивке JSON құжатын жүктеу. Әрі қарай зерттеу осы бағытта жүреді.

Қорытынды. Жоғарыда сипатталған деректерді өңдеудің графикалық веб-интерфейсінің жауапты моделін құру технологиясы серверлік қосымшаны әзірлеу болып табылады.

Гетерогенді деректер үшін ақпаратты өңдеу интерфейсін құрудың ұсынылған әдісі әзірленген графикалық веб-интерфейстерді ақпараттық ресурстың белгілі бір құрылымына автоматты түрде бейімдеуге мүмкіндік береді. Арнайы графикалық XML редакторының дамыған моделі кез-келген XML деректерін бағдарламалық кодты өзгертпестен

кез-келген құрылымның бастапқы деректерін түрлендіру арқылы өңдеуге мүмкіндік береді. Сонымен қатар, пайдаланушылар оны XML форматындағы файлдармен жұмыс істеу үшін редактор түрінде тәуелсіз функционалды жүйе ретінде пайдалана алады.

Құрастырылған XML жазбаларының адаптивті графикалық редакторының моделі ZOOSPACE FIC IVT SRAS (Есептеу технологиялары институты, Новосибирск) платформасының ZOOSPACE –W ішкі жүйесіне енгізілді. Қазіргі уақытта авторлар XML форматында ұсынылған библиографиялық деректерді толық функционалды өңдеуге мүмкіндік беретін библиографиялық МБ XML – серверін құрастыру, сонымен бірге қолдау көрсетілетін XSD және JSON элементтерінің тізімін кеңейту бөлігінде модельдің функционалдығын арттыру бойынша жұмыс жүргізуде.

Information about the authors:

Mukhitova Aigul – Ph.D student, specialty “Information systems”, researcher, al-Farabi Kazakh National University, *mukhitova.aigul@gmail.com*, <https://orcid.org/0000-0002-4081-7694>;

Yerimbetova Aigerim – Ph.D, Associate Professor, Institute of Information and Computational Technologies, Almaty, Kazakhstan, *aigerian@mail.ru*, <https://orcid.org/0000-0002-2013-1513>;

Barakhnin Vladimir – Doctor of Technical Sciences, Professor, Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia, *bar@ict.nsc.ru*, <https://orcid.org/0000-0003-3299-0507>;

Daiyrbayeva Elmira – research fellow, Institute of Information and Computational Technologies CS MES RK; assist. prof., International Educational Corporation, Almaty, Kazakhstan, *nurbekkyzy_e@mail.ru*, <https://orcid.org/0000-0002-4255-5456>;

Adalbek Alibek – Senior Lecturer, L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan, *q_al@mail.ru*.

ӘДЕБИЕТТЕР:

Баазизи М. - А., Н. Бидуа, Д. Колазцо, Н. Малла, М. Саакян (2011). XML жаңартуын оңтайландыруға арналған Проекция: мәліметтер базасын кеңейту бойынша 14-ші халықаралық конференция материалдары, EDBT’11, Уппсала, Швеция, 21-24 наурыз, 307-318 бет (ағылшын тілінде).

Бидоит Н. Бидуа, Д. Колазцо, К. Сартиани, А. Солимандо, Ф. Ульяна (2015). Андромеда: үлкен XML құжаттарының сұраныстары мен жаңартуларын өңдеу жүйесі, Т. Морзи, П. Вальдуриес, Л. Беллатрес (ред.), мәліметтер базасы мен ақпараттық жүйелердегі жаңа тенденциялар: ADBIS 2015 BigDap қысқаша баяндамалары мен

семинарлары, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, in: компьютерлік және ақпараттық ғылымдардағы коммуникациялар, 539 том, Springer International Publishing, Швейцария, 18-228 беттер (ағылшын тілінде).

Бушу Б., Дуарте Д. (2007). XML схемасының эволюциясына көмектесу. 22-ші Бразилиялық мәліметтер базасы симпозиумының материалдары, SBBD'07, Жуан песоа, Параиба, Бразилия, 15-19 қазан, 270-284 беттер (ағылшын тілінде).

Ванг Ж.-Р. Жанг Х.-Л. (2005). Жоғары деректер моделіне негізделген XML декларативті жаңарту тілі, J.Comput. Sci. Технол. 20 (3), 373–377 беттер (ағылшын тілінде).

Ванг Ф., Заниоло С., Жоу З. (2008). ArchIS: транзакциялар кезіндегі уақыт базасының жүйелеріне XML-ге негізделген тәсіл. VLDB J. 17 (6) (2008) 1445-1463 беттер (ағылшын тілінде).

Вонг Р.К. (2001). Үлкен XML дерекқорларын сұрау және жаңарту үшін кеңейтілген XQL. 2001 жылғы ACM құжаттарды әзірлеу симпозиумының материалдары, DocEng '01, Атланта, Джорджия, АҚШ, 09-10 қараша, 95-104 беттер (ағылшын тілінде).

Даргам Ж., Алти З., Карам М. (2008). XML құжаттарын олардың жарамдылығын бұзбай жаңарту: Интернет және веб-қосымшалар мен қызметтер бойынша 3-ші Халықаралық конференцияның материалдары, ICiW'08, Афина, Греция, 8-13 маусым, 342-347 беттер (ағылшын тілінде).

Жижимов О.Л. (2015). Zoospace платформасының Explain қызметтері және пайдаланушы интерфейсі, CEUR семинар материалдары, №.1536:30-36. ISSN 1613-0073. - <http://ceur-ws.org/Vol-1536/paper4.pdf> (55145) (орыс тілінде).

Кавалерья Ф., Гюеринни Г., Месити М., Олибони Б. (2011). XML құжаттар мен схеманы жаңарту операцияларын азайту бойынша. Data Engineering 27-ші халықаралық конференция, ICDE'11 семинар, Hannover, Germany, 11-16 сәуір, бет.77-86 беттер (ағылшын тілінде).

Кирчер Л., Гроссниклаус М., Грюн К., Шолл М. (2005). Pre /Dist /Size XML кодтаудағы тиімді құрылымдық жаппай жаңартулар: IEEE-дің деректерді әзірлеу жөніндегі 31-ші халықаралық конференциясының материалдары, ICDE '15, Сеул, Оңтүстік Корея, 13-17 сәуір, 447-458 беттер (ағылшын тілінде).

Клеттке М., Мейер Х., Хенсель Б. (2005). Эволюция: XML жаңартудың кері жағдайы, ICDE семинарларының материалдары, 2005 ж., Токио, Жапония, 5-8 сәуір, 2005 ж., 1279 бет (ағылшын тілінде).

Мухитова А., Жижимов О.Л. (2018). Гетерогенді ақпараттық жүйелер үшін адаптивті әкімшілік графикалық веб-интерфейстерде деректерді енгізу және өңдеу моделін құруда XML технологияларын қолдану. Ақпараттық салалардағы аналитика және деректерді басқару: XX халықаралық конференция DAMDID/RCDL, Мәскеу, Ресей, Ресей Ғылым академиясының Орталық ғылыми-зерттеу орталығы. - Б.157-160. - ISBN: 978-5-519-65438-8 (ағылшын тілінде).

Ни В., Линг Т. (2007). XML деректерін графикалық тілдер арқылы жаңарту. Оқу құралдары, плакаттар, панельдер және өнеркәсіптік материалдар тұжырымдамалық модельдеу жөніндегі 26-шы халықаралық конференция, ER'2007, Окленд, Жаңа Зеландия, 5-9 қараша, 209-214 беттер (ағылшын тілінде).

Норваг К., Лимстранд М., Миклебуст Л. (2003). ТехOR: мәліметтер базасының объектілік-реляциялық жүйесіндегі XML уақытша дерекқоры. Андрей Ершовты еске алуға арналған 5-ші конференцияның материалдары, PSI'2003, Академгородок, Новосибирск, Ресей, 9-12 шілде, ішінде: LNCS, 2890 том, 520-530 беттер (ағылшын тілінде).

О'Коннор, Роантри М. (2010). XML жаңарту механизмдерінің қажетті қасиеттері. 2010 жылғы EDBT/ICDT семинарларының материалдары, Лозанна, Швейцария, 22-26 наурыз, №23 мақала (ағылшын тілінде).

Пардеде Е., Рахауи Ж., Танияр Д. (2005). XML жаңарту кезінде тұжырымдамалық шектеулерді сақтау. *Int J. Web Inf. Syst.* 1 (2), 65-82 беттер (ағылшын тілінде).

Пардеде Е., Рахауи Ж., Танияр Д. (2008). XML дерекқорындағы XML деректерін жаңартуды басқару, *J. Comput. Sci* Жүйесі. 74 (2). 170-195 беттер (ағылшын тілінде).

Роддик Ж.Ф., L. Liu, M.T. Özsü (Eds.), (ред.). (2018). Схемалардың нұсқаларын басқару. Мәліметтер базасының энциклопедиясы, Нью-Йорк, Нью-Йорк, АҚШ, <http://dx.doi.org/10.1007/978-1-4614-8265-9> (ағылшын тілінде).

Солимандо А., Делзанно Г., Геррини Г. (2012). XML құжаттарының бейімделуіне статикалық талдау. *er 2012 CMS, ECDMNoCoDA, MoDIC, MORE-BI, RIGiM, SeCoGIS, WISM, Флоренция, Италия, 15-18 қазан, 57-66 беттер* (ағылшын тілінде).

Татаринов И., Айвс З.Г., Халеви А.Ю., Уэлд Д.С. (2001). XML жаңартуы. ACM SIGMOD халықаралық деректерді басқару конференциясының материалдары, ACM SIGMOD'01 конференциясы, Санта-Барбара, Калифорния, АҚШ, 21-24 мамыр, 413-424 беттер (ағылшын тілінде).

REFERENCES:

Baazizi M.-A., Bidoit N., Colazzo D., Malla N., Sahakyan M. (2011). Projection for XML update optimization, in: *Proceedings of the 14th International Conference on Extending Database Technology, EDBT'11, Uppsala, Sweden, 21–24 March*, pp. 307–318.

Bidoit N., Colazzo D., Sartiani C., Solimando A., Ulliana F. (2015). Andromeda: A system for processing queries and updates on big XML documents, in: T. Morzy, P. Valduriez, L. Bellatreche (Eds.), *New Trends in Databases and Information Systems: ADBIS 2015 Short Papers and Workshops BigDap, DCSA, GID, MEBIS, OASIS, SW4CH, WISARD*, in: *Communications in Computer and Information Science*, vol. 539, Springer International Publishing, Switzerland, pp. 218–228.

Bouchou B., Duarte D. (2007). Assisting XML schema evolution that preserves validity, in: *Proceedings of the 22nd Brazilian Symposium on Databases, SBBD'07, João Pessoa, Paraíba, Brasil, 15–19 October*, pp. 270–284.

Cavaliere F., Guerrini G., Mesiti M., Oliboni B. (2011). On the reduction of sequences of XML document and schema update operations, in: *Workshops Proceedings of the 27th International Conference on Data Engineering, ICDE'11 Workshops, Hannover, Germany, 11–16 April*, pp. 77–86.

Dargham J., Alti Z., Karam M. (2008). Updating XML documents without breaking their validity, in: *Proceedings of the 3th International Conference on Internet and Web Applications and Services, ICIW'08, Athens, Greece, 8–13 June*, pp. 342–347.

Kircher L., Grossniklaus M., Grun C., Scholl M.H. (2015). Efficient structural bulk updates on the Pre/Dist/Size XML encoding, in: *Proceedings of the 31st IEEE International Conference on Data Engineering, ICDE'15, Seoul, South Korea, 13–17 April*, pp. 447–458.

Klettke M., Meyer H., Hänsel B. (2005). Evolution: The other side of the XML update coin, in: *Proceedings of the ICDE Workshops 2005, Tokyo, Japan, 5–8 April*, paper 1279.

Ni W., Ling T.W. (2007). Update XML data by using graphical languages, in: *Proceedings of the tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling, ER'2007, Auckland, New Zealand, 5–9 November*, pp. 209–214.

O'Connor M.F., Roantree M. (2010). Desirable properties for XML update mechanisms, in: Proceedings of the 2010 EDBT/ICDT Workshops, Lausanne, Switzerland, 22–26 March, Article No. 23.

Pardede E., Rahayu J.W., Taniar D. (2005). Preserving conceptual constraints during XML updates, *Int. J. Web Inf. Syst.* 1 (2) 65–82.

Pardede E., Rahayu J.W., Taniar D. (2008). XML data update management in XML-enabled database, *J. Comput. System Sci.* 74 (2) 170–195.

Norvag K., Limstrand M., Myklebust L. (2003). TexOR: Temporal XML database on an object-relational database system, in: Proceedings of the 5th Andrei Ershov Memorial Conference, PSI'2003, Akademgorodok, Novosibirsk, Russia, 9–12 July, in: LNCS, vol. 2890, pp. 520–530, Revised papers.

Mukhitova A., Zhizhimov O.L. (2018). Application of XML-technologies in the Construction of Input and Editing Data Model in the Adaptive Administrative Graphical Web Interfaces for Heterogeneous Information Systems // Data Analytics and Management in Data Intensive Domains: XX International Conference DAMDID/RCDL'2018 (October 9-12, 2018, Moscow, Russia): Conference Proceedings / Edited by Leonid Kalinichenko, Yannis Manolopoulos, Sergey Stupnikov, Nikolay Skvortsov, Vladimir Sukhomlin. - Moscow: FRC CSC RAS. - P.157-160. - ISBN: 978-5-519-65438-8.

Roddick J.F., Liu L., Özsu M.T. (Eds.) (2018). Schema Versioning, second ed., *Encyclopedia of Database Systems*, New York, NY, USA, <http://dx.doi.org/10.1007/978-1-4614-8265-9>.

Solimando A., Delzanno G., Guerrini G. (2012). Static analysis of XML document adaptations, in: Proceedings of the ER 2012 Workshops CMS, ECDMNoCoDA, MoDIC, MORE-BI, RIGiM, SeCoGIS, WISM, Florence, Italy, 15–18 October, pp. 57–66.

Tatarinov I., Ives Z.G., Halevy A.Y., Weld D.S. (2001). Updating XML, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, ACM SIGMOD'01 Conference, Santa Barbara, CA, USA, 21–24 May, pp. 413–424.

Wang F., Zaniolo C., Zhou X (2008). ArchIS: an XML-based approach to transaction-time temporal database systems, *VLDB J.* 17 (6) 1445–1463.

Wong R.K. (2001). The extended XQL for querying and updating large XML databases, in: Proceedings of the 2001 ACM Symposium on Document Engineering, DocEng'01, Atlanta, Georgia, USA, 09–10 November, pp. 95–104.

Zhizhimov O.L. (2015). Explain Services on ZooSPACE Platform and Adaptive User Interfaces // CEUR Workshop Proceeding. - Vol.1536. - P.30-36. - ISSN 1613-0073. - <http://ceur-ws.org/Vol-1536/paper4.pdf> (55145).

Wang G.-R., Zhang X.-L. (2005). Declarative XML update language based on a higher data model, *J. Comput. Sci. Technol.* 20 (3) 373–377.

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

Volume 2, Number 342 (2022), 112–129
<https://doi.org/10.32014/2022.2518-1726.133>

УДК 50.47.29

**Б.Б. Оразбаев¹, Ж.Ж. Молдашева^{1*}, В.И. Гончаров²,
К.Н. Оразбаева³**

¹Л.Н.Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан;

²Томск политехникалық ұлттық зерттеу университеті, Томск, Ресей;

³Қазақ экономика, қаржы және халықаралық сауда университеті,
Нұр-Сұлтан, Қазақстан.

E-mail: zhadira1985@mail.ru

МАГИСТРАЛДЫ ҚҰБЫРЛАРМЕН МҰНАЙ ТАСМАЛДАУДЫ ДИАГНОСТИКАЛАУ ЖӘНЕ БАСҚАРУ ЖҮЙЕЛЕРІ

Аннотация. Зерттеудің өзектілігі мұнайды магистральдық құбырлар арқылы тасымалдау процестерін басқару кезінде бастапқы ақпараттың кейбір бөлігінің анық еместігін ескере отырып, мұнай құбыры қондырғыларының оңтайлы жұмыс режимдерін анықтау және таңдау қажеттілігімен негізделеді. Осыған байланысты технологиялық объектілердің тиімді жұмыс режимдерін, мысалы, бұлыңғыр ортада жиі сипатталатын ыстық мұнай құбыры жүйесінің мұнай жылыту станциясы, анық емес жиындардың математикалық аппараты негізінде көп өлшемді таңдау мәселелерін тұжырымдау және шешу өзекті ғылыми-практикалық міндет болып табылады. Математикалық модельдерді әзірлеу және бастапқы ақпараттың анық еместігі жағдайында магистральдық мұнай құбырларының мұнайды жылыту станциясының жұмыс режимдерін оңтайландыру.

Нысанның анық емес шығу параметрлері жағдайында модельдерді синтездеу әдісі жасалды, оның көмегімен магистральдық мұнай құбырының мұнай жылыту станциясының анық емес модельдері жасалды. Магистральды мұнай құбырының мұнайын қыздыру станциясының

тиімді жұмыс режимдерін анық емес ортада көп критериалды таңдау мәселесі ресімделді және модификация мен оңтайлылықтың әртүрлі принциптерінің үйлесімі негізінде формальды есептің математикалық тұжырымдары алынды. Негізгі критерий әдістері мен идеалды нүктенің анық емес ортада жұмыс істеу үшін модификациялау және бейімдеу арқылы бастапқы анық емес ақпаратты қолдана отырып, объектінің жұмыс режимдерін көп өлшемді таңдаудың тұжырымдалған мәселесін шешудің эвристикалық әдісі жасалды. Бұлыңғыр ортада көп өлшемді таңдаудың ұсынылған эвристикалық әдісі шешім қабылдаушының тәжірибесі мен білімін қолдануға негізделген. Шешім қабылдаушының бастапқы ақпараттың болуы мен қол жетімділігіне байланысты режимді таңдау мәселесін шешудің неғұрлым ыңғайлы әдісін таңдау мүмкіндігін қамтамасыз ету үшін әртүрлі оңтайлылық принциптерін және олардың комбинациясын қолдану арқылы таңдау мәселелерін тұжырымдау және шешу ұсынылады. Ұсынылған тәсіл Өзен-Атырау-Самара магистральдық мұнай құбырына Атырау пунктінің мұнайды жылыту станциясының жұмыс режимдерін көп критериалды таңдау міндетін қою және шешу кезінде іске асырылды. Алынған нәтижелер қойылған міндеттерді шешуге ұсынылған тәсілдің тиімділігін растады.

Түйін сөздер: магистральдық мұнай құбыры, мұнай тасымалдау, басқару, Мұнай айдау станциялары, жылыту пештері, диагностикалау, автоматтандырылған басқару жүйесі.

**Б.Б. Оразбаев¹, Ж.Ж. Молдашева^{1*}, В.И. Гончаров²,
К.Н. Оразбаева³**

¹Евразийский национальный университет имени Л.Н. Гумилева,
Нур-Султан, Казахстан;

²Томский национальный политехнический университет, Томск, Россия;

³Казахский университет экономики, финансов и международной
торговли, Нур-Султан, Казахстан.
E-mail: zhadira1985@mail.ru

ДИАГНОСТИРОВАНИЕ СИСТЕМЫ УПРАВЛЕНИЯ ТРАНСПОРТИРОВКИ НЕФТИ ПО МАГИСТРАЛЬНЫМ ТРУБОПРОВОДАМ

Аннотация. Актуальность исследования обосновывается тем, что при управлении процессами транспортировки нефти по магистраль-

ным трубопроводам возникает необходимость определения и выбора оптимальных режимов работы узлов нефтепроводов с учетом нечеткости некоторой части исходной информации. В связи с этим решение задачи многокритериального выбора эффективных режимов работы станции подогрева нефти для системы трубопроводов горячей нефти, которая часто описывается в нечеткой среде, на основе аппарата теорий нечетких множеств, является актуальной научно-практической задачей. Разработан метод синтеза моделей в условиях нечетких выходных параметров объекта, с помощью которого построены нечеткие модели исследуемой станции подогрева нефти магистрального нефтепровода. На основе модификации и сочетания различных принципов оптимальности получены математические решения задачи многокритериального выбора эффективных режимов работы станции подогрева нефти в нечеткой среде. Путем модификации и адаптации принципов гарантированного результата и равенства в нечеткой среде разработан эвристический метод решения сформулированной задачи выбора режимов работы объекта с использованием исходной нечеткой информации. Предлагаемый эвристический метод многокритериального выбора в нечеткой среде основан на использовании опыта и знаний лица, принимающего решения. Предлагаемый подход реализован при постановке и решении задач многокритериального выбора режимов работы станции подогрева нефти Атырау магистрального нефтепровода Узень-Атырау-Самара. В результате применения предлагаемого метода достигнуто повышение степени выполнения нечеткого ограничения воздействия на окружающую среду на 2 %, а также повышены оптимальные значения эксплуатационных параметров объекта: температура снижена на 1,85% (5,67 К), давление – на 0,04% (кПа), расход топлива – на 2,9% (0,0002 кг/с). Полученные результаты подтвердили эффективность предлагаемого подхода к решению поставленных задач.

Ключевые слова: магистральный нефтепровод, транспортировка нефти, управление, станции перекачки нефти, печи подогрева, диагностирование, автоматизированная система управления.

**B.B. Orazbayev¹, Zh.Zh. Moldasheva^{1*}, B.I. Goncharov²,
K.N. Orazbayeva³**

¹L.N.Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan;

²Tomsk Polytechnic University, Tomsk, Russia;

³Kazakh University of Economics, Finance and International Trade,
Nur-Sultan, Kazakhstan.

E-mail: *zhadira1985@mail.ru*

DIAGNOSTICS AND SYSTEMS OF OIL TRANSPORTATION THROUGH MAIN PIPELINES

Abstract. The relevance of the study is substantiated by the fact that when managing the processes of oil transportation through main pipelines, it becomes necessary to determine and select the optimal operating modes of the oil pipeline units, taking into account the fuzziness of some part of the initial information. In this regard, solving the problem of multi-criteria selection of effective operating modes for an oil heating station for a hot oil pipeline system, which is often described in a fuzzy environment, based on the apparatus of fuzzy set theories, is an urgent scientific and practical problem. A method for the synthesis of models in the conditions of fuzzy output parameters of the object has been developed, with the help of which fuzzy models of the investigated oil heating station of the main oil pipeline have been built. Based on the modification and combination of various optimality principles, mathematical formulations of the problem of multi-criteria selection of effective operating modes for an oil heating station in a fuzzy environment are obtained. By modifying and adapting the principles of guaranteed results and equality in a fuzzy environment, a heuristic method has been developed for solving the formulated problem of selecting object's operation modes using the initial fuzzy information. The proposed heuristic method for multi-criteria selection in a fuzzy environment is based on the use of the experience and knowledge of the decision-maker. The proposed approach is implemented in the formulation and solution of the problem of multi-criteria selection of operating modes of the oil heating station in Atyrau of the Uzen-Atyrau-Samara main oil pipeline. As a result of the application of the proposed method, an improvement in the degree of fulfillment of a fuzzy restriction on environmental impact was achieved by 2%, as well as the optimal values of the operating parameters of the object were improved: the temperature was reduced by 1.85% (5.67 K), pressure – by 0.04% (kPa) and fuel consumption – by 2.9% (0.0002 kg/s). The

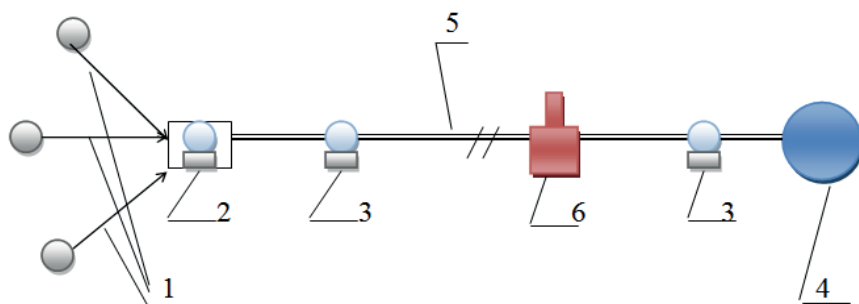
obtained results have confirmed the effectiveness of the proposed approach to solving the assigned tasks

Key words: main oil pipeline, oil transportation, control, oil transfer stations, preheating furnaces, diagnostics, automated control system.

Кіріспе. Магистралды мұнай құбырлары айтарлықтай үлкен қашықтықта таратылған және күрделі жоғары механикаландырылған және автоматтандырылған гидравликалық жүйе болып табылады. Мұндай жүйелер қуатты сорғы станцияларымен (мұнай айдау станциялары), қыздыру станцияларымен (тұтқыр, жоғары парафинді мұнайларды айдау үшін, яғни «ыстық» мұнай құбырлары жағдайында), құбырдың сызықтық бөлігімен, технологиялық байланыстар, телемеханика мен автоматтандыру құрылғыларымен, өртке қарсы құрылғылармен жабдықталған (Вайншток, 2002:407; Zhenpei, 2016:167)

Магистралды мұнай құбырлары жүйелерінің негізгі элементтері, яғни, мұнай құбырларының негізгі агрегаттары төменде 1-суретте келтірілген. Магистралды мұнай құбырларымен мұнай тасымалдау тәжірибесінде мұнай құбырлары жүйесін диагностикалау және ондағы мұнай айдау үдерістері мен мұнай құбырлары негізгі агрегаттарын олардың қоршаған ортаға әсерін минимизациялай отырып тиімді басқару қазіргі таңда өзекті мәселе болып табылады (Богданов, 2011; Zharbasbayev, 2004: 485-492)

Бұл жұмыстың мақсаты магистралды мұнай құбырларымен мұнай тасымалдау үдерістері мен агрегаттарын диагностикалау және басқару жүйелерін зерттеу, мұнай құбырлары негізгі агрегаттарын айқын емес ортада басқару есебін тұжырымдау және шешу тәсілін ұсыну болып табылады.



1 – мұнай көзі мен мұнай құбырының басты жабдықтарын байланыстыратын жеткізуші мұнай құбырлары; 2 – басты айдау станциялары; 3 – аралық айдау станциялары; 4 – мұнай құбырынан өнімдерді қабылдайтын ақырғы пункт; 5 – мұнай құбырының сызықтық құрылысы (құбыр желісі); 6 – «ыстық» мұнай құбырлары үшін мұнай қыздыру станциялары.

Сурет 1. Мұнай құбырлары жүйесінің негізгі элементтері

Зерттеу есептері. Жұмыс мақсатына жету үшін келесі зерттеу есептері қойылып, оларды шешу жолдары қарастырылады: өндірістік шығындарды қысқартуға, технологиялық агрегаттардың пайдалық әсер коэффициенттерін арттыруға, энергия тұтыну мен мұнай шығынын төмендетуге, қол еңбегін шектеуге жағдай жасайтын мұнай айдау технологиялық үрдістерін автоматтандыруға бағытталған магистралды мұнай құбырларын диагностикалау және мұнай айдау үрдістері мен технологиялық агрегатырын тиімді басқару мәселелерін зерттеу; магистралды мұнай құбырларының негізгі агрегаттарының жұмыс режимдерін оптимизациялау мен оларды басқару есептерін тұжырымдау және математикалық қойылымын алу; қойылымы алынған есепті бастапқы ақпараттың айқын еместігі жағдайында шешу тәсілін құрып, сипаттау.

Сонымен жалпы жағдайда зерттеу есебін айқын емес ортада келесідей тұжырымдауға болады:

1. $\mu_0(x) = (\mu_0^1(x), \dots, \mu_0^m(x))$ - мұнай құбырлары жүйесі жұмысының тиімділігін бағалайтын нормалданған критерийлер векторы болсын; $\varphi_q(x) \succ b_q$, $q = \overline{1, L}$ - айқын емес шектеулер делік, ал олардың орындалуын сипаттайтын тиістілік функцияларын $\mu_q(x)$, $q = \overline{1, L}$ деп белгілейік (Dubois, 2011:184, Orazbayev и т.д. 2017:945–952.) Сондай-ақ критерийлер мен шектеулердің өз-ара маңыздылықтарын бейнелейтін салмақ векторлары: $\gamma = (\gamma_1, \dots, \gamma_m)$ және $\beta = (\beta_1, \dots, \beta_L)$ белгілі, не оларды шешім қабылдаушы тұлға (ШҚТ), эксперт-мамандар көмегімен анықтауға болады деп қабылдаймыз.

Сонда магистралды мұнай құбырларының тиімді жұмыс режимдерін айқын емес ортада анықтау және мұнай айдау үрдістерін тиімді басқару есептерін жалпы түрде келесідей жазуға болады (Оразбаев, 2016:398)

$$\max_{x \in X} \mu_0^i(x), \quad i = \overline{1, m} \quad (1)$$

$$X = \{x : \arg \max_{x \in \Omega} \mu_q(x), \quad q = \overline{1, L}\} \quad (2)$$

Практикада, өндірістік жағдайларда мұндай есептер әдетте айқынсыздықпен, көпкритерийлікпен және ол критерийлер белгілі бір деңгейде қарама-қайшылықпен сипатталады. Сондықтан, тұжырымдалған есептің нақты қойылымын алу үшін түрлі компромисстік схемаларды айқын емес ортаға жұмыс жасау үшін түрлендіру, яғни бейімдеу қажет (Orazbaev, 2016:398)

Зерттеу есептерін шешу тәсілдері. Зерттеу есептерін шешу үшін, мысалы магистралдық мұнай құбырлары жұмысын диагностикалау

және мұнай айдау үрдістері мен нысандарын басқару үшін, технологиялық агрегаттар жұмыстары мен жағдайларын диагностикалау тәсілдері (Sabzi, 2017:145–163; Оразбаев, 2017:480) және технологиялық үрдістерістер мен нысандарды басқару тәсілдері (Крестин, 2016:376; Емельянов, 2017:120) қолданылады.

Магистралды құбырлар арқылы мұнай тасымалдау технологиялық агрегаттары мен үрдістерін критерийлер векторы бойынша және бас-тапқы ақпараттардың, не олардың бір бөлігі айқын емес болған жағ-дайларында оптимизациялау және тиімді басқару есептерін тұжы-рымдап, математикалық қойылымдарын алу үшін жүйелік талдау тәсіл-дері, эксперттік тәсілдер (Оразбаев, 2017:945–952), айқын емес жиындар теориясына (АЕЖТ) негізделген тәсілдер (Zharbasbayev, 2004:485-492; Dubois, 2011:184; Рей, 1983:368) қолданылған. Бұл жұмыста эксперттік бағалауларды жүргізу үшін компьютерлік желіге негізделген Дельфи тәсілі (Dubois, 2011;184) пайдаланған.

Алынған нәтижелер – магистралды құбырлармен мұнай тасы-малдау нысандарын диагностикалау, оларды тиімді басқару тәсілі.

Мұнайды тасымалдау нысандарын диагностикалау және бас-қару жүйелері. Өндіріс шығындарының қысқаруы, агрегаттардың пайдалы әсер коэффициенттерін артуы, энергия тұйыну мен мұнай шығындарының төмендеуі, қол еңбегінің қысқаруы магистралды мұнай құбырларының негізгі технологиялық нысандарын (сорғы стан-циялары, құбырдың сызықтық бөлігі, қыздыру пештері т.б.), мұнай айдау үрдістерін автоматтандыру, оларға автоматтандыру құралдары мен жүйелерін ендіру арқылы жүзеге асады. Бұл кезде автоматтан-дырылған басқару жүйелерінің жұмыс жасауының маңызды шарты басқару нысанының жағдайы мен барлық өндірістік нысандардың бай-ланыстарын дұрыс бейнелейтін ақпаратты алу болып табылады.

Автоматтандырылған басқару жүйелерінің жұмыс сапасы өлшеу құ-ралдарының метрологиялық сапаларына, реттеу және орындау құрыл-ғыларының статистикалық және динамикалық қасиеттеріне айтар-лықтай байланысты болады (Sabzi, 2017:145–163).

Жалпы жағдайда нысан жағдайы шығыс шамасымен – Y , ал басқару құрылғысының кірісіне, Y шығыс параметрінің қажетті мәні жайында ақпараты анықталған, X – кіріс әсері беріледі. Нысанның кірісіне бас-қару құрылғысынан U – басқару әсері басқару құрылғысына енді-рілген алгоритмге сәйкес құрылады және үйлесімсіздікті, оны туын-датқан шаманы жоюға бағытталады. Магистралды құбырлардың Авто-маттандырылған басқару жүйелерінде (АБЖ) басқару әсері ретінде

көп жағдайларда сұйықтық қысымы немесе шығыстар (өнімділік) шамаларының өзгеруі қолданылады.

АБЖ түсетін ақпаратты алудың негізгі құралы ретінде, яғни, басқару құрылғысына берілетін X, Y және U қшамаларын өлшеу үшін сезімтал құрылғылар – өлшеу құрылғылары мен ақпараттық-өлшеу жүйелері қолданылады. Мысалы, өлшеу құрылғысы ретінде шығын мен қысым датчиктері пайдаланылады. Қазіргі таңда жартылай өткізгіштік қысым датчиктері электроника саласында ең динамикалық дамыған бағыт болып саналады, ал қысым датчиктері нарығында көшбасшылық Honeywell фирмасының еншісінде (Оразбаев 2017:480). Кез-келген Honeywell қысым датчигінің негізі тензосезімтал элемент – кремний мембранасы бетінде өңделген жырашақта имплантанған көпірлік схема бойынша қосылған төрт бірдей пьезорезистор.

Заманауи АБЖ бірінші кезекте диагностикалау және болжау жүйелерін қажет етеді. Қазіргі кезде технологиялық нысан үрдістің бұзылуын дер кезінде анықтауға мүмкіндік беретін көптеген түрлі құрылғылармен жабдықталғаны белгілі. (Дорф, 2014:832) жұмыста магистралды құбырдан өнімнің ағуын анықтау жүйесі келтірілген, ол жүйе келесі тәсілдемеге негізделген: бақыланатын участкелер шеттерінде, ағыс пайда болған сәтте пайда болатын ажырату толқындары тіркеледі, ол тәсілге айтарлықтай жоғары сезімталдық ($1,5 - 3\% Q_{ном}$), тіркеу уақытын (5 – 10 мин) және ағыс орнын анықтаудың төменгі ауытқуын (100 км учесткеде 500 – 1000 м.) береді. Тәсіл қысымды тұрақты мониторингтеуді талап етеді. Бұл тәсілді қолданғанда, ағынның тұтастығы үзілген жағдайында құбырдың өз бетімен ағатын бөлігінде сигнал іс жүзінде толықтай шеттеленеді және оны тіркеу мүмкіндігі болмайды. Ал құбырда бөтен заттардың болуы, шұғыл бұрылыстары, соққы толқындардың шебін бұрмалайтын толқындардың шағылысын тудырады, бұл да тіркеу мүмкіндігін төмендетеді. Бұл жұмыста мониторингті деректерді алмастыруды бақыланатын нысан ерекшеліктеріне өздік баптау мүмкіндігі бар интеллектуалды құрылғылар көмегімен екі деңгейлі схема негізінде қамтамасыз ететін тәсіл ұсынылған. Бұл тәсілді қолдану мұнай айдау технологиялық үрдістерін басқару жүйелерінің жалған іске қосылуын төмендетуге, қолданатын техникалық құралдардың аздығы есебінен, ақпаратты дұрыс беру мен жылдам өңдеу мүмкіндігін туғызады.

Сонымен құбырлардан мұнайдың ағуын анықтау жүйелерінің барлық белгілі мониторингтеу (динамикалық) тәсілдері айдаудың бақыланатын параметрлері мен шынайы уақыт масштабында айдау про-

цесстерін модельдеу нәтижесінде алынған, есептеу көрсеткіштерімен салыстыруға негізделген. Мұнайдың ағуы жайлы қортындылардың дұрыстығы айтарлықтай деңгейде қолданылатын мұнай құбыры модельдеріне байланысты болады.

Магистралды құбырлардан мұнай ағуын дистанциондық анықтаудың басқа әдісі де белгілі (Джексон, 2007:384), ол мұнай құбыры трассасының жылу өрісінің аэротүсіріліміне, оның шектік ашықтық (яркость) мәнін анықтауға, аномалды температурасы бар жеке участкелердің орналасуын анықтауға, сондай-ақ жеке участкелердің жылу өрісінің жарықтылығы мәнін тіркеуге негізделген. Қосымша мұнай құбыры трассасын мұнайдың газдық фракциясы мен бұл компоненттермен сіңіруден бос, алайда оларды айтпарлықтай жақын аймақта орналасқан толқынның тіреу ұзындығында негізгі компоненттерімен сауле шығаруын сіңіретін үштен артық зерттелетін толқын ұзындығының төсеніш бетін лазерлік алдын ала тексеру (зондтау) жүргізіледі. Әр зерттелетін толқын ұзындығына және ыдырау элементтерінің әрқайсысына төсеніш бетінің салыстырмалы ашықтығының кескінін, толқынның әр ұзындығының төсеніш бетінің ашықтығының толқынның тіреу ұзындығында төсеніш беттің ашықтығы қатынасы ретінде анықтайды. Содан кейін салыстырмалы ашықтық логарифмдерін және олардың орташа мәндерін анықтайды. Құбырдан мұнай ағуының орны бірінші толқынға салыстырмалы ашықтық кескіні барлық бақылау учесткесі үшін орташа мәні берілген шектік мәннен өзгеше болатын, ал салыстырмалы ашықтық кескіндері логарифмдері үш толқын үшін $(1 \pm 0,2):(1,4 \pm 0,2):(1,2 \pm 0,2)$ пропорцияларын құрайтын аномалды температурасы бар учесткенің орналасуы бойынша анықтайды.

Құбырдағы сұйықтық қысымы мәнін басқарудың келесі тәсілін қолдануға болады: құбырдың бір шетінен кеңістік электромагниттік толқынды тудырады, ал соңғы пунктте оның қарқындығы тіркеледі, кенеттен өзгеруі бойынша тасымалданатын сұйықтықтың тұтастығының бұзылуы, онда газ-ауалық жиналудың, не бөтен қоспалардың пайда болуы жайында қорытынды жасалады. Алынған уақыттық көрсеткіштер бойынша құбырда жиналған бөтен қоспалардың орны мен көлемі анықталады.

«IMS» компаниялар тобы қысым толқынын деңгейлестіру жүйесін, сондай-ақ өткізгіштік қабілетін, мұнай құбырлары жүйелерінің сенімділігі мен ресурсын айтарлықтай арттыруға мүмкіндік беретін, гидросокқыдан қорғау жүйесін құрып, ендірілген. Жүйелердің жұмыс жасау принципі, шығын басқару жүйесі арқылы бақыланатын және

түзетілетін, жұмыстық сұйықтықты реттеу (түсіру) клапандары арқылы уақытымен түсіруге негізделген. Аталған жүйелер «Балтнефтепровод» ААҚ, «Транссібір Магистралды мұнай құбырлары» ААҚ, «Орал-сібірмұнайқұбыры» және басқа кәсіпорындарды табысты қолданылуда.

Қазіргі таңда, магистралды мұнай құбырын пайдалану участкесін тиімді басқаруды қамтамасыз ететін Диспетчерлік басқару және деректерді жинау (SCADA) жүйелері кеңінен таралған. Нақты уақыт ауқымында SCADA-жүйелер барлық мұнай айдау технологиялық параметрлерінің өзгеруін қадағалауға мүмкіндік береді. Заманауи SCADA технологияларын, телемеханика және автоматтандырылған басқару жүйелерін қолдану мұнай мен мұнай өнімдері құбырлары технологиялық режимдерінің ауытқуларын, функционалдық диагностикалау есептерін шешуге мүмкіндік береді. Әлемдік нарықта ең әйгілі SCADA-жүйелер: Factory Link (жасаушы US DATA Co, АҚШ), Genesis (Iconics, АҚШ), Sitex (Jade Software, Англия), TraceMode (AdAstra Research Group Ltd., Ресей), SIMATIC WinCC (Siemens AG, Германия) т.б.

Мұнай құбырларын пайдалану кезінде жүретін үрдістерді басқару мен құбырлар жағдайын дұрыс бақылауды жүзеге асыратын автоматтандырылған басқару жүйесін құру сұрақтарын қарастырайық.

Құбыр жағдайын бағалауға қажетті бастапқы деректерді: сыртқы және ішкі диагностика; компрессорлық станция, қыздыру станциясы, сызықтық өндірістік басқару немесе тұтастай тасымалдау кәсіпорны деңгейінде біріктірілген АБЖ көмегімен алынған, тасымалданатын мұнайдың параметрлері мен қасиеттері жайында ақпаратты жинау; металграфикалық зерттеулер; картографикалық материалдарды және жобалау-құрылыстық құжаттарды талдау нәтижелерінде анықтауға болады. Мұнай құбырларын сыртқы диагностикалау, атап айтқанда, топырақтың табиғи жылжуы мен құбырлардың жылу деформациялары нәтижесінде құбырдың жобалық орналасуынан ауытқуын бағалауға мүмкіндік береді. Басқару жүйесін жобалау үшін SCADA-жүйелері немесе жүйе икемділігін айтарлықтай арттыруға мүмкіндік беретін, сондай-ақ тұрақты бақылау және басқару үрдістерін жеңілдететін, жобалау жүйесіне жоғары деңгейлік меншікті интерпретациялау тілі қолданылады. Шешім қабылдау жылдамдығын жүйеге меншікті WEB және WAP интерфейстерін ендіру жолымен қамтамасыз етуге болады.

Мұнай құбырларындағы апаттық жағдайды ерте болжау үшін бұзбайтын ішкі құбырлық бақылау тәсілдерін (Бусыгин, 1999:30-31) қолдануды ұсынуға болады. Жүйенің көпфункционалылығы келесідей қамтамасыз етіледі: қолданыстағы не жобадағы мұнай құбырының

немесе оның участкесінің тұрақты қолданатын моделі құрылады. Содан кейін пайдаланушы жүйеге келіп түсетін параметрлер жиынтығын, сондай-ақ жүйенің есептеу параметрлері жиынтығын анықтайды. Кешеннің деректер және білімдер базасында жинақталған деректер, мұнай құбырының немесе оның участкесінің үш өлшемді моделінде бейнеленеді, ал ол магистралды мұнай құбырлары жағдайы жайында деректерді түсінуді айтарлықтай жеңілдетеді.

Өнеркәсіптік және магистралды құбырларда АБЖ интеллектуалды басқару жүйелерін құру принциптеріне сәйкес жасақтау қазіргі уақытта өте өзекті бағыт болып табылады (Zhapbasbayev, 2004:485-492; Orazbayev, 2017:945–952). Айқын емес басқару және болжау тәсілі басқару мақсаттарын жақсы теңдестіру бағалауын жүзеге асыруға мүмкіндік береді. Мұнай айдау станцияларында орналасқан контроллерлер, әлсін-әлсін мұнай айдау процесіндегі шынайы параметрлердің модельдік мәндерінен ауытқуларын бағалайды және айдаудың берілген технологиялық режимін қамтамасыз ету үшін басқару командасын құрастырады. Көппараметрлік адаптивті модельді қолдану апаттық жағдайларда шешім қабылдау жылдамдығын арттыруға, мұнай құбырларындағы үрдістердің даму үрдісін қадағалауға мүмкіндік береді. Мұндай интеллектуалды жүйелер мен модельдер SCADA-жүйелер жинаған мұнай айдау технологиялық параметрлерін пайдаланушыға түсінікті түрде интерпретациялау және диагностика нәтижелерін талдауға ыңғайлы формада көрсете алады. Айқын артықшылықтары мен үлкен болашағына қарамастан жаңа SCADA-жүйелерді құру үлкен материалдық, қаржылық, уақыт шығындарын талап етеді, ал бұл жағдай қолда бар инструментарийді пайдалануға мәжбүр етеді.

Мұнай құбырларында жинақталған газдарды диагностикалау үшін түрлі тәуелсіз бағалау алгоритмдері ұсынылған: қысым импульсы өту уақыты бойынша; газ жағдайы заңына сәйкес жинақталған газдар көлемінің өзгеруі бойынша; өтімділікті байланысты жинақталған газдар ұзындығының өзгеруі бойынша (Бусыгин, 1999:30-31). Алайда бұл алгоритмдерді қолдану қосымша материалдық және уақыттық шығындарды талап етеді, себебі, ол үшін арнаулы жабдықтар мен алынған деректерді талдау мен өңдеу үшін үлкен қосымша уақыт шығындар керек.

Мұнай құбыры маңында салынған кабельдің кіріс кедергісінің өзгеруі бойынша құбырда ауаның жинақталған ауа орындарын анықтау тәсілі белгілі (Алеев, 2001:2). Кабель кедергісі бақыланатын сұйықтықтың

құбырдың ашылу және газдық-ауалық тығынның пайда болу орнында сіңуіне байланысты өзгереді. Алайда тәсілді жүзеге асыру үшін құбыр бойында төселген тұрақты қолданыстағы кабель қажет, ал ол үлкен материалдық шығындарға алып келеді. Сонымен қатар ол сенімділіктің қажетті деңгейін қамтамасыз ете алмайды, себебі ағып кету болмағанда жинақталу көлемі мен орнын анықтауға мүмкіндік бермейді.

Мұнай құбырын іске қосқанда, сондай-ақ оны пайдалану кезінде жинақталған газдарды жою айтарлықтай қиындықтар туғызады. Бұл кезде келесі шараларды қолдану ұсынылады:

1) айдау өнімділігін «шығару» жылдамдығынан артатын мәнге дейін арттыру;

2) жинақталған қоспалар толықтай ерігінше айдау қысымын ұзақ уақытқа көтеру;

3) сателлиттік желідегі газды толықтай шығару үшін автоматтық вантуздарды қолдану;

4) тазалау құрылғылары мен «жұмсақ» айырғаштарды ендіру.

Соңғы әдіс ең тиімді әдіс ретінде қарастырылады, алайда жылдам диагностикалау құрылдары және жинақталған қоспалардың динамикасының негізделген моделі жоқ болса, айтарлықтай және әрқашан да орынды емес қосымша шығындарға алып келеді.

Магистралды мұнай құбырын толтырғанда трассаның дөңестеу жеріндегі ауаны шығару үшін втоматтық вантуздар орналастырылады, сондай-ақ олар мұнай құбырын пайданау кезінде құбырда газ мен ауаның жинақталуың болдырмау үшін де қолданылады.

Ғылыми тұрғыдан ең негізделген әдістерге 1- және 2- әдістер жатады. Алайда, құбырдың жарылуына алып келу мүмкіндігі болғандықтан және құбырлардың 60% тозуын ескере отырып, оларды практикада қолдану айтарлықтай тәуекелдік тудырады.

Сонымен, мұнай құбырларымен тасымалдаудың тиімділігіне әсер ететін қиыншылықтарды анықтаудың белгілі динамикалық тәсілдері айдаудың бақыланатын параметрлері мен олардың, нақты уақыт ауқымында үрдісті модельдеу арқылы алынған, есептік мәндерін салыстыруға негізделген. Қиыншылықтың бар болуы жайында қорытындылардың дұрыстығы қолданатын құбыр модельдеріне байланысты болады. Диагностикалау есептерін шешу үшін қазіргі уақытта, құбырлардың қалыпты және қиыншылық жағдайларда жұмысы жайында деректер негізінде құрылатын түрлі модельдер қолданылады. Мысалы, егер құбырдың қалыпты жұмысы кезінде оның қысымының төмендеуі белгілі болса, онда өлшенген қысымның түсіп кетуінің ауытқуы

бойынша қиыншылық бар болуы жайында жорамалдауға болады. Сондықтан қиындықтарды дер кезінде диагностикалау үшін қалыпты жұмыс жағдай үшін жақсы есептеу моделі қажет болады.

Енді, қойылған зерттеу есебін (1)(2) нақтылап, басты критерий тәсілі негізінде мұнай құбырлары агрегаттарының жұмыс режимдерін басқару есебін айқын есем ортада тұжырымдайық және оны шешу тәсілін ұсынайық. Яғни критерийлер векторынан бір басты критерий таңдап алу мүмкін деп есептейміз, ал қалған жеке критерийлерді шектеулер құрамына ендіреміз және басты критерий тәсілуі идеясын айқын еместікке бейімдейміз. Сонда магистралды мұнай құбырын басқару кезінде оның тиімді жұмыс режимін анықтау есебін m критерийлі және L шектеулі жағдайда келесідей нақтылап, жаза аламыз:

$$\max_{x \in X} \mu_0^1(x), \quad (3)$$

$$X = \{x : x \in \Omega \wedge \arg(\mu_q(x) \geq \mu_q^R) \wedge \arg(\mu_0^i(x) \geq \mu_{0R}^i), q = \overline{1, L}, i = \overline{2, m}\}. \quad (4)$$

Келтірілген есептің шешімі шешім қабылдаушы тұлға (ШҚТ, эксперт) ендіретін критерийлер мен шектеулердің шектік мәндеріне, яғни $\mu_1^R, \dots, \mu_L^R; \mu_{0R}^2, \dots, \mu_{0R}^m$ байланысты болады.

Қойылған (3)-(4) есебін шешу үшін ұсынылған тәсілдің негізгі кезеңдерін келтірейік.

Тәсіл БК+БК.

1. Жеке критерийлер және шектеулер үшін басымқылар қатарын, яғни $I_C = \{1, \dots, m\}$ және $I_R = \{1, \dots, L\}$ ендіру (басты критерий басымқы 1 ие болу керек).

2. ШҚТ, эксперттер көмегімен шектеуге ендірілген жеке критерийлер мен шектеулердің шектік мәндерін анықтау: $\mu_{0R}^i, i = \overline{2, m}; \mu_q^R, q = \overline{1, L}$.

3. Эксперттік бағалау негізінде айқын емес параметрлердің термжиыны анықтау және айқын емес шектеулердің орындалу тиістілік функцияларын тұрғызу: $\mu_q(x) \geq \mu_q^R, q = \overline{1, L}$ и $\mu_0^i(x) \geq \mu_{0R}^i, i = \overline{2, m}$.

4. Қойылған шектеулерді ескеретін X жиынында (4) $\mu_0^1(x)$ - басты критерийді (3) максимизациялау және ағымдағы шешімдерді шығару: $x(\mu_{0R}^i, \mu_q^R); \mu_0^1(x(\mu_{0R}^i, \mu_q^R)), \dots, \mu_0^m(x(\mu_{0R}^i, \mu_q^R)); \mu_1(x^*(\mu_{0R}^i, \mu_q^R)), \dots, \mu_L(x^*(\mu_{0R}^i, \mu_q^R)), i = \overline{2, m}, q = \overline{1, L}$.

5. Алынған ағымдағы шешімдерді ШҚТ ұсыну. Егер ағымдағы шешімдер ШҚТ-ны қанағаттандырмаса, онда ШҚТ μ_{0R}^i, μ_q^R мәндерін өзгертеді және 3 кезеңге шешімді іздеуді жалғастыру үшін қайтып келеді. Ағымдағы шешімдер ШҚТ-ны қанағаттандырса, тиімді шешімдерді шығару үшін 6-кезеңге өту.

6. Шешімді іздеуді тоқтату және ШҚТ қабылдаған соңғы шешімдерді шығару: кіріс, режимдік параметрлердің оптималды $x^*(\mu_{0R}^i, \mu_q^R)$; мәндері, олар критерийлердің тиімді мәндерін $\mu_0^1(x^*(\mu_{0R}^i, \mu_q^R)), \dots, \mu_0^m(x^*(\mu_{0R}^i, \mu_q^R))$ және шектеулердің максималды орындалу тиістілік функцияларын $\mu_1(x^*(\mu_{0R}^i, \mu_q^R)), \dots, \mu_L(x^*(\mu_{0R}^i, \mu_q^R))$ қамтамасыз етеді.

ШҚТ тағайындаған критерийлер мен шектеулердің шектік мәндерінің: $\mu_{0R}^i, \mu_q^R, i = \overline{2, m}, q = \overline{1, L}$ негізделуін арттыру үшін, түрлі шектеулік мәндерді тағайындау және алынған нәтижелерді ШҚТ-мен талдау және μ_{0R}^i, μ_q^R жаңа мәндерін таңдау диалогты процедураларын құруға болады.

Қорытынды. Магистралды құбырлармен мұнай тасымалдау сұрақтары талданып, оның технологиялық агрегаттары жағдайын, түрлі туындайтын қиындықтарды диагностикалау және мұнай айдау үрдістерін басқару мәселелері зерттеліп, оларды шешу жолдары қарастырылған. Мұнай құбырлары жүйесінің құрылымы мен негізгі элементтері сипатталған. Мұнай құбырлары мен онда жүретін мұнай айдау үрдістерін басқаруда қолданатын заманауи автоматтандырылған басқару жүйелері, бірінші кезекте диагностикалау және болжау жүйелерін, қажет етеді. Сондықтан жұмыста магистралды мұнай құбырлары негізгі агрегаттарының жағдайы мен жинақталған газ, ауа қоспаларын, құбырдан ағу жағдайларын анықтау, диагностикалау тәсілдері мен құралдары сипатталып талданған. Магистралды құбырлардан мұнайдың ағуын анықтау жүйелерінің барлық белгілі мониторингтеу (динамикалық) тәсілдері айдаудың бақыланатын параметрлері мен шынайы уақыт масштабында айдау процесстерін модельдеу нәтижесінде алынатын есептеу көрсеткіштерімен салыстыруға негізделген.

Айқын емес басқару және болжау тәсілін қолдану мұнай айдау үрдістерін интеллектуалдық басқару жүйелерін құруға негіз болатыны тұжырымдалған. Магистралды мұнай құбыры пайдалану участкесін тиімді басқаруды қамтамасыз ететін Диспетчерлік басқару және деректерді жинау (SCADA) жүйелері, SCADA технологияларын және автоматтандырылған басқару жүйелерін мұнай құбырлары технологиялық режимдерінің ауытқуларын, функционалдық диагностикалау есептерін шешуге қолдану сұрақтары талқыланды. Мұнай құбырын іске қосқанда, сондай-ақ оны пайдалану кезінде жинақталған газдарды жою үшін қолдануға болатын негізгі шаралар ұсынылған.

Магистралды мұнай құбырларының тиімді жұмыс режимдерін айқын емес ақпараттық ортада анықтау арқылы мұнай айдау үрдістерін тиімді

басқару есептерінің жалпы түрі және айқынсыз ортаға бейімделген басты критерий тәсілі негізінде нақтыланған жазбасы алынып, алынған есепті шешу үшін ШҚТ-ны қатыстыруға негізделген эвристикалық тәсіл жасақталып, ұсынылған. Ұсынылған тәсілдің негізгі кезеңдерінің сипаттамасы берілген. Жұмыста жасақталып, қойылған магистралды мұнай құбырларын басқаруда құбыр технологиялық агрегаттарының тиімді жұмыс режимдерін анықтау (3)-(4) есебін шешу үшін ұсынылған БК+БК тәсілінде басты критерий тәсілі критерийлер үшін және түрленген күйде айқын емес шектеулер үшін екі рет қолданылады.

Information about the authors:

Moldasheva Zh.Zh. – graduate student at the Department of Information Systems, L.N. Gumilyov Eurasian National University, Satpayev str., Nur-Sultan, Kazakhstan. Tel.: +7 (705)802-1151. E-mail: zhadira1985@mail.ru, ORCID: <https://orcid.org/0000-0002-0559-3410>;

Orazbayev B.B. – doctor of engineering sciences, Professor at the Department of Systems of analysis and management, L.N.Gumilyov Eurasian National University, Satpayev str., Nur-Sultan, Kazakhstan. Tel.: +7 (701) 781-9639. E-mail: batyr_o@mail.ru, ORCID: <http://orcid.org/0000-0003-2387-9946>;

Goncharov B.I. – Consultant Professor, Department of Automation and Robotics, National Research Tomsk Polytechnic University. ORCID: 0000-0002-1249-1981;

Orazbayeva K.N. – doctor of engineering sciences Professor at the Department of Management and marketing, Kazakh University of Economics, Finance and International Trade, Zhubanov str., Nur-Sultan, Kazakhstan. Tel.: +7 (701) 174-2473. E-mail: kulman_o@mail.ru, ORCID: <https://orcid.org/0000-0002-1741-7553>.

ӘДЕБИЕТТЕР

Вайншток С.М. Трубопроводный транспорт нефти: учебник / С.М. Вайншток, -М.: ООО «Недра-Бизнесцентр», 2002. Т.1. -407 с.

Zhenpei Li, Jinfei Wang, R. Brook, R. Easton. *Promoting data requirement for the oil & gas pipeline integrity management* // Oil Gas European Magazine. 2016, Vol.132. -P. 167-193.

Богданов Р.М., Лукин С.В. Определение ряда оптимальных режимов работы магистральных трубопроводов при выбранных критериях оптимальности (ОРОР-МТ)»/ Патент No 011611173, зарегистр. в Реестре программ для ЭВМ по заявке No2010617845 от 4.02.2011 г.

Czuprat O. Simulation Method Aids Pigging Operation in Subsea Waxy Crude Oil Pipelines // Oil Gas European Magazine. Vol.132, 2016. –P.200-223.

Zhapbasbayev U.K. and et. al. Investigation of Effect of Depressant Additives on

Hydraulic Resistance and Heat Exchange in Oil Mixture Flow // *Mathematics and Computers in Simulation*. Amsterdam: Elsevier, 2004. №67. -P. 485-492.

Dubois D. The role of fuzzy sets indecision sciences: Old techniques and new directions // *Fuzzy Sets and Systems*. 2011. V. 184. P. 3.

B. Orazbayev, E. Ospanov, N. Kissikova, N.Mukataev, K.Orazbayeva. Decision-making in the fuzzy environment on the basis of various compromise schemes // *Procedia Computer Science* 120, 2017, -P.945–952.

Оразбаев Б.Б. Методы моделирования и принятия решений для управления производством в нечеткой среде: монография / Б.Б. Оразбаев. ЕНУ, -Астана, 2016, - 398 с.

Orazbaev B.B., Orazbayev K.N., Orazbayeva L.T., Kurmangaziyeva, V.E. Makhatova Multi-criteria optimisation problems for chemical engineering systems and algorithms for their solution based on fuzzy mathematical methods. *EXCLI Journal* 2015; vol.14: P.984-998.

Ушаков В.М. Неразрушающий контроль и диагностика нефтегазового оборудования, технологических агрегатов магистральных нефтепроводов: монография / В.М. Ушаков. -Вологда: Инфра-Инженерия, 2014. -318 с.

Крестин Е.А. Диагностика машин и оборудования: Учебное пособие / Е.А. Крестин, И.Е. Крестин. -СПб.: Лань, 2016. -376 с.

Емельянов С.Г. Автоматизированные нечетко-логические системы управления: монография / С.Г. Емельянов, В.С. Титов, М.В. Бобырь. -М.: Инфра-М, 2017. -120 с.

Рей У. Методы управления технологическими процессами / Пер с англ. -М.: Мир, 1983. -368 с.

Sabzi H.Z. Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: A case study // *Expert systems with applications*. 2017. -V. 82. № 3. -P. 145–163.

Оразбаев Б.Б. Теория и практика методов нечетких множеств: учебник / Б.Б. Оразбаев. -Алматы: изд. МОН РК Бастау, -480 с.

Дорф Р. Современные системы управления: монография /Р.Дорф. –М.: Лаборатория базовых знаний, 2014. –832 с.

Джексон Р.Г. Новейшие датчики: учебник /Р.Г. Джексон. –М.: Техносфера, 2007. -384 с.

Бусыгин Г.Н. Эксплуатация систем контроля утечек в ОАО «Уралтранснефтепродукт» // *Транспорт и хранение нефтепродуктов: НТС*. –М.: ЦНИИТЭнефтехим, 1999. - № 9. -С. 30-31.

Патент 2073816 РФ. Способ дистанционного обнаружения утечек нефти из магистрального трубопровода /Алеев Р.М., Алешко Е.И., Чепурский В.Н.; опубл. 2001. – 2 с.

Николаев Д.А. Системы 3D моделирования в рамках мониторинга и оценки состояния линейной части магистральных трубопроводов.//*Труды международной научно-практ. конф.* –Самара: 2002. - С.70–77.

Маслов Л.С. Удаление воды и воздуха из трубопроводов в пусковой период // *Строительство трубопроводов*. 2010. № 9. -С.12–15.

Патент 55-6856 Японии. Способ обнаружения утечки жидкости из трубопровода.; опубл. 2015. – 2 с.

Б.Б. Оразбаев, Ж.Ж. Молдашева, К.Н. Оразбаева, В.Е. Махатова, Л.Т. Курмангазиева, А.Е. Габдулова Разработка математических моделей и оптимизация режимов работы станции подогрева нефти магистральных нефтепроводов в условиях нечеткости исходной информации. *Eastern-European Journal of Enterprise Technologies*, 6(2(114)), 147-162. Doi:<https://doi.org/10.15587/1729-4061.2021.244949>

REFERENCES

- Vainshtok S.M. Pipeline oil transport: textbook / S.M. Vainshtok, -M.: LLC “Nedra-Businesscenter”, 2002. Vol.1. -407 p.
- Zhenpei Li, Jinfei Wang, R. Brook, R. Easton. Promoting data requirement for the oil & gas pipeline integrity management // Oil Gas European Magazine. 2016, Vol.132. -P. 167-193.
- Bogdanov R.M., Lukin S.V. Determination of a number of optimal modes of operation of trunk pipelines under selected optimality criteria (ORORMT) // Patent No. 011611173, registered. in the Register of computer programs according to application No2010617845 dated 4.02.2011 .
- Czuprat O. Simulation Method Aids Pigging Operation in Subsea Waxy Crude Oil Pipelines // Oil Gas European Magazine. Vol.132, 2016. –P.200-223.
- Zhapbasbayev U.K. and et. al. Investigation of the Effect of Depressant Additives on Hydraulic Resistance and Heat Exchange in Oil Mixture Flow // Mathematics and Computers in Simulation. Am-sterdam: Elsevier, 2004. No.67. -pp. 485-492.
- Dubois D. The role of fuzzy sets indecision sciences: Old techniques and new directions // Fuzzy Sets and Systems. 2011. V. 184. P. 3.
- B. Orazbayev, E. Ospanov, N. Kissikova, N.Mukataev, K.Orazbayeva. Decision-making in the fuzzy environment on the basis of various compromise schemes // Procedural Computer Science 120, 2017, -p.945-952.
- Orazbayev B.B. Methods of modeling and decision-making for production management in a fuzzy environment: monograph / B.B. Orazbayev. ENU, -Astana, 2016, - 398 p.
- Orazbaev B.B. Orazbayev, K.N. Orazbayeva, L.T. Kurmangaziyeva, V.E. Makhatova Multi-criteria optimisation problems for chemical engineering systems and algorithms for their solution based on fuzzy mathematical methods. EXCLI Journal 2015; vol.14: P.984-998.
- Ushakov V.M. Non-destructive testing and diagnostics of oil and gas equipment, technological units of trunk oil pipelines: monograph / V.M. Ushakov. -Vologda: Infra-Engineering, 2014. -318 p.
- Krestin E.A. Diagnostics of machines and equipment: A textbook / E.A. Krestin, I.E. Krestin. -St. Petersburg: Lan, 2016. -376 p.
- Emelyanov S.G. Automated fuzzy logic control systems: monograph / S.G. Emelyanov, V.S. Titov, M.V. Bobyr. -M.: Infra-M, 2017. -120 p.
- Ray U. Methods of technological process control / Per from English -M.: Mir, 1983. -368 p.
- Sabzi H.Z. Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: A case study // Expert systems with applications. 2017. -V. 82. № 3. -P. 145–163.
- Orazbayev B.B. Theory and practice of fuzzy set methods: textbook / B.B. Orazbayev. -Almaty: ed. MES RK Bastau, -480 p.
- Dorf R. Modern control systems: monograph /R.Dorf. –M.: Laboratory of Basic knowledge, 2014. -832 p.
- Jackson R.G. The latest sensors: textbook / R.G. Jackson. –M.: Technosphere, 2007. -384 p.
- Busygin G.N. Operation of leak control systems in OJSC Uraltransnefteprodukt // Transportation and storage of petroleum products: NTS. –M.: Tsniiteneftekhim, 1999. - No. 9. -pp. 30-31.

Patent 2073816 of the Russian Federation. A method for remote detection of oil leaks from the main pipeline / Aleev R.M., Aleshko E.I., Chepursky V.N.; publ. 2001. – 2 p.

Nikolaev D.A. 3D modeling systems in the framework of monitoring and assessment of the condition of the linear part of the main pipelines. //Proceedings of the International Scientific and Practical Conference -Samara: 2002. - pp.70-77.

Maslov L.S. Removal of water and air from pipelines during the start-up period // Construction of pipelines. 2010. No. 9. -p.12-15.

Patent 55-6856 of Japan. Method for detecting liquid leakage from the pipeline.; publ. 2015. -2 p.

B.B. Orazbaeva, Zh.Zh. Moldasheva, K.N. Orazbaeva, V.E. Makhatova, L.T. Kurmangazieva, A.E. Gabdulova Development of mathematical models and optimization of operating modes of the oil heating station of main oil pipelines in conditions of initial information fuzziness. Eastern-European Journal of Enterprise Technologies, 6(2(114)), 147-162. Doi:<https://doi.org/10.15587/1729-4061.2021.244949>.

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

ISSN 1991-346X

Volume 2, Number 342 (2022), 130–141

<https://doi.org/10.32014/2022.2518-1726.134>

УДК – 004

HTAMP - 50.01.75

Б.Б. Тастемір

Алматы технологиялық университеті, Алматы, Қазақстан.

E-mail: tbeka97@gmail.com

ЭЛЕКТРОНДЫҚ ПОШТА СПАМДЫ СҮЗГІЛЕУГЕ АРНАЛҒАН RANDOM FORESTS МАШИНАЛЫҚ ОҚЫТУ ӘДІСІ

Аннотация. Электрондық пошта спамы-әлемдегі әрбір электрондық пошта пайдаланушысының күнделікті кездесетін негізгі мәселелерінің бірі. Күн сайын электрондық поштаны пайдаланушылар робот-бағдарламалық агенттер автоматты түрде жасайтын анонимді мекенжайлардан жаңа мазмұндағы жүздеген спам-хаттарды алады. Қара және ақ тізімдерді (домендер, IP мекенжайлары, пошта мекенжайлары) пайдалану сияқты қалаусыз поштаны сүзудің дәстүрлі әдістері спам-хабарламалар қаупін төмендетуде өте тиімсіз болып шықты. Бұл жоғары сенімді электрондық пошта спам сүзгілерін ойлап табу қажеттілігіне әкелді. Жақында машиналарды оқыту тәсілі спам хаттарын анықтау және сүзу үшін сәтті қолданылды. Бұл мақалада спам электрондық пошталарын тиімді жіктеу үшін кездейсоқ орман машинасын оқыту алгоритмін қолдану ұсынылады. Негізгі мақсат - болжаудың дәлдігі мен мүмкіндіктері аз қалаусыз пошта сүзгісін жасау. Хам, спам және қарапайым электрондық пошталардың 5180 электрондық поштасынан тұратын Enron жалпыға қол жетімді мәліметтер жиынтығынан random forests алгоритмі арқылы алынған және қолданылған 99,92% жіктеу дәлдігі, жалған позитивтердің өте төмен жиілігі (0,01) және шынайы позитивтердің өте жоғары жиілігі 0,999. Барлық эксперименттер WEKA деректерді модельдеу және машиналық оқыту ортасында жүзеге асырылады. Қолданыстағы электрондық пошта спамдарын сүзгілеу әдістерінің көпшілігі спам жіберушілермен күнделікті жіберілетін

кейбір спамдарды тиімді өңдей алмайды. Себебі спам жіберушілер спам сүзгісі арқылы анықтаудан жалтарудың күрделірек әдістерін ойлап табуды жалғастырды. Спамерлердің жаңа техниканы үздіксіз қолдануымен электрондық поштаны спамды сүзу зерттеушілер үшін қызу зерттеу саласына айналды. Бұл зерттеуде біз электрондық пошта спамдарын тиімді және тиімді сүзгілеуге арналған Random Forests алгоритмін ұсындық. Алгоритмнің тиімділігі мен тиімділігін анықтау үшін дәлдік, TPR, FPR, дәлдік және F-өлшемі арқылы Enron спам деректер жиынындағы RFs алгоритмінің өнімділігін бағалады. Біз RF электрондық пошта пайдалану шыларының кіріс жәшігіндегі спам хабарламаларының көлемін одан әрі азайту үшін пошта серверінде немесе пошта клиенті жағында қабылдануы мүмкін перспективалы алгоритм екенін айта отырып қорытындылаймыз.

Түйін сөздер: машиналық оқыту, спамды сүзу, кездейсоқ ормандар, нейрондық желілер, тірек векторлық Машиналар, аңғал Байес.

Б.Б. Тастемир

Алматинский технологический университет, Алматы, Казахстан.

E-mail: tbeka97@gmail.com

МЕТОД МАШИННОГО ОБУЧЕНИЯ RANDOM FORESTS ДЛЯ ФИЛЬТРАЦИИ НЕЖЕЛАТЕЛЬНОЙ ПОЧТЫ

Аннотация. Спам по электронной почте является одной из основных проблем, с которыми ежедневно сталкивается каждый пользователь электронной почты в мире. Ежедневно пользователи электронной почты получают сотни спам-писем нового содержания с анонимных адресов, которые автоматически генерируются программными агентами-роботами. Традиционные методы фильтрации спама, такие как использование черных и белых списков (домены, IP-адреса, почтовые адреса), оказались совершенно неэффективными в снижении угрозы спам-сообщений. Это привело к необходимости изобретения высоконадежных фильтров электронной почты для спама. В последнее время подход машинного обучения успешно применяется для обнаружения и фильтрации спам-писем. В этой статье предлагается использовать алгоритм машинного обучения случайного леса для эффективной классификации спам-сообщений электронной почты. Основная цель — разработать спам-фильтр электронной почты с большей точностью предсказания и меньшим количеством функций. Из общедоступного

набора данных Enron, состоящего из 5180 электронных писем, как спама, так и обычных писем, был извлечен набор характерных признаков спама (из литературы), которые были применены с помощью алгоритма случайных лесов с результирующей точностью классификации 99,92%, очень низкий ложноположительный результат (0,01) и очень высокий истинно положительный уровень 0,999. Все эксперименты проводятся в среде моделирования интеллектуального анализа данных и машинного обучения WEKA. Большинство существующих методов фильтрации спама в электронной почте неэффективно справляются с частью ежедневно рассылаемого спамерами спама. Это связано с тем, что спамеры продолжают изобретать более изощренные способы избежать обнаружения с помощью спам-фильтров. С постоянным использованием спамерами новых методов фильтрация электронной почты от спама стала горячей областью исследований для исследователей. В этом исследовании мы предложили алгоритм Random Forests для эффективной и действенной фильтрации спама в электронной почте. Чтобы определить эффективность и действенность алгоритма, Enron оценила производительность алгоритма RFs в наборе данных о спама, используя точность, TPR, FPR, точность и F-размер. В заключение мы скажем, что RF является многообещающим алгоритмом, который можно использовать на почтовом сервере или на стороне почтового клиента для дальнейшего уменьшения количества спам-сообщений в папке входящих сообщений пользователей электронной почты.

Ключевые слова: машинное обучение, фильтрация спама, случайные леса, нейронные сети, опорные векторные машины, наивный Байес.

B.B. Tastemir

Almaty Technological University, Almaty, Kazakhstan.

E-mail: *tbeka97@gmail.com*

RANDOM FORESTS MACHINE LEARNING TECHNIQUE FOR EMAIL SPAM FILTERING

Abstract. Email spam is one of the major problems that every email user in the world faces on a daily basis. Every day, email users receive hundreds of spam emails with new content from anonymous addresses that are automatically generated by robotic software agents. Traditional spam filtering methods such as blacklisting and whitelisting (domains, IP addresses, email addresses) have proven to be completely ineffective in

reducing the threat of spam messages. This necessitated the invention of highly reliable spam email filters. Recently, a machine learning approach has been successfully applied to detect and filter spam emails. This article proposes to use a random forest machine learning algorithm to effectively classify spam emails. The main goal is to develop an email spam filter with more predictive accuracy and fewer features. From Enron's public dataset of 5180 emails, both spam and regular, a spam feature set (from the literature) was extracted and applied using a random forest algorithm with a resulting classification accuracy of 99.92%, very poor. a false positive (0.01) and a very high true positive of 0.999. All experiments are carried out in the WEKA data mining and machine learning simulation environment. Most existing email spam filtering methods do not effectively handle some of the daily spam sent by spammers. This is because spammers have continued to invent more sophisticated ways to avoid detection through spam filters. With the constant use of new techniques by spammers, spam filtering of e-mail has become a hot field of research for researchers. In this study, we proposed a Random Forests algorithm for efficient and effective filtering of email spam. To determine the efficiency and effectiveness of the algorithm, Enron evaluated the performance of the RFs algorithm in the spam data set using accuracy, TPR, FPR, accuracy, and F-size. We conclude by saying that RF is a promising algorithm that can be adopted on the mail server or on the mail client side to further reduce the amount of spam messages in the inbox of e-mail users.

Key words: machine learning, spam filtering, random forests, neural networks, reference vector machines, Naive Bayes.

Кіріспе. Жақында спам деп аталатын сұраныссыз коммерциялық жаппай электрондық пошталар интернетте үлкен проблемаға айналды. Алаяқтық хаттарды жіберетін спамер әртүрлі веб-сайттарды, вирустарды және зиянды бағдарламаларды пайдаланып электрондық пошта мекенжайларын жинайды. Спам интернет пайдаланушыларына сақтау сыйымдылығын және желінің өткізу қабілеттілігін арттыруға кедергі келтіреді. Компьютерлік желілерде спам-хабарламалардың үлкен көлемінің болуы электрондық пошта серверінің жадын, өткізу қабілеттілігін, процессорды өңдеу жылдамдығын және пайдаланушы уақытын тиімді пайдалануға зиян келтіреді (Fonseca және т.б., 2016). Есептер спам-хабарламалар дүние жүзіндегі электрондық пошта трафигінің 77%-дан астамына жауап беретінін көрсетті (Касперский, 2017). Спам электрондық пошталар 419 интернет-хабарламаның және бейтаныс адамдарды пайдаланушы аты мен құпия

сөз, Банк растау нөмірі (BVN) және несие сияқты құпия ақпаратты шығаруға тарту мақсатында электрондық хаттарды жіберудің басқа алаяқтық тәжірибесінің құрбаны болған пайдаланушылар үшін өте тітіркендіргіш және жағымсыз. карта нөмірлері. Әдебиеттерде электрондық поштаны спам сүзгілеудің әртүрлі тәсілдерін ұсынған бірнеше жұмыс жарияланды. Электрондық хаттарды спам немесе спам емес деп жіктеу үшін сәтті қолданылды. Бұл әдістерге ықтималдық, шешім ағашы, жасанды иммундық жүйе (Бахгат және басқалар, 2016), тірек векторлық машинасы (SVM) (Bougila және Ayaugi, 2009), жасанды нейрондық желілер (ANN) (Сао және басқалар, 2004) және жағдайға негізделген әдіс (Fdez-Riverola, 2007).

Электрондық хаттардағы белгілі бір мүмкіндіктерді (әдетте спам электрондық пошталарында жиі қолданылатын кілт сөздер) анықтау мүмкіндігі бар мазмұнға негізделген сүзу әдісін қолдану арқылы спам хаттарды сүзу үшін осы машинада оқыту әдістерін пайдалануға болатыны көрсетілді. Электрондық хаттарда бұл мүмкіндіктердің пайда болу жиілігі шекті мәнмен өлшенген кезде электрондық поштаның спам ретінде жіктелу ықтималдығын анықтайды. Шекті мәннен асатын электрондық пошта хабарлары спам ретінде жіктеледі (Mason, 2003).

Материалдар мен әдістер. Электрондық поштаны спамды сүзу әдістерінің көпшілігі мәтінді санаттау тәсілдерін пайдаланады. Демек, спам сүзгілері нашар жұмыс істейді және спам хаттардың пайдаланушылардың кіріс жәшігіне түсуіне тиімді жол бермейді. Бұл жұмыс электрондық пошталардан маңызды мүмкіндіктерді шығару және электрондық хаттарды ветчина, спам немесе қалыпты деп жіктеу үшін Кездейсоқ ормандар (RF) алгоритмін пайдалану ережелерін пайдаланады. Enron спам деректер жинағы эталондық деректер жинағы ретінде пайдаланылды. Random Forests машиналық оқыту алгоритмі WEKA көмегімен модельденді (Wang, 2005). WEKA-да деректерді алдын ала өңдеу, жіктеу, регрессия, кластерлеу және ассоциация ережелері үшін пайдалануға болатын машиналық оқыту алгоритмдерінің жиынтығы бар. WEKA жүйесінде енгізілген машиналық оқыту әдістері әртүрлі нақты әлемдегі мәселелерді шешуге көмектеседі. Құралдар жинағы зерттеушілер мен әзірлеушілерге машиналық оқытудың әртүрлі алгоритмдерімен тәжірибе жасау, олардың үлгілерін құру және бағалау үшін жақсы анықталған құрылымды ұсынады. Барлық эксперименттер AMD A 10-7300 Radeon R6, 10 есептеу ядросы 4C+6G, 1,90 ГГц, 8,00 ГБ жедел жады бар машинада жүргізілді.

Random Forest. Random Forests (RFs) ансамбльді оқытудың классикалық мысалы және деректерді жіктеу мәселелерін шешуге

жарамды регрессия әдісі. Breiman and Cutler (2007) RFs алгоритмін ұсынды. Алгоритм шешім ағаштарының көмегімен деректерді әртүрлі класстарға жіктейді. Жаттығу кезеңінде кейбір шешім ағаштары жасалады және кейінірек жіктеу тапсырмалары үшін пайдаланылады. Бұл жеке ағаштардың сайланған сыныбын ескере отырып жұмыс істейді және ең көп дауыс жинаған сынып соңғы нәтиже ретінде қарастырылады. RF алгоритмі жылдар бойы өте танымал болды және ол адам әрекетінің әртүрлі салаларындағы ұқсас мәселелерді шешу үшін қолданылады (Fette et al., 2007), (Koprinska, et al., 2007) және (Whittaker et al., 2010).). Кездейсоқ ормандардың бірнеше артықшылығы бар, мысалы: жіктеу қатесінің азаюы және кейбір басқа машиналық оқыту әдістерімен салыстырғанда жақсы f-баллдары. Оның үстіне, оның өнімділігі әдетте SVMs сияқты жақсы немесе одан да жоғары. Ол мәндері жоқ теңгерімсіз деректер жиындарын тиімді өңдей алады. Ол жетіспейтін деректердің болжалды мәнін есептеудің және деректердің елеулі бөлігі жетіспейтін жағдайларда деректердің дәлдігін сақтаудың тиімді алгоритмі ретінде қызмет етеді. RF үшін оқу уақыты әдетте SVM және нейрондық желілермен салыстырғанда қысқа (бірақ бұл жеке енгізуге байланысты). RF дәлдігі бойынша қолданыстағы машиналық оқыту алгоритмдерінің көпшілігінен жақсырақ. Оның үлкен дерекқорлардағы өнімділігі өте жақсы. Ол жүздеген мың кіріс айнымалыларын тиімді өңдей алады. RF орманды өңдеу кезінде ұжымдық қатенің ішкі бейтарап болжамын жасайды. Ол бейтарап деректер жиыны бар популяция класындағы қателерді жеңілдететін тәсілді ұсынады. RF таңбаланбаған деректерді тиімді өңдеу мүмкіндігі бар, бұл таңбаланбаған деректерді кластерлеудің өте қолайлы әдісі. Кездейсоқ ормандар күрделі емес және ол бақылаулар санымен салыстырғанда азырақ параметрлерді пайдаланады. RF пайдаланушыға мүмкіндігінше көп ағаштарды жоғары жылдамдықпен өсіруге мүмкіндік береді. RF ормандағы жеке ағаштардың жанында кіріс векторын қосу арқылы кіріс векторынан жаңа деректерді жіктейді. Әрбір ағаш өзінің классификациясын орындайды, ол әдетте сол сынып үшін ағаш «дауыстары» ретінде белгілі. Орман ормандағы жалпы ең көп дауыс жинаған сыныпты таңдайды. Ағаштарды өсірудің қадамдары төменде сипатталған:

1. N бар деректерден ауыстыруға болатын N даналарды кездейсоқ түрде көрсететін оқу даналарының саны делік. Мұндай мысалдар ағашты өсіруге арналған жаттығулар жиынтығы ретінде пайдаланылады.

2. P кіріс айнымалылары бар делік, $p \ll P$ саны сәйкес келетін әрбір

түйін үшін P айнымалылары P ішінен кездейсоқ таңдалатын және p нүктесіндегі ең жақсы бөлік түйінді бөлу үшін пайдаланылады, осылайша p енді болады. орман өсірудің барлық кезеңінде тұрақты құндылық.

3. Ағаштарды кесуге тыйым салынады, өйткені әрбір ағаш мүмкін болатын ең үлкен деңгейге дейін өсіріледі.

Ағаш қателік деңгейі аз болса, күшті классификатор деп аталады. Оның үстіне ормандағы әрбір ағаштың шоғырлануы артқан сайын орманның қателік деңгейі төмендейді. p мәнін азайту орманның қатынасын және күшін өзара төмендетеді, ал p мәнін арттыру әдетте өте кең болып табылатын p -тің ең жақсы шекарасының аймағында да артады. p мәнін қапшықтан тыс (OOB) қатесі арқылы есептеуге болады (сонымен қатар қаптан тыс бағалау ретінде белгілі) шек ішіндегі p мәнін дереу табуға болады. Бұл кездейсоқ ормандар оның дәл бапталуына аздап сезімтал болатын жалғыз сандық фактор.

Төмендегі алгоритм орман ағаштарын жасау үшін қажетті қадамдарды қысқаша сипаттады.

RF алгоритмін іске қосу

Кіріс: X : түйіндер саны

N : функциялар саны

Y : өсетін ағаштар саны

Шығыс: G : ең көп дауыс жинаған класс

While Тоқтату критерийлері дұрыс емес do:

Жаттығу деректерінен Y кездейсоқ түрде өздігінен басталатын S үлгісін таңдаңыз

Төмендегі қадамдарды қолдана отырып, таңдалған өздігінен басталатын S үлгісінен R_i ағашын жасаңыз:

(1) N ішінен кездейсоқ n мүмкіндікті таңдаңыз; мұндағы $n \ll N$

(2) n мүмкіндіктер арасындағы d түйіні үшін ең жақсы бөлу нүктесін есептеңіз

(3) Негізгі түйінді оңтайлы бөлу арқылы екі ұрпақ түйініне бөліңіз

(4) Түйіндердің максималды саны (x) жасалғанша 1-3-қадамдарды орындаңыз Y рет 1-4 қадамдарды қайталау арқылы орманыңызды жасаңыз

EndWhile

Әрбір жасалған ағаштар үшін өнім шығару $\{R_t\} 1Y$

Түбір түйінінен бастап әрбір жасалған ағаштар үшін жаңа үлгіні пайдаланыңыз Үлгіні жапырақ түйініне сәйкес келетін сыныпқа тағайындаңыз.

Әрбір ағаштың дауыстарын немесе нәтижелерін біріктіріңіз

Ең көп дауыс жинаған сыныпты шығарыңыз (G).

End RF алгоритмі

Эксперимент үшін қолданылатын мәліметтер жиынтығы. Enron спам деректер жинағы біздің тәжірибеміз үшін пайдаланылды (Копринска, және т.б., 2007). Бұл зерттеуде Enron корпорациясының Enron спам деректер жинағы пайдаланылады. Үш қалтада деректер жинағы ретінде 5180 электрондық пошта бар: қалыпты үшін норма, спам емес үшін ветчина және спам электрондық пошта үшін спам. Enron-да 5180 дана, 3672 ветчина, 8 норма және 1500 спам электрондық поштасы бар. Деректер жинағының мүмкіндіктері келесідей:

i. Электрондық хаттарда белгілі бір сөз немесе таңба қайталанып отырды.

ii. Орындау ұзақтығы атрибуттары (55-57) дәйекті бас әріптер тізбегінің ұзындығын өлшейді.

Деректерді Қалыпқа Келтіру Процесі. Біздің эксперименттерімізде пайдаланылған бастапқы деректер жинағы 5180 мәтіндік файлдан тұрады. Бұл файлдардағы деректер қалыпқа келтірілмеген. Бұл WEKA-ға кіріс ретінде қызмет ету үшін оларды қалыпқа келтіру керек дегенді білдіреді. Оқыту үшін WEKA-ға бермес бұрын, барлық деректерді бір .arff файлына түрлендіру қажет. Бұған қол жеткізу үшін WEKA пәрмен жолы интерфейсінде келесі пәрменді қолданамыз.

```
“java weka.core.converters.TextDirectoryLoader -dir D:/Enron > D:/Spam_mails.arff”
```

Қалыпқа келтіру процесінен кейін қалыпқа келтірілген файл алдын ала өңдеу үшін WEKA-ға берілді.

Белгілерді алу. Алдын ала өңдеу кезеңінен кейін мүмкіндікті шығару келеді. Мүмкіндіктерді шығару - оқу жинағында кездесетін терминдердің ішкі жиынын таңдау және мәтінді жіктеуде мүмкіндіктер ретінде тек осы жиынды пайдалану процесі. Бұған кейбір ережелер жиынтығын қолдану арқылы қол жеткізіледі. Мүмкіндіктерді шығару тиімді сөздіктің көлемін азайту арқылы жіктеуішті оқытуды және қолдануды тиімдірек етеді. Сондай-ақ әдетте шу мүмкіндіктерін жою арқылы жіктеу дәлдігін арттырады. Спамды сүзгілеу үшін пайдаланған кейбір маңызды электрондық пошта мүмкіндіктеріне мыналар жатады: Хабардың негізгі мазмұны және тақырыбы, Хабардың көлемі, Сөздердің пайда болу саны, Хабардағы мағыналық сәйкессіздік үлгілерінің саны, Алушының жасы, Жынысы мен елі, Алушының жауап бергені, Ересектерге арналған мазмұн, Хабарлама мазмұнындағы сөздер қапшығы, Домен атауы, IP мекенжайы, Негізгі бөлімде көбірек бос жолдар.

Нәтижелер және талқылау. Бұл бөлімде орындалған тәжірибелердің нәтижелері берілген. Деректер жиынын жіктеу және бағалау

үшін «Кездейсоқ орман» алгоритмі қолданылды, біз 10 еселік кросс-валидация сынағы қолданылды, бұл модельді оқыту үшін бастапқы жиынтықты оқу үлгісіне бөлетін болжамды модельдерді бағалауда қолданылатын тәсіл және оны бағалау. Біріншіден, деректер жиынын оқыту әрбір хабарлама тақырыбын талдау, кілт сөздерді және ақтізім/қара тізімді тексеру арқылы алынған мүмкіндік векторларымен орындалды. Оқытылған үлгілердің өнімділігі оның жіктелу дәлдігі үшін 10 еселік айқаспалы валидация арқылы бағаланады. Жіктеу дәлдігі электрондық поштаның спам классификациясының өнімділік көрсеткіштерінің бірі болып табылады. Ол сынақ деректер жинағындағы дұрыс жіктелген даналардың саны мен сынақ жағдайларының жалпы санының қатынасы ретінде өлшенеді. Спамды сүзгілеуде жалған негативтер кейбір спам хаттардың спам емес деп қате жіктелгенін және пайдаланушының кіріс жәшігіне кіруіне рұқсат етілгенін білдіреді. Жалған позитив спам емес электрондық пошталардың қате спам ретінде жіктелгенін және спам қалтасына жылжытылғанын немесе жойылғанын білдіреді. Көптеген пайдаланушылар үшін жарамды электрондық пошталарды қате түрде спам ретінде жіктеу олардың кіріс жәшігіндегі спам хаттарды алудан өте қымбат болуы мүмкін. Жалған оң көрсеткіш электрондық пошта спам сүзгісінің тиімділігін бағалауда қолданылатын өнімділік көрсеткіштерінің бірі болып табылады. Төмендегі 1-суретте WEKA модельдеу ортасындағы нәтиженің скриншоты көрсетілген.

```

Classifier output
-----
@data

Classifier Model
RandomForest

Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S -2118349930 -do-not-check-capabilities

Time taken to build model: 45.2 seconds
=== Evaluation on training set ===

Time taken to test model on training data: 4.25 seconds

=== Summary ===

Correctly Classified Instances      5176          99.9228 %
Incorrectly Classified Instances     4             0.0772 %
Kappa statistic                    0.9981
Mean absolute error                 0.0296
Root Mean Squared Error             0.04
Relative absolute error              10.7404 %
Root relative squared error         16.1506 %
Total Number of Instances           5180

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Correctly Classified Instances      0.999  0.001  1.000    0.999    0.999    0.998    1.000    1.000    ham
Incorrectly Classified Instances     1.000  0.000  1.000    1.000    1.000    1.000    1.000    1.000    spam
Weighted Avg.                      0.999  0.001  0.998    0.999    0.999    0.998    1.000    1.000

=== Confusion Matrix ===
 a  b  c  <-- classified as
3669  0  3  |  a = ham
  0  8  0  |  b = norm
  1  0 1499 |  c = spam
    
```

Сурет-1 - Enron спам электрондық поштасының деректер жиыны үшін Random Forests жіктеу шығысының скриншоты

Тиімділік. Бұл бөлімде біз барлық машиналық оқыту классификаторларының тиімділігін үлгіні жасауға кететін уақыт, дұрыс жіктелген даналар, қате жіктелген даналар және жіктеу дәлдігі тұрғысынан бағалаймыз. Нәтижелер 1-кестеде көрсетілген.

1-кесте - RFS алгоритмінің жұмысын бағалау

Бағалау критерийлері	
Модельді (модельдерді) жасауға жұмсалған уақыт	17,75
Дұрыс жіктелген даналар	5176
Қате жіктелген даналар	4
Дәлдік (%)	99,92

Біз қарастырып жатқан машиналық оқыту алгоритмдерінің өнімділігін әділ және жақсырақ бағалау үшін бұл жұмыста модельдеу қатесі де ескеріледі. Бұл алгоритмдердің тиімділігі келесі терминдер арқылы бағаланады: Карра статистикасы (KS), Орташа абсолютті қате (MAE), Орташа түбірлік қате (RMSE), Салыстырмалы абсолютті қате (RAE), Түбір салыстырмалы квадраттық қате (RRSE). KS, MAE және RMSE сандық мәндерде. RAE және RRSE пайызбен берілген. Нәтижелер 2-кестеде көрсетілген.

Кесте 2 - RFS алгоритмдерін оқыту және модельдеу қатесі

Бағалау критерийлері	RFS
Каппа статистикасы (KS)	0,9981
Орташа абсолютті қате (MAE)	0,0296
Орташа квадраттық қате (RMSE)	0,06
Салыстырмалы абсолюттік қателік (RAE) %	10.7404
Квадраттың орташа квадраттық абсолютті қатесі (RRAE) %	16.1506

Тиімділік. Болжамдық модельді жасағаннан кейін RFs алгоритмінің тиімділігі төмендегі 3-кестеде көрсетілгендей бағаланды.

Кесте 3 - TPR, FTR, дәлдік және F-Score негізіндегі RFS алгоритмдерінің жұмысын бағалау

Техника	TPR	FPR	Дәлдігі	F-есебі	Классы
RFS	0.999	0.001	1.000	0.998	Хам
	1.000	0.000	1.000	1.000	Қарапайым
	0.999	0.001	0.998	0.998	Спам

1-кестеден оның моделін құру үшін RF шамамен 17,75 секунд қажет болды. RFs 99,92% жіктеу дәлдігіне ие. Сондай-ақ

нәтижелерден RF өте жоғары дұрыс жіктелген даналар және қате жіктелген даналардың өте аз саны бойынша тамаша орындағаны анық. 2-кестеде бейнеленген оқыту және модельдеу қатесі RF тамаша классификация нәтижесін (0,9992%) және өте төмен қателік көрсеткішін (0,0296) бергенін көрсетеді. Модель жасалғаннан кейін келесі қадам қарастырылатын алгоритмдердің тиімділігін анықтау үшін алынған нәтижелерді талдау болып табылады. 3-кестеде RF-ның ветчина, норма және спам кластары үшін TPR, FTR, Precision және F-Score терминдерінде өте жақсы нәтиже бар екені көрсетілген. Төменде 4-кестеде классификаторлардың өнімділігін бағалаудың практикалық әдісін қамтамасыз ететін RF алгоритмінің шатастыру матрицалары берілген, кестенің әрбір жолы сыныптың нақты жылдамдығын білдіреді, ал әрбір баған болжамды көрсетеді.

Кесте 4 - RF алгоритмі үшін шатастыру матрицасы

	Хам	Қарапайым	Спам	Class
RFs	3669	0	3	Хам
	0	8	0	Қарапайым
	1	0	1499	Спам

Жоғарыдағы 4-кестеден RF 3672 дананың 3669 данасын (шынында ветчина болып табылатын 3669 ветчина данасы және шын мәнінде спам болып табылатын 1 спам данасы) және қате болжанған 3 дананы (спам ретінде болжанған ветчина класының 3 данасы) дәл болжайды. Біздің эксперименттерімізден оның жіктелу дәлдігін, TPR, FPR, дәлдік және F-бағалылығын ескере отырып, RF тиімділік пен тиімділік тұрғысынан тамаша орындағаны анық. Ол сондай-ақ 1500 дананың 1499 данасын (шынымен спам болып табылатын 1499 спам данасы және шын мәнінде спам болып табылатын 1 ветчина данасы) және қате болжанған 3 данасын (ветчина ретінде болжанған спам класының 3 данасы) дұрыс болжайды.

Information about the author:

B.B. Tastemir – Almaty Technological University, Almaty, Kazakhstan,
E-mail: tbeka97@gmail.com, ORCID ID - 0000-0002-3869-1973.

ӘДЕБИЕТТЕР

Breiman L., Cutler A. (2007). Random forests-classification description, Department of Statistics Homepage, 2007, <http://www.stat.berkeley.edu/~breiman/RandomForests/cchome.htm>.

- Cao Y., Liao X., Li Y. (2004). An e-mail filtering approach using neural network, In International Symposium on Neural Networks, Springer Berlin Heidelberg, 688-694.
- Dhanaraj KR, Palaniswami V (2014). Firefly and Bayes Classifier for Email Spam Classification in a Distributed Environment. Australian Journal of Basic and Applied Sciences, 8(17):118-130.
- Fdez-Riverola F., Iglesias E.L., Diaz F., Méndez J.R., Corchado J.M. (2007). SpamHunting: An instance-based reasoning system for spam labelling and filtering, Decision Support Systems, 43(3):722-736.
- Fonseca D.M., Fazzion O.H., Cunha E., Las-Casas I., Guedes P.D., Meira W., Chaves M. (2016). Measuring Characterizing, and Avoiding Spam Traffic Costs. IEEE Internet Computing, 99.
- Karthika R., Visalakshi P. (2015). A Hybrid ACO Based Feature Selection Method for Email Spam Classification. WSEAS Transaction on Computers, 14, pp. 171-177.
- Kaspersky lab Spam Report (2017). Visited on May 15, 2018 https://www.securelist.com/en/analysis/204792230/Spam_Report_April_2012, 2012.
- Koprinska I., Poon J., Clark J., Chan J. (2007). Learning to classify e-mail, Information Sciences, 177(10): 2167–2187.
- Sharma A. and Suryawansi A. (2016). A Novel Method for Detecting Spam Email using KNN Classification with Spearman Correlation as Distance Measure. International Journal of Computer Applications, 136 (6):28-34.

REFERENCES

- Breiman L., Cutler A. (2007). Random forests-classification description, home page of the Department of Statistics, 2007, <http://www.stat.berkeley.edu/~Breiman/randomforests/cchome.htm>.
- Cao Yi, Liao X., Li Yi (2004). An approach to email filtering using a neural network, at the International Symposium on neural networks, Springer Berlin Heidelberg, 688-694.
- Dhanaray kr, Palaniswami in (2014). Firefly and Bayesian email classifier. Classification in distributed environments. Australian Journal of fundamental and Applied Sciences, 8 (17): 118-130.
- Fdes-Riverola F., Iglesias El, Diaz F., Mendes Jr., Corchado Jr.Moscow (2007). Spam hunting: a reasoned thinking system for tagging and filtering spam, a decision-making support system, 43 (3): 722-736.
- Fonseca D.M., Facion O., Kunya E., Las Casas I., Guedes P.D., Meiram J., Chavez M. (2016). Measuring characteristics and preventing spam traffic losses. IEEE Internet Computing, 99.
- Kartika R., Visalakshi P. (2015). A hybrid method for selecting ACO-based functions to classify email spam. Wseas transactions on computers, pp. 14, 171-177.
- Kaspersky Lab spam report (2017). Visited on May 15, 2018 https://www.securelist.com/en/analysis/204792230/Spam_Report_April_2012, 2012.

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

Volume 2, Number 342 (2022), 142–152
<https://doi.org/10.32014/2022.2518-1726.135>

UDC 004.588

A. Urynbassarova^{1,2*}, D. Urynbassarova^{3,4}, E. Al-Hussam⁵

¹Narxoz University, Almaty, Kazakhstan;

²L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan;

³Beijing Institute of Technology, Beijing, China;

⁴Beijing Key Laboratory of Fractional Signals and Systems,
Beijing, China;

⁵TikTok, ByteDance, Dubai, United Arab Emirates.

E-mail: *8azhb@mail.ru*

WEBSITE FOR THE LATIN SCRIPT OF THE KAZAKH LANGUAGE

Abstract. The state language of the Republic of Kazakhstan is the Kazakh language, and the written form is the Kazakh alphabet based on Cyrillic script. Recently, by the decision of the Kazakhstan government Kazakh alphabet starts to move from Cyrillic script to Latin script. To the best of our knowledge, Kazakh language courses are available on a few websites, but there is no website that teaches the Latin script of the Kazakh language. In this article, by using HTML, CSS, and JavaScript, we will develop an effective and convenient website for people of all ages, which will explain the rules of the Latin script of the Kazakh language, and teach them to read and write Latin script. We will demonstrate the effectiveness of the aforementioned programs while preparing the website. Moreover, we will create a converter that converts words or text from Cyrillic script to Latin script. Also, we depict its design and illustrate program coding. In the end, we will describe the activation of the website.

Key words: Kazakh language, Latin script, Cyrillic script, website, JavaScript, HTML, CSS.

А. Урынбасарова^{1,2*}, Д. Урынбасарова^{3,4}, Э. Ал-Хуссам⁵

¹Нархоз университеті, Алматы, Қазақстан;

²Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан;

³Бейжің технологиялық институты, Бейжің, Қытай;

⁴Бөлшек сигналдар мен жүйелердің Бейжіңдік негізгі зертханасы,
Бейжің, Қытай;

⁵ТикТок, ByteDance, Дубай, Біріккен Араб Әмірліктері.
E-mail: *sazhb@mail.ru*

ҚАЗАҚ ТІЛІНІҢ ЛАТЫН ГРАФИКАСЫНА АРНАЛҒАН ВЕБ-САЙТ

Аннотация. Қазақстан Республикасының мемлекеттік тілі – қазақ тілі, ал жазба түрі – кирилл графикасына негізделген қазақ алфавиті. Жақында Қазақстан үкіметінің шешімімен қазақ алфавиті кирилл графикасынан латын графикасына ауысуын бастады. Біздің білуімізше, қазақ тілі курстары бірнеше веб-сайттарда бар, бірақ қазақ тілінің латын графикасын оқытатын бірде-бір веб-сайт жоқ. Бұл мақалада HTML, CSS және JavaScript бағдарламаларын қолдану арқылы біз барлық жастағы адамдар үшін қазақ тілінің латын графикасының ережелерін түсіндіретін, оқуға және жазуға үйрететін тиімді, әрі ыңғайлы веб-сайтты ұсынамыз. Веб-сайтты құруда біз жоғарыда аталған бағдарламалардың қолданысының тиімділігін көрсетеміз. Сонымен қатар, сөзді немесе мәтінді кирилл графикасынан латын графикасына алмастыратын конвертер жасаймыз. Әрі веб-сайттың дизайнын бейнелейміз және бағдарлама кодтарын жазамыз. Соңында веб-сайттың белсендірілуін сипаттаймыз.

Түйін сөздер: Қазақ тілі, латын графикасы, кирилл графикасы, веб-сайт, JavaScript, HTML, CSS.

А. Урынбасарова^{1,2*}, Д. Урынбасарова^{3,4}, Э. Ал-Хуссам⁵

¹Университет Нархоз, Алматы, Казахстан;

²Евразийский национальный университет имени Л.Н. Гумилева,
Нур-Султан, Казахстан;

³Пекинский технологический институт, Пекин, Китай;

⁴Пекинская ключевая лаборатория дробных сигналов и систем,
Пекин, Китай;

⁵ТикТок, ByteDance, Дубай, Объединенные Арабские Эмираты;
E-mail: *sazhb@mail.ru*

ВЕБ-САЙТ ЛАТИНСКОЙ ГРАФИКИ КАЗАХСКОГО ЯЗЫКА

Аннотация. Государственный язык Республики Казахстан – казахский язык, письменная форма – казахский алфавит на основе кириллицы. В последнее время решением правительства Казахстана казахский алфавит начинает переходить с кириллицы на латиницу. Насколько нам известно, курсы казахского языка доступны на нескольких веб-сайтах, но нет ни одного веб-сайта, на котором можно самостоятельно обучиться латинскому графику казахского языка. В этой статье мы с помощью HTML, CSS и JavaScript разработаем эффективный и удобный веб-сайт для людей любого возраста, который объяснит правила алфавита латинской графики казахского языка, а также научит читать и писать используя латинские графики. Мы продемонстрируем эффективность вышеупомянутых программ при подготовке веб-сайта. Кроме того, мы создадим конвертер, который конвертирует слово или текст с кириллицы на латиницу. Также мы изобразим дизайн веб-сайта и проиллюстрируем кодирование программы. Затем опишем активацию веб-сайта.

Ключевые слова: Казахский язык, латиница, кириллица, веб-сайт, JavaScript, HTML, CSS.

Introduction. A language is a communication tool, which functions in society in two different forms: oral and written. The Kazakh language belongs to the Turkic language family (Alkim, et al., 2019). In general, the Kazakh language script consists of 5 stages: ancient Turkic script, Arabic script, direct writing, Latin script, Cyrillic script (Pascal, 2020; Ledic, et al., 2019). At first ancient Turkic script was used; evidence of this is Balasagun's work "Kutadgu Bilik". Later, with the adoption of Islam, the Arabic script was used. From 1913 Kazakh alphabet switched to direct writing and from 1929 to the Latin script. Ten years later, it changed to Cyrillic, and till now,

Cyrillic script is in use. Currently, the Kazakh alphabet is slowly moving from Cyrillic script to the Latin script (Gradual transition of the Kazakh alphabet to the Latin alphabet).

Due to the development of information technology, a variety of study opportunities are emerging. According to the latest data, the Internet is actively used by more than 70% of Kazakhstan's population (Toybazarova, et al., 2018). It is easy to find different websites to learn a new language. To the best of our knowledge, there are few websites that teach the Kazakh language. In particular, no website teaches the Latin script of the Kazakh language. In this article, by using HyperText Markup Language (HTML) (Ashley, 2020), Cascading Style Sheets (CSS) (Ashley, 2020), and JavaScript (Frisbie, 2019), (Mason, 2020), (Flanagan, 2011), we will develop a website for people of all ages, which will teach users to read and write the Latin script of the Kazakh alphabet. Also, we describe the convenience of the selected programs and explain their usage. Then, we show how to develop a website, depict its design (Visser, et al., 2019; Procter, et al., 2001; Corry, et al., 1997) and program coding.

Research materials and methods. In this article, HTML, CSS, and JavaScript are selected as the main components that make up a website. Each of them has its own role.

HTML is the language that forms the backbone of a web page. All paragraphs, sections, images, headings, and text are written in HTML format. Content appears on the website in the order in which it is written in HTML. The advantages of using HTML include:

- widely used with a large number of resources;
- works individually in each browser;
- has a clear and consistent source code.

For more advanced features, HTML can be combined with CSS and JavaScript.

CSS manages the style and additional aspects of the layout. CSS is used to create a website design, such as colors, fonts, columns, borders, etc. For example, by employing CSS, we can change the color and size of the heading. Moreover, CSS can be used for animation effects.

JavaScript makes the website lively and interactive. JavaScript is used by all modern web browsers (Internet Explorer, Google Chrome, Firefox) and can run on any operating system (Windows, Linux, Mac). That's why we use JavaScript to create this website. JavaScript can do the following in a browser:

- can add new HTML to a web page, change the content provided, and change styles.

- can provide feedback on user actions
- can send requests via the network to remote servers, download files.
- can receive and set cookies, ask questions to website visitors, display messages.
- can remember all data related to the client.

Results. Designing of website. Website design is a labor-intensive task. While choosing fonts and colors of the website, we have considered social and pedagogical aspects.

First, we use Sublime Text 3 to program HTML and CSS, open the files index.html and style.css. Because the site is designed for people of different ages, neutral colors were chosen during the design. The dimensions of the site are designed to be approximately 800x600 and higher resolutions. It is better to put a logo at the top of the website. The logo was created using CorelDraw (Figure 1).



Figure 1. Logo “Latyn grafikasy”

We then place a menu with four parts, namely Home, Information, Rules, and Contacts. By default, we place the menu in the middle and set it with the same color as the logo (Figure 2).



Figure 2. Website menu

Next, we will create a converter that translates from Cyrillic to Latin. When creating a converter, we add several things: keyboard, place to input text, clear button, text conversion button, a place to display the text after conversion, and a button that allows copying converted text. Due to the fact that some users may not have a keyboard, we will create a special keyboard. Below that, we create a place to enter a word or text. Figure 3 shows the output text after conversion. By clicking the Copy button we can copy the word or text.

After completing the design of the Home page, we will start creating the Information page. It contains information related to the Latin script of the Kazakh language. For example, the documents for the translation of the Kazakh alphabet into Latin script.

The third menu is the Rules page, it shows the Kazakh language rules

based on the Latin alphabet (Figure 4). Through the Rules page, site users can quickly and efficiently learn the rules of the Latin alphabet. We use tags to create tables. In the beginning, we make a list with the names of the rules. Clicking on any of those creates a link to the rules and examples. After learning the rules, it is better to test a user’s knowledge, where tests are optional. Therefore, after each set of rules, we referred to the test tasks.

We have created a menu Contact page to contact us if some users of the website face trouble using the website. This page contains phone numbers and an e-mail link.

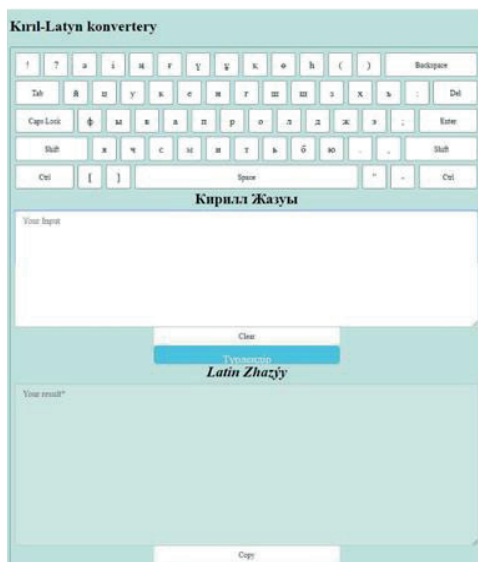
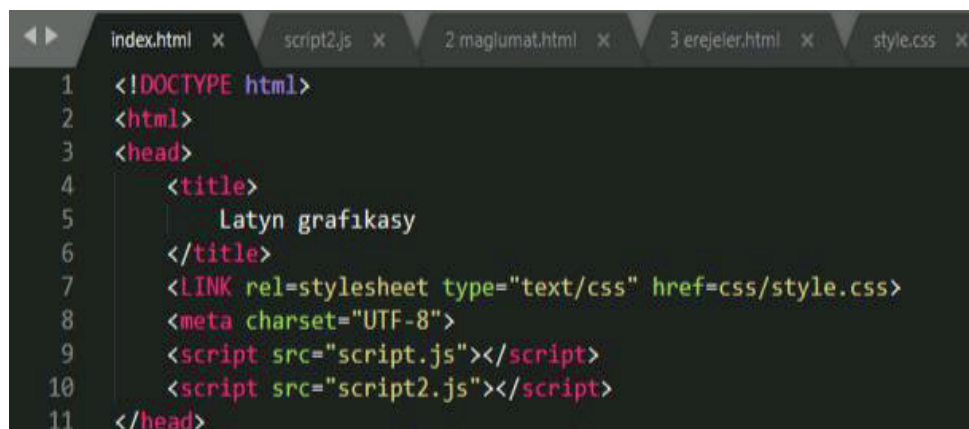


Figure 3. The design of the Home page



Figure 4. Design of the upper part of the Rules page

1. **Activation of website.** In the previous section, we created the design of the website. But the website is not ready to use, it needs to be activated. More precisely, we should let it work. For that, we will use JavaScript. We open a new file script2.js. Then, we enter script2.js in the index.html file (Figure 5).



```
1 <!DOCTYPE html>
2 <html>
3 <head>
4   <title>
5     Latyn grafikasy
6   </title>
7   <LINK rel=stylesheet type="text/css" href=css/style.css>
8   <meta charset="UTF-8">
9   <script src="script.js"></script>
10  <script src="script2.js"></script>
11 </head>
```

Figure 5. Insert a script2.js file into the index.html file

The keyboard is designed. Now we need to write the appropriate code and link the keyboard to the text input area so that users can enter text by pressing the keyboard as below

```
function btn(e) {let letter = e.innerHTML;
                  document.getElementById('sourceStr').value += letter
                  if(window.shift){upperLowerLetters()
                  window.shift = false}}
```

By writing the code as shown above, we make sure that when we press the keys, the sourceStr lands where we enter the text. However, it only works when we press the letters, symbols, Space, and Enter keys. In addition, we have Backspace, Shift, CapsLock, and so on. For each aforementioned keys, we need to write a different code. For example, for the Backspace key, we write the following code function on Press Backspace () {let sourceStr=document.get Element ById ('sourceStr');

```
sourceStr.value=sourceStr.value.substring (0,sourceStr.value.lenght-1);}
```

The Shift and CapsLock keys are the same. The difference is that if we can capitalize only one letter by pressing Shift, and by pressing CapsLock we can capitalize the whole text/word we would like to type. Their code can be seen as follows function onPressShift() {if (!window.shift) {upperLowerLetters('upper')}

```

        window.shift = true;}
    else{upperLowerLetters()
        window.shift = false;}
    function onPressCaps() {if (!window.caps)
{upperLowerLetters('upper')
        window.caps = true;}
    else{upperLowerLetters()
        window.caps = false;}

```

Next, we write a code that determines when to write in uppercase and lowercase (Figure 6). After that, we activate Space and Tab keys as function

```

onPressSpace() {document.getElementById('sourceStr').value += " "}
function onPressTab() {document.getElementById('sourceStr').value
+= " "}

```

```

41 function upperLowerLetters(type='lower'){
42     if(type==='upper') {
43         let keys = document.getElementsByClassName("key");
44         for (let i = 0; i < keys.length; i++) {
45             let key = keys[i]
46             if (key.classList.length <= 1) {
47                 key.innerHTML = key.innerHTML.toUpperCase()
48             }
49         }
50     }
51     else {
52         let keys = document.getElementsByClassName("key");
53         for (let i = 0; i < keys.length; i++) {
54             let key = keys[i]
55             if (key.classList.length <= 1) {
56                 key.innerHTML = key.innerHTML.toLowerCase()
57             }
58         }
59     }
60 }

```

Figure 6. Uppercase and Lowercase

To delete the word or text, we have entered the Clear button as follows

```

<button type="button" class="btn btn-default btn-xs" onclick="
reset(); focus(); value="clear"> Clear </button>

```

After we activate the button to convert the entered word or text to the Latin script from Cyrillic script. We call this button Convert (Түрлендір).

```

<button class="btn btn-info btn-lg" type="button" onclick="
convert(this)"> Түрлендір</button>

```

Now we show how we change letters of the Cyrillic script of the Kazakh language to Latin script. By following the new alphabet, we write down all the letters (Figure 7).

```

61 function convert(e){
62     const dictionary = {
63         "й" : 'i', "ц" : 'ts', "у" : 'y', "к" : 'k', "е" : 'e',
64         "н" : 'n', "г" : 'g', "ш" : 'sh', "щ" : 'sh', "э" : 'z',
65         "х" : 'h', "ь" : '', "ф" : 'f', "ы" : 'y', "в" : 'v',
66         "а" : 'a', "п" : 'p', "р" : 'r', "о" : 'o', "л" : 'l',
67         "д" : 'd', "ж" : 'j', "э" : 'e', "я" : 'ia', "ч" : 'ch',
68         "с" : 's', "м" : 'm', "и" : 'i', "т" : 't', "ь" : '',
69         "б" : 'b', "ю" : 'iy', "э" : 'a', "і" : 'i', "ң" : 'n',
70         "ғ" : 'g', "ү" : 'u', "қ" : 'q', "ө" : "o",
71         "һ" : 'h',
72         "Й" : 'I', "Ц" : 'Ts', "У" : 'Y', "К" : 'K', "Е" : 'E',
73         "Н" : 'N', "Г" : 'G', "Ш" : 'Sh', "Щ" : 'Sh', "Э" : 'Z',
74         "Х" : 'H', "Ь" : '', "Ф" : 'F', "Ы" : 'Y', "В" : 'V',
75         "А" : 'A', "П" : 'P', "Р" : 'R', "О" : 'O', "Л" : 'L',
76         "Д" : 'D', "Ж" : 'J', "Э" : 'E', "Я" : 'Ia', "Ч" : 'Ch',
77         "С" : 'S', "М" : 'M', "И" : 'I', "Т" : 'T', "Ь" : '',
78         "Б" : 'B', "Ю" : 'Iy', "Э" : 'A', "І" : 'I', "Ң" : 'N',
79         "Ғ" : 'g', "Ү" : 'u', "Қ" : 'Q', "Ө" : "o",
80         "һ" : 'H',
81         '': ''
82     }

```

Figure 7. Translation of letters from Cyrillic to Latin

We write the code so that the entered word/text falls into the text input area, and the converted word/text is visible from the text output area (Figure 8).

```

83 let text = document.getElementById('sourceStr').value;
84 let transcript = ""
85 for(let letter of text){
86     console.log(letter, letter in dictionary)
87     if(letter in dictionary){
88         transcript += dictionary[letter]
89     }
90     else transcript+=letter
91 }
92 document.getElementById("outputStr").value = transcript;
93 }

```

Figure 8. The output of converted word/text

Now we need a special button to copy all the text at once. We call that button Copy (Figure 9).

```

function copy() {
    const el = document.getElementById("outputStr")
    el.select();
    document.execCommand('copy');
}

```

Figure 9. Activation of the button Copy

Discussions. Design and colors of this website are chosen based on national preferences. More elements can be added to this website. Using a website that teaches Latin scripts of the Kazakh language offers the following features:

- provide all the necessary information, including the accepted documents by the government;
- teach different rules of the alphabet;
- able to create various self-assessment tests;
- translate the necessary words or texts from Cyrillic to Latin.

Conclusions. In the rapid development of information technology, there is no website that teaches Latin scripts of the Kazakh language. This article describes the development of a website for people of different ages, who can quickly and effectively learn the Latin scripts of the Kazakh language, convert texts from Cyrillic to Latin script and learn the rules of the Latin script. The website, created using JavaScript, HTML and CSS, solves many problems and allows users to learn the alphabet using the Internet's capabilities, taking into account the interests, needs, and competencies of the individual.

Information about authors:

Altyn Urynbassarova – received her specialist degree in Applied Mathematics from Aktobe State University named after K. Zhubanov, Aktobe, Kazakhstan, in 2008. Currently, she is pursuing a Ph.D. degree at Narxoz University, Almaty, Kazakhstan, and working as a Senior Lecturer at L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan. Her research interests are in the areas of information technology and signal processing; *8azhb@mail.ru*; 0000-0002-6256-8641;

Didar Urynbassarova – received her B.Ed. degree (2011) in Mathematics from Aktobe State Pedagogical Institute, Aktobe, Kazakhstan, and her M.Sc. degree (2013) in Mathematics from Aktobe State University named after K. Zhubanov, Aktobe, Kazakhstan, and Ph.D. degree (2018) in Mathematics from Beijing Institute of Technology, Beijing, China. She is currently a Postdoctoral Researcher at the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. Her research interests are in the areas of mathematical signal processing; *didaruzh@mail.ru*; 0000-0001-6261-5586.

Ebrahim Hazzaa Ali Hassan Al-Hussam – received his specialist degree (2007) in IT from Taiz University, Taiz, Yemen, and his M.Sc. degree (2015) in Software Engineering from Beijing Institute of Technology, Beijing, China. Currently, he works at TikTok, ByteDance, Dubai, United Arab Emirates; *em.alhossam@gmail.com*; 0000-0001-6176-2664.

REFERENCES

- Alkim E., Çebi Y. (2019) Machine translation infrastructure for Turkic languages (MT-Turk), *The International Arab Journal of Information Technology*, 16(3):380-388.
- Ashley D. (2020) HTML Pages and CSS. In: *Foundation Dynamic Web Pages with Python*, Berkeley, Canada: Apress.
- Corry M.D., Frick T.W., Hansen L. (1997) User-centered design and usability testing of a web site: An illustrative case study. *ETR & D*, 45(4):65-76.
- Gradual transition of the Kazakh alphabet to the Latin alphabet, (In Russ. and Kaz.) <http://adilet.zan.kz/rus/docs/U1700000569>
- Flanagan D. (2011) *JavaScript: The Definitive Guide*, 6th ed., O'Reilly & Associates.
- Frisbie M. (2019) *JavaScript in HTML*. In *Professional JavaScript for Web Developers*, John Wiley & Sons, Inc.
- Ledic A., Rosandić G. (2019) Comparing Latin with Cyrillic scripts in handwriting identification, *Policija i sigurnost*, 28:27-43.
- Mason P. (2020) *JavaScript*. In: *SAS Stored Processes*, Berkeley, Canada: Apress.
- Pascal A.D. (2020) Cyrillic writing system: from Slavic to Romanian, *Proceedings of SPSTL SB RAS*. 3:5-10 (In Russ.).
- Procter C., Symonds J.A. (2001) Designing for website usability. *Australasian Journal of Information Systems*, 9(1):92-101.
- Toybazarova N.A., Nazarova G. (2018) The modernization of education in Kazakhstan: trends, perspective and problems. *Bulletin of National Academy of sciences of the Republic of Kazakhstan*. 6(376):104-114 <https://doi.org/10.32014/2018.2518-1467.33>
- Visser M., Mark W., Hoving M. (2019) Designing effective websites In book: *Digital Marketing Fundamentals*.

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

ISSN 1991-346X

Volume 2, Number 342 (2022), 153–168

<https://doi.org/10.32014/2022.2518-1726.136>

МРНТИ 50.53.15

УДК 004.932

Э.Э. Эльдарова^{1*}, В.В. Старовойтов², К.Т. Искаков¹

¹Евразийский национальный университет имени Л.Ю. Гумилева,
Нур-Султан Казахстан;

²Национальная академия наук Беларуси, Минск, Беларусь.
E-mail: doctorphd_eldarova@mail.ru

УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА КОНТРАСТНО ИСКАЖЕННЫХ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

Аннотация. Контрастность изображения должна быть достаточной для четкой визуализации изображения. Потому что информативность напрямую связана с контрастностью изображения. Из-за неравномерного освещения и низкой контрастности информационное содержание изображения не передается должным образом, что ограничивает его использование в реальных приложениях. В этой статье проводятся всесторонние исследования и анализируются некоторые традиционные и новейшие методы для улучшения изображений при слабом освещении. Это поможет оценить плюсы и минусы различных методов улучшения контрастных изображений. Представлены экспериментальные результаты, полученные с помощью методов улучшения изображений: Gamma correction, BIMEF, CegaHE, CRM, LIME, MF, HE, MSRCP, SRIE, CEFPBHE, EEBHEGF, 2DHE, MBLLN, KinD, RetinexNet, GLADNet, ZeroDCE. Чтобы оценить результаты исследуемых методов использовали NR меры MDM и NIQMC, которые экспериментально подтверждены их высоким соответствием с субъективными оценками экспертов. Для сравнения с эталоном используем оценку SSIM. В качестве экспериментальных данных используются контрастно искаженные изображения наборов данных LOL и CCID2014. После проведения обширных экспериментов и анализа результатов улучшения слабоконтрастных

изображений, можно заметить, что каждый подход дает хорошие результаты в одних условиях, но имеет ограничения в других. Хотя, в большинстве случаев наблюдаются высокие показатели для методов основанных на машинном обучении, таких как RetinexNet, GLADNet, ZeroDCE, MBLLEN, KinD. Все рассмотренные алгоритмы имеют определенные недостатки, такие как потеря деталей, искажение цвета или высокая вычислительная сложность. Следовательно, рекомендуется выбрать наиболее подходящий алгоритм улучшения контраста, исходя из конкретных требований приложения.

Ключевые слова: цифровое изображение, безэталонная оценка (NR), объективные показатели, контраст, улучшения контраста.

Э.Э. Эльдарова^{1*}, В.В. Старовойтов², К.Т. Искаков¹

¹Л.Ю. Гумилев атындағы Евразиялық ұлттық университеті,
Нұр-Сұлтан Қазақстан;

²Беларусь Ұлттық ғылым академиясы, Минск, Беларусь.
E-mail: doctorphd_eldarova@mail.ru

БҰРМАЛҒАН КОНТРАСТТЫ ЦИФРЛЫҚ БЕЙНЕНІҢ ВИЗУАЛДЫ САПАСЫН ЖАҚСARTУ

Аннотация. Кескінді анық көрсету үшін кескіннің контрасты жеткілікті болуы керек. Өйткені ақпараттық мазмұн кескіннің контрастына тікелей байланысты. Біркелкі емес жарықтандыру және төмен контраст салдарынан кескіннің ақпараттық мазмұны дұрыс берілмейді, бұл оны нақты қолданбаларда пайдалануды шектейді. Бұл мақалада төмен контрасты цифрлық бейнелердің сапасын жақсартудың кейбір дәстүрлі және соңғы әдістеріне талдау жасалып, зерттеулер жүргізіледі. Бұл, өз кезегінде, контрасты кескіндерді жақсартудың әртүрлі әдістерінің оң және теріс жақтарын бағалауға көмектеседі. Гамма түзетуі, BIMEF, SegaHE, CRM, LIME, MF, HE, MSRCP, SRIE, CEFPBHE, EEBHEGF, 2DHE, MBLLEN, KinD, RetinexNet, GLADNet, ZeroDCE тәрізді бейнені жақсарту әдістерін қолдану арқылы алынған эксперименттік нәтижелер ұсынылған. Зерттелетін әдістердің нәтижелерін бағалау үшін сарапшылардың субъективті бағаларымен жоғары сәйкестігі тәжірибе жүзінде расталған NR өлшемдері MDM және NIQMC қолданылды. Эталонмен салыстыру үшін біз SSIM бағалауын қолданамыз. Эксперименттік деректер ретінде LOL және CCID2014 деректер жиынының контрасты бұрмаланған бейнелері пайдаланылады. Төмен контрасты бейнелерді

жақсарту нәтижелеріне кеңейтілген эксперимент пен талдаудан кейін әрбір тәсіл кейбір жағдайларда жақсы нәтиже беретінін, бірақ өзге жағдайларда шектеулер бар екенін көруге болады. Дегенмен де, көптеген жағдайларда RetinexNet, GLADNet, ZeroDCE, MBLen, KinD сияқты машиналық оқытуға негізделген әдістердің көрсеткіштері жоғары болды. Қарастырылған барлық алгоритмдердің белгілі бір кемшіліктері бар, мысалы, бөлшектердің жоғалуы, түстердің бұрмалануы немесе жоғары есептеу күрделілігі. Сондықтан қолданбаның нақты талаптарына негізделген контрастты жақсарту алгоритмін таңдау ұсынылады.

Түйін сөздер: сандық бейне, эталонсыз бағалау (NR), объективті көрсеткіштер, контраст, контрасты жақсарту.

E.E. Eldarova^{1*}, V.V. Starovoytov², K.T. Iskakov¹

¹L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan;

²National Academy of Sciences of Belarus, Minsk, Belarus.

E-mail: doctorphd_eldarova@mail.ru

IMPROVED VISUAL QUALITY OF CONTRAST DISTORTED DIGITAL IMAGES

Abstract. The contrast of the image must be sufficient to render the image clearly. Because the information content is directly related to the contrast of the image. If the image has uneven illumination and/or low contrast, the information content of the image will not be transmitted properly, which limits its use in real applications. This article conducts in-depth research and analyzes some of the traditional and latest techniques for image enhancement in low light. This will help to evaluate the pros and cons of various methods for improving contrast images. Experimental results obtained using image enhancement methods are presented: Gamma correction, BIMEF, CegaHE, CRM, LIME, MF, HE, MSRCP, SRIE, CEFPBHE, EEBHEGF, 2DHE, MBLen, KinD, RetinexNet, GLADNet, ZeroDCE. To evaluate the results of the studied methods, the NR measures MDM and NIQMC used, which were experimentally confirmed by their high agreement with the subjective assessments of experts. For comparison with the benchmark, we use the SSIM score. Contrast distorted images of the LOL and CCID2014 datasets are used as experimental data. After extensive experimentation and analysis of low-contrast image enhancement results, it can be seen that each approach gives good results in some conditions, but has limitations in others. In most cases, there are high rates for methods based on machine learning,

such as: RetinexNet, GLADNet, ZeroDCE, MBLen, KinD. All considered algorithms have certain drawbacks, such as: loss of detail, color distortion, or high computational complexity. Therefore, it is recommended to choose the most appropriate contrast enhancement algorithm based on the specific requirements of the application.

Key words: digital image, no-reference assessment (NR), objective metrics, contrast, contrast enhancements.

Введение. Повышение контрастности является сложной областью и играет жизненно важную роль в приложениях обработки изображений, таких как медицина, наблюдения и области машинного зрения (Muslim, et al, 2019, Khan, et al, 2018, Munir, et al 2018, Khan, et al, 2020). Контрастность изображения должна быть достаточной для четкой визуализации изображения. Потому что информативность напрямую связана с контрастностью изображения. Однако в процессе получения изображения часто существуют некоторые неконтролируемые факторы, приводящие к различным дефектам изображения. В частности, в условиях плохой освещенности, например, в помещении, в ночное время или в пасмурные дни, свет, отраженный от поверхности объекта, может быть слабым; Из-за неравномерного освещения и низкой контрастности информационное содержание изображения не передается должным образом, что ограничивает его использование в реальных приложениях. В этой главе рассматриваются основные методы улучшения изображения при слабом освещении, разработанные за последние десятилетия.

Материалы и методы. Ученые во всем мире предложили множество алгоритмов улучшения изображений, снятых в условиях низкой освещенности, для улучшения видео и изображений при слабом освещении с разных точек зрения. Анализируя множество исследований от простого до новейших методов (Muslim, et all, 2019), используемыми для повышения яркости можно разделить на два класса: традиционные методы и методы машинного обучения. Эти методы могут быть далее разделены на разные подклассы в соответствии с различиями в их принципах.

Традиционные методы улучшения контраста подразделяются на методы пространственной и частотной областей, основанные на операциях с пикселями. В методах пространственной области операции непосредственно применяются к изображению с помощью алгоритмов, которые обычно основаны на содержимом уровня серого. Исследователи обычно использовали методы преобразования серого, методы на основе гистограммы и методы ретинекса для повышения контрастности.

Метод преобразования серого – это алгоритм улучшения изображения в пространственной области, основанный на принципе преобразования значений серого, отдельных пикселей в другие значения серого с помощью математической функции [Голуб], который обычно называют подходом, основанным на отображении. Такой метод улучшает изображение, изменяя распределение и динамический диапазон значений серого в пикселях. К основным подклассам метода этого типа относятся линейные и нелинейные преобразования. Подробная информация написана в статьях (Shukla, et al, 2017, Park, et al, 2018, Wang, et all, 2020).

Гистограмма изображения представляет собой распределение значений интенсивности пикселей. Это дает представление об изменении пикселей изображения, чтобы визуализировать качество до естественного вида. Если значения пикселей изображения равномерно распределены по всем возможным уровням серого, то изображение демонстрирует высокую контрастность и большой динамический диапазон. На основе этой характеристики алгоритм выравнивания гистограммы (histogram equalization (HE)) использует кумулятивную функцию распределения (cumulative distribution function (CDF)) для настройки выходных уровней серого, чтобы иметь функцию плотности вероятности, которая соответствует равномерному распределению. Таким образом, скрытые детали в темных областях могут снова появиться, а визуальный эффект входного изображения может быть эффективно улучшен (Qi, et al, 2021). Основываясь на гистограмме, методы улучшения классифицируются как глобальное выравнивание гистограммы, локальное выравнивание гистограммы и сопоставление гистограммы (Celik, 2014, Celik, 2012). Методы выравнивания глобальной гистограммы подразделяются на методы, основанные на бигистограмме и двухмерной гистограмме. В методах на основе бигистограмм используется разделение гистограмм (HS) и выравнивание субгистограмм. HE широко используется в обработке медицинских изображений, сопоставлении изображений и поиске. Он расширяет и сглаживает динамический диапазон изображения. Однако иногда он дает размытый вид и имеет тенденцию к смещению средней яркости. HE не очень хорошо подходит для бытовой электроники из-за эффектов насыщения в небольших визуальных областях. Эффект насыщения вызывает деградацию изображения, неестественность изображений и потерю характеристик изображения. Для решения вышеуказанных проблем необходимо сохранение средней яркости. Чтобы преодолеть проблемы исследователи предлагают множество методов, такие как BBHE (1997), DSIHE (1999), MMBEHE (2003), RMSHE (2003), RSIHE (2007), BHE-PL (2009),

SSDHE (2014), ESIHE (2014), MMHE (2014), AIEBHE (2014), NHE (2015), ACMHE (2015), BHETPL (2017), VHEUMBH (2017), SEFPBHE (2018), VBBPDHE (2018), EEBHEF (2019), RMMGHT (2019). Методы выравнивания по бигистограмме обеспечивают повышение контрастности наряду с некоторыми дополнительными функциями.

Несколько методов повышения контраста основаны на двухмерной гистограмме. Двумерная (2D) гистограмма состоит из набора бинов которые подсчитывают количество интенсивностей, попадающих в заданную область, тогда как одномерная (1D) гистограмма дает подробную информацию только о распределении интенсивностей. Методы на основе двумерных гистограмм учитывают распределение интенсивностей вместе с их пространственным расположением, благодаря чему методы на основе 2D-гистограмм создают визуально приятные изображения по сравнению с методами на основе 1D-гистограмм (Wang, et al, 2020, Qi, et all, 2021, Celik, 2014, Celik, 2012). Широко используемыми методами основанными на двухмерной гистограмме являются 2DHE, SECEDCT, RSECEDCT, RESECEDT (Qi, et all, 2021, Celik, 2014, Celik, 2012).

Теория ретинекса, созданная Лэндом и Макканном, основана на восприятии цвета человеческим глазом и моделировании цветовой инвариантности (Hussei, et al, 2019). Суть этой теории состоит в том, чтобы определить отражательную природу объекта путем удаления из изображения эффектов освещающего света. Согласно теории ретинекса, зрительная система человека обрабатывает информацию определенным образом во время передачи визуальной информации, тем самым устраняя ряд неопределенных факторов, таких как интенсивность источника света и неравномерность света. Следовательно, сохраняется только информация, отражающая существенные характеристики объекта, такие как коэффициент отражения (Park, et al, 2017, Tanaka, et al, 2017). Многие исследователи предложили эффективные алгоритмы улучшения изображения, основанные на теории Retinex. Из них широко используемые методы SSR, MSR, MSRCR, KBR, MSRCR, NPE (Wang, et all, 2020). Алгоритмы Retinex могут быть легко реализованы. Эти методы позволяют не только повысить контрастность и яркость изображения, но и имеют очевидные преимущества с точки зрения улучшения цветности изображения. Однако эти алгоритмы используют шаблон гауссовской свертки для оценки освещенности и не имеют возможности сохранять края; следовательно, они могут привести к появлению ореолов в некоторых областях с резкими границами или сделать все изображение слишком ярким.

Методы улучшения изображения, основанные на частотной области, преобразуют изображение в частотную область для фильтрации с помощью анализа Фурье, а затем оно преобразуется обратно в пространственную область. Типичные методы частотной области включают методы гомоморфной фильтрации (homomorphic filtering (HF)) [Park, et al, 2018] и вейвлет-преобразования (wavelet transform (WT)) (Łoza, et al, 2013). Алгоритмы HF могут улучшать неровные области, генерируемые светом, сохраняя при этом информацию о контурах изображения. Однако такой алгоритм требует двух преобразований Фурье, т. е. одной экспоненциальной операции и одной логарифмической операции для каждого пикселя изображения; следовательно, его вычислительная сложность велика. WT представляет собой математическое преобразование, которое использует группу функций, называемых вейвлет-функциями, для представления или аппроксимации сигнала. WT можно использовать не только для вычисления локальных характеристик сигналов во временной и частотной областях, но и для проведения многомасштабного анализа функций или сигналов с помощью таких операций, как масштабирование и преобразование. Таким образом, с помощью методов WT был достигнут большой прогресс в повышении контрастности изображения. В алгоритме улучшения изображения на основе WT входное изображение сначала разлагается на низкочастотные и высокочастотные компоненты изображения, затем компоненты изображения на разных частотах улучшаются отдельно, чтобы выделить детали изображения. Основная идея метода вейвлет-анализа заключается в применении вейвлет-разложения к исходному изображению для получения вейвлет-коэффициентов для разных поддиапазонов, корректировке этих вейвлет-коэффициентов, а затем применении обратного преобразования к новым коэффициентам для получения обработанного изображения. Такой алгоритм улучшения изображения может улучшать изображение в нескольких масштабах на основе WT. Считается, что в условиях низкой освещенности большее влияние оказывают высокочастотные компоненты изображения, которые обычно концентрируются на краях изображения и в контурных областях [Sun, et al, 2017]. Следовательно, алгоритм на основе WT будет усиливать высокочастотные компоненты входного изображения и подавлять его низкочастотные компоненты. В частности, комплекс двойного дерева WT обычно позволяет достичь удовлетворительных результатов. Алгоритмы на основе частотной области могут эффективно выделять детали изображения за счет улучшения вейвлет-коэффициентов, но они

также могут увеличивать шум в изображении. Как и другие методы преобразования в частотной области, эти методы улучшения изображения требуют больших объемов вычислений, а выбор параметров преобразования часто требует ручного вмешательства.

Большинство существующих методов улучшения изображения при слабом освещении основаны на моделях, а не на данных. Только в последние годы методы, основанные на машинном обучении для улучшения изображений, начали появляться в значительном количестве (Wang, et al,2020, Gómez, et al, 2019).

Благодаря инновациям в области глубокого обучения и подходов, основанные на сверточных нейронных сетях (CNN) (Gómez, et al, 2019), широко используются учеными для решения таких задач, как туманное изображение и супер разрешение. Архитектуры на основе CNN часто используются для улучшения изображения при слабом освещении. Многие конвейеры используют парное контролируемое обучение путем синтеза изображений при слабом освещении и использования их исходных изображений в качестве основных правд. GAN с неконтролируемым подходом также широко используются для улучшения (Wang, et al,2020, Gómez, et al, 2019). Сбор парных данных крайне непрактичен и является основным ограничением для контролируемого обучения и, следовательно, требует искусственного синтеза изображений при слабом освещении. GAN имеют преимущество, когда дело доходит до сбора данных, но требуют больших вычислительных ресурсов по сравнению с CNN.

В этой статье проводятся всесторонние исследования и анализируются некоторые традиционные и новейшие методы для улучшения изображений при слабом освещении. Это поможет оценить плюсы и минусы различных методов улучшения контрастных изображений. Представлены экспериментальные результаты, полученные с помощью методов улучшения изображений: Gamma correction, BIMEF, CegaHE, CRM, LIME, MF, HE, MSRCP, SRIE, CEFPBHE, EEBHEGF, 2DHE, MBLLN, KinD, RetinexNet, GLADNet, ZeroDCE (Рис.3). Их объективные оценки перечислены в таблицах 7, где наилучший результат каждой метрики выделен жирным шрифтом.

В качестве экспериментальных данных используются контрастно искаженные изображения наборов данных. LOL и CCID2014. LOL – это первый набор данных, содержащий пары изображений, взятых из реальных сцен для улучшения качества при слабом освещении. LOL содержит 500 пар изображений при слабом/нормальном освещении.

щении, сделанные снимки из различных сцен, например, домов, университетских городков, клубов, улиц (рис1.)

CCID2014 – это специальный набор данных контрастно-искаженных изображений. Он содержит 655 контрастно-искаженных изображений пяти типов. Гамма-перенос, выпуклые и вогнутые дуги, кубические и логистические функции, среднее смещение и составная функция используются для создания этих пяти типов искажений.

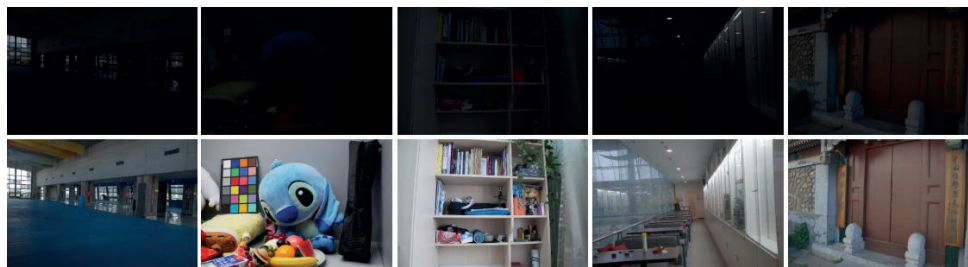


Рисунок 1. Несколько примеров пар изображений при слабом (верхний ряд) и нормальном (нижний ряд) освещении в наборе данных LOL.

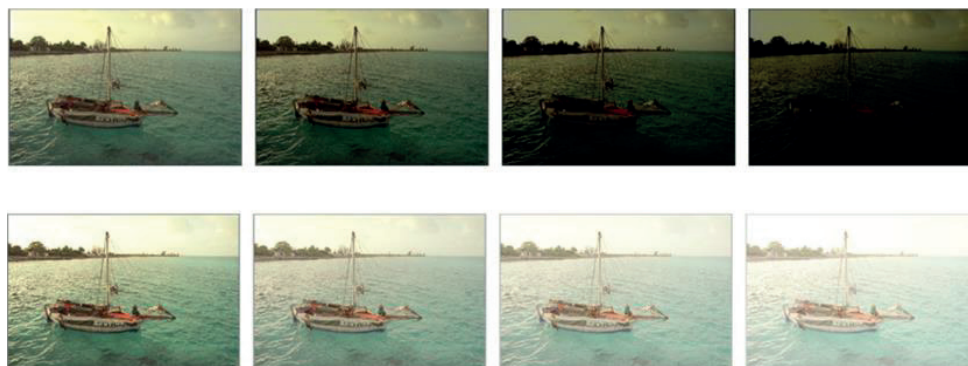


Рисунок 2. Примеры изображений из базы данных CCID2014.

Результаты и обсуждение. Поскольку, при слабом освещении эталонное изображение недоступно, большинство методов, подходящих для оценки улучшения изображения при слабом освещении, основаны на показателях NIQA. Наиболее распространенные показатели NIQA включают среднее значение (MV), стандартную разность (STD), средний градиент (AG), информационную энтропию (IE) и Minkowski Distance based Metric (MDM) и другие. Кроме того, существует несколько универсальных методов оценки качества изображения, в том числе BRISQUE, BTMQI, NIQMC и другие.

Методы объективной оценки имеют различные преимущества такие

как, простота расчетов, быстрота выполнения, простота количественного расчета на основе построенной модели, высокая стабильность; Чтобы сравнить результаты NIQA меры оценки контрастных изображений, а также их согласованность с субъективной оценкой, из литературы были отобраны следующие меры, которые экспериментально подтверждены их высоким соответствием с субъективными оценками экспертов: CEIQ, NIQMC, MDM, CLRIQA, BREN, NCC. Подробное описание указанных мер приведены в статьях (Muslim, et al, 2019, Khan, et al, 2018, Munir, et al 2018, Khan, et al, 2020). В таблице I представлены результаты производительности между сравниваемым NR-IQA, с точки зрения SRC и PLCC для набора данных CCID2014. Для набора данных LOL эксперимент не проводился, так как имеет MOS оценки.

Таблица 1. Сравнение оценок качества изображений из базы данных CCID2014. Лучшая оценка выделена жирным шрифтом

	CEIQ	NIQMC	MDM	CLRIQA	BREN	NCC
SRC	0.816	0.844	0.872	0,821	0.761	0.787
PLCC	0.79	0.81	0.82	0.80	0.73	0.75

Сравнивая оценки NR-IQA, можно заметить, что в данном эксперименте для оценки качества контрастно-искаженных изображений более эффективными являются методы MDM и NIQMC. Для сравнения с эталоном используем оценку SSIM. Она рассматривает деградацию изображения как, изменение структурной информации, а также учитывает условия маскирования, как яркости, так и контраста.

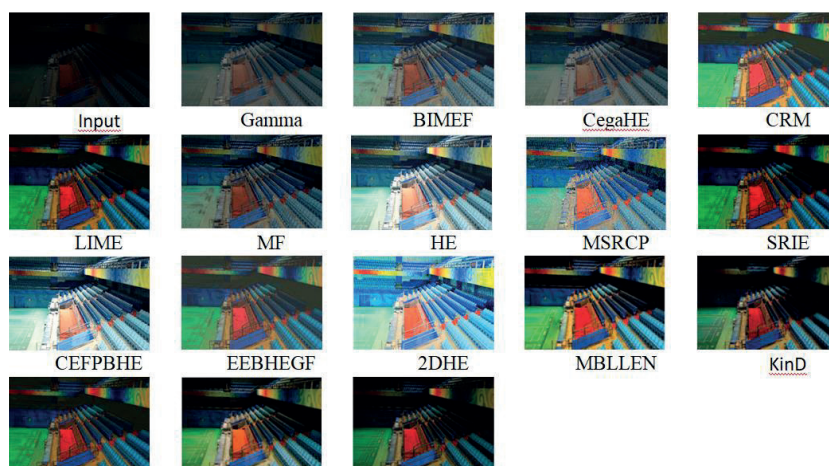


Рисунок 3. Результаты применения разных методов обработки на слабоконтрастных изображениях “img1.png” из базы LOL



Рисунок 4. Результаты применения разных методов обработки на слабоконтрастных изображениях “img2.png” из базы CCID2014

Таблица 2. Объективные оценки для “img1.png” и “img2.png” с использованием различных показателей NIQA

меры	“img1.png”			“img2.png”		
	MDM	NIQMC	SSIM	MDM	NIQMC	SSIM
Gamma	2,57	2,47	0,73	3,72	2,92	0,72
BIMEF	2,91	2,98	0,65	3,62	3,15	0,79
CegaHE	2,8	2,34	0,61	3,68	3,35	0,66
CRM	2,8	2,24	0,66	3,62	3,4	0,71
LIME	3,01	3,57	0,75	4,01	3,62	0,78
MF,	2,98	3,2	0,73	3,90	3,31	0,74
HE	2,89	3,41	0,65	3,76	3,14	0,76
MSRCP	3,1	3,51	0,73	3,84	3,02	0,78
SRIE	2,5	2,59	0,78	3,64	2,95	0,74
CEFPBHE	2,41	3,38	0,61	2,43	2,37	0,69
EEBHEGF	2,56	3,46	0,68	2,47	2,53	0,77
2DHE	2,46	3,52	0,58	2,51	2,43	0,51
MBLLEN	3,96	3,52	0,66	3,87	3,42	0,77
KinD	4,68	3,69	0,67	4,86	4,81	0,71
RetinexNet	4,04	3,98	0,78	3,85	2,98	0,68
GLADNet	4,19	3,99	0,73	4,91	4,77	0,87
ZeroDCE	4,72	3,83	0,75	3,81	3,78	0,73

Таблица 3. Среднее значения объективной оценки для набора данных LOL и CCID2014 с использованием различных показателей NIQA

меры	LOL			CCID2014		
	MDM	NIQMC	SSIM	MDM	NIQMC	SSIM
Gamma	2,48	2,38	0,64	3,63	2,83	0,63
BIMEF	2,82	2,89	0,56	3,53	3,06	0,70
CegaHE	2,71	2,25	0,52	3,59	3,26	0,57
CRM	2,71	2,15	0,57	3,53	3,31	0,62
LIME	2,92	3,48	0,66	3,92	3,53	0,69
MF	2,89	3,11	0,64	3,81	3,22	0,65
HE	2,80	3,32	0,56	3,67	3,05	0,67
MSRCP	3,01	3,42	0,64	3,75	2,93	0,69
SRIE	2,41	2,50	0,69	3,55	2,86	0,65
CEFPBHE	2,32	3,29	0,52	2,34	2,28	0,60
EEBHEGF	2,47	3,37	0,59	2,38	2,44	0,68
2DHE	2,37	3,43	0,49	2,42	2,34	0,42
MBLLEN	3,87	3,43	0,57	3,78	3,33	0,68
KinD	4,59	3,60	0,58	4,77	4,72	0,62
RetinexNet	3,95	3,89	0,69	3,76	2,89	0,59
GLADNet	4,10	3,90	0,64	4,82	4,68	0,78
ZeroDCE	4,63	3,74	0,66	3,72	3,69	0,64

Таблица 4. Сравнение временной сложности (единица измерения: секунды)

Методы	time	Методы	time
Gamma	0,041	CEFPBHE	0,27
BIMEF	0,176	EEBHEGF	0,29
CegaHE	0,063	2DHE	0,15
CRM	0,16	MBLLEN	0,175
LIME	0,226	KinD	0,165
MF	0,45	RetinexNet	0,097
HE	0,14	GLADNet	0,094
MSRCP	0,172	ZeroDCE	0,0012
SRIE	7,082		

Как показано на рисунках 3-4, все эти методы улучшения изображения в некоторой степени улучшают визуальный эффект исходного изображения. Детали становятся более четкими при использовании методов Gamma, BIMEF и CegaHE, но общий уровень яркости темный. Методы HE, 2DHE, CEFPBHE могут осветлить все изображение, но оттенок резко меняется для изображения “img1.png”, что приводит к потере реального цвета исходной сцены. На “img1.

png” яркость изображения после обработки методом SRIE, LIME невелика, но на рис2 имеются явные отличия и эффект восстановления тона выше. Методы MBLLN, MF, LIME и CRM демонстрируют хорошее улучшение цвета и деталей, а их визуальные эффекты явно превосходят HE, 2DHE, Gamma, VIMEF. Так же заметно значительное улучшение контрастности выходного изображения при использовании метода бигистограммы SEFPBHE и EEBHEGF. Края сохраняются на изображениях, полученных методом EEBHEGF, благодаря использованию управляемого фильтра. Сравнение методов на основе машинного обучения показало, что MBLLN на основе слияния имел самое большое время работы и показал чрезмерное усиление в более ярких областях. Основываясь на теории Retinex, метод KinD обработал изображения значительно лучше, чем RetinexNet, но продемонстрировал потерю деталей и сглаживание в более темных областях, хотя сохранил их структуру. RetinexNet поддерживала точные цветовые описания, но вносила значительное количество шума. Метод GLADNet, несмотря на простую архитектуру, показал превосходные показатели и сохранил естественность изображения. ZeroDCE – самый быстрый метод, хотя и не показал высоких показателей, но давал визуально привлекательные изображения с хорошим сходством характеристик и постоянством цвета.

В таблице 3 представлены средние значения объективной оценки для наборов данных LOL и CCID2014 с использованием различных показателей NR IQA. Эти данные показывают, что разные метрики присваивают разные показатели одному и тому же алгоритму улучшения изображения и что интерпретации результатов оценки в некоторых случаях полностью противоположны. Возможно причина в том, что метрики NR IQA, используемые в этой оценке, учитывают различные аспекты изображения, полученного после улучшения. Хотя, в большинстве случаев наблюдаются высокие показатели для методов основанных на машинном обучении, таких как RetinexNet, GLADNet, ZeroDCE, MBLLN, KinD.

Заключение. После проведения обширных экспериментов и анализа результатов улучшения слабоконтрастных изображений, можно заметить, что каждый подход дает хорошие результаты в одних условиях, но имеет ограничения в других. Методы преобразования серого имеют низкую вычислительную сложность, но могут легко потерять детали. Методы на основе гистограммы обеспечивают быстрое улучшение контрастности и детализации изображений, но могут

вызвать потерю цвета и создание шума. Методы Retinex эффективно обрабатывают цветные искажения, но имеют высокую сложность из-за гауссовой фильтрации. Методы машинного обучения могут добиться отличной производительности, но имеют сильную зависимость от набора данных и высокую вычислительную сложность в процессе обучения. Все рассмотренные алгоритмы имеют определенные недостатки, такие как потеря деталей, искажение цвета или высокая вычислительная сложность. Следовательно, рекомендуется выбрать наиболее подходящий алгоритм улучшения контраста, исходя из конкретных требований приложения.

Information about authors:

Eldarova E.E. – PhD doctoral student, L.N. Gumilyov Eurasian National University, 87013748754, doctorphd_eldarova@mail.ru;

Starovoytov V.V. – doctor of technical sciences, professor, ²National Academy of Sciences of Belarus, valerystar@mail.ru;

Iskakov K.T. – doctor of Physical and Mathematical Sciences, professor, L.N. Gumilyov Eurasian National University, kazizat@mail.ru.

ЛИТЕРАТУРА:

Celik T. (2012) Two-dimensional histogram equalization and contrast enhancement. *Pattern Recognition*, 45(10), 3810-3824.

Celik T. (2014). Spatial entropy-based global and local image contrast enhancement. *IEEE Transactions on Image Processing*, 23(12), 5298-5308.

Gómez P., Semmler M., Schützenberger A., Bohr C., & Döllinger M. (2019). Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Medical & biological engineering & computing*, 57(7), 1451-1463.

Hussein R.R., Hamodi Y.I. & Sabri R.A. (2019) Retinex theory for color image enhancement: a systematic review. *International Journal of Electrical & Computer Engineering* (2088-8708), 9(6).

Khan M.A., Javed K., Khan S.A., Saba T., Habib U., Khan J.A. & Abbasi A.A. (2020) Human action recognition using fusion of multiview and deep features: an application to video surveillance. *Multimedia tools and applications*, 1-27.

Khan S.A., Hussain S., Xiaoming S. & Yang S. (2018) An effective framework for driver fatigue recognition based on intelligent facial expressions analysis. *Ieee Access*, 6, 67459-67468.

Khan S.A., Ishtiaq M., Nazir M. & Shaheen M. (2018) Face recognition under varying expressions and illumination using particle swarm optimization. *Journal of computational science*, 28, 94-100.

Łoza A., Bull D.R., Hill P.R. & Achim A.M. (2013). Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients. *Digital Signal Processing*, 23(6), 1856-1866.

Munir A., Hussain A., Khan S.A., Nadeem M. & Arshid S. (2018) Illumination

invariant facial expression recognition using selected merged binary patterns for real world images. *Optik*, 158, 1016-1025.

Muslim H.S.M., Khan S.A., Hussain S., Jamal A. & Qasim H.S.A. (2019) Knowledge-based image enhancement and denoising approach. *Computational and Mathematical Organization Theory*, 25(2), 108-121.

Park S., Kim K., Yu S. & Paik J. (2018) Contrast enhancement for low-light image enhancement: A survey. *IEIE Transactions on Smart Processing & Computing*, 7(1), 36-48.

Park S., Kim K., Yu S. & Paik J. (2018). Contrast enhancement for low-light image enhancement: A survey. *IEIE Transactions on Smart Processing & Computing*, 7(1), 36-48.

Park S., Yu S., Moon B., Ko S. & Paik J. (2017) Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2), 178-184.

Qi Y., Yang Z., Sun W., Lou M., Lian J., Zhao W. & Ma Y. (2021) A comprehensive overview of image enhancement techniques. *Archives of Computational Methods in Engineering*, 1-25.

Shukla K.N., Potnis A. & Dwivedy P. (2017) A review on image enhancement techniques. *Int. J. Eng. Appl. Comput. Sci*, 2(07), 232-235.

Sun T., Jung C., Ke P., Song H. & Hwang J. (2017, December). Readability enhancement of low light videos based on discrete wavelet transform. In *2017 IEEE International Symposium on Multimedia (ISM)* (pp. 342-345). IEEE.

Tanaka H., Waizumi Y. & Kasezawa T. (2017) Retinex-based signal enhancement for image dark regions. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 205-209). IEEE.

Wang W., Wu X., Yuan X. & Gao Z. (2020) An experiment-based review of low-light image enhancement methods. *Ieee Access*, 8, 87884-87917.

Голуб Ю.И. & Старовойтов Ф.В. (2019) Исследование локальных оценок контраста цифровых изображений при отсутствии эталона. *Системный анализ и прикладная информатика*, (2), 4-11.

REFERENCES:

Celik T. (2012) Two-dimensional histogram equalization and contrast enhancement. *Pattern Recognition*, 45(10), 3810-3824.

Celik T. (2014). Spatial entropy-based global and local image contrast enhancement. *IEEE Transactions on Image Processing*, 23(12), 5298-5308.

Golub Y.U.I. & Starovoytov F.V. (2019) Issledovaniye lokal'nykh otsenok kontrasta tsifrovyykh izobrazheniy pri otsutstvii etalona. *Sistemnyy analiz i prikladnaya informatika*, (2), 4-11 (in Russ.).

Gómez P., Semmler M., Schützenberger A., Bohr C. & Döllinger M. (2019). Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Medical & biological engineering & computing*, 57(7), 1451-1463.

Hussein R.R., Hamodi Y.I. & Sabri R.A. (2019) Retinex theory for color image enhancement: a systematic review. *International Journal of Electrical & Computer Engineering* (2088-8708), 9(6).

Khan M.A., Javed K., Khan S.A., Saba T., Habib U., Khan J.A. & Abbasi A.A. (2020) Human action recognition using fusion of multiview and deep features: an application to video surveillance. *Multimedia tools and applications*, 1-27.

Khan S.A., Hussain S., Xiaoming S. & Yang S. (2018) An effective framework for

driver fatigue recognition based on intelligent facial expressions analysis. *Ieee Access*, 6, 67459-67468.

Khan S.A., Ishtiaq M., Nazir M. & Shaheen M. (2018) Face recognition under varying expressions and illumination using particle swarm optimization. *Journal of computational science*, 28, 94-100.

Loza A., Bull D.R., Hill P.R. & Achim A.M. (2013). Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients. *Digital Signal Processing*, 23(6), 1856-1866.

Munir A., Hussain A., Khan S.A., Nadeem M. & Arshid S. (2018) Illumination invariant facial expression recognition using selected merged binary patterns for real world images. *Optik*, 158, 1016-1025.

Muslim H.S.M., Khan S.A., Hussain S., Jamal A. & Qasim H.S.A. (2019) Aknowledge-based image enhancement and denoising approach. *Computational and Mathematical Organization Theory*, 25(2), 108-121.

Park S., Kim K., Yu S. & Paik J. (2018) Contrast enhancement for low-light image enhancement: A survey. *IEIE Transactions on Smart Processing & Computing*, 7(1), 36-48.

Park S., Kim K., Yu S. & Paik J. (2018). Contrast enhancement for low-light image enhancement: A survey. *IEIE Transactions on Smart Processing & Computing*, 7(1), 36-48.

Park S., Yu S., Moon B., Ko S. & Paik J. (2017) Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2), 178-184.

Qi Y., Yang Z., Sun W., Lou M., Lian J., Zhao W. & Ma Y. (2021) A comprehensive overview of image enhancement techniques. *Archives of Computational Methods in Engineering*, 1-25.

Shukla K.N., Potnis A. & Dwivedy P. (2017) A review on image enhancement techniques. *Int. J. Eng. Appl. Comput. Sci*, 2(07), 232-235.

Sun T., Jung C., Ke P., Song H. & Hwang J. (2017, December). Readability enhancement of low light videos based on discrete wavelet transform. In *2017 IEEE International Symposium on Multimedia (ISM)* (pp. 342-345). IEEE.

Tanaka H., Waizumi Y. & Kasezawa T. (2017) Retinex-based signal enhancement for image dark regions. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 205-209). IEEE.

Wang W., Wu X., Yuan X. & Gao Z. (2020) An experiment-based review of low-light image enhancement methods. *Ieee Access*, 8, 87884-87917.

МАЗМҰНЫ

Т.И. Ганиева, Н.С. Семенов, С.Р. Семенов ЖАҒАНДЫҚ ҚОҒАМНЫҢ АҚПАРАТТЫҚ ИНФРАҚҰРЫЛЫМЫ САЛАСЫНДАҒЫ АҚПАРАТТЫҚ ҚАТЫНАСТАРДЫҢ КИБЕРҚАУПСІЗДІГІ.....	5
Е.С. Голенко, А.А. Исмаилова, А.С. Жумаханова «GENE ONTOLOGY» БАЗАСЫН ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ ҮЛГІЛЕРІН ПАЙДАЛАНА ОТЫРЫП АҚУЫЗ ФУНКЦИЯЛАРЫН БОЛЖАУ.....	19
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали СУ ЭКОЖҮЙЕЛЕРІН ЗЕРТТЕУДІҢ АҚПАРАТТЫҚ ТАЛДАУ ЖҮЙЕСІН ӨЗІРЛЕУ.....	39
А.А. Мырзатай, Л.Г. Рзаева, Г. Абитова, М.А. Жакенов ОҚИҒАЛАРДЫ БОЛЖАУ ЖҮЙЕЛЕРІНІҢ КІРІСТЕРІН ЖҮЙЕЛЕУ ҮШІН LAN МОНИТОРИНГ ЖҮЙЕСІН ЕНГІЗУ ЖӘНЕ ПАЙДАЛАНУ.....	54
Ж.С. Иксебаева, К. Жетписов, Ж.М. Муратова ГАНТ ДИАГРАММАСЫН ҚҰРУДЫҢ АҚПАРАТТЫҚ ЖҮЙЕСІ.....	64
Қ.Т. Қырғызбай, Е.Х. Какимжанов, Ж.М. Сагинтаев ГАЖ-ТЕХНОЛОГИЯЛАРЫ НЕГІЗІНДЕ АЛМАТЫ ОБЛЫСЫН АГРОКЛИМАТТЫҚ АУДАНДАСТЫРУ.....	76
А.А. Мухитова, А.С. Еримбетова, В.Б. Барахнин, Э.Н. Дайырбаева, А. Адалбек РЕЛЯЦИЯЛЫҚ ЖӘНЕ УАҚЫТҚА ТӘУЕЛДІ XML-ДЕРЕКТЕР ҚОРЫНДАҒЫ XML-ДЕРЕКТЕРДІ ӨНДЕУДІҢ ЗАМАНАУИ ӘДІСТЕРІ....	92
Б.Б. Оразбаев, Ж.Ж. Молдашева, В.И. Гончаров, К.Н. Оразбаева МАГИСТРАЛДЫ ҚҰБЫРЛАРМЕН МҰНАЙ ТАСМАЛДАУДЫ ДИАГНОСТИКАЛАУ ЖӘНЕ БАСҚАРУ ЖҮЙЕЛЕРІ.....	112
Б.Б. Тастемір ЭЛЕКТРОНДЫҚ ПОШТА СПАМДЫ СҮЗГІЛЕУГЕ АРНАЛҒАН RANDOM FORESTS МАШИНАЛЫҚ ОҚЫТУ ӘДІСІ.....	130
А. Урынбасарова, Д. Урынбасарова, Э. Ал-Хуссам ҚАЗАҚ ТІЛІНІҢ ЛАТЫН ГРАФИКАСЫНА АРНАЛҒАН ВЕБ-САЙТ.....	142
Э.Э. Эльдарова, В.В. Старовойтов, К.Т. Искаков БҰРМАЛҒАН КОНТРАСТТЫ ЦИФРЛЫҚ БЕЙНЕНІҢ ВИЗУАЛДЫ САПАСЫН ЖАҚСARTУ.....	153

СОДЕРЖАНИЕ

Т.И. Ганиева, Н.С. Семенов, С.Р. Семенов КИБЕРБЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ОТНОШЕНИЙ В СФЕРЕ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ ГЛОБАЛЬНОГО ОБЩЕСТВА.....	5
Е.С. Голенко, А.А. Исмаилова, А.С. Жумаханова ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКОВ ПРИ ПОМОЩИ БАЗЫ ДАННЫХ «GENE ONTOLOGY» И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ.....	19
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали К РАЗРАБОТКЕ ИНФОРМАЦИОННОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ ИССЛЕДОВАНИЯ ВОДНЫХ ЭКОСИСТЕМ.....	39
А.А. Мырзатай, Л.Г. Рзаева, Г. Абитова, М.А. Жакенов ВНЕДРЕНИЕ И ИСПОЛЬЗОВАНИЕ СИСТЕМ МОНИТОРИНГА ЛВС ДЛЯ СИСТЕМАТИЗИРОВАНИЯ ВХОДНЫХ ДАННЫХ СИСТЕМ ПРОГНОЗИРОВАНИЯ ИНЦИДЕНТОВ.....	54
Ж.С. Иксебаева, К. Жетписов, Ж.М. Муратова ИНФОРМАЦИОННАЯ СИСТЕМА ПОСТРОЕНИЯ ДИАГРАММЫ ГАНТА.....	64
Қ.Т. Қырғызбай, Е.Х. Какимжанов, Ж.М. Сагинтаев АГРОКЛИМАТИЧЕСКОЕ РАЙОНИРОВАНИЕ АЛМАТИНСКОЙ ОБЛАСТИ С ПРИМЕНЕНИЕМ ГИС-ТЕХНОЛОГИЙ.....	76
А.А. Мухитова, А.С. Еримбетова, В.Б. Баракнин, Э.Н. Дайырбаева, А. Адалбек СОВРЕМЕННЫЕ МЕТОДЫ ОБРАБОТКИ XML-ДАННЫХ В РЕЛЯЦИОННЫХ И ВРЕМЕННЫХ XML-БАЗАХ ДАННЫХ.....	92
Б.Б. Оразбаев, Ж.Ж. Молдашева, В.И. Гончаров, К.Н. Оразбаева ДИАГНОСТИРОВАНИЕ СИСТЕМЫ УПРАВЛЕНИЯ ТРАНСПОРТИРОВКИ НЕФТИ ПО МАГИСТРАЛЬНЫМ ТРУБОПРОВОДАМ.....	112
Б.Б. Тастемир МЕТОД МАШИННОГО ОБУЧЕНИЯ RANDOM FORESTS ДЛЯ ФИЛЬТРАЦИИ НЕЖЕЛАТЕЛЬНОЙ ПОЧТЫ.....	130
А. Урынбасарова, Д. Урынбасарова, Э. Ал-Хуссам ВЕБ-САЙТ ЛАТИНСКОЙ ГРАФИКИ КАЗАХСКОГО ЯЗЫКА.....	142
Э.Э. Эльдарова, В.В. Старовойтов, К.Т. Искаков УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА КОНТРАСТНО ИСКАЖЕННЫХ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ.....	153

CONTENTS

T.I. Ganieva, N.S. Semenov, S.R. Semenov CYBERSECURITY OF INFORMATION RELATIONS IN THE FIELD OF INFORMATION INFRASTRUCTURE OF A GLOBAL SOCIETY.....	5
Y.S. Golenko, A.A. Ismailova, A.S. Zhumakhanova PREDICTING PROTEIN FUNCTIONS USING THE «GENE ONTOLOGY» DATABASE AND MACHINE LEARNING MODELS.....	19
R.M. Moldasheva, A.A. Ismailova, A.K. Zhamangara, A.M. Zadagali ABOUT DEVELOPMENT OF AN INFORMATION ANALYTICAL SYSTEM FOR THE STUDY OF AQUATIC ECOSYSTEMS.....	39
A.A. Myrzatay, L.G. Rzayeva, G. Abitova, M.A. Zhakenov THE IMPLEMENTATION AND THE USE OF THE LAN MONITORING SYSTEMS FOR SYSTEMATISATION OF THE INPUT DATA OF THE INCIDENT FORECASTING SYSTEMS.....	54
Zh.S. Ixebayeva, K. Jetpisov, Zh.M. Muratova INFORMATION SYSTEM FOR CONSTRUCTING GANTT CHARTS.....	64
K.T. Kyrgyzbay, E.Kh. Kakimzhanov, Jay Sagin AGRO-CLIMATIC ZONING OF ALMATY REGION USING GIS TECHNOLOGIES.....	76
A.A. Mukhitova, A.S. Yerimbetova, V.B. Barakhnin, E. Daiyrbayeva, A. Adalbek MODERN METHODS OF PROCESSING XML DATA IN RELATIONAL AND TEMPORARY XML DATABASES.....	92
B.B. Orazbayev, Zh.Zh. Moldasheva, B.I. Goncharov, K.N. Orazbayeva DIAGNOSTICS AND SYSTEMS OF OIL TRANSPORTATION THROUGH MAIN PIPELINES.....	112
B.B. Tastemir RANDOM FORESTS MACHINE LEARNING TECHNIQUE FOR EMAIL SPAM FILTERING.....	130
A. Urynbassarova, D. Urynbassarova, E. Al-Hussam WEBSITE FOR THE LATIN SCRIPT OF THE KAZAKH LANGUAGE.....	142
E.E. Eldarova, V.V. Starovoytov, K.T. Iskakov IMPROVED VISUAL QUALITY OF CONTRAST DISTORTED DIGITAL IMAGES.....	153

**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 29.06.2022.

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

9,0 п.л. Тираж 300. Заказ 1.