

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN



**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION
AND COMMUNICATION TECHNOLOGIES**

2022 (3) 2
Сәуір-маусым

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

БАС РЕДАКТОР:

Хикметов Аскар Кусупбекович — басқарма төрағасы, Халықаралық ақпараттық технологиялар университетінің ректоры, физика-математика ғылымдарының кандидаты (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Ақпараттық жүйелер» кафедрасының проректоры (Қазақстан)

ҒАЛЫМ ХАТШЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, «Халықаралық ақпараттық технологиялар университеті» АҚ, Ғылыми-зерттеу жұмыс департаментінің директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

Разак Абдул — PhD, Халықаралық ақпараттық технологиялар университетінің профессоры (Қазақстан)

Лучио Томмазо де Паолис — Саленто университетінің (Италия) инновациялар және технологиялық инженерия департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

Лиз Бэкон — профессор, Абертэй университеті вице-канцлердің орынбасары (Ұлыбритания)

Микеле Пагано — PhD, Пиза университетінің профессоры (Италия)

Отелбаев Мухтарбай Отелбасвич — физика-математика ғылымдарының докторы, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Дайнеко Евгения Александровна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің Жаһандық серіктестік және қосымша білім беру жөніндегі проректоры (Қазақстан)

Дузбаев Нуржан Токсуажавич — PhD, Халықаралық ақпараттық технологиялар университетінің Цифрландыру және инновациялар жөніндегі проректоры (Қазақстан)

Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының профессоры (Қазақстан)

Сейлова Нүргүл Абдуллаевна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Компьютерлік технологиялар және киберқауіпсіздік» факультетінің деканы (Қазақстан)

Мухамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Цифрлық трансформациялар» факультетінің деканы (Қазақстан)

Ыдырыс Айжан Жұмабайқызы — PhD, Халықаралық ақпараттық технологиялар университетінің «Математикалық және компьютерлік модельдеу» кафедрасының меңгерушісі (Қазақстан)

Шильдибеков Ерлан Жаржанович — PhD, Халықаралық ақпараттық технологиялар университетінің «Экономика және бизнес» кафедрасының меңгерушісі (Қазақстан)

Аманжолова Сауле Токсановна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Киберқауіпсіздік» кафедрасының меңгерушісі (Қазақстан)

Ниязгулова Айгүл Аскарбековна — филология ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Медиакоммуникациялар және Қазақстан тарихы» кафедрасының меңгерушісі (Қазақстан)

Айтмағамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Радиотехника, электроника және телекоммуникация» кафедрасының профессоры (Қазақстан)

Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университетінің қауымдастырылған профессоры (Қазақстан)

Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры (Қазақстан)

Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

Тадеш Валлас — PhD, Адам Мицкевич атындағы университеттің проректоры (Польша)

Мамырбаев Өркен Жұмажанұлы — Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялары институты директорының ғылым жөніндегі орынбасары (Қазақстан)

Бушуев Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның «УКРНЕТ» жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сәулет университетінің «Жобаларды басқару» кафедрасының меңгерушісі (Украина)

Белошицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Астана IT университетінің деректер жөніндегі есептеу және ғылым кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Ералы Диана Русланқызы — «Халықаралық ақпараттық технологиялар университеті» АҚ (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Меншіктенуші: «Халықаралық ақпараттық технологиялар университеті» АҚ (Алматы қ.)

Қазақстан Республикасы Ақпарат және әлеуметтік даму министрлігінің Ақпарат комитетінде – **20.02.2020** жылы берілген.

№ KZ82VPY00020475 мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: ақпараттық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологияларға арналған.

Мерзімділігі: жылына 4 рет.

Тиражы: 100 дана

Редакцияның мекенжайы: 050040, Алматы қ-сы, Манас к-сі, 34/1, 709-кабинет, тел: +7 (727) 244-51-09).

E-mail: ijict@iitu.edu.kz

Журнал сайты: <https://journal.iitu.edu.kz>

© Халықаралық ақпараттық технологиялар университеті АҚ, 2022

© Авторлар ұжымы, 2022

ГЛАВНЫЙ РЕДАКТОР:

Хикметов Аскар Кусупбекович — кандидат физико-математических наук, председатель правления - ректор Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

УЧЕНЫЙ СЕКРЕТАРЬ:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Лучино Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Дайнеко Евгения Александровна — PhD, ассоциированный профессор, проректор по глобальному партнерству и дополнительному образованию Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета цифровых трансформаций Международного университета информационных технологий (Казахстан)

Ыдырыс Айжан Жумабаевна — PhD, ассистент профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шилдибеков Ерлан Жаржанович — PhD, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Аманжолова Сауле Токсановна — кандидат технических наук, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, заведующая кафедрой медиакоммуникаций и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Zufарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Алмисреб Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас — PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белоощипская Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Ералы Диана Русланқызы — АО «Международный университет информационных технологий» (Казахстан).

Международный журнал информационных и коммуникационных технологий

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № КЗ82VPY00020475, выданное от 20.02.2020 г.

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем.

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

E-mail: ijict@iitu.edu.kz

Сайт журнала: <https://journal.iitu.edu.kz>

© АО Международный университет информационных технологий, 2022

© Коллектив авторов, 2022

EDITOR-IN-CHIEF:

Khikmetov Askar Kusupbekovich — Candidate of Physical and Mathematical Sciences, Chairman of the Board, Rector of International Information Technology University (Kazakhstan)

DEPUTY CHIEF DIRECTOR:

Kolesnikova Katerina Viktorovna — Doctor of Technical Sciences, Vice-Rector of Information Systems Department, International Information Technology University (Kazakhstan)

SCIENTIFIC SECRETARY:

Ipalakova Madina Tulegenovna — Candidate of Technical Sciences, Associate Professor, Director of the Research Department, International University of Information Technologies (Kazakhstan)

EDITORIAL BOARD:

Razaq Abdul — PhD, Professor of International Information Technology University (Kazakhstan)

Lucio Tommaso de Paolis — Director of Research and Development, AVR Laboratory, Department of Innovation and Process Engineering, University of Salento (Italy)

Liz Bacon — Professor, Deputy Director, and Deputy Vice-Chancellor of the University of Abertay. (Great Britain)

Michele Pagano — Ph.D., Professor, University of Pisa (Italy)

Otelbaev Mukhtarbay Otelbayuly — Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of the Republic of Kazakhstan, Professor of the Department of Mathematical and Computer Modeling of International Information Technology University (Kazakhstan)

Rysbayuly Bolatbek — Doctor of Physical and Mathematical Sciences, Professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Daineko Yevgeniya Alexandrovna — PhD, Associate Professor, Vice-Rector for Global Partnership and Continuing Education, International Information Technology University (Kazakhstan)

Duzbaev Nurzhan Tokkuzhaevich — Candidate of Technical Sciences, Vice-Rector for Digitalization and Innovations, International Information Technology University (Kazakhstan)

Sinchev Bakhtgeray Kuspanuly — Doctor of Technical Sciences, Professor of the Department of Information Systems, International Information Technology University (Kazakhstan)

Seilova Nurgul Abdullaevna — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Mukhamedieva Ardak Gabitovna — Candidate of Economic Sciences, Dean of the Faculty of Digital Transformations, International Information Technology University (Kazakhstan)

Idyrys Aizhan Zhumabaevna — PhD, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Shildibekov Yerlan Zharzhanuly — PhD, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Amanzholova Saule Toksanovna — Candidate of Technical Sciences, Head of the Department of Cyber Security, International Information Technology University (Kazakhstan)

Niyazgulova Aigul Askarbekovna — Candidate of Philology, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Aitmagambetov Altai Zufarovich — Candidate of Technical Sciences, Professor of the Department of Radioengineering, Electronics and Telecommunication, International Information Technology University (Kazakhstan)

Almisreb Ali Abd — PhD, Associate Professor, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, Associate Professor, Department of Information systems, International Information Technology University (Kazakhstan)

Young Im Choo — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, University of Dr. Litt Adam Miskevich in Poznan (Poland)

Mamyrbayev Orken Zhumazhanovich — PhD in Information Systems, Deputy Director for Science, Institute of Information and Computing Technologies CS MSHE RK (Kazakhstan)

Bushuyev Sergey Dmitriyevich — Doctor of Technical Sciences, Professor, Director of Удoктoр тeхнических наук, профессор, директор Ukrainian Association of Project Management UKRNET, Head of Project Management Department, Kyiv National University of Construction and Architecture (Ukraine)

Beloshitskaya Svetlana Vasilyevna — Doctor of Technical Sciences, Associate Professor, Professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

EXECUTIVE EDITOR

Eraly Diana Ruslankyzy — International Information Technology University (Kazakhstan)

«International Journal of Information and Communication Technologies»

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Owner: International Information Technology University JSC (Almaty).

The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.

Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies

Periodicity: 4 times a year.

Circulation: 100 copies.

Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz

Journal website: <https://journal.iitu.edu.kz>

© International Information Technology University JSC, 2022

© Group of authors, 2022

МАЗМҰНЫ

БАҒДАРЛАМАЛЫҚ ҚАМТАМАНЫ ӨЗІРЛЕУ ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ

Жақсылық Г.Б., Пашенко Г.Н.

МЕДИЦИНАЛЫҚ МЕКЕМЕҢІҢ АҚПАРАТТЫҚ ЖҮЙЕСІН ЖАСАУ
ЖӘНЕ ЗЕРТТЕУ.....8

Тукенова Г.С.

ТЕЛЕКОММУНИКАЦИЯ ИНДУСТРИЯСЫНДА БҰЗЫЛУДЫ БОЛЖАУ ҮШІН
МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ.....19

Буравов А.А., Дузбаев Н.Т.

ПРАКТИКАЛЫҚ ТАПСЫРМАЛАРДЫ АВТОМАТТЫ ТЕКСЕРУ ЖӘНЕ
ОНЛАЙН ОҚЫТУ ТӘСІЛДЕРІ.....26

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУІПСІЗДІК

Жұматай Н.Е.

STARLINK ГЕОСТАЦИОНАРЛЫ ЕМЕС ЖЕРІК ЖЕЛІЛЕРІНІҢ KAZSAT-2
ГЕОСТАЦИОНАРЛЫ СПУТНИКТИК ЖЕЛІСІНЕ ӨСЕРІН ТАЛДАУ.....37

Абдуллаева А.С., Әйтiм Ә.Қ., Тян А.В.

4G ЖЕЛІСІН 5G-ГЕ КӨШІРУ. 5G ЭКОЖҮЙЕСІНІҢ ИННОВАЦИЯЛЫҚ
ӨЛЕУЕТІ.....47

Намиялы А.Е., Валиев Б.Б., Сагымбекова А.О., Әділ А.Ж.

КИБЕРҚАУІПСІЗДІКТІ ЗЕРТТЕУ ҮШІН СЕНТИМЕНТАЛДЫ ТАЛДАУДЫ
ҚОЛДАНУ.....59

ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР

Абдуллаева А.С., Тян А.В., Айтим А.К.

ЛОГИСТИКАЛЫҚ ЖҮЙЕНІҢ ҚАЖЕТТІЛІГІН ТАЛДАУ ЖӘНЕ ТИІМДІЛІКТІ
АРТТЫРУ МАҚСАТТАРЫН БЕЛГІЛЕУ.....67

Әйтiм Ә.Қ.

СЕМАНТИКАЛЫҚ ІЗДЕУ НӘТИЖЕЛЕРІН ЖЕТІЛДІРУ ҮШІН ТАБИҒИ
ТІЛДЕРДІ ӨҢДЕУ МОДЕЛДЕРІ.....82

МАТЕМАТИКАЛЫҚ ЖӘНЕ КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ

Қадырбаева Ж.М., Абилкаир Д.С., Масалимов Б.С.

ҮШ НҮКТЕЛІ ШАРТЫ БАР ЕЛЕУЛІ ТҮРДЕ ЖҮКТЕЛГЕН ДИФФЕРЕНЦИАЛДЫҚ
ТЕҢДЕУЛЕР ЖҮЙЕСІНІҢ САНДЫҚ ШЕШІМІ ТУРАЛЫ.....92

Сүлейменова А.Р., Саябаева А.Ж., Молдагулова А.Н.

ҚАРЖЫ САЛАСЫНДАҒЫ ТӘУЕКЕЛДЕРДІ ТАЛДАУ ӘДІСТЕРІН ҚАРЖЫ
САЛАСЫНДАҒЫ ҚАРЖЫЛЫҚ БЫҚТИМАЛДЫҚ ҮЛГІЛЕРІН
ПАЙДАЛАНҒАН ЗЕРТТЕУ.....103

СОДЕРЖАНИЕ

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

Жаксылык Г.Б., Пашенко Г.Н.

РАЗРАБОТКА И ИССЛЕДОВАНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ
ДЛЯ МЕДИЦИНСКОГО УЧРЕЖДЕНИЯ.....8

Туменова Г.С.

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗА ОТТОКА
В ТЕЛЕКОММУНИКАЦИОННОЙ ОТРАСЛИ.....19

Буравов А.А., Дузбаев Н.Т.

ПОДХОДЫ К АВТОМАТИЧЕСКОЙ ПРОВЕРКЕ ПРАКТИЧЕСКИХ ЗАДАНИЙ
В МООС И ОНЛАЙН-ОБУЧЕНИИ.....26

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

Жұматай Н.Е.

АНАЛИЗ ВЛИЯНИЯ НЕГЕОСТАЦИОНАРНЫХ СПУТНИКОВЫХ СЕТЕЙ
STARLINK НА ГЕОСТАЦИОНАРНУЮ СПУТНИКОВУЮ СЕТЬ KAZSAT-2.....37

Абдуллаева А.С., Айтим А.К., Тян А.В.

ПЕРЕХОД СЕТИ 4G НА 5G. ИННОВАЦИОННЫЙ ПОТЕНЦИАЛ
ЭКОСИСТЕМЫ 5G.....47

Намиялы А.Е., Валиев Б.Б., Сагымбекова А.О., Әділ А.Ж.

ИСПОЛЬЗОВАНИЕ АНАЛИЗА ТОНАЛЬНОСТИ ДЛЯ ИЗУЧЕНИЯ
КИБЕРБЕЗОПАСНОСТИ.....59

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

Абдуллаева А.С., Тян А.В., Айтим А.К.

АНАЛИЗ НЕОБХОДИМОСТИ ЛОГИСТИЧЕСКОЙ СИСТЕМЫ
И ПОСТАНОВКА ЗАДАЧ ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ.....67

Айтим А.К.

МОДЕЛИ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ УЛУЧШЕНИЯ
СЕМАНТИЧЕСКИХ РЕЗУЛЬТАТОВ ПОИСКА.....82

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Кадирбаева Ж.М., Абилкаир Д.С., Масалимов Б.С.

О ЧИСЛЕННОМ РЕШЕНИИ СИСТЕМ СУЩЕСТВЕННО НАГРУЖЕННЫХ
ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ТРЕХТОЧЕЧНЫМ УСЛОВИЕМ.....92

Сулейменова А.Р., Саябаева А.Ж., Молдагулова А.Н.

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА РИСКА С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ
ВЕРОЯТНОСТИ ДЕФолТА В ФИНАНСОВОЙ ОТРАСЛИ.....103

CONTENTS

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

Zhaksylyk G.B., Pachshenko G.N.

DEVELOPMENT AND RESEARCH OF INFORMATION SYSTEM FOR
A MEDICAL INSTITUTION.....8

Tukenova G.S.

USING MACHINE LEARNING FOR CHURN PREDICTION IN THE
TELECOM INDUSTRY.....19

Buravov A.A., Duzbayev N.T.

APPROACHES TO AUTOMATIC CHECKING OF PRACTICAL ASSIGNMENTS
IN MOOCS AND ONLINE LEARNING.....26

INFORMATION AND COMMUNICATION NETWORKS AND CYBERSECURITY

Zhumatay N.E.

ANALYSIS OF THE IMPACT OF NON-GEOSTATIONARY SATELLITE NETWORKS
STARLINK ON THE GEOSTATIONARY SATELLITE NETWORK KAZSAT-2.....37

Abdullayeva A.S., Aitim A.K., Tyan A.V.

TRANSITION FROM 4G LTE TO 5G. INNOVATIVE POTENTIAL
OF THE 5G ECOSYSTEM.....47

Namiyaly A.E., Valiyev B.B., Sagymbekova A.O., Adil A.Zh.

UTILIZING SENTIMENT ANALYSIS FOR CYBER SECURITY LEARNING.....59

SMART SYSTEMS

Абдуллаева А.С., Тянь А.В., Айтим А.К.

ANALYSIS OF THE NECESSITY OF A LOGISTICS SYSTEM AND SETTING
GOALS TO INCREASE EFFICIENCY.....67

Aitim A.K.

MODELS OF NATURAL LANGUAGE PROCESSING FOR IMPROVING
SEMANTIC SEARCH RESULTS.....82

MATHEMATICAL AND COMPUTER MODELING

Kadimbayeva Zh.M., Abilkair D.S., Massalimov B.S.

ON THE NUMERICAL SOLUTION OF SYSTEMS OF ESSENTIALLY LOADED
DIFFERENTIAL EQUATIONS WITH A THREE-POINT CONDITION.....92

Suleimenova A.R., Sayabayeva A.Zh., Moldagulova A.N.

RESEARCH ON RISK ANALYSIS METHODS USING MODELS OF DEFAULT
PROBABILITY IN THE FINANCIAL INDUSTRY.....103

**БАҒДАРЛАМАЛЫҚ ЖАСАҚТАМАНЫ ӨЗІРЛЕУ
ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ**

**РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ
И ИНЖЕНЕРИЯ ЗНАНИЙ**

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 8–18

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.001>

УДК 681.5

**DEVELOPMENT AND RESEARCH OF INFORMATION
SYSTEM FOR A MEDICAL INSTITUTION**

G.B. Zhaksylyk, G.N. Pachshenko*

Galina N. Pachshenko — Т.Ф.К., associate professor, Information Systems Department, International Information Technology University, +7 777 261 7322

E-mail: gauhario98@gmail.com;

Gaukhar B. Zhaksylyk — master's student, International Information Technology University, +7 775 0654105

E-mail: galina_pashenko@mail.ru.

© G.B. Zhaksylyk, G.N. Pachshenko, 2022

Abstract. The development of medical equipment and technologies in general and medical information systems in particular is associated with changing needs of the healthcare industry, including the steady growth of knowledge in the field of medicine, the complication of methods of examination, diagnosis and treatment. The article discusses the existing medical information systems and highlights the features of these systems. The analysis of processes and analysis of the algorithm of existing information systems is given. This article is devoted to the development of a prototype of a medical information system that allows you to automate relationships within medical institutions. As a result of the analysis, an optimal model of an information system for medical institutions was formed. The description of the developed medical information system is given.

Keywords: medical information system, information system, public health information management system



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

For citation: G.B. Zhaksylyk, G.N. Pachshenko. Development and research of information system for a medical institution // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 8–18 (In Russ.). DOI: 10.54309/IJICT.2022.10.2.001.

МЕДИЦИНАЛЫҚ МЕКЕМЕНІҢ АҚПАРАТТЫҚ ЖҮЙЕСІН ЖАСАУ ЖӘНЕ ЗЕРТТЕУ

Г.Б. Жақсылық, Г.Н. Пащенко*

Пащенко Галина Николаевна — Candidate of Technical Sciences, ассоциацияланған профессор ақпараттық жүйелер кафедрасы Халықаралық ақпараттық технологиялар университеті, +7 777 261 7322

E-mail: gauhario98@gmail.com;

Жақсылық Гаухар Бакытжанқызы — магистрант, Халықаралық ақпараттық технологиялар университеті, +7 775 065 4105

E-mail: galina_pashenko@mail.ru.

© Г.Б. Жақсылық, Г.Н. Пащенко, 2022

Аннотация. Жалпы медициналық техника мен технологиялардың және әсіресе медициналық ақпараттық жүйелердің дамуы денсаулық сақтау саласының қажеттіліктерінің өзгеруімен, соның ішінде медицина саласындағы білімнің тұрақты өсуімен, тексеру, диагностикалау және емдеу әдістерінің күрделенуімен байланысты. Мақалада қолданыстағы медициналық ақпараттық жүйелер талқыланып, осы жүйелердің ерекшеліктері көрсетілген. Процестерді талдау және қолданыстағы ақпараттық жүйелердің алгоритмін талдау берілген. Бұл мақала медициналық мекемелердегі қарым-қатынастарды автоматтандыруға мүмкіндік беретін медициналық ақпараттық жүйенің прототипін әзірлеуге арналған. Талдау нәтижесінде медициналық мекемелердің ақпараттық жүйесінің оңтайлы моделі қалыптастырылды. Жасалған медициналық ақпараттық жүйенің сипаттамасы берілген.

Түйін сөздер: медициналық ақпараттық жүйе, ақпараттық жүйе, қоғамдық денсаулық сақтау ақпаратын басқару жүйесі

Дәйексөз үшін: Г.Б. Жақсылық, Г.Н. Пащенко. Медициналық мекеменің ақпараттық жүйесін жасау және зерттеу // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 8–18 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.001.



РАЗРАБОТКА И ИССЛЕДОВАНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ МЕДИЦИНСКОГО УЧРЕЖДЕНИЯ

Г.Б. Жаксылык, Г.Н. Пащенко*

Пащенко Галина Николаевна — кандидат технических наук, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий, +7 777 261 7322

E-mail: gauhario98@gmail.com;

Жаксылык Гаухар Бакытжанкызы — магистрант Международного университета информационных технологий, +7 775 065 4105

E-mail: galina_pashenko@mail.ru.

© Г.Б. Жаксылык, Г.Н. Пащенко, 2022

Аннотация. Развитие медицинской техники и технологий в целом и медицинских информационных систем в частности связано с изменением потребностей отрасли здравоохранения, включая неуклонный рост знаний в области медицины, усложнение методов обследования, диагностики и лечения. В статье рассматриваются существующие медицинские информационные системы и выделяются особенности данных систем. Приводится анализ процессов и разбор алгоритма существующих информационных систем. Данная статья посвящена разработке прототипа медицинской информационной системы, позволяющей автоматизировать взаимоотношения внутри медицинских учреждений. В результате анализа сформирована оптимальная модель информационной системы для медицинских учреждений. Приводится описание разработанной медицинской информационной системы.

Ключевые слова: медицинская информационная система, информационная система, система управления информацией общественного здравоохранения

Для цитирования: Г.Б. Жаксылык, Г.Н. Пащенко. Разработка и исследование информационной системы для медицинского учреждения // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 8–18 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.001.

Введение

Особенностью основных услуг общественного здравоохранения является предоставление точных и полных медицинских услуг на протяжении всей жизни для жителей или мигрантов в юрисдикции, в то время как информационные технологии могут сделать услуги общественного здравоохранения более эффективными и полными. С созданием системы управления информацией в области общественного здравоохранения электронные медицинские карты жителей и различные последующие данные могут управляться и использоваться более эффективно; таким образом, санитарное обслуживание первичных медицинских



учреждений улучшится в дальнейшем, что может облегчить административному отделу здравоохранения вести статистику и анализ состояния обслуживания различных первичных медицинских учреждений.

Медицинская информационная система представляет собой автоматизированную систему, целью которой является автоматизация всех задач, выполняемых в медицинских учреждениях. Система позволяет автоматизировать широкий спектр действий, таких как: запись на прием, уведомления о записи, создание разного рода отчетов, контроль за лечением пациентов.

Использование медицинских информационных систем позволяет:

- Рационально использовать и планировать рабочее время обслуживающего персонала.
- Полностью избавиться от бумажной документации, что позволило снизить количество врачебных ошибок.
- Повысить скорость оказания медицинской помощи за счет оперативного доступа к информации о пациенте.
- Исключить дублирование информации и направлений исследований.
- Планировать нагрузку и распределение пациентов по медицинскому оборудованию.

Методы исследования

В статье используется метод сравнительного анализа, применяемый к существующим медицинским информационным системам. Сравнение информационных систем проводится по нескольким критериям. Полученные данные анализируются и выделяются особенности данных систем. Конечным итогом анализа является сформированный список необходимых функциональных требований к разрабатываемой медицинской информационной системе.

Результаты исследования

В первичных медицинских учреждениях ведущие технологии в процессах исследований, разработки и внедрения системы управления информацией службы общественного здравоохранения включают электронные медицинские карты, технологию виртуальной частной сети и технологию резервного копирования в реальном времени для хранения данных, которые могут помочь системе реализовать свои жизненные функции и обеспечить штатную работу системы (Жуманбаева и др., 2020: 25–32).

Информационные системы здравоохранения могут использоваться всеми работниками здравоохранения, от пациентов до медсестёр, врачей и должностных лиц общественного здравоохранения. Они собирают данные и обрабатывают их таким образом, чтобы их можно было использовать для принятия решений в области здравоохранения.

Этапы развития МИС и анализ

Все этапы развития МИС приведены в Таблице 1:

Таблица 1 - Развитие медицинских информационных систем в мире

Годы, автор, научное сообщество	События
1959г. Ледли, американский учёный, профессор физиологии и биофизики	Опубликовал статью о диагностическом процессе принятия решений в журнале Science, провел анализ методов обработки принятия решений в условиях риска и неопределенности, первым предложил всестороннее обсуждение поддержки принятия медицинских решений на компьютерной основе
1960-е гг.	Разработаны больничные информационные системы на основе ЭВМ, которые могли бы интегрировать сведения о пациентах в базу данных
1965 г. Больница Эль-Камино, Калифорния	Создание одной из первых клинически-ориентированных информационных систем здравоохранения Technicon Medical Information System (TMIS)
1967 г.	Создание первой медицинской информационной системы для интеграции клинических накопленных данных в систему поддержки принятия клинических решений Health Evaluation through Logical Processing (HELP)
1967 г.	Основана Международная ассоциация медицинской информатики, МАМИ (International Medical Informatics Association, IMIA) – независимая организация, тесно взаимодействующая с Всемирной организацией здравоохранения (World Health Organization, WHO). Сейчас МАМИ играет глобальную роль в приложении информатики и технологий в здравоохранении. Ассоциация охватывает все континенты земного шара, более 50 научных организаций, свыше 50 000 человек
1968 г. Окто Барнетт, признанный основатель медицинской информатики, Лаборатория кибернетики Массачусетского госпиталя	Создана система электронных историй болезни COSTAR (Computer Stored Ambulatory Record)
1970-е гг.	Использование мини-ЭВМ позволило вводить сведения о пациентах в истории болезни и сохранять их. Создание системы Decentralized Hospital Computer Program (DHCP)
1976 г. Лоуренс Вид и Ян Шульц, Медицинский центр в университете в Вермонте	Создана PROMIS (Problem-Oriented Medical Information System) – компьютеризованная система продвинутого подхода к клиническому обучению и уходу за пациентами
1980 г. Эдвард Шортлиф, американский ученый, пионер в области использования искусственного интеллекта в медицине	Создал первую образовательную программу в области биомедицинской информатики в Стэндфордском университете
1980-е гг.	Подсистемы крупных медицинских информационных систем были интегрированы в центральную систему управления баз данных
1986 г.	Создано Европейское общество искусственного интеллекта в медицине (European Society for Artificial Intelligence in Medicine, AIME)
1990-е гг.	Спектр услуг ухода за пациентами был расширен по объему и сложности, новые информационные системы, предлагаемые больницам, обладали дополнительными возможностями. Влияние Интернета расширило глобальный обмен клиническими данными и медицинскими знаниями



1990 г.	МАМИ опубликовала первые международные рекомендации по образованию в области медицинской информатики. Эти рекомендации были широко использованы и переведены на многие языки. Данный документ является явным признаком международного присутствия и зрелости медицинской информатики как научной дисциплины
2000-е гг.	Распределенные информационные системы позволили врачам использовать автоматизированные рабочие места для введения и обработки результатов анализов по всей базе данных медицинского учреждения
2010-е гг.	Всемирная беспроводная связь с применением облачных технологий позволила объединить хранилища данных медицинских центров на национальном уровне; появилась мобильная медицинская помощь mobile e-health care

Примеры информационных систем здравоохранения включают:

Электронная медицинская карта — это цифровая запись медицинской информации. Он содержит всю информацию, которую вы найдете в бумажной диаграмме, и многое другое. Электронная медицинская карта может включать прошлую историю болезни, жизненные показатели, записи о прогрессе, диагнозы, лекарства, даты иммунизации, аллергии, лабораторные данные и отчеты о визуализации. Он также может содержать другую важную информацию, например информацию о страховании, демографические данные и даже данные, импортированные с персональных оздоровительных устройств.

Программное обеспечение для управления практикой. Программное обеспечение для управления практикой помогает работникам сферы администрирования в медицинских услугах управлять повседневными операциями, такими как планирование и выставление счетов. Системы управления необходима для автоматизации многих административных задач.

Программное обеспечение главного указателя пациентов (Master Patient Index – MPI) относится к основному указателю пациентов, реестру пациентов и реестру клиентов. Это система электронной базы данных, которая содержит демографическую информацию о пациентах. Преимущества MPI заключаются в устранении дублирования регистрационных записей пациентов и ведении центрального реестра всех пациентов.

Порталы для пациентов. Порталы пациентов позволяют пациентам получать доступ к своим личным медицинским данным, таким как информация о приеме, лекарствах и результатах лабораторных исследований, через подключение к Интернету. Некоторые порталы для пациентов обеспечивают активное общение со своими врачами, запросы на пополнение рецепта и возможность записи на приём.

Удаленный мониторинг пациента (Remote patient monitoring – RPM) — это подкатегория телемедицины для ухода на дому, которая позволяет пациентам использовать мобильные медицинские устройства и технологии для сбора данных о состоянии здоровья, сгенерированных пациентом и отправки их медицинским работникам. Общие физиологические данные, которые можно

собрать с помощью программ RPM, включают показатели жизненно важных функций, вес, артериальное давление и частоту сердечных сокращений. После сбора данные пациента отправляются в кабинет врача с помощью специальной компьютерной системы телемедицины или программного приложения, которое можно установить на компьютер, смартфон или планшет.

Архитектура информационной системы — это понятие, определяющее модель, структуру, выполняемые функции и взаимосвязь компонентов информационной системы (Pachshenko, 2019: 29–35). В такой централизованной архитектуре все компоненты системы расположены в централизованном ядре системы.

Работа и доступ с прикладным ПО осуществляется по модели SaaS. В этом случае вам нужно только браузер для работы с программой (Айтбекова и др., 2021: 25–35). В лечебном учреждении будет построена медицинская сеть, к которой рабочие места различных специалистов, подключено различное диагностическое оборудование. Данные загружаются в централизованное хранилище ядра. Такая архитектура не обрабатывает медицинские данные. Возможно предоставление доступа и работы с программным обеспечением авторизованных пользователей и через мобильные устройства. Взаимодействие с другими информационными системами.

После проведенного сравнительного анализа следует выделить следующие основные модули и функции, которые следует внедрить в разрабатываемую медицинскую информационную систему. Автоматизированные системы необходимы для защиты данных от несанкционированной передачи третьим лицам (Имангалиева и др., 2021: 47–54).

Информационная система состоит из трех больших модулей: поликлиника, стационар и диагностические подразделения (рис. 1). Данные модули выделены по принципу прохождения лечения пациентом. Больной может приходить на амбулаторное лечение в поликлинику, а при необходимости лечь в стационарное отделение. В итоге мы получаем два связанных, но непересекающихся этапа лечения (так как пациент не может одновременно находиться на амбулаторном и стационарном лечении), что удобно для моделирования двух подсистем. При переносе системы на поликлиники или же больницы без амбулаторного этапа лечения подобное разделение поможет легко отделить ненужные модули и масштабировать систему.

Некоторые диагностические подразделения могут работать без прямого контакта с пациентами (например, клиничко-биохимическая лаборатория может работать с анализами, пришедшими из других медицинских учреждений). Данная возможность реализована с помощью отдельных интерфейсной части и набора сущностей, взаимодействующих с остальной системой как внешний блок. Эта заявочная система может размещаться как на внутреннем, так и на внешнем сервере, что позволяет дать доступ к ней максимальному количеству медицинских учреждений. К тому же немногие медицинские центры и больницы имеют диагностические блоки в своем составе, так что выделение этих подразделений в отдельный модуль является естественным.



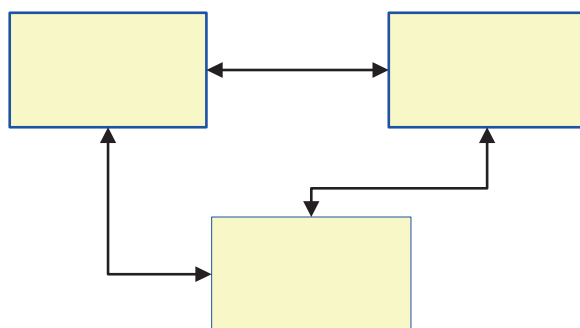


Рисунок 1 - Основные модули системы

При рассмотрении взаимодействия модулей можно выделить также приемное отделение (рис. 2), которое не имеет прямого отношения к лечебному процессу, а осуществляет прием населения и регистрирует все перемещения пациентов в ходе лечебного процесса. Например, на амбулаторном этапе пациента могут направить на госпитализацию. А после госпитализации порекомендовать процедуры для скорейшего восстановления здоровья, которые нужно проходить в поликлинике. Все переходы назначения и действия врачей регистрируются системой.

Конечно, модульность не должна означать разделение информации о пациенте. Разделения на модули подразумевают лишь минимизацию связей между ними, для возможности оперировать ими отдельно.

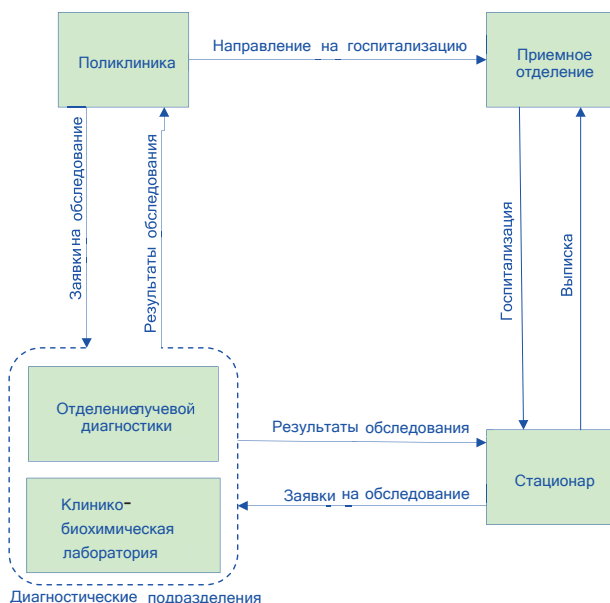


Рисунок 2 - Взаимодействие модулей

Идентификация, аутентификация и контроль доступа субъектов:

- В систему.
- В компьютерные терминалы.
- В программы.
- В тома, поля, записи.

Регистрация и учет:

- Вход (выход) в (из) системы.
- Выдача печатных (графических) выходных документов.
- Запуск (завершение) программ и процессов (задач).
- Доступ программ и субъектов доступа к защищенным файлам, включая их создание и удаление, передачу внутри системы.
- Программы доступа субъектов доступа к терминалам, компьютерам, узлам компьютерной сети, связи, каналам, внешним компьютерным устройствам, программам, томам, каталогам, файлам, записям, полям записи.
- Очистка (обнуление, обезличивание) освободившихся областей оперативной памяти компьютеров и внешних накопителей.

Подсистема целостности:

- Обеспечение целостности программного обеспечения и обрабатываемой информации.
- Физическая безопасность компьютерного оборудования и носителей информации.
- Периодическое тестирование системы защиты информации от несанкционированного доступа.
- Наличие средств восстановления системы защиты информации от несанкционированного доступа.

Кроме внедрения вышеописанных функций, необходимо учесть технологии, описанные ниже.

Электронные медицинские карты (EHR): Электронная медицинская карта относится к сбору всей личной медицинской информации и необходимых личных данных пациента. EHR использует современные компьютеры для сохранения и использования информации. Наиболее важной целью EHR является обмен информацией. Во-вторых, он защищает данные от потери и сохраняет точность данных, которые не могут быть легко изменены.

Технология виртуальной частной сети (VPN): технология VPN может легко подключить любые медицинские учреждения, учреждения общественного здравоохранения и системы общественного здравоохранения, к которым необходимо получить доступ в этом районе или даже за его пределами. Пользователи в локальной интрасети используют интранет для подключения, а удаленные пользователи полагаются на Интернет для

доступа к интрасети, используя технологию VPN для обеспечения информационного взаимодействия.

Хранение данных и технология синхронизации в реальном времени: ядром общедоступного хранилища медицинских данных является синхронизация. Когда основному модулю сохранения необходимо изменить сохраненные данные системы, данные будут сохранены в кэш-модуле, а затем изменены (Pachshenko, 2019: 29–35). Он не хранится непосредственно на диске; в то же время модуль сохранения резервной копии отправляет сообщение, которое можно изменить. После изменения данных в хранилище резервных копий основное хранилище и хранилище резервных копий одновременно вводят данные на диск.

Обсуждение результатов

Разрабатываемая система для медицинских учреждений учитывает все вышеперечисленные особенности медицинских информационных систем. Основной целью разрабатываемой медицинской информационной системы является повышение качества оказываемых медицинских услуг за счет:

- Автоматизации деятельности медицинского персонала.
- Полный переход с бумаги на цифру.
- Управленческий надзор за деятельностью медицинских учреждений.
- Внедрение современных информационно-коммуникационных систем.

Как и любые инновации, внедряемые в повседневную жизнь, медицинские информационные системы имеют ряд преимуществ, помогающих правильно организовать работу на разных уровнях медицины.

Для персонала:

- Оперативное получение данных о результатах диагностических и лабораторных исследований.
- Техническая поддержка схем лечения, включая диагностические, лечебные и лечебные рекомендации.
- Сбор и обработка заявок на диагностические исследования и медицинские приемы.
- Контроль выполнения медицинских назначений медицинским персоналом.

Для пациентов:

- Повышение качества обслуживания за счет информационного взаимодействия врачей и специалистов, задействованных в лечении.

Заключение

В данной работе проведен анализ существующих медицинских информационных систем. Также были детально рассмотрены функции, которые необходимо внедрить в медицинскую информационную систему. Разработанная информационная система будет полезна для медицинских

сотрудников и пациентов. Система имеет преимущества перед приведёнными выше аналогами и имеет оптимизированную систему доступа к данным. Разработанная информационная система поможет значительно повысить качество предоставляемых услуг в сфере медицины.

ЛИТЕРАТУРЫ

Andrikov D.A., Kuchin A.S. (2021). Development of a prototype of a medical information system for a clinical diagnostic center, *Procedia Computer Science*. — Volume 186. — 2021. — Pp 287–292.

Айтбекова М.Б., Пашченко Г.Н. (2021). Разработка информационной системы для расчетов рейтингов успеваемости студентов // *International journal of information and communication technologies*. — Том 2. — Выпуск 1. — Алматы. — март 2021. — С. 25–35.

Жуманбаева С.К., Пашченко Г.Н. (2020). Разработка информационной системы для обработки научных трудов // *Новости науки Казахстана, научно-технический журнал*, № 1 (143). — Алматы. — 2020. — С. 25–32.

Имангалиева А.А., Пашченко Г.Н. (2021). Проектирование и разработка информационной системы для управления научно-образовательной деятельностью университета // *International journal of information and communication technologies*. — Том 2. — Выпуск 1. — Алматы. — март 2021. — С. 47–54.

Pachshenko G. (2019). Statistical methods for business and enterprise control. // *Новости науки Казахстана*. — Алматы, 2019. — № 3. — С. 29–35.

Pachshenko G. Statistical methods for business and enterprise control // *Новости науки Казахстана*. — Алматы, 2019. — № 3. — С. 29–35.

REFERENCES

Aitbekova M.B., Pashchenko G.N. (2021). Development of an information system for calculating student performance ratings // *International journal of information and communication technologies*. — Volume 2. — Issue 1. — Almaty. — March 2021. — Pp. 25–35.

Andrikov D.A., Kuchin A.S. (2021). Development of a prototype of a medical information system for a clinical diagnostic center, *Procedia Computer Science*. — Volume 186. — 2021. — Pp 287–292.

Imangalieva A.A., Pashchenko G.N. (2021). Design and development of an information system for managing the scientific and educational activities of the university // *International journal of information and communication technologies*. — Volume 2. — Issue 1. — Almaty. — March 2021. — Pp. 47–54.

Zhumanbaeva S.K., Pashchenko G.N. (2020). Development of an information system for processing scientific papers // *Science News of Kazakhstan, scientific and technical journal*. — No. 1 (143). — Almaty, 2020. — Pp. 25–32.

Pachshenko G. (2019). Statistical methods for business and enterprise control. // *Новости науки Казахстана*. — Алматы, 2019. — № 3. — Pp. 29–35.

Pachshenko G. Statistical methods for business and enterprise control // *Новости науки Казахстана*. — Алматы, 2019. — № 3. — Pp. 29–35.



INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 19–25

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.002>

УДК 004.896

USING MACHINE LEARNING FOR CHURN PREDICTION IN THE TELECOM INDUSTRY

G.S. Tukenova, A.N. Moldagulova*

Moldagulova Aiman Nikolaevna — Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Software Engineering of the Kazakh National Research Technical University named after K.I. Satpayev;

Tukenova Gulmira Symbatkyzy — second-year master student of the specialty "Business Analytics" of the International University of Information Technologies.

© G.S. Tukenova, A.N. Moldagulova, 2022

Abstract. This article describes how effectively a deep learning approach can be used for the churn forecasting process in the telecommunications industry with greater accuracy and less processing time. It is observed that data mining systems are gradually succeeding in predicting customer churn over the previous couple of years. Developing a powerful churn forecasting model is a critical task that involves a lot of research directly from recognizable proof of ideal performance, from the vast amount of customer information available to the selection of a successful information mining system that matches the list of possibilities.

Keywords: churn prediction, data mining, machine learning, big data, data processing.

For citation: G.S. Tukenova, A.N. Moldagulova. Using machine learning for churn prediction in the telecom industry // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 19–25 (In Russ.). DOI: *10.54309/IJICT.2022.10.2.002*.

ТЕЛЕКОММУНИКАЦИЯ ИНДУСТРИЯСЫНДА БҰЗЫЛУДЫ БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ

Г.С. Туkenова, А.Н. Молдагулова*

Молдағұлова Айман Николаевна — физика-математика ғылымдарының кандидаты, доцент, Қ.И.Сәтбаев атындағы Қазақ ұлттық ғылыми-зерттеу техникалық университетінің бағдарламалық қамтамасыз ету инженериясы кафедрасының меңгерушісі;



Тукунова Гүлмира Сымбатқызы — Халықаралық ақпараттық технологиялар университетінің «Бизнес-аналитика» мамандығының екінші курс магистранты.

© Г.С. Тукунова, А.Н. Молдагулова, 2022

Аннотация. Бұл мақалада телекоммуникация индустриясында үлкен дәлдікпен және аз өндеу уақытымен жұмыссыздықты болжау процесі үшін терең оқыту тәсілін қаншалықты тиімді пайдалануға болатыны сипатталған. Деректерді іздеу жүйелері соңғы екі жыл ішінде тұтынушылардың жоғалуын болжауда біртіндеп табысқа жеткені байқалады. Кепілісті болжаудың қуатты моделін әзірлеу — мінсіз өнімділіктің танылатын дәлелінен бастап, тұтынушы ақпаратының үлкен көлемінен мүмкіндіктер тізіміне сәйкес келетін табысты ақпаратты іздеу жүйесін таңдауға дейін тікелей көптеген зерттеулерді қамтитын маңызды міндет.

Түйін сөздер: бұзылуды болжау, деректерді өндіру, машиналық оқыту, үлкен деректер, деректерді өндеу

Дәйексөз үшін: Г.С. Тукунова, А.Н. Молдагулова. Телекоммуникация индустриясында бұзылуды болжау үшін машиналық оқытуды пайдалану // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 19–25 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.002

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗА ОТТОКА В ТЕЛЕКОММУНИКАЦИОННОЙ ОТРАСЛИ

Г.С. Тукунова, А.Н. Молдагулова*

Молдагулова Айман Николаевна — кандидат физико-математических наук, ассоциированный профессор, заведующая кафедрой «Кафедра Программной инженерии» Казахского национально-исследовательского технического университета имени К.И. Сатпаева;

Тукунова Гүлмира Сымбатқызы — магистрант второго курса специальности «Бизнес-аналитика» Международного университета информационных технологий.

© Г.С. Тукунова, А.Н. Молдагулова, 2022

Аннотация. В этой статье описывается насколько эффективно подход глубокого обучения может быть использован для процесса прогнозирования оттока в телекоммуникационной отрасли с большей точностью и меньшим временем обработки. Наблюдается, что системы интеллектуального анализа данных постепенно преуспевают в прогнозировании оттока клиентов за предыдущие пару лет. Разработка мощной модели прогнозирования оттока является критически важной задачей, которая включает в себя множество исследований непосредственно от узнаваемого доказательства идеальных показателей, от огромного объема доступной информации о клиентах до выбора успешной системы интеллектуального анализа информации, соответствующей списку возможностей.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Ключевые слова: прогнозирование оттока, интеллектуальный анализ данных, машинное обучение, большие данные, обработка данных

Для цитирования: Г.С. Туkenова, А.Н. Молдагулова. Использование машинного обучения для прогноза оттока в телекоммуникационной отрасли // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 19–25 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.002.

Введение

Поскольку повседневная жизнь, по-видимому, все больше зависит от использования телекоммуникационных продуктов и услуг, ожидается, что в ближайшие годы мировой рынок телекоммуникаций будет расти феноменальными темпами. Глобальный рынок телекоммуникаций постоянно трансформируется благодаря постоянным инновациям и разработкам, происходящим последовательно и быстрыми темпами. Таким образом, для роста телекоммуникационного рынка лояльность клиента является ключевым фактором. На высоко конкурентном рынке операторы связи могут сохранить свои ценности, установив близость с абонентами, наряду с заранее установленными ожиданиями доверия, стимулирующего лояльность. Удержание существующих клиентов и привлечение новых клиентов — мощное оружие на современном рынке. Однако привлечение новых клиентов обходится дороже, чем удержание существующих. В телекоммуникационном секторе ежедневно генерируется огромный объем данных из-за обширной клиентской базы. Но в настоящее время в данной отрасли наблюдается большой отток клиентов. Это приводит к необходимости эксперта по предметной области для прогнозирования оттока клиентов, а также для выяснения основных причин отказа от услуг телекоммуникаций, что будет основой для работы по удержанию клиента.

Создание эффективной политики удержания является важной задачей для предотвращения оттока клиентов. Зная существенные факторы оттока из данных клиентов, можно повысить производительность, рекомендовать соответствующие рекламные акции группе потенциальных клиентов с оттоком на основе аналогичных моделей поведения и значительно улучшить маркетинговые кампании компании. Различные телекоммуникационные компании предлагают передовую тактику, чтобы на ранней стадии спрогнозировать отток клиентов. Традиционно для прогнозирования оттока клиентов применялись различные типы подходов к машинному обучению, такие как Decision Tree, Random Forest и т.д.

Телекоммуникационные отрасли собирают обширную информацию о клиентах, например, профилирование клиентов, список звонков, используемый трафик, активное время в сети и т.д. Принимая во внимание исторический фон клиентов есть возможность отличить уйдет ли клиент или нет. Была использована идея искусственного интеллекта для реализации ожиданий оттока средств массовой информации. В нижеописанной части кратко описана проделанная работа по прогнозированию оттока.

Сопутствующие работы

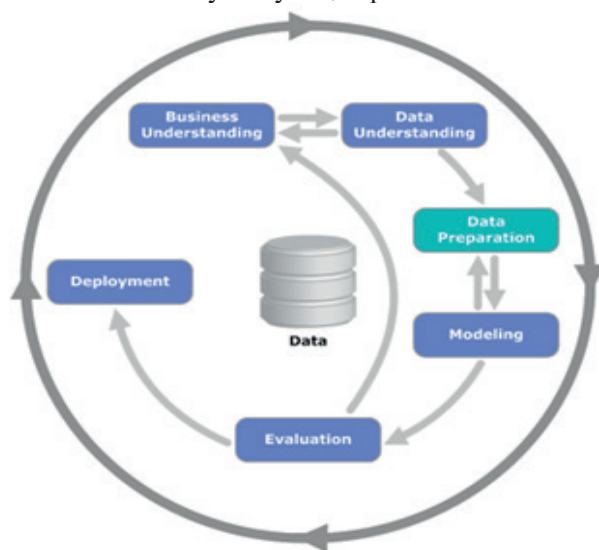


Рисунок 1 - Стандартный процесс интеллектуального анализа данных

Многие подходы применялись для прогнозирования оттока в телекоммуникационных компаниях. В большинстве этих подходов используется машинное обучение и интеллектуальный анализ данных. Большая часть связанной работы была сосредоточена на применении только одного метода интеллектуального анализа данных для извлечения знаний, а другие были сосредоточены на сравнении нескольких стратегий для прогнозирования оттока. На рисунке 1 показан стандартный процесс интеллектуального анализа данных. Было установлено, что эти симуляции оказались успешными в положительном при прогнозировании оттока для телекоммуникационной модели сбора данных (<https://basegroup.ru/community/articles>, 2018). Для отбора дискриминационных элементов использовались методы наименьшего избытка и наиболее экстремальной уместности. Оборудование различных базовых классификаторов подключается в качестве индикаторной процедуры. Для убедительных прогнозов использовались Rotation Forest и KNN, в которых использовалась львиная доля голосов. Аналогичным образом было рассмотрено предсказание оттока для обслуживания клиентов с использованием подхода генетического расчета (Максименко и др., 2017: 134–136). Для каждого класса были созданы различные проекты с использованием стратегии Adaboost. Эти проекты были использованы для прогнозирования с использованием более высокой доходности, исходя из взвешенной общей доходности проектов. Для проверки точности прогноза был использован метод перекрестного утверждения с 10 перекрытиями, и была найдена территория с показателем изгиба 0,89.

Гаврил и др. (Brandusoiu и др., 2020: 97–100) представили передовую методологию интеллектуального анализа данных для прогнозирования оттока

клиентов с предоплатой с использованием набора данных для сведений о звонках 3333 клиентов с 21 функцией и зависимого параметра оттока с двумя значениями: Да/Нет. Некоторые функции включают информацию о количестве входящих и исходящих сообщений и голосовой почты для каждого клиента. Автор применил алгоритм анализа главных компонент «PCA» для уменьшения размерности данных. Для прогнозирования коэффициента оттока использовались три алгоритма машинного обучения: нейронные сети, машина опорных векторов и байесовские сети. Автор использовал AUC для измерения производительности алгоритмов. Значения AUC составляли 99,10 %, 99,55 % и 99,70 % для байесовских сетей, нейронных сетей и метода опорных векторов соответственно. Набор данных, использованный в этом исследовании, небольшой, и пропущенных значений не было.

Различные исследования изучали проблему несбалансированных наборов данных, когда классы ушедших клиентов меньше, чем классы активных клиентов, поскольку это серьезная проблема в проблеме прогнозирования оттока. Амин и др. (Amin и др., 2020: 7940–57) сравнили шесть различных методов выборки для избыточной выборки в отношении проблемы прогнозирования оттока клиентов в телекоммуникациях. Результаты показали, что алгоритмы (MTDF и генерация правил на основе генетических алгоритмов) превзошли другие сравниваемые алгоритмы передискретизации.

Хе и др. (He и др., 2017: 92–4) предложили модель прогнозирования на основе алгоритма нейронной сети для решения проблемы оттока клиентов в крупной китайской телекоммуникационной компании, которая насчитывает около 5,23 млн клиентов. Стандартом точности предсказания был общий показатель точности, который достиг 91,1 %.

Материалы и методы

Были использованы все методы в предложенной стратегии beat desire на языке программирования Python. В Python есть несколько встроенных библиотек для работы с данными, например, scikit-learn, pandas, numpy для различных задач интеллектуального анализа. Весь рабочий процесс был выполнен на IPython Notebook. Для создания значимых структур нейронного каркаса была использована библиотека Keras (открытая библиотека, обеспечивающая взаимодействие с искусственными нейронными сетями). Набор данных был получен из источника Kaggle — данные об оттоке клиентов Telco (<https://www.kaggle.com/datasets/blatchar/telco-customerchurn?resource=download>). Были собраны 18 различных факторов о клиенте, такие как 1) Многолинейность, 2) Срок действия контракта с клиентом, 3) есть ли у клиента техническая поддержка и т.д. Набор данных разделен на обучающий и тестовый наборы. Набор готовых данных содержит группы оттока среди 7000 клиентов.

Результаты и обсуждение

В наборе данных 216 клиентов, у которых отсутствуют данные более чем по семи факторам. Поэтому эти клиенты были исключены из набора данных. Кроме того, есть несколько клиентов, у которых отсутствовали данные по паре переменных.

Использовалась стратегия опережающего заполнения, создавая в прошлом законный стимул к отсутствующему полю, к этим факторам. Из 18 индикаторных факторов было исключено свойство «регион клиентского администрирования», несущественное для ожидания. Оставшиеся переменные были в качестве вклада в предлагаемые модели глубокого обучения. В кратком изложении имеется 7000 клиентских записей, каждая из которых имеет 18 характеристику, из которых 1800 являются отточившимися (26 %). Исходя из анализа данных:

- % Пенсионеров – только 16 % клиентов являются пенсионерами. Таким образом, большинство наших клиентов в данных - молодые люди;
- Партнер – около 50 % клиентов имеют партнера;
- Статус иждивенца – только 30 % от общего числа клиентов имеют иждивенцев;
- Телефонная связь – около 90,3 % клиентов имеют телефонную связь;
- Безбумажный биллинг – около 59,2 % клиентов выставляют безбумажные счета;
- Клиенты, у которых нет Партнеров, с большей вероятностью уйдут;
- Клиенты без иждивенцев также с большей вероятностью уйдут;
- У клиентов с несколькими линиями процент оттока немного выше.

Построение модели

Следующим этапом после извлечения необходимых основных моментов из наборов данных является разработка отображения. Проблема прогнозирования демонстрируется как проблема группировки двух классов. Три системы для обоих наборов данных были подготовлены. Для утверждения выполнения этих классификаторов использовался стратифицированный 10-перекрывающийся обход данных. Была выбрана подгонка стратифицированного хода, поскольку оба набора данных неравномерны, например, неравное количество сбрасывающих и не сбрасывающих. Стратегия стратифицированного перекрестного одобрения гарантирует, что уровень тестов для каждого класса является сравнительным по сгибам.

Заключение

В работе были предложены простые модели, основанные на правилах принятия решений, и сложные модели классификации для прогнозирования оттока. Хотя эти стратегии эффективны при выполнении задачи по прогнозированию оттока, они требуют ручной процедуры проектирования компонентов, которая утомительна и склонна к ошибкам. В момент, когда результат не получен в нужное время, мы не можем предпринять основные действия, чтобы воздержаться от взбалтывания, поэтому нам нужен еще более логичный ответ для воздержания от взбалтывания. В этой статье мы продемонстрировали пробное выполнение использования моделей глубокого обучения, чтобы отказаться от процедуры ручного проектирования элементов. Мы создали глубокие модели нейронной системы для задачи по оттоку клиентов. Исследования были направлены с использованием набора данных телекоммуникационной компании. Результаты наших испытаний показывают, что модели глубокого обучения работают так же в том же классе, что и обычные классификаторы, например, SVM и Random Forest.



REFERENCES

- Amin A., Anwar S., Adnan A., Nawaz M., Howard N., Qadir J., Hawalah A., Hussain A. (2020). Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. — IEEE Access. 2020. 4: 7940–57.
- Adnan A., Changez K., Imtiaz A., Sajid A. (2019). Customer Churn Prediction in Telecommunication Industry, - Pakistan: With and without Counter-Example. Institute of Management Sciences: — 2019. — Pp. 214–216.
- “Applying Data Mining to Insurance Customer Churn Management”, RezaAllahyariSoeini and Keyvan 8. Han and Kamber. Data Mining: Concepts and Techniques. Second Morgan Kaufman Publisher, — 2018.
- Ahn J., Han S., Lee Y. (2018). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry, Science direct. // Telecommunications Policy. — 2018. — № 30 (10-11).
- Base Group Labs — технологии анализа данных. Статьи. [Электронный ресурс]. Режим доступа: <https://basegroup.ru/community/articles/> (дата обращения: 15.03.2018).
- Brandusoiu I., Todorean G., Ha B. (2020). Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. — 2020. — Pp. 97–100.
- Maksimenko Z.V., Hafizova R.K., Yanyшева E.R. (2017). On the issue of modeling and forecasting customer outflow // Beginning in science: materials of the IV International Scientific and Practical Conference of schoolchildren, students, undergraduates and postgraduates.Ufa: AETERNA. — 2017. — Pp. 134–136.
- He Y., He Z., Zhang D. (2017). A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery. — Vol. 1. — 2017. — Pp. 92–4.
- Ponomarev A.A. (2016). Options for using big data in the telecommunications business. – M: Computer tools in education. — 2016. —No. 4. — Pp. 3–8.
- “Predicting Customer Churn in Mobile Telephony Industry Using Robabilistic Classifiers in Data Mining”,Clement Kirui1, Li Hong, Wilson Cheruiyot and Hillary Kirui.
- Telco Customer Churn – <https://www.kaggle.com/datasets/blastchar/telco-customerchurn?resource=download>.



INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 3. Is. 2. Number 10 (2022). Pp. 26–36
Journal homepage: <https://journal.iitu.edu.kz>
<https://doi.org/10.54309/IJICT.2022.10.2.003>

УДК 004.42

APPROACHES TO AUTOMATIC CHECKING OF PRACTICAL ASSIGNMENTS IN MOOCS AND ONLINE LEARNING

A.A. Buravov, N.T. Duzbayev*

Alexey A. Buravov — master student of the «Computer engineering» department, International Information Technology University
ORCID 0000-0002-6930-4219. E-mail: 24793@iitu.edu.kz;
Nurzhhan T. Duzbayev — PhD, associated professor, vice-Rector for Digitalization and Innovation, International Information Technology University
ORCID 0000-0002-7989-9463. E-mail: n.duzbayev@edu.iitu.kz.

© A.A. Buravov, N.T. Duzbayev, 2022

Abstract. One of the most important components of distance learning for software developers is practice. For effective learning and scaling, an instructor needs to check student's code automatically. This paper presents an actual overview of the present state of online e-learning platforms, Massive Open Online Courses (MOOCs), and other online learning services and solutions with automatic checking of students' practical assignments. It also provides an extensive theoretical framework for automatic compilation, code validation, architecture, scalability, and fault tolerance.

Keywords: online courses, MOOC, automatic code checking, automatic grading, online learning, practical code assignments

For citation: A.A. Buravov, N.T. Duzbayev. Approaches to automatic checking of practical assignments in moocs and online learning // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 26–36 (In Russ.). DOI: [10.54309/IJICT.2022.10.2.003](https://doi.org/10.54309/IJICT.2022.10.2.003).

ПРАКТИКАЛЫҚ ТАПСЫРМАЛАРДЫ АВТОМАТТЫ ТЕКСЕРУ ЖӘНЕ ОНЛАЙН ОҚЫТУ ТӘСІЛДЕРІ

A.A. Буравов, Н.Т. Дузбаев*

Буравов Алексей Александрович — «Есептеуіш техника» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті
ORCID 0000-0002-6930-4219. E-mail: 24793@iitu.edu.kz;



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Дузбаев Нуржан Токкужаевич — PhD, Халықаралық ақпараттық технологиялар университеті, байланысты профессоры, цифрландыру және инновациялар жөніндегі проректоры
ORCID 0000-0002-7989-9463. E-mail: n.duzbayev@edu.iitu.kz.

© А.А. Буравов, Н.Т. Дузбаев, 2022

Аннотация. Бағдарламалық жасақтаманы әзірлеушілер үшін қашықтан оқытудың маңызды құрамдастарының бірі тәжірибе болып табылады. Тиімді оқу және масштабтау үшін нұсқаушы студенттің кодын автоматты түрде тексеруі керек. Бұл мақалада желідегі қазіргі жағдайдың нақты шолуы берілген электрондық оқыту платформалары, Жаппай Ашық Онлайн Курстары (MOOCs) және басқа да онлайн оқыту қызметтері мен студенттердің практикалық тапсырмаларын автоматты түрде тексеретін шешімдер. Ол сондай-ақ автоматты жинақтау, кодты тексеру, архитектура, масштабтау және ақауларға төзімділік үшін кең теориялық негізді қамтамасыз етеді.

Түйін сөздер: онлайн курстар, MOOC, кодты автоматты тексеру, автоматты бағалау, онлайн оқыту, практикалық код тапсырмалары

Дәйексөз үшін: А.А. Буравов, Н.Т. Дузбаев. Практикалық тапсырмаларды автоматты тексеру және онлайн оқыту тәсілдері // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 26–36 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.003.

ПОДХОДЫ К АВТОМАТИЧЕСКОЙ ПРОВЕРКЕ ПРАКТИЧЕСКИХ ЗАДАНИЙ В MOOCS И ОНЛАЙН-ОБУЧЕНИИ

А.А. Буравов*, Н.Т. Дузбаев

Буравов Алексей Александрович — магистрант кафедры «Компьютерная инженерия», Международный университет информационных технологий
ORCID 0000-0002-6930-4219. E-mail: 24793@iitu.edu.kz;

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям, Международный университет информационных технологий
ORCID 0000-0002-7989-9463. E-mail: n.duzbayev@edu.iitu.kz.

© А.А. Буравов, Н.Т. Дузбаев, 2022

Аннотация. Одним из важнейших компонентов дистанционного обучения для разработчиков программного обеспечения является практика. Для эффективного обучения и масштабирования преподавателю необходимо автоматически проверять код ученика. В этой статье представлен фактический обзор текущего состояния онлайн-платформы электронного обучения, Массовые Открытые Онлайн-Курсы (MOOCs) и другие сервисы и решения онлайн-обучения с автоматической проверкой практических заданий студентов. Он также предоставляет обширную теоретическую основу для автоматической компиляции, проверки кода, архитектуры, масштабируемости и отказоустойчивости.



Ключевые слова: онлайн-курсы, МООС, автоматическая проверка кода, автоматическая оценка, онлайн-обучение, практические задания по коду.

Для цитирования: А.А. Бураков, Н.Т. Дузбаев. Подходы к автоматической проверке практических заданий в моос и онлайн-обучении // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 26–36 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.003.

Introduction

Online learning is an integral part of IT-industry as community needs more software developers, architects, analysts, etc. as the foundation for future digitalization. Online learning has gained particular relevance during the massive transition of most industries to remote work during the COVID epidemic (Patricia, 2020). Many services and platforms provide online education for IT-specialists. However, practice remains one of the most important components of teaching programming. The student must carry out and repeat all theoretical stages of learning in real programs (Keuning et al., 2016). In addition, one important aspect of the practical classes is the familiarity with similar solutions of other students, which encourages competition and improved results (Shaw, 2012). In classical education formats (schools, courses, universities), this aspect is provided by the work of teachers, manual scoring, checking and correcting the student's work. However, in the online format, this component becomes a real problem, since manually checking a student's practical assignments requires a lot of the teacher's resources and is pricy (Drummond et al., 2014: 785–790).

The lack of automated practice tests prevents the online learning platform from scaling effectively, and in some cases increases student attrition before completing the course or at its initial stage (Coussement et al., 2020). Different platforms solve this problem in different ways – for example, on Udemy there is no possibility to create practical assignments at all, so the quality of mastering materials on Udemy suffers (Wrigley et al., 2018). Other platforms provide the possibility to work on practical assignments, but with some specific types of assignments, such as JavaScript or Linux Bash. Some information systems like Moodle can be connected to an external checking system (Del Fatto et al., 2016).

Another important point is that for a better quality of learning, practical assignments should be varied and have individual components for each student (Hozak, 2020). In some areas of development, the situation is complicated by local features. For example, when testing mobile applications, one needs to run them on a real device or emulator and check the behavior logically (Bruzual et al., 2020). A similar situation occurs when checking programming assignments in distributed environments — it is necessary to check the behavior of each node launched by students (Maicus et al., 2019).

Existing platforms and services will be analyzed and reviewed to research existing methods and tools for organizing automated practice assignments.

Contributions are:

- a review of current approaches to automated code testing and evaluation;



- a current review and analysis of learning platforms with the function of automatic checking of assignments. There is no current review or analysis of those platforms. This analysis can become a foundation for further research on this topic;
- an assessment of these platforms according to different criteria like the quality of automatic code checking, availability of manual checking, built-in correction for students and others. At the moment there is no current review considering solutions with automatic code checking in the context of these criteria. This assessment also includes a description of the features and disadvantages of the platforms in question. Further research in this area may use this analysis as a reference and a model.

The rest of the paper is organized as follows:

Section II presents the problem identification and its significance. Section III discusses the related work devoted to the existing methods. Section IV presents existing approaches to checking practical code assignments. Section V presents review of existing MOOC and open-source solutions with the feature of automatic code evaluation. Section VI presents discussion of results and Section VII is conclusion.

Materials and methods

The main problem is to identify automatic e-learning platforms. As there is a lack of a universal tools for automatically checking practical assignments in online coding educational systems.

Furthermore, there is no current detailed review of existing automatic checking methods and tools, which should also include online-services. There are many online platforms and learning services, including open-source ones (e.g., Moodle). However, only a small number of them have practical assignments, where a student can code after completing a theoretical unit. Most of them contain only lecture modules and closed multiple choice tests.

This problem has a great negative effect, which is that a student who starts learning programming on online platforms cannot immediately check and apply the knowledge he or she has acquired. Also, the platform and instructors cannot verify how well the student has mastered certain theoretical modules. The problem has great importance, because the field of education for software engineers has an accelerated tendency to move to a distance format, which accelerated even more during the COVID-19 pandemic (Patricia, 2020).

There are several possible suggestions for solving this problem. For example, simplifying manual checking of practical assignments by an instructor via email or git-hosting. It is also possible to create courses on platforms with the ability to automatically check, however each platform has its own limitations and disadvantages. Optimistic solution consists of some universal system which is open and understandable to instructors and students and allow checking.

This section provides important characteristics of the relevant methods. Staubitz et al. (Staubitz et al., 2016). discussed the practical programming exercises and automated assessment. The proposed work used some typical scenarios of user experience in online courses. The authors showed the basic low-level integrations and some parts of code checking architecture, however, they failed to provide detailed practical examples.

Glassman, et al. (Glassman et al., 2015: 1–35) reviewed different approaches of code reviewing and logical tree building. This work showed examples of detailed statistical analysis of evaluating student's code. The authors provided detailed review of visualization and code comparison practices in the field of automatic code checking. However, they didn't provide any examples of architecture of these systems.

Lowis, et al. (Löwis et al., 2015) discussed social parameters of MOOC users, statistics of continuous learning curve and support stats.

Királya, et al. (Király et al., 2017) described creating a platform for online courses with automatic Java code checking using a test framework. They reviewed methods for building an evaluation system based on test libraries and frameworks.

Nickchen and Mertsching (Nickchen et al., 2016: 482–488) analyzed tools for automatic checking of Octave3D tasks. They demonstrated the operation of automatic task checking for external software. But the authors failed to analyze the architecture of the system.

Rim (Rim, 2017) looks at a system for building online courses. He showed main features of online learning platform for children. But he failed in description of examples of custom practical assignments or a system for professional software developers.

English (English, 2004) described powerful solution for Graphical User Interface (GUI) software tasks written on Java. He showed main features of this solution: automatic checking of GUI-programs and automatic code validation. But he failed in description of tasks on other programming languages besides Java.

Derval et al. (Derval et al., 2015: 86–91) described architecture and principle of external grading system called INGIInious. They provide analysis of main solution's features such as editing code in a browser, connection to edX courses as backend checking system and support of many programming languages like Python, C++, Java, Scala. But authors didn't provide any examples of checking user-interface tasks.

Pieterse (Pieterse, 2013: 45–56) provided analysis of online checking system for educational department. Author underscored important system's features like uploading directory of source files and using makefiles. But described system was developed only for C language and doesn't have any frontend editor.

Wang et al. (Wang et al., 2021) describe techniques for peer-reviewing students when solving practical programming problems, as well as the relationship between commits in the version control system (Git) and assessments of reviewers. Now cross-review is often used as a main or additional technique in some Coursera courses.

Approaches to automatic code checking

This section provides our classification of methods for testing and checking program code in general cases. All approaches below use term “grader”, referring to both the testing system and its environment.

- Compilation and communication with result via standard in and out system channels or console input and output. This method is one of the oldest approaches to testing solutions written in any programming language. Advantages: simplicity. Disadvantages: the need to configure compilation and assembly separately for each programming language, the complexity of building complex scenarios, probability of



breakage of the grader logic while reading diagnostic messages or logs of the invoked code.

- **File input/output.** In this method grader compiles solution, uses some files as input, also checks files with predefined names as output of code. This approach could be used in combination of approach a. This approach allows to test work with larger amounts of data than when working through standard input-output streams (a).

- **Web requests, API-testing.** With this approach, the code under test is compiled and run as a web server. Next, the grader sequentially sends requests to this web server via HTTP/HTTPS, TCP, WebSocket or other protocols and compares the responses with the expected values. Advantages – simplicity, less dependence on the language. Disadvantages – this approach is not applicable in all cases, quite often there is a need to test the visual interface, or other code outputs. Also, student must have the skills to write a web server, and this is usually not the initial stage of training to apply this approach.

- **Testing with own tools of a language.** Most programming languages have their own tools for writing internal tests. For example, they are Jest for JavaScript, testify and test packages for Golang, JUnit for Java, and so on. In this approach tutor should write test cases on same programming languages as the learning language. The advantages of using it are wide support for language functions, the ability to quickly write test cases. Disadvantages — binding to a specific language, the need to write everything anew for each new programming language (Akahane et al., 2015: 1–6).

- **UI/visual testing.** This approach is needed when students write tasks with some graphic results as solution. For example, webpages, desktop or mobile applications. As part of this approach, an automated grader can check html tags in a rendered web page, compare pixel colors at certain coordinates, emulate keystrokes and mouse clicks. Advantages — independence from the programming language, the application is tied much to real applications. Disadvantages – the complexity of setting up test cases, many opportunities to deceive the grader, for example, by drawing pixels or tags specifically for it, but not completing the rest of the assignment (Király et al., 2017).

- **Containerization.** In this approach application is compiled and moved to container (for example — Docker). On next steps automated grader could launch console commands, create and read files, make web requests, check output webpages or interface, mixing previous approaches. Advantages – universality (for example, student could write Dockerfile and instructions to compilation), no need to modify the environment and test logics for each programming language, and the ability to mix other approaches. Disadvantages – increased resource consumption, and possible errors with the removal and cleaning of containers (Maicus et al., 2019).

Mooc and open-source solutions

One of the most popular Massive Open Online Courses (MOOCs) Coursera¹ is an online learning platform where the most advanced and popular universities can create their own courses. It is possible to create both open and closed tests with variants and tasks with automatic code checking. Of course, this is a proprietary closed system, but some of the grading system code is open source². This system is written in Python and uses Docker containers to execute students' code. Accordingly, the course infrastructure team should write their solution in Python language and publish it using the built-in

1 <https://www.coursera.org>

2 <https://github.com/coursera/courseraprogramming>

command-line utilities. At the same time, the Coursera's frontend interface does not look so advanced and, in fact, is a normal universal interface for downloading files or archives.

Another popular platform, edX³, also has the ability to check student's assignments automatically. To check them the course's instructors must connect an external grader through a special interface — xqueue⁴. Your grading server must accept JSON-requests and respond to them with messages of a special format. Unlike Coursera, it is entirely up to you to manage resources and provide the verification itself; the platform actually provides only an external graders interface. No interface or editor for students is provided here either.

The next platform, Hexlet⁵, is a relatively young proprietary online learning platform designed primarily for the Russian-speaking segment. At the same time, it provides the same opportunities both for regular automatic checking of student assignments and for manual checking of individual project assignments by selected instructors. One of the advantages of the platform is a powerful online editor with code highlighting. A student can write code in the selected language, view the results of a web server, Linux container, edit any allowed files within his/her container, and call network commands. When checking the code, the platform shows the error's specific place and the full test code so that the student can get familiarized with it, which almost eliminates the cases of a student's misunderstanding of the checking rules and the peculiarities of the grading system (which happens on other platforms). Not the main but rather important advantage of the platform is the code style checker's functionality. For example, for practical exercises in the JavaScript language, the system, after validating the code output result, also checks the code style using the ESLint plugin. However, since the system is proprietary, all assignments and courses are posted only by the company's employees, so the external tutor cannot create custom courses.

The Codio⁶ platform is specially designed for creation of tutors' and universities' own computer science courses. It provides a convenient infrastructure for students and tutors, including an online code editor. It also provides the infrastructure for creating, editing, and running coding assignments for students. Special attention should be paid to the specialized interface for editing such tasks by the tutor. This interface allows you to create standard tasks and tests for Java, Python, and any other languages by executing commands from the command line and providing output HTML code tests.

The Stepic⁷ platform is also designed to host and provide courses for universities and individual instructors. Depending on the tariff plan, it provides the possibility of manual and automatic checks, including editing tests in the Java language in a local editor using a special library EduTools (written in Java). However, these possibilities are rather limited and have several disadvantages. Automatic code checking is limited

3 <https://www.edx.org>

4 <https://github.com/edx/xqueue>

5 <https://hexlet.io>

6 <https://www.codio.com>

7 <https://stepik.org>



to tasks for the Linux terminal and tasks in the Java language. Also, the platform does not have a convenient interface for students to edit the code.

The INGINious⁸ platform is an open solution focused on self-building infrastructure and already includes all the necessary components for developing and maintaining online courses of a university, company or instructor. It is written in Python and uses Docker containers to run student's code inside a secured environment.

INGInious provides a backend system with flexible and secure architecture which manages running and interaction with Docker and checking code, and a frontend for students and tutors (Derval et al., 2015: 86–91). The frontend part allows tutors to create and modify code tasks but doesn't have a convenient code editor. Also, a serious advantage of the system is the ability to integrate it with other platforms, such as edX or Moodle. Among the shortcomings, one can single out a not very clear interface and insufficiently detailed documentation. However, these disadvantages can be eliminated as the solution code is open source.

LearnDB⁹ is also an interesting example of a specialized online learning system. This platform is specially designed for the study of SQL, which determines its features. The platform has a powerful command editor with syntax highlighting in the browser, a test base, and a practical task for each passed module. The platform's key advantage is a well-thought-out organization for performing practical tasks with databases, which other online platforms do not have. Since the platform is proprietary, there is no way to integrate or create your own courses. There are also no practical assignments in other languages or frameworks.

Codecademy¹⁰ is a popular well-automated platform for learning programming in various languages and frameworks. On this platform, special attention is paid to practical assignments. After each micromodule, the student must immediately solve several tasks with increasing difficulty. In this case, tasks have a different level of participation. In some assignments, students only have to change a certain symbol. In other more complicated tasks students should write full-fledged code that solves a specific task. The built-in editor has unique functionality for highlighting the necessary sections of the code and interactive tutorials. Simultaneously, the platform also offers projects with manual checks for training (after the student reaches a certain level).

Codegym¹¹ is another platform for teaching development in a specific language. Java programming courses and assignments are well designed and integrated with gamification. This platform is intended mainly for beginners, and middle developers, i.e., assignments are designed to learn the syntax and principles. Another interesting feature is the format of mini projects. The platform has mini projects where the student must solve any small practical problem (program the movement, the ATM logic, and so on).

Each of the considered solutions has its own pros and cons. However, the best

8 <https://github.com/UCL-INGI/INGInious>

9 <https://learndb.ru>

10 <https://www.codecademy.com>

11 <https://codegym.cc>



solutions based on the review results are Hexlet, Codio and INGINious. Hexlet has features built and developed with the highest quality, especially the mechanism for creating and checking containers, various types of tasks, the ability to change and launch almost all entities inside an isolated container (which contributes to the speed of learning and the ability to try new techniques even outside the task). Codio and INGINious allow the instructor to create their own courses to develop almost any practical assignment. At the same time, INGINious is an open-source solution, which allows you to level its shortcomings and adjust it to your infrastructure.

Based on the review and analysis, the main features of a solution with automatic code checks can be merged and described as following:

- Solution should have feature of automatic code checking (by compilation and verification). Hexlet, Codio and CodeCademy show the best support of this feature in a review.
- Solution should have possibility of human manual checking for some tasks (for example, for projects) for best flexibility of learning process. This will adapt e-learning model to all cases and tasks. Hexlet, CodeCademy and Codio show best support of human-checking feature.
- Support as many programming languages as possible. From reviewed list Coursera and edX support almost all variants of languages but that is because these systems move compilation logic to external graders or developers from tutor's side.

The solutions under consideration currently support a different number of languages, as shown at Figure 1.

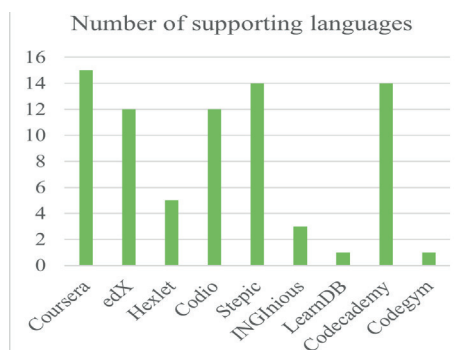


Figure 1 – Number of supporting languages for each solution

- An external editor is not such a common feature, but its presence significantly improves user experience. The student can edit the code directly in the browser and use additional features to explore the environment. Also, an external editor can be made on the base of a desktop program. This feature is best represented in the Hexlet, CodeCademy and LearnDB platforms.

- The code style checking functionality is rather optional. However, this is a good addition if the platform provides courses for novice developers and professionals. As the basic rules of syntax become more obvious with an in-depth study of technology, it

becomes more important to write code according to the standards and guidelines. This functionality is best represented on the LearnDB and Hexlet platforms.

- The backend architecture is, although externally invisible, a significant parameter of the platform as a whole. Automatic compilation and code validation are very resource-intensive tasks. The architecture must be flexible and scalable as needed. At the same time, the backend must remain well protected by the responsive core of the platform. The INGINious's backend is organized in the best way because it is divided into components and allows to scale containers quickly.

- It can also be concluded that the number of supported languages (L criteria) correlates well with the functionality of practical exercises (PT criteria).

- **Results and discussion**

- One of the main problems of online programming education is the lack of practical assignments. As a result, the quality of education suffers directly; students learn the material worse and have significant problems in applying real work skills. The review and analysis of platforms with automatic code checking function, made in this work, brings up to date information about the current situation in this area.

- The result and quality of the online programming learning platform depend not only on practical assignments' functionality. There should also be some or all of the features of manual teacher review, extended support for languages and frameworks, built-in code editor, code style checker, and the ability to scale flexibly and quickly.

- One of the main difficulties was a large number of platforms and the difficulties in assessing their components and functionality. Also, some platforms are known only locally and do not work in the global market; nevertheless, having a good implementation of automatic checks could be overlooked.

- Individual low-level components and libraries for checking and compiling the code have not been considered. The analysis did not include separate techniques and methods for the automatic compilation of code in an isolated environment. These topics will be discussed and analyzed in the future works.

Conclusion

The paper introduces an overview of existing online e-learning platforms with automatic code checking. It reviews and analyzes the existing solutions.

As a result, learning platform for programming could have one or more features like automatic code checking, manual checking, many supported languages and frameworks, and a convenient external editor for students, flexible and easy-scaling backend architecture, code-style checking, and automated peer-review.

In the future, we are interested in analyzing existing low-level libraries, tools for automatic UI tasks evaluation and other techniques for automatic code compilation.

REFERENCES

Akahane Y., Kitaya H., Inoue U. (2015). «Design and evaluation of automated scoring java programming assignments», in 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu. — Jun. 2015. — Pp. 1–6. — Doi: 10.1109/snspd.2015.7176255.

Bruzual D., Montoya Freire M.L., Di Francesco M. (2020). «Automated assessment of android



exercises with cloud-native technologies», annu. Conf. Innov. Technol. Comput. Sci. Educ. Iticse. — Issue section 5. — Pp. 40–46. — 2020. — Doi: 10.1145/3341525.3387430.

Coussemont K., Phan M., De Caigny A., Benoit D.F., Raes A. (2020). «Predicting student dropout in subscription-based online learning environments: the beneficial impact of the logit leaf model», *decis. Support syst.* — V. 135. — Issue may. — Pp. 113–325. — 2020. — Doi: 10.1016/j.dss.2020.113325.

Drummond A., Lu Y., Chaudhuri S., Jermaine C., Warren J., Rixner S. (2015). «Learning to grade student programs in a massive open online course», *proc. - ieee int. Conf. Data min. Icdm*, n. 2015-january. — issue January. — Pp. 785–790. — 2014. — Doi: 10.1109/icdm.2014.142.

Del Fatto V., Dodero G., Gennari R. (2016). «How measuring student performances allows for measuring blended extreme apprenticeship for learning bash programming», *comput. Hum. Behav.* — V. 55. — Pp. 1231–1240. — 2016. — Doi: 10.1016/j.chb.2015.04.007.

Derval G., Gego A., Reinbold P., Frantzen B., Van Roy P., De Louvain U. (2015). «Automatic grading of programming exercises in a mooc using the ingenious platform», *proc. Third eur. Moocs stakehold. Summit.* — V. 1. — issue 12. — Pp. 86–91. — 2015.

English J. (2004). «Automated assessment of gui programs using jewl», *sigcse bull. Assoc. Comput. Mach. Spec. Interest group comput. Sci. Educ.* — V. 36. — Issue 3. — Pp. 137–141. — 2004. — Doi: 10.1145/1026487.1008033.

Glassman E.L., Scott J., Singh R., Guo P.J., Miller R.C. (2015). «Overcode: visualizing variation in student solutions to programming problems at scale», *acm trans. Comput.-hum. Interact.* — V. 22. — Issue 2. — Pp. 1–35. — 2015. — Doi: 10.1145/2699751.

Hozak K. (2020). «online software using randomization and automated grading to help teach linear programming with excel and solver in business courses», *j. Educ. Bus.* — Pp. 1–8. — 2020. — Doi: 10.1080/08832323.2020.1821345.

Keuning H., Jeuring J., Heeren B. (2016). «Towards a systematic review of automated feedback generation for programming exercises», *annu. Conf. Innov. Technol. Comput. Sci. Educ. Iticse.* — № 11–13. — July. — 2016. — Pp. 41–46. — 2016. — Doi: 10.1145/2899415.2899422.

Király S., Nehéz K., Hornyák O. (2017). «Some aspects of grading java code submissions in moocs», *res. Learn. Technol.* — V. 25. — Issue 1063519. — Pp. 1–16. — 2017. — Doi: 10.25304/rlt.v25.1945.

Maicus E., Patterson S., Peveler M., Cutler B. (2019). «Autograding distributed algorithms in networked containers», *sigcse 2019 - proc. 50th acm tech. Symp. Comput. Sci. Educ.* — Pp. 133–138. — 2019. — Doi: 10.1145/3287324.3287505.

Staubitz T., Klement H., Renz J., Teusner R., Meinel D.C. (2016). «Towards practical programming exercises and automated assessment in massive open online courses», *proc. 2015 ieee int. Conf. Teach. Assess. Learn. Eng. Tale 2015.* — Issue December. — Pp. 23–30. — 2016. — Doi: 10.1109/tale.2015.7386010.

Shaw S.R. (2012). «A study of the relationships among learning styles, participation types, and performance in programming language learning supported by online forums», *comput. Educ.* — V. 58. — Issue 1. — Pp. 111–120. — 2012. — Doi: 10.1016/j.compedu.2011.08.013.

Löwis M., Staubitz T., Teusner R., Renz J., Meinel C., Tannert S. (2015). «Scaling youth development training in it using an xmooc platform», *proc. - front. Educ. Conf. Fie.* — 2015. — Doi: 10.1109/fie.2015.7344145.

Nickchen D., Mertsching B. (2016). «Combining mathematical revision courses with hands-on approaches for engineering education using web-based interactive multimedia applications», *procedia - soc. Behav. Sci.* — V. 228. — Issue June. — Pp. 482–488. — 2016. — Doi: 10.1016/j.sbspro.2016.07.074.

Rim H. (2017). «A study on teaching using website „code.org“ in programming education based on computational thinking», *j. Korea multimed. Soc.* — V. 20. — Issue 2. — Pp. 382–395. — feb. — 2017. — Doi: 10.9717/kmms.2017.20.2.382.

Pieterse V. (2013). «Automated assessment of programming assignments», *3rd comput. Sci. Educ. Res. Conf. Comput. Sci. Educ. Res.* — V. 3. — Issue April. — Pp. 45–56. — 2013.

Patricia A. (2020). «College students' use and acceptance of emergency online learning due to covid-19», *int. J. Educ. Res. Open.* — Pp. 100011. — 2020. — Doi: 10.1016/j.ijedro.2020.100011.

Wang S., Jager L.R., Kammers K., Hadavand A., Leek J.T. (2021). «Linking open-source code commits and mooc grades to evaluate massive online open peer review», *arxiv210412555 cs stat.* — apr. — 2021. — viewed: february 27. — 2022 [online]. Available at: <http://arxiv.org/abs/2104.12555>

Wrigley C., Mosely G., Tomitsch M. (2018). «Design thinking education: a comparison of massive open online courses», *she ji.* — V. 4. — Issue 3. — Pp. 275–292. — 2018. — Doi: 10.1016/j.sheji.2018.06.002.



**АҚПАРАТТЫҚ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР
ЖӘНЕ КИБЕРҚАУІПСІЗДІК
ИНФОКОММУНИКАЦИОННЫЕ СЕТИ
И КИБЕРБЕЗОПАСНОСТЬ
INFORMATION AND COMMUNICATION NETWORKS
AND CYBERSECURITY**

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 37–46

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.004>

**ANALYSIS OF THE IMPACT OF NON-GEOSTATIONARY SATELLITE
NETWORKS STARLINK ON THE GEOSTATIONARY SATELLITE
NETWORK KAZSAT-2**

N.E. Zhumatay, A.Z. Aitmagambetov*

Nazerke E.Zhumatay — Master's student, RET specialty, International Information Technology University;

Aitmagambetov A.Z. — International Information Technology University.

© N.E. Zhumatay, A.Z. Aitmagambetov, 2022

Abstract. The article analyzes the impact of small Starlink satellites in non-geostationary satellite orbit on the GSO KazSat-2 satellite network of the Republic of Kazakhstan. The planned and developed constellations of small satellites in a non-geostationary satellite orbit are studied, the characteristics of the Starlink non-geostationary satellite constellation technology from SpaceX are considered. Calculations of EMC of non-GSO starlink on the GSO network of the Republic of Kazakhstan KazSat-2 were visualized.

Keywords: non-geostationary orbital system, electromagnetic compatibility, Starlink satellite constellation

For citation: N.E. Zhumatay, A.Z. Aitmagambetov. Analysis of the impact of non-geostationary satellite networks starlink on the geostationary satellite network kazsat-2 // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 37–46 (In Russ.). DOI: [10.54309/IJICT.2022.10.2.004](https://doi.org/10.54309/IJICT.2022.10.2.004).



STARLINK ГЕОСТАЦИОНАРЛЫ ЕМЕС ЖЕРІК ЖЕЛІЛЕРІНІҢ KAZSAT-2 ГЕОСТАЦИОНАРЛЫ СПУТНИКТІК ЖЕЛІСІНЕ ӘСЕРІН ТАЛДАУ

Н.Е. Жұматай, А.З. Айтмағамбетов*

Жұматай Назерке Еркінқызы — Халықаралық ақпараттық технологиялар университетінің РЭТ-201М группасының магистранты;

Айтмағамбетов А.З. — Халықаралық ақпараттық технологиялар университеті.

© Н.Е. Жұматай, А.З. Айтмағамбетов, 2022

Аннотация. Мақалада геостационарлық емес жерсеріктік орбитада шағын Starlink спутниктерінің Қазақстан Республикасының GSO KazSat-2 спутниктік желісіне әсері талданады. Геостационарлық емес спутниктік орбитада шағын спутниктердің жоспарланған және әзірленген шоқжұлдыздары зерттелді, SpaceX-тен Starlink геостационарлық емес спутниктік шоқжұлдыз технологиясының сипаттамалары зерттеледі. Қазақстан Республикасының KazSat-2 GSO желісіндегі GSO емес Starlink ЭМС есептеулері visualyse көмегімен жасалды.

Түйін сөздер: геостационарлық емес орбиталық жүйе, электромагниттік үйлесімділік, Starlink спутниктік жүйесі

Дәйексөз үшін: Н.Е. Жұматай, А.З. Айтмағамбетов. Starlink геостационарлық емес жерік желілерінің kazsat-2 геостационарлық спутниктік желісіне әсерін талдау // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 37–46 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.004.

АНАЛИЗ ВЛИЯНИЯ НЕГЕОСТАЦИОНАРНЫХ СПУТНИКОВЫХ СЕТЕЙ STARLINK НА ГЕОСТАЦИОНАРНУЮ СПУТНИКОВУЮ СЕТЬ KAZSAT-2

Н.Е. Жұматай, А.З. Айтмағамбетов*

Жұматай Назерке Еркінқызы, магистрант группы РЭТ-201М, Международного университета информационных технологий;

Айтмағамбетов А.З. — Международного университета информационных технологий.

© Н.Е. Жұматай, А.З. Айтмағамбетов, 2022

Аннотация. В статье ведется анализ влияния малых спутников Starlink негеостационарной спутниковой орбите на спутниковую сеть ГСО KazSat-2 Республики Казахстан. Исследованы планируемые и разрабатываемые группировки малых спутников на негеостационарной спутниковой орбите, изучены характеристики технологии негеостационарной спутниковой группировки Starlink



от SpaceX. Произведены расчеты ЭМС НГСО starlink на ГСО сеть РК KazSat-2 по visualyse .

Ключевые слова: негеостационарная орбитальная система, электромагнитная совместимость, спутниковая группировка Starlink.

Для цитирования: Н.Е. Жұматай, А.З. Айтмагамбетов. Анализ влияния негеостационарных спутниковых сетей starlink на геостационарную спутниковую сеть kazsat-2 // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 37–46 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.004.

Введение

Радиочастотный спектр и спутниковые орбиты являются общими и ограниченными ресурсами, поэтому регулирование, координация и управление использованием спектра и спутниковых орбит являются неизбежной необходимостью (Контроль за использованием спектра, 2011).

Вследствие постоянно растущего спроса проблема рационального использования, совместного использования и защиты ограниченных общих ресурсов становится серьезной проблемой в национальном и международном масштабе.

В последние пару лет мы стали свидетелем растущего интереса к предоставлению услуг широкополосной связи с негеостационарных орбитальных систем (НГСО). Проблема развертывания многоспутниковых НГСО систем состоит в том, что на сегодня нет единой методологии обеспечения совместимости между НГСО системами, имеющими множество (сотни или тысячи) спутников в своей орбитальной группировке и, как правило, глобальную зону обслуживания. Таким образом, решение вопроса электромагнитной совместимости таких НГСО систем, с учетом наличия переменных во времени взаимных помех, возложено на заявляющие администрации. Кроме того, в связи с обострением проблемы электромагнитной совместимости (ЭМС) возрастает актуальность решения задач радиоконтроля и радиомониторинга.

Целью является исследование электромагнитной совместимости развертываемых спутниковых сетей НГСО Starlink и спутниковой сети ГСО KazSat-2 Республики Казахстан в Ku-диапазоне.

Для достижения указанной цели в работе поставлены следующие задачи:

1. Исследование планируемых и разрабатываемых группировок малых спутников на негеостационарной спутниковой орбите
2. Изучение характеристик технологии негеостационарной спутниковой группировки Starlink от SpaceX
3. Расчет влияния НГСО starlink на ГСО сеть РК KazSat-2 по visualyse

Исследование планируемых и разрабатываемых группировок малых спутников на негеостационарной спутниковой орбите.

Идея обеспечения интернета из космоса с помощью больших созвездий спутников LEO вновь обрела популярность в последние годы. В период с 2014 по 2016 год возникла новая волна предложений о создании крупных низкоорбитальных

Low Earth Orbit созвездий спутников для обеспечения глобальной широкополосной связи. В общей сложности 11 компаний обратились в Федеральную комиссию по связи Federal Communications Commission с просьбой развернуть крупные созвездия на негеостационарных спутниковых орбитах Non-Geostationary satellite orbits в качестве средства предоставления широкополосных услуг.

Прогнозируется, что рост спутниковой индустрии в этот период будет происходить в основном за счет развития и развертывания сетей группировки малых спутников НГСО. Действительно, ожидается, что в обозримом будущем мировой рынок коммерческих ГСО заказов будет оставаться на уровне примерно от 8 до 10 спутников в год, тогда как рынок малых спутников связи НГСО будет расти со среднегодовым темпом роста 15 % (Butash, 2018).

Эти новые технологии варьируются от 2 спутников, предложенных Space Norway, до 4 425 спутников, предложенных SpaceX. Из-за большого количества спутников в этих созвездиях было придумано название “мега-созвездия” для обозначения этих новых предложений. Список наиболее известных группировок малых спутников НГСО, планируемых в настоящее время, разрабатываемых для развертывания в течение следующего десятилетия, приведен в таблице 1. К сожалению, из-за запатентованного характера этих систем и невероятно конкурентной среды, в которой работают их разработчики, детали конструкции редко раскрываются в открытой литературе.

Таблица 1 – «Планируемые и / или разрабатываемые группировки малых спутников на негеостационарной спутниковой орбите (НГСО) за 2010 г. – апрель 2021 г.»

Спутниковые группировки	Параметры конструкции системы	Состояние
Amazon Kuiper	3236 спутников: 784, 1296 и 156 спутников на высоте 590, 610 и 630 км, соотв., в 98 плоскостях	В разработке
Boeing V-band	2956 спутников	Не разглашается
LeoSat	78–108 спутников с оптическими ISL на высоте 1400 км	Отменено
OneWeb	648 спутников в 18 плоскостях на высоте 1200 км	Банкрот/Реструктуризация/ Запущено 182 спутника на сегодняшний день
O3b	20 спутников в экваториальной плоскости на высоте 8063 км	Готовый к эксплуатации
O3b mPOWER	20 спутников в 2 плоскостях на высоте 8063 км	В разработке
Samsung	4600 спутников	Не разглашается
SpaceX Starlink	4425 спутников с оптическими ISL в 83 плоскостях на высоте 1000 км + 1600 спутников на высоте 550 км (для ускорения развертывания)	1383 запущено на сегодняшний день, 1318 на орбите
SpaceX VLEO (very-LEO)	7518 спутников на высоте 346 км + 30 000 спутников	В разработке

Изучение характеристик технологии негеостационарной спутниковой группировки Starlink от SpaceX



Созвездие Starlink Ku+Ka-диапазона от SpaceX (<http://www.rcsc.kz/Infrastructure/KSat2>) включает в себя 4425 спутников, которые будут распределены по нескольким наборам орбит. Основная группировка, которая будет развернута первой, состоит из 1600 спутников, равномерно распределенных в 32 орбитальных плоскостях на расстоянии 1150 км с наклоном 53° (синий). Остальные 2825 спутников последуют во вторичном развертывании и будут распределены следующим образом: набор из 32 плоскостей с 50 спутниками на расстоянии 1110 км и наклоном $53,8^\circ$ (оранжевый), набор из 8 орбитальных плоскостей с 50 спутниками на расстоянии 1130 км и наклоном 74° (пурпурный), набор из 5 плоскостей с 75 спутниками на расстоянии 1275 км и наклоном 81° (черный) и набор из 6 орбитальных плоскостей с 75 спутниками на расстоянии 1325 км и наклоном 70° (желтый). На рисунке 1 изображена схема мега-созвездия Starlink.

Различные наборы орбит представлены разными цветами.

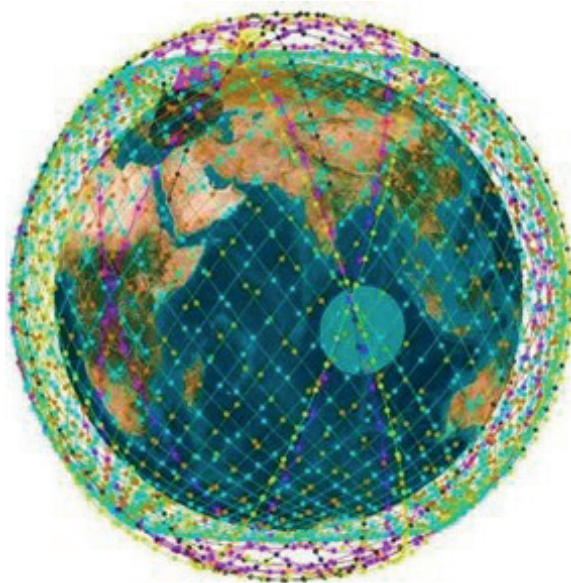


Рисунок 1 – Схема созвездия системы Starlink [10]

SpaceX внесла изменения в свою заявку в 2018 году, понизив рабочую орбиту спутников до 550 км (SATMOD2019083000087. – FCC, 30.08.2019). В таблице 2 приведен состав группировки, согласно последней заявки SpaceX в FCC от 17 апреля 2020 года (SATMOD2020041700037. – FCC, 17.04.2020) (заявка на данный момент одобрена FCC от 27 апреля 2021 года (<https://spacenews.com/fcc-approves-starlink-license-modification>)).

Таблица 2 – «Состояние группировки Starlink на 24 марта 2021 года и заявка на данный момент одобрена FCC»

Высота орбиты, км	Наклонение орбиты, градусов	Орбитальных плоскостей	Спутников плоскости	Всего спутников	Выведено на орбиту	Сведено с орбиты
550	53	72	22	1584	1373	65
540	53.2	72	22	1584	0	
570	70	36	20	720	0	
560	97.5	6	58	348	0	
560	97.5	4	43	172	10	
Всего				4408	1383	65

Каждый спутник будет нести на борту усовершенствованную цифровую полезную нагрузку, содержащую фазированную решетку, которая позволит каждому из лучей индивидуально управлять и формировать форму. Минимальный угол возвышения для пользовательского терминала составляет 40° , в то время как общая пропускная способность на спутник предполагается равной 17–23 Гбит / с, в зависимости от характеристик пользовательских терминалов. Кроме того, спутники будут также иметь оптические межспутниковые линии связи для обеспечения непрерывной связи, предоставления услуг над морем и смягчения последствий помех.

Наземный сегмент будет состоять из 3 различных типов элементов: станций слежения, телеметрии и команд, антенн шлюзов и пользовательских терминалов. С одной стороны, станции ТТ&С будут малочисленны и распределены по всему миру, а их антенны будут иметь диаметр 5 м. С другой стороны, как шлюзы, так и пользовательские терминалы будут основаны на технологии фазовых решеток. Starlink SpaceX планирует иметь очень большое количество шлюзовых антенн, распределенных по всему миру рядом с точками обмена данными в Интернете или совместно с ними.

Система Starlink будет использовать Ku-диапазон для связи пользователей, а шлюзовая связь будет осуществляться в Ka-диапазоне. В частности, полосы частот 10,7–12,7 ГГц и 14,0–14,5 ГГц будут использоваться для нисходящей и восходящей связи пользователей соответственно, в то время как полосы частот 17,8–19,3 ГГц и 27,5–30,0 ГГц будут использоваться для нисходящей и восходящей связи шлюза соответственно.

Расчет влияния НГСО сетей Starlink на геостационарную спутниковую сеть Республики Казахстан Kazsat-2

В данной главе приведены результаты расчета влияния спутниковой сети НГСО ФСС «Steam-2» Starlink на спутниковую сеть ГСО ФСС «KAZSAT2A». Расчеты проводились на ПО Visualyse.

Спутниковая система оператора SpaceX будет работать в рамках спутниковой сети Starlink. Данная сеть будет развертываться в несколько этапов. На первом этапе планируется запустить 1600 КА. Именно для 1600 КА и была проведена



симуляция влияния линии DA2831 спутниковой сети Steam-2 на линию TKU спутниковой сети KAZSAT2A

Документы, на основании которых проводились расчеты:

Регламент Радиосвязи Статья 22 таблица 22-1А- Уровень Э.П.П.М (EPFD) для полосы частот 10.7-11.7 ГГц не должен превышать -175.4 дБ (Вт/м²) на эталонную ширину полосы 40 КГц в течении 0% времени (Регламент, 2020);

Регламент Радиосвязи Статья 22 таблица 22-1В- Уровень Э.П.П.М (EPFD) для полосы частот 17.8-18.6 ГГц не должен превышать -161.4 дБ (Вт/м²) на эталонную ширину полосы 1 МГц в течении 0% времени (Регламент, 2020);

Регламента Радиосвязи Статья 5 (5.441, 5.484А, 5.487А, 5.516) гласит, что в определенных диапазонах частот: «негеостационарные спутниковые системы фиксированной спутниковой службы не должны требовать защиты от геостационарных спутниковых сетей фиксированной спутниковой службы функционирующих в соответствии с Регламентом Радиосвязи» (Регламент, 2020);

Согласно Рекомендации БР-МСЭ «Rec. ITU-R S.1323-2» (GSO/FSS; non-GSO/FSS; non-GSO/MSS feeder links):

соотношение помеха/шум (I/N) не должно превышать 6 % (-12,21 дБ) дольше 10% времени;

НГСО ФСС должны нести ответственность не более чем за 10 % времени за ухудшение отношения C/N.

Использована Методика из Рекомендации МСЭ-R S.1503-3, описывающий реализацию расчета на программном средстве ПО Visualyse.

ПО Visualyse продукт компании Transfinite Systems Limited, расположенной в Лондоне, Англия. Transfinite Systems Limited принимает участие с 1994 года в разработке технических, нормативных и политических вопросов, связанных с радиочастотным спектром, в основном в рамках МСЭ-R. Visualyse применяется для исследований и моделирования спутниковых систем и наземных сетей всех видов и во всех диапазонах частот. В качестве инструмента для изучения сценариев. Независимо от того, являются ли спутники GSO, NGSO, НЕО или на любом другом типе околоземной орбиты, Visualyse может моделировать их поведение. Он представляет собой механизм для расчета уровней несущей, уровней помех и уровней шума в радиоперелиниях. Он выдает числа C/I, C/N, C/N+I, pfd, EPFD и I/N и статистику практически для любого сценария совместного использования спектра или анализа помех.

При моделировании на ПО Visualyse на рисунке 2 земная станция ГСО ФСС «KAZSAT2A» была размещена в городе Нур-Султан.

На сегодняшний день было произведено 26 запусков космических аппаратов системы Starlink, на орбите находится 1553 спутников. Частотные присвоения сети STEAM-2 для нисходящей линии связи 10 700–13 250 МГц и для восходящей линии связи 13 750–14 500 МГц для связи с терминалами пользователей.

Параметры спутниковых сетей, используемые в расчете представлены в таблице 3. Рассматриваемый период 150 дней.

Таблица 3 – «Параметры спутниковых сетей, используемые в расчете»

STEAM-2	
Параметр	значение
Кол-во спутников в сети	4408
Кол-во спутников, рассматриваемые в расчете	1584
Наклонение	53°
Высота	550 км
Луч	DA2831↓
Несущая	11,1825 ГГц
Полоса	250 МГц
Коэффициент усиления антенны на борту спутников	28,3 дБ
Подводимая мощность	17 дБ
KAZSAT2A	
Параметр	значение
Орбитальная позиция	86,5°
Луч	TKU↓
Несущая	11,125 ГГц
Полоса	35 МГц
Коэффициент усиления антенны на борту спутников	34,5 дБ
Подводимая мощность	-8,4 дБ

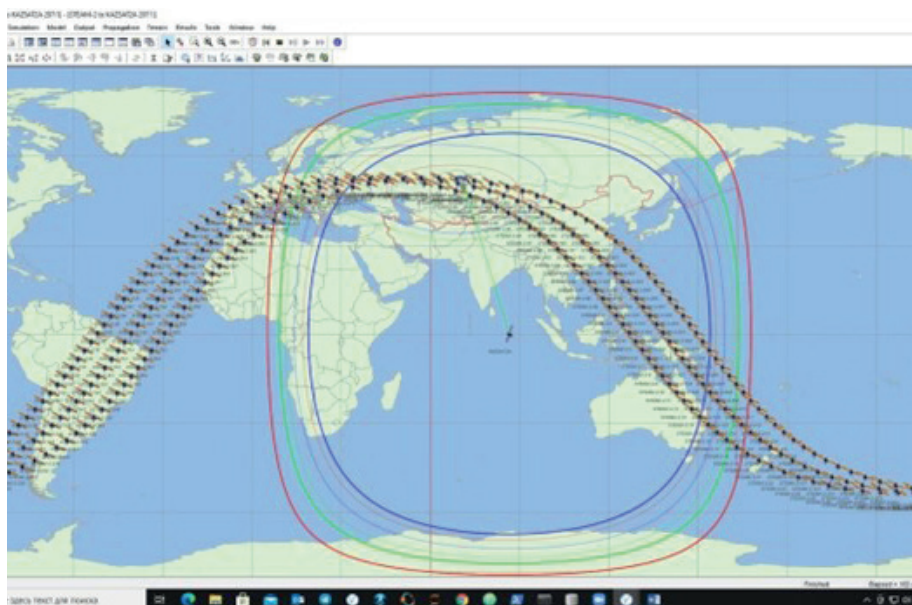


Рисунок 2 – Скриншот с ПО Visualyse при симуляции сети «STEAM-2» Starlink и KazSat-2

Результаты расчетов влияния Луча STEAM-2 DA2831 Е с шириной полосы пропускания 250 МГц и подводимой выходной мощностью 17 дБ для одной ЗС KAZSAT2A представлены в таблице 4.

Таблица 4 – «Результаты расчета влияния сети STEAM-2 для одной ЗС KAZSAT2A»

STEAM-1 (11,1825 Гц) на KAZSAT2A (11,125ГГц)				
Параметр	Наихудший уровень	Средний уровень	Пороговый уровень	Время, в течение которого имеется превышение
I/N	50,8	-28,1	-12,2 дБ	54,17 %
EPFD	24,69	-134,5	-161.4 дБ (Вт/м ²)	54,17 %

Интерпретация результатов симуляции.

Ввиду значительного превышения допустимого значения долговременных помех влияние сети НГСО ФСС STEAM-2 на ГСО ФСС KAZSAT2A является неприемлемым, электромагнитная совместимость не обеспечивается.

Заключение

Таким образом, в настоящей работе исследованы планируемые и разрабатываемые группировки малых спутников на негеостационарной спутниковой орбите, изучены характеристики технологии негеостационарной спутниковой группировки Starlink от SpaceX. Было проведено моделирование спутниковых сетей НГСО OneWeb и Starlink с ГСО ФСС KazSat-2 Республики Казахстан на ПО Visualyse на основе анализа нормативных документов МСЭ-R, касающиеся сосуществования систем НГСО/ГСО ФСС в диапазоне Ku и Ka, согласно Рекомендации ITU-R «Rec. S.1503» и методологии расчета для реализации в программном средстве. Проведенное моделирование сети «STEAM-2» Starlink из 1584 спутников с мощностью излучения 7 дБ и шириной полосы 250 МГц показало значения I/N с наихудшим уровнем 50,8 дБ и усредненным уровнем равный – 28,1 дБ, с 54,17 % времени превышения порогового значения. Значения EPFD равны с наихудшим уровнем 24,69 дБ (Вт/м²) и усредненным уровнем -134,5 дБ (Вт/м²), с 54,17 % времени превышения порогового значения.

Полученный результат исследования и проведенные расчеты значений I/N, EPFD методом геометрии наихудшего случая показывают значительное превышение допустимого значения долговременных помех, влияние сети НГСО ФСС Starlink на ГСО ФСС KazSat-2 является неприемлемым. Электромагнитная совместимость этих систем не обеспечивается.

REFERENCES

Butash T. (2018). Industry disruption, transformation continue into fourth year. – AIAA Aerospace America, December 2018.

Information about the technical characteristics of KazSat-2. Website address: <http://www.rcsc.kz/Infrastructure/KSat2>. ITU-R Report SM.2257-2 (06/2014). Handbook "Control over the use of Spectrum" /2011 edition of the ITU Radio Bureau.

Information about the approval of SpaceX's application to change the working height. Website address: <https://spacenews.com/fcc-approves-starlink-license-modification/> Date of application: 27.04.2021.

ITU. Radio Regulations. — Article 22 "Space services". — Geneva, 2020.

ITU. Radio Regulations. — Article 5 "Frequency allocation". — Geneva 2020.

ITU-R. Recommendation ITU-R S.1323–2. Maximum permissible levels of interference in a satellite network (GSO/FSS; non-GSO/FSS; non-GSO/MSS feeder links)* in the fixed-satellite service caused by other codirectional FSS networks below 30 GHz. — 2002.



SpaceX. Application for Fixed Satellite Service by Space Exploration Holdings, LLC SAT-MOD-20190830-00087 / SATMOD2019083000087. – FCC, 30.08.2019.

SpaceX. Application for Fixed Satellite Service by Space Exploration Holdings, LLC SAT-MOD-20200417-00037 / SATMOD2020041700037. – FCC, 17.04.2020.

Pekhterev S. Low-orbit satellite groupings: StarLink. Website address: <https://www.cableman.ru/article/nizkoorbitalnye-sputnikovye-gruppirovki-starlink> .



INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 47–58

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.005>

УДК 004.42

UDK 004.72

TRANSITION FROM 4G LTE TO 5G. INNOVATIVE POTENTIAL OF THE 5G ECOSYSTEM

A.S. Abdullayeva, A.K. Aitim, A.V. Tyan*

Asel S. Abdullayeva — master of technical sciences, senior lecturer of the "Information Systems" department, International Information Technology University

ORCID: 0000-0001-5188-3008;

Aigerim K. Aitim — master of technical sciences, senior lecturer of the "Information Systems" department, International Information Technology University

ORCID: 0000-0003-2982-214X. E-mail: a.aitim@iitu.edu.kz;

Alexandra V. Tyan — master of economical sciences, lecturer of the "Information Systems" department, International Information Technology University.

© A.S. Abdullayeva, A.K. Aitim, A.V. Tyan, 2022

Abstract. This article is based on an analysis of the new generation of 5G network. The smooth transition from the fourth generation to the fifth and the need for a 5G network for connection and data transmission are considered. Also, the impact of the new generation network on the global economy. Transition problems and technical needs for network implementation are analyzed. Everyday products and communications, whether images, audio, or video, are increasingly dependent on high-speed, 24/7 Internet connections. To keep up with the ever-increasing demand, the 5th generation of wireless network technology was introduced; better known as 5G. Perseverance in technological innovation contributes to a timely economic recovery and will continue to promote prosperity in the post-pandemic era. 5G advances mobile communications from connecting people to people and people to information to a single communication network connecting everyone and everything. This paper examines the evolution and rise of wireless technologies of all generations, highlights the importance of revolutionary 5G networks, examines the main technologies that make it possible to apply them, explores their trends and challenges, explores their applications in various industrial sectors.

Keywords: ecosystem, 5G, speed, data transfer, internet, efficiency

For citation: A.S. Abdullayeva, A.K. Aitim, A.V. Tyan. Transition from 4g lte to 5g. innovative potential of the 5g ecosystem // INTERNATIONAL JOURNAL OF

INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2.
Number 10. Pp. . 47–58 (In Russ.). DOI: 10.54309/IJICT.2022.10.2.005.

4G ЖЕЛІСІН 5G-ГЕ КӨШІРУ. 5G ЭКОЖҮЙЕСІНІҢ ИННОВАЦИЯЛЫҚ ӘЛЕУЕТІ

А.С. Абдуллаева, Ә.Қ. Әйтiм, А.В. Тян*

Абдуллаева Асель Сейдуллаевна — техника ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0001-5188-3008;

Әйтiм Әйгерiм Қайратқызы — техника ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0003-2982-214X;

Тян Александра Владимировна — экономика ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының оқытушысы, Халықаралық ақпараттық технологиялар университеті.

© А.С. Абдуллаева, Ә.Қ. Әйтiм, А.В. Тян, 2022

Аннотация. Бұл мақала 5G желісінің жаңа буынын талдауға негізделген. Төртінші ұрпақтан бесіншіге бірқалыпты өту және қосылу, деректерді беру үшін 5G желісінің қажеттілігі қарастырылады. Сондай-ақ, жаңа буын желісінің жаһандық экономикаға әсері қарастырылған. Өтпелі мәселелер мен желіні енгізудің техникалық қажеттіліктері талданады. Күнделікті өнімдер мен коммуникациялар, кескіндер, аудио немесе бейне жоғары жылдамдықты, тәулік бойы жұмыс істейтін интернет қосылымдарына көбірек тәуелді. Үнемі өсіп келе жатқан сұранысқа ілесу үшін сымсыз желі технологиясының 5-ші буыны енгізілді. Технологиялық инновациялардағы табандылық экономиканы уақтылы қалпына келтіруге ықпал етеді және пандемиядан кейінгі дәуірде өркендеуді жалғастыруға септігін тигізеді. 5G ұялы байланысты адамдарды адамдарға және адамдарды ақпаратқа біріктіретін бір байланыс желісіне дейін жетілдіреді. Соған қоса, бұл мақала сымсыз технологиялардың барлық буындарының эволюциясы мен өсуін қарастырады, революциялық 5G желілерінің маңыздылығын көрсетеді, оларды қолдануға мүмкіндік беретін негізгі технологияларды салыстырып, олардың тенденциялары мен қиындықтарын қарастырып, олардың әртүрлі өнеркәсіп салаларында қолданылуын зерттейді.

Түйін сөздер: экожүйе, 5G, жылдамдық, деректерді беру, интернет, тиімділік

Дәйексөз үшін: А.С. Абдуллаева, Ә.Қ. Әйтiм, А.В. Тян. 4G желісін 5g-ге көшіру. 5g экожүйесінің инновациялық әлеуеті // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. . 47–58 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.005.



ПЕРЕХОД СЕТИ 4G НА 5G. ИННОВАЦИОННЫЙ ПОТЕНЦИАЛ ЭКОСИСТЕМЫ 5G

А.С. Абдуллаева, А.К. Айтим, А.В. Тян*

Абдуллаева Асель Сейдуллаевна — магистр технических наук, сениор-лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0001-5188-3008;

Әйтiм Әйгерiм Қайратқызы — магистр технических наук, сениор-лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0003-2982-214X;

Тян Александра Владимировна — магистр экономических наук, лектор кафедры «Информационных систем», Международный университет информационных технологий.

© © А.С. Абдуллаева, Ә.Қ. Айтим, А.В. Тян, 2022

Аннотация. Эта статья основана на анализе нового поколения сети 5G. Рассмотрены плавность перехода с четвертого поколения на пятую и надобность сети 5G для подключения и передачи данных. Также, влияние сети нового поколения на мировую экономику. Анализируются проблемы перехода и технические потребности для имплементации сети. Повседневные продукты и средства коммуникации, будь то изображения, аудио или видео, все больше и больше зависят от высокоскоростного круглосуточного подключения к Интернету. Чтобы не отставать от постоянно растущего спроса, было представлено 5-е поколение беспроводной сетевой технологии; более известный как 5G. Настойчивость в технологических инновациях способствует своевременному восстановлению экономики и будет продолжать способствовать процветанию в постпандемическую эпоху. 5G продвигает мобильную связь от соединения людей с людьми и людей к информации к единой сети связи, соединяющей всех и все. В этой статье рассматривается эволюция и рост беспроводных технологий всех поколений, подчеркивается важность революционных сетей 5G, исследуются основные технологии, которые позволяют их применять, исследуются их тенденции и проблемы, исследуется их применение в различных промышленных секторах.

Ключевые слова: экосистема, 5G, скорость, передача данных, интернет, эффективность

Для цитирования: А.С. Абдуллаева, А.К. Айтим, А.В. Тян. Переход сети 4g на 5g. инновационный потенциал экосистемы 5g // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. . 47–58 (на русском языке). DOI: 10.54309/IJICT.2022. 10. 2.005.

Introduction

Over the years, technology has revolutionized our world and everyday life. Other than that, technology has made amazing tools and resources for people who have the



information we need at our fingertips whenever we need it. Modern technologies have paved the way for functional devices such as smart watches and telephones. Comps are being made faster, more portable, and stronger than ever before. Thanks to all these revolutions, technology has also made our life easier, faster than any other and more joyful. Thus, with every passing year, fresh technologies frame us more and more and make life comfortable. One of these technologies is the new generation 5G network.

5G is the 5th generation mobile network. This is a fresh mass stereotype of wireless communication later 1G, 2G, 3G and 4G networks. At this moment, the 4G network is used all over the world. 4G has contributed to the ingenious development of interdependent economic sectors of the industry. 4G movement is centered around 3 leading pillars. The 1st pillar focuses on owning a phone in either hand, or the buyers taking control of their own lives with the support of a five-inch screen. In the given figure 1 the second concerns the rethinking of providers of mobile communications offerings, redesigning their networks from voice to data-focused communications offerings with universal coverage and capacity, essentially allowing the pace to keep pace with increasing demand for traffic (Eric Dahlmann et al., 2016: 210–218). After all, centralized, hyperscale data centers host the applications and content used on our phones.

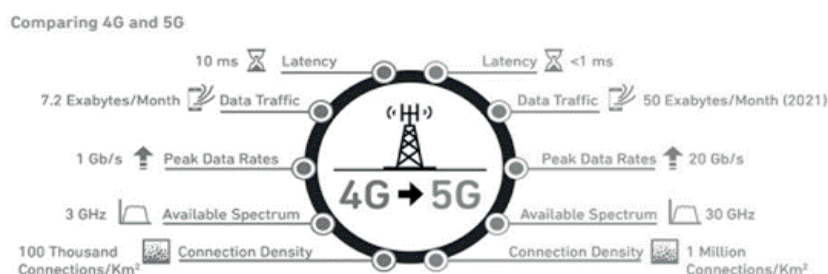


Figure 1 - Transition from 4G to 5G

Now there are 7 billion devices connected to the Online in the world, and by 2025 there will be within 21 billion of them. Almost all of these fresh devices will power and control our homes, urban infrastructure, vehicles and almost everything else; this network is widely popular as the Online of Things (IoT). This is touted as one of the proper big digital revolutions, and due to the fundamental need to keep power on to the critical devices that control our security, network resonance time or latency will need to be improved.

Compared to 4G, 5G promises to increase the data transfer rate from 1 Gb/s to 20 Gb/s thanks to this improvement, users will have access to information and information in much more short resonance time, in fact, which is considered a necessary improvement, especially for critical support services, combat and critical response teams. In the future, the development of 5G will be applied in many ways, using its superior data rate. For example, open-source video games are predicted to become obsolete as consoles because users rely on cloud technology to access games and software. Due to the reduction in waiting time and ease of access, and thanks to the fact that any remote impact has

become softer and clearer thanks to 5G technology, video games and video consoles that are popular now will disappear, in fact, potentially changing the entire financial structure of the world. industry. Beyond the social domain, the possibilities of 5G technology are limitless. From autonomous cars to smarter communities, remote medicine and surgery, or Industry 4.0, all sectors of the economy have the potential to see big changes with the introduction of 5G technology.

Materials and methods

Analysis of the impact of the 5G network on the global economy in the future

An actual study of the 5G economy has found that the absolute financial impact of 5G is likely to be sold globally by 2035, supporting a sweeping range of economic sectors and potentially providing up to \$13.1 trillion worth of products and offerings.

This impact is much greater than that of past generations of grids. Requests for the development of a fresh 5G network still apply not only to the classic players of mobile networks, but also to these branches, like the auto industry.

The study also demonstrated that the 5G value chain itself (including OEMs, carriers, content developers, app creators, and consumers) has the capacity to support up to 22.8 million workspaces on its own, or more than one job per Beijing resident China. And there are many emerging and fresh applications that will be included in the future.

In the defense industry illustrated in Figure 2, 5G will be the best answer to military needs; looking for advanced instrumentation, hybrid meshes, active comparability, cyber security, and resiliency, 5G can offer it all, moving from one system for every need to one system for all your needs (Eric Dahlmann et al., 2018: 29–35). Rapid data delivery will result in reduced latency, unrivaled situational awareness during your most significant missions, and improved communication between who, who is in the spaces and at headquarters. The remote control will be improved with the shortest waiting period, drone control will be clearer, and the integrated association for the fighter will be improved.

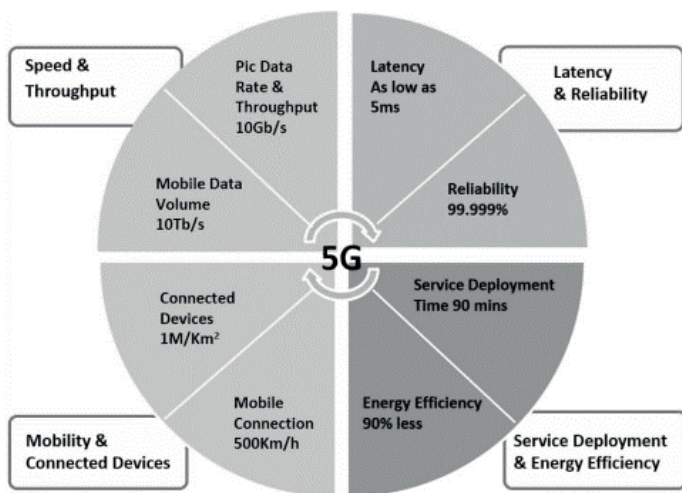


Figure 2 - Benefits of 5G sustainability

Technical difficulties with the implementation of the 5G network

5G wireless development increases performance improves efficiency which most guarantees fresh user extensions and merges fresh branches. Devices running with the inclusion of 5G will experience high battery drain, leading to the need for more ideal battery technologies. Increasing the speed is still associated with other trade-offs. Speed ups and latency reductions have been achieved through changes to 5G radio; one of the changes is related to the introduction of more high frequency bands. However, these high-frequency bands are not bad only at relatively short distances and are simply blocked by houses. Very well suited for manufacturing automation, but in towns, operators will need significantly more 5G radios to cover a particular area, and in real time it does not work at all in rural areas. While 5G will increase the adoption of IoT devices, the faster 5G turn-on rates are likely to carry a greater risk of cyberattacks, leaving hackers more likely to target more devices through loose links between devices connected to the Internet. As the number of IoT devices grows, security challenges grow; devices connected to 5G networks pose a threat to hackers downloading and extracting information, including personal data, at a much faster rate than ever before.

Monitoring the demand for traffic in the next 10 years illustrates the increase of 1000 scales and more than 100 billion inclusions to the Online of things.

This poses gigantic challenges for future mobile communication technologies after 2020, vision given in Figure 3 (Henrik Asplund et al., 2020: 115–120). Buyers are strongly asking for high-speed data transmission at low rates. 5G aims to conclude these conflicting claims by 2020. 5G was named by the State Infrastructure Commission in the report "5G Infrastructure Requirements in the UK" as "ultra-fast, ultra-reliable, ultra-high throughput capacity transmitting with ultra-low latency". (2016). The objects that can be seen with 5G technology include significantly better inclusion and coverage values.

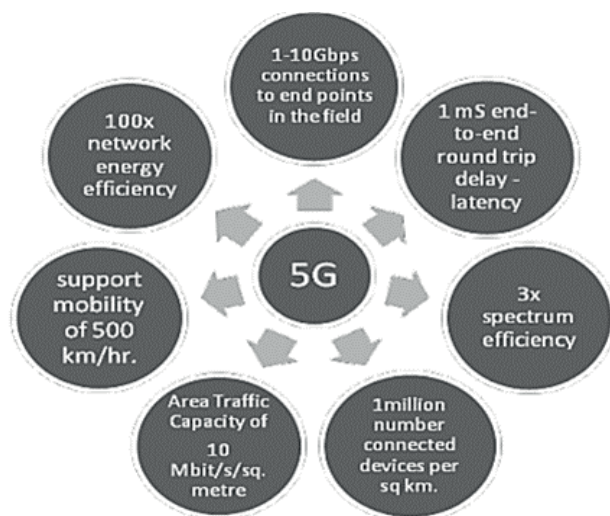


Figure 3 - Visions for the Fifth Generation

Implementation of fresh service values for 5G network implementation

Just during the introduction of 4G, telecommunications firms realized that they needed different infrastructure values to support different classes of service. With the absolute adoption of the first perfect set of 5G specifications in June 2018, 5G has begun to provide 3 service values, any of which can be customized to suit the specific requirements of customer business models (Vlasios Tsiatsis et al., 2018: 178–189):

The Enhanced Mobile Broadband Association (eMBB) aims to serve more dense urban centers with downlink speeds approaching 1 Gbps (gigabits per second) indoors and 300 Mbps (megabits per second) outdoors. This will be achieved by installing very high frequency millimeter wave (mmWave) antennas throughout the landscape - on lampposts, the sides of houses, tree branches, where electrical towers are and, in one new application, elite city buses. Because any of these antennas, if used in the subway, will cover an area, probably no more than a baseball, a hundred square meters, but it has the opportunity to be, and thousands of these antennas are needed for the painstaking service of any densely populated center of a metropolis. And because most of them would not be omnidirectional -their maximum beamwidth would be only within 4 degrees - mm-wave antennas would display signals from buddy-buddy mirrors until they eventually reached their own intended customers. For the largest number of suburban and rural areas, eMBB will aim to replace the current 4G LTE system with a fresh network of lowest power omnidirectional antennas providing a 50 Mbps downlink (Panagiota et al., 2015).

Massive Machine Type Communications (mMTC) enables machine-to-machine (M2M) and Internet of Things (IoT) applications that a new wave of wireless users can expect from their own network without burdening other classes. suggestions. IoT is a relatively recent developing engineering. Over the past few years, devices using IoT have become more extensive, deeper and more accessible. Detectors and markers are rapidly getting cheaper. Readers and detectors consume less energy, become more and more mental, work faster and at greater distances and are ready to overcome with interference. This means more superior system performance, more extensive capabilities for the use of sensors and tags with a huge data size, and even more simple integration into existing systems without reprogramming. Experts in M2M and logistics say that 2G proposals are actually perfect for the narrow bandwidths required by their signaling devices, and in fact that later generations have actually worsened this proposal by introducing fresh delay informants. MMTC will look to restore this level of service by implementing a broken level of service for devices that require downlink throughput of up to 100 Kbps (kilobits per second, right there with phone modems), but with low latency in the range of 10 milliseconds (ms) (Akyildiz et al., 2014: 201–210).

Ultra Reliable Low Latency Communication (URLLC) will serve mission-critical communication needs where throughput capability is not, for example, as important as the speed, in particular, the end-to-end hitch is 1 ms or less. This will become a degree relevant to the category of autonomous vehicles, where there is literally no time to reach a conclusion to react to a likely tragedy. URLLC has the ability to practically make

5G competitive with satellite, revealing the possibility - still under discussion between telcos of 5G to replace GPS for geolocation.

In July 2020, in the heat of the pandemic, the 3GPP organization completed Release 16, which includes provisions for:

Vehicle-to-Everything (V2X) communications, which will connect low-latency communication links between moving vehicles (especially with independent driving systems) and cloudy data centers, in fact, which will allow most of the software for managing and servicing moving vehicles to work from stationary mode, data centers are completed and serviced. Satellite access, which has the ability to plug in the possibility of satellite transmission to fill gaps in underserved or geographically remote areas.

Virtualization in 5G networks

Network Virtualization (NFV) and Software Defined Networking (SDN) are separate technologies, but they are best considered together as one complements the other.

Software-Defined Networks (SDN) can become an effective technology to reduce the amount of equipment and simplify the maintenance of infrastructure. This technology contributes to digital transformation and the transfer of services to cloud technologies. The basis of the principle is the remote control of the network and data transmission devices, in other words, programmatically.

The principle of network management implies separation by management functionality and separation by packet forwarding, and network planning and traffic coordination is carried out programmatically. It is this approach that transfers the intelligent leaving of devices to a dedicated server, and as a result, simplifies and reduces the cost of using network elements (routers and switches).

The idea behind SDN is:

- separating the management of network equipment from the management of data transmission through the creation of special software;
- transition from managing individual instances of network equipment to managing the network as a whole;
- creation of an intelligent program-controlled interface between network applications and the transport environment of the network.

The SDN architecture consists of three layers:

- the lower level is the infrastructure level, which is a set of network devices (switches and a data transmission channel);
- the middle layer is the control layer, which includes the network operating system, which provides applications with network services and a software interface for managing network devices and the network;
- the upper level is the level of network applications, this level is necessary for flexible and efficient network management.

Software-defined networks have a number of advantages:

- by accelerating the movement of traffic, network performance increases;
- logical centralization of management functions;
- replacement of complex and specific hardware with virtualization solutions and software;



- independence of software and hardware development of the network.

However, as with any technology, not everything is perfect. Software-defined networks have their perceived disadvantages:

- channel of interaction between the controller and switches;
- docking with the traditional network;
- SDN security;
- ensuring the protection of the main elements of the SDN network;
- controllers and interaction between them;
- ensuring the safety of controller-switch protocols;
- ensuring the security of the SDN-network software and controller applications.

Virtualization of network functions NFV (Network Functions Virtualization) allows you to virtualize the functions of network elements of mobile operators, at the same time it allows you to implement a "network on demand". NFV involves the separation of hardware and computing parts. Private hardware is responsible for network functions. This technology makes it possible to virtualize these functions, i.e. takes you to the software level. Thus, if in the traditional model each network function requires its own equipment, then in the NFV model, one physical foundation is enough. All data will be processed and stored in a virtual environment ("Cloud"). And the function of transmitting user traffic will remain with the classic equipment.

This approach to the organization of fifth generation networks is in line with the emerging trend in wireless communication, namely convergence. Convergence implies the integration of separate network objects into a single computing complex. This is also important for "smart" devices in order to exchange information online.

Network Functions Virtualization has a number of advantages:

- improves network performance while optimizing operators' costs by reducing the cost of purchasing and maintaining equipment, as well as lower network power consumption;
- Based on NFV, it is easier to create and launch new services.

First of all, this technology will be in demand among telecom operators who are preparing to launch a high-speed generation of 5G mobile communications. With NFV, operators will be able to enjoy rich multimedia content and new services.

An important role is played by the Cloud RAN cloud infrastructure, which is part of the NG-RAN radio access network, replacing the traditional radio access network (RAN) in fifth generation networks. This network is based on the implementation of base station models for 5G networks - gNB using NFV virtualization technology and has a number of advantages:

- ensuring the implementation of software-defined multi-standard base stations NR/E-UTRA;
- flexibility and scalability of virtual solutions;
- placement of base station equipment in a data center with a high degree of reliability and security.

In addition, Cloud RAN manages a large number of radio modules and covers a

large area of radio coverage, thereby ensuring optimal use of the frequency resource, the effective operation of interference compensation algorithms ICIC (Inter-Cell Interference Coordination), CoMP (Coordination of reception and transmission for servicing a subscriber by several cells), and Intra-RAT-handovers (ensures the preservation of communication when the subscriber moves between cells).

SDN and NFV technologies aim to optimize network performance, but use different methods to achieve the goal. SDN involves separating the network management and transport plane to provide centralized network management for more efficient use of network resources. NFV, in turn, moves network functions (such as DNS, DHCP, etc.) into a virtual environment to optimize the network, separating them from the hardware.

These technologies can be implemented both separately and together with each other.

The main advantages of SDN and NFV technologies for telecom operators are:

- simplification and centralization of management, increasing business efficiency, reducing operating costs;

- rapid deployment of new services;

- Creation of new markets through the transition to cloud services:

- 1) shift of the operator business from providing communication channels to providing cloud services;

- 2) operators can provide data center infrastructure as a service (IaaS) with the integration of communication channel resources and cloud IT resources.

- efficient use of telecommunications network resources by centralizing resource management, virtualization of data center resources.

OpenFlow technology OpenFlow is a protocol for managing the process of processing data transmitted over a data transmission network by routers and switches, which implements the technology of a software-defined network. Its development began as part of a research project, the purpose of which was to develop network architectures and protocols in a real environment.

The basic idea of OpenFlow is simple, based on the observation that, despite the differences between hundreds of switch and router models, they all contain a transfer table that defines the basic function of data transfer - for each incoming packet, forward it as quickly as possible to a specific outgoing interface.

Each entry in such an OpenFlow transmission table is a "rule" and is associated with a "flow" of data (flow). A stream is defined by a packet header — for example, a combination of source and destination MAC addresses, IP addresses, and port numbers. Not all elements of this combination need to be defined - for example, a flow can be defined as all traffic to some host.

The next element of the table is the "action" (action), which determines the required processing of stream packets:

- Forward the packet to the switch port.

- Send the packet to the controller over a secure channel. The controller is the control center of the network, so if the flow has just been formed and the controller does not yet know what to do with it, the first packet of the unknown flow is sent to the controller to determine the rules and actions for this flow.

- Drop the package. As a rule, this action is configured manually, but can also be automatically configured, for example, in the fight against hacker attacks.
- Sending the packet to standard processing.

Each OpenFlow switch contains several flow tables, each such table contains several flow entries. This model opens up unique possibilities. An OpenFlow switch can be anything from a simple switch to a firewall and router.

Using the protocol, the controller exchanges information with the OpenFlow switch controlled by it. The main task of such a switch is to transfer a packet from one port to another without delay, with some processing in accordance with the rules. At the same time, the switch informs the controller about its capabilities and configuration, and also informs about changes in its state. The controller, or rather the application that uses it, is tasked with determining the overall picture and configuring the switches. They define the network topology and optimal routes. With the help of the controller, the necessary rules are set, the status of devices is monitored, traffic is monitored and statistical data is collected.

When a package is received, metadata and other fields of the package are extracted from it. Then the fields are compared with the entries in the flow table according to the corresponding "protocol" field. The result with the highest priority is the best and processing of such a packet starts first. Each entry in the table has its own priority, and its level determines how the packet will be processed further. Once a rule is selected, a specific action is applied to the packet. Possible action can be send to port, send to controller, or drop.

While OpenFlow is not capable of handling the SDN software side, it does an excellent job of providing simple basic hardware management functions - switches. Manufacturers of network equipment, and primarily manufacturers of switches, declare their support for OpenFlow, the software interface from the controller to applications is either not available at all, or is the manufacturer's own development.

Conclusion

In the beginning, the whole meaning of the "G" in wireless standards was to highlight the ease of transition between one wireless delivery system and a newer one - or at least make the transition painless. 5G is bringing with it several simultaneous revolutions, and all of them must pass without further hitches.

A converged service can lead to converged carriers. In much of the continental part of the world, an online broadband provider is still considered an Internet cable TV provider. And these cases are protected by state-regulated monopolies. The 5G Wireless Association aims to level the playing field by putting competition between operators both in the field of broadband access to the Internet, for example, and in the field of cable TV.

Small cell infrastructure could change the scenery. To lower costs for 5G operators, 5G allows for the smallest volume transmitters that consume less power but cover significantly smaller service areas than conventional 4G towers. The carrier will need more of them-by one estimate, four hundred and one more tower than is deployed in

real time, but it is likely that they are more than any other included with the landscape. It is assumed that a small 5G cell can freeze with the same simple appearance in urban areas as lampposts and graffiti.

The global technology economy has the potential to be reconstructed. Once completed, the 5G transition project will represent a major overhaul of the communications infrastructure like no other in the situation. The magnitude of the transition from 4G to 5G could be described as history if the telegraph industry merged at the end of the 19th century to agree on a phased transition to fax. The true root cause of this shift is not so much to freeze sooner, but to make the wireless industry sustainable in the long term, because the 4G transmission pattern is approaching instability sooner than industry experts predicted.

REFERENCES

- Akyildiz I.F., Nie S., Lin S.C., Chandrasekaran M. (2016). 5G roadmap: 10 key enabling technologies. *Comput Netw*, pp.201-210, 2014
- Eric Dahlmann, Stefan Parkwall, Johan Skold. 4G, LTE-Advanced Pro and the road to 5G. Academic press; 3rd Edition August 2. — Pp. 210–218. — 2016
- Eric Dahlmann, Stefan Parkwall, Johan Skold (2018). 5G NR: the next generation of wireless access technology. Academic press. — 1st Edition August 31. — Pp. 29–35. — 2018
- Henrik Asplund, David Asteli, Peter von Butovich, Thomas Chapman, Matthias Frenne, Farshid Gasemzade, Mons Hagstrom (2020). Advanced antenna systems for 5G network deployment: bridging the gap between theory and practice. — Pp.115–120. — 2020
- Vlasios Tsiatsis, Stamatis Karnouskos, Ian Holler, David Boyle, Katherine Mulligan (2018). Internet of Things: Technologies and Applications for a New Era of Intelligence. — Pp.178–189. — 2018
- Panagiota D. Giotopoulou (2015). The evolution of mobile communications: Moving from 1G to 5G, and from human-to-human to machine-to-machine communications, National and Kapodistrian University of Athens, School of Science. — Pp. 164–179. — November 2015.

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 59–66

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.006>

УДК 004.42

УДК 004.056.57

UTILIZING SENTIMENT ANALYSIS FOR CYBER SECURITY LEARNING

A.E. Namiyaly, B.B. Valiyev, A.O. Sagymbekova, A.Zh. Adil*

Azamat E. Namiyaly — Master's student, Department of Information Systems, International Information Technology University;

Bakhytzhhan B. Valiyev — Master's student, Department of Computer Engineering, International Information Technology University;

Azhar O. Sagymbekova — Senior Lecturer, Department of Cybersecurity, International Information Technology University;

Altynai Zh. Adil — Master's student, Department of Computer Engineering, International Information Technology University.

© A.E. Namiyaly, B.B. Valiyev, A.O. Sagymbekova, A.Zh. Adil, 2022

Abstract. The main goal of this research article is to apply modern theoretical and practical knowledge in cybersecurity and develop a website by conducting sentiment analysis for cybersecurity learning on the developed platform. Electronic educational resources, information protection technologies, and data integrity preservation have been investigated to create the final product. The work has resulted in a secure platform for increasing user exposure in IT using the key machine learning models and hashing algorithms.

Keywords: Sentiment Analysis, Cyber Security, Learning, Website, Bag-of-words, Words, Server Log Files, SQL Injection, Cross-site-scripting, Genetic Algorithm, Database

For citation: A.E. Namiyaly, B.B. Valiyev, A.O. Sagymbekova, A.Zh. Adil. Utilizing sentiment analysis for cyber security learning // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 59–66 (In Russ.). DOI: [10.54309/IJICT.2022.10.2.006](https://doi.org/10.54309/IJICT.2022.10.2.006).



КИБЕРҚАУІПСІЗДІКТІ ЗЕРТТЕУ ҮШІН СЕНТИМЕНТАЛДЫ ТАЛДАУДЫ ҚОЛДАНУ

А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ*

Намиялы Азамат Ержанович — «Ақпараттық жүйелер» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті;

Валиев Бахытжан Бауржанович — «Компьютерлік инженерия» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті;

Сагымбекова Ажар Орынғалиевна — «Киберқауіпсіздік» кафедрасының сениор лекторы;

Әділ Алтынай Жанарбекқызы — «Компьютерлік инженерия» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті.

© А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ, 2022

Аннотация. Ғылыми мақаланың негізгі мақсаты - киберқауіпсіздікте заманауи теориялық және практикалық білімдерді қолдану және платформада киберқауіпсіздікті үйрену үшін сентименталды талдау арқылы вебсайтты құру. Ақырғы өнімді жасау үшін электронды білім беру ресурстары, ақпаратты қорғау технологиялары мен деректердің тұтастығын сақтау әдісі зерттелді. Барлық жұмыстардың нәтижесінде ақпараттық технологиялар саласындағы пайдаланушылардың машиналық оқытудың ең маңызды модельдері мен хэштеу алгоритмдерін қолдана отырып, экспозициясын жоғарылату үшін қауіпсіз платформа жасалды.

Түйін сөздер: сентиментті талдау, киберқауіпсіздік, оқу, вебсайт, сөз қабы, лог-файлы, SQL-инъекция, сайтаралық скриптинг, генетикалық алгоритм, дерекқор

Дәйексөз үшін: А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ. Киберқауіпсіздікті зерттеу үшін сентименталды талдауды қолдану// ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 59–66 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.006.

ИСПОЛЬЗОВАНИЕ АНАЛИЗА ТОНАЛЬНОСТИ ДЛЯ ИЗУЧЕНИЯ КИБЕРБЕЗОПАСНОСТИ

А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ*

Намиялы Азамат Ержанович — магистрант кафедры «Информационные системы», Международный университет информационных технологий;

Валиев Бахытжан Бауржанович — магистрант кафедры «Компьютерная инженерия», Международный университет информационных технологий;

Сагымбекова Ажар Орынғалиевна — сениор лектор кафедры «Кибербезопасность», Международный университет информационных технологий;

Әділ Алтынай Жанарбекқызы — магистрант кафедры «Компьютерная инженерия», Международный университет информационных технологий.

© А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ, 2022



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Аннотация. Основная цель научной статьи - применить современные теоретические и практические знания в области кибербезопасности и разработать веб-сайт с использованием анализа настроений для изучения кибербезопасности на платформе. Электронные образовательные ресурсы, технологии защиты информации и сохранения целостности данных были исследованы для создания конечного продукта. В результате всей работы была разработана безопасная платформа для увеличения воздействия на пользователей в области информационных технологий с использованием самых ключевых моделей машинного обучения и алгоритмов хеширования.

Ключевые слова: анализ тональности, кибербезопасность, изучение, вебсайт, мешок слов, лог-файлы, SQL-инъекции, межсайтовый скриптинг, генетический алгоритм, базы данных

Для цитирования: А.Е. Намиялы, Б.Б. Валиев, А.О. Сагымбекова, А.Ж. Әділ. Использование анализа тональности для изучения кибербезопасности // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 59–66 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.006.

Introduction

Nowadays, the ability and knowledge of securing information technologies from intruders are highly prized. The ongoing epidemics forced many individuals to move away from classic real-world services and switch to online analogues (<https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html>). The increase in online activity has led to frequent encounters with criminals, such as scammers (<https://www.cmu.edu/iso/aware/coronavirus-alerts/coronavirus-scam-awareness.html>). Computer-aided learning development, which was conducted by our team, appears to be an effective measure in enhancing the awareness of citizens striving for security.

The replacement of all human activity by information technology requires readiness against intruder attacks. The awareness-raising safety training aims to help avoid any malicious attack on an information system by providing the information needed. The main goal is to develop a web portal to raise citizen awareness and help prepare for information security. Citizens looking for information technology security will be able to use awareness-raising safety training.

Materials and methods

Temporal sentiment analysis. Sentiment time series have been widely used for event prediction, such as political election prediction (Bermingham, 2011), and disaster event detection (Sakaki, 2010). Similarly, in cyberattack scenarios, the goal is to predict the sentiment score of individual messages as well as temporal sentiment changes over time. The goal is to provide insight into how social media users reverse the polarity of sentiment towards public events. Second, understand how sentiment can indicate upcoming attack events. Third, assess the consequences of an attack after its occurrence. To answer these questions, data D about relevant social media posts are collected by

querying relevant keywords. It occurs on social media during a specific time range τ , which covers the time interval before, during, and after the attack event. Our team plots the sentiment time series $S = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m)$ in chronological order using a previously learned sentiment predictor. We can also group and average sentiments at varying granular levels over time, such as on a per-day basis, to study sentiment variation.

Sentiment analysis is a type of textual order that manages emotional explanations. It is also known as assessment mining because it measures feelings to learn about open recognition. Opinion exploration and assessment mining are similar and are used in reverse throughout the process. It uses Natural Language Processing (NLP) to collect and validate assessment or conclusion words. SA is refined as recognizing people's views on a subject and its main points. The justification for the fame of assessment mining is that people like to consult with others to make intelligent contributions. Revealing emotional points of view in a large amount of useful information is an area of interest in information mining and NLP.

Enterprise systems generate many communication logs. They are used to track Internet traffic directed to a specific source. They create many logs, which are then collected and stored. System-generated logs show running system states and contain important information generated by the system to support diagnostics. Therefore, log analysis is a vital technique for detecting vulnerability scans, XSS, and SQLI attacks.

However, manually analysing a huge number of logs is unrealistic and impractical. Therefore, it is important to use machine learning to automate log analysis.

The objectives of this study can be summarized as follows:

- Analyse log data and detect scanning vulnerabilities.
- Detect XSS and SQLI attacks and check access log files for detection.

Sentiment trends predict cyberattacks before they happen. To study the behaviour of cyberattacks, historical data from the Apache log files is considered to develop a logistic regression classifier for each type of event. The data is divided into a training set and a test set. The training set includes the first 80 % of the time covered by the data, and the test set includes the remaining 20 %. In each test, logistic regression models predict whether each type of attack will occur. The results are shown in Figure 1. The model is much better at predicting XSS attacks or SQL injection attacks. It contrasts with the regular traffic type, with very high accuracy and recall scores for both.

```

*****accuracy_score*****
0.8984
*****confusion_matrix*****
[[556 10  4]
 [ 15 250 53]
 [ 18  27 317]]
*****classification report*****

```

	precision	recall	f1-score	support
Normal	0.94	0.98	0.96	570
SQL Injection	0.87	0.79	0.83	318
XSS attack	0.85	0.88	0.86	362
accuracy			0.90	1250
macro avg	0.89	0.88	0.88	1250
weighted avg	0.90	0.90	0.90	1250

Figure 1 - Results for predicting attacks using sentiment sensor



The logical design for using sentiment analysis for cybersecurity education is illustrated in Figure 2. The presented flowchart describes the entire process of the system, starting from the starting point and ending with the completion process.

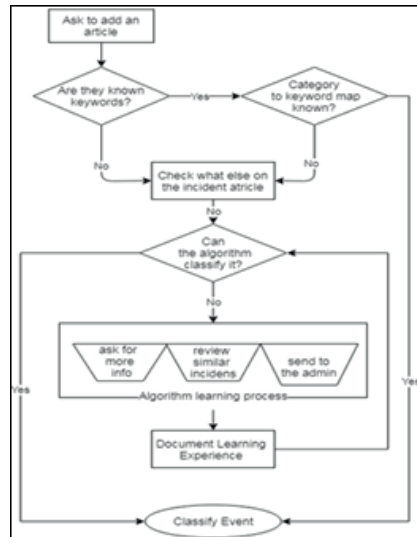


Figure 2 - Block-scheme of article adding process

The user starts the application process by rendering it. The system uploads articles and courses to the system. If the user is authorized, several functions are available: adding materials to favorites; participating in a discussion on a forum; and adding the user's article. Adding user materials is filtered and, provided that the material is suitable, the article is sent to the administrator for re-checking. If these processes are correct, the article is accepted, and the process ends. If the user is not authorized, the system displays only course materials and articles.

Figure 2 shows the article checking process by the system. First, a response to a user adding an article is sent to the system. If the article has known keywords and if the category of the keyword is known, the event is classified successfully. This process takes place during the filtering process. If there are no known keywords, the process starts to check for the presence of other components in the incident article and tries to categorize them. In case of failure, the learning process of the algorithm is started, which includes requesting additional information, viewing similar incidents, and sending them to the administrator. The learning process is then documented and stored.

The three main components that make up the system are the user interface or front-end, the server, and the database, as shown in the UML of the system in Figure 3. The user sends a request to the server through the interface to add a

new article. The request is processed and then saved in the database. After this stage, the server sends the result of the request to the user. The further extraction process is carried out by optimizing the user query using a genetic algorithm, which is extremely useful for rapidly growing databases.

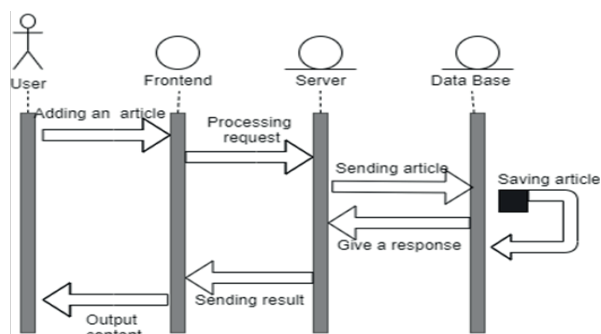


Figure 3 – UML diagram of the system

Structuring data. Analysing data requires it be in certain formats. In our case, the web logs were converted to data frames for easy analysis as shown in Figure 4. Dataframe columns refer to a host IP address, the date including day, month, year and time information, date offset, location in file system, status of success, and size.

	Host	Client_ID	User_Name	Date	Offset	Method	Loc	HTTPVer	Status_Code	Size
0	207.46.13.214	-	-	18/Aug/2020:02:33:55	-700	GET	/robots.txt	HTTP/1.1	200	377
1	117.107.147.153	-	-	18/Aug/2020:02:35:33	-700	GET	/	HTTP/1.1	200	12883
2	117.107.147.153	-	-	18/Aug/2020:02:35:34	-700	GET	/css/style.css	HTTP/1.1	404	328
3	172.245.228.28	-	-	18/Aug/2020:02:35:59	-700	GET	/	HTTP/1.1	200	12839
4	172.245.228.28	-	-	18/Aug/2020:02:36:00	-700	GET	/datasets%20description/pe_malware/virusshare...	HTTP/1.1	200	1607

Figure 4 – Data converted to a data frame

Results and discussion

Attacks launched on servers. XSS attacks work by embedding script tags in URLs/HTTP requests and tricking unsuspecting users into clicking on them. It ensures that malicious JavaScript is executed on the victim's machine (<https://www.sans.org/reading-room/whitepapers/logging/detecting-attacks-web-applications-log-files-2074>).

These attacks exploit the trust between the user and the server and the fact that there is no I/O validation on the server. Since it may reject JavaScript or other active code characters. Here we focus on the most common injection, SQL injection. For SQL injections to work, the attacker must go beyond the original SQL statement. This is usually done with a single quote (') or double dash (/). The single quote acts as a delimiter for the SQL query; double dash is a comment character in Oracle and MS SQL.

Feature building. Once the attacks were identified using regular expression patterns, they were compared against the corresponding data frames, resulting in a new column "State". It identifies the attack as "Xss_attack" OR "Sqli_attack" and normal as "Normal".

After creating new functions, i.e., URL_Length and number of parameters, the DataFrame was coded using Pandas layouts for method and state code columns. The logs were analysed in different ways and the results are reported according to the logs. We have applied a set of common algorithms, such as linear regression, support vector machine, etc and tested their performance.

It was observed that the Gradient Boosting and Random Forest Classifier performed the best algorithm for log attack classification.

Multiple accuracy metrics can be generated from the confusion matrix:

TP – True Positive; TN – True Negative; FP – False Positive; FN – False Negative.

- Accuracy = (all correct / all) = $TP + TN / TP + TN + FP + FN$. This is the general accuracy of the model. It is not useful, particularly when the classes are not equally represented i.e., the data is imbalanced in terms of the target classes.

- Precision = (true positives / predicted positives) = $TP / TP + FP$. Precision can be interpreted as the proportion of objects called positive by the classifier. At the same time, real positive, and recall shows what proportion of objects from all objects of a positive class the algorithm found.

- Recall = (true positives / all actual positives) = $TP / TP + FN$. Sensitivity states, out of all the actual malignant cases, how many cases were correctly identified as malignant. This is a class-level metric. Precision and Recall are critical class-level metrics. However, maximizing precision can pull down the value of recall, and vice versa (Shabbir, 2020).

Results according to the partially shown list in Figure 5 are: Normal – 90.6 %; SQL injection – 4.3 %; XSS attack – 5.1 %. It indicates that majority of the traffic belongs to the normal class, i.e. has no harm. However, remaining 9.4 % contain SQL injection and cross scripting patterns.

дата	HTTPver	Хост	логи	метод	Размер	Статус код	URL_Length	Результат
30-Jan-2020-02:42:36	HTTP/1.1	207.46.13.30	/self/logs/error.log.2018-04-14.gz	GET	4113	200	34	Normal
30-Jan-2020-02:42:40	HTTP/1.1	40.77.167.122	/self/logs/error.log.2018-10-10.gz	GET	5617	200	34	Normal
30-Jan-2020-02:42:57	HTTP/1.1	207.46.13.30	/self/logs/access.log.2019-03-03.gz	GET	24253	200	35	Normal
30-Jan-2020-02:46:08	HTTP/1.1	103.225.13.243	/security-data-analysis/lab_1/conn.log.zip	GET	14491960	200	42	Normal
30-Jan-2020-02:55:37	HTTP/1.0	46.149.86.51	/	GET	12623	200	1	Normal
30-Jan-2020-02:55:57	HTTP/1.0	46.149.86.51	/honeypot/honeybot%20-%20howto%20-%20to%20use%20a%20python.pdf	GET	1892040	200	56	SQL Injection
30-Jan-2020-02:56:21	HTTP/1.0	119.40.82.210	/ftp://download.telab.for.se/dataset/	GET	257	404	37	XSS attack
30-Jan-2020-02:57:52	HTTP/1.1	47.111.229.152	/static/gttignore	GET	294	404	18	Normal
30-Jan-2020-02:57:47	HTTP/1.1	47.111.229.152	/static/js/admin_common.js	GET	294	404	27	SQL Injection
30-Jan-2020-02:58:53	HTTP/1.1	157.55.39.6	/robots.txt	GET	288	200	11	Normal

Figure 5 - Plot of the score of attacks in log files

Conclusion

The main aim of the project was to utilize sentiment analysis for cyber security learning. It is done to enhance the awareness of the citizens who attain to be secured. The educational platform was created, and it contains articles and video demonstrations about how to be protected from the various type of network attacks. The authorized users are able to share their own articles, which are controlled by the sentiment analysis model and automating this process, thus it became reachable to find out suitable information.

The application can be used to improve efficiency in analysing cybersecurity in everyday situations. For example, to find the latest cybersecurity news and analyse log files for containing any attacks, such as SQL injection or XSS attack.

The web platforms become important avenues for social communication. Thus, the study can serve as an indicator for measuring the extent to which a consumer is truly concerned about the cybersecurity issues and an identifier of the online behaviour effect. These results can help cybersecurity researchers plan and use web platforms to increase user awareness of cybersecurity issues in the future.

REFERENCES

- Bermingham A., Smeaton A. (2020). On using twitter to monitor political sentiment and predict election results. In: SAAIP'11 1. Ella Koeze and Nathaniel Popper. The New York Times. [Electronic resource]. *The Virus Changed the way we Internet*. — April 2020. — URL: <https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html>
- Carnegie Mellon University. [Electronic resource]. *Coronavirus (COVID-19) Scams*. December, 2020. URL: <https://www.cmu.edu/iso/aware/coronavirus-alerts/coronavirus-scam-awareness.html>
- Meyer, "Detecting Attacks on Web Applications from Log Files," — 2008. [Online]. Available: <https://www.sans.org/reading-room/whitepapers/logging/detecting-attacks-web-applications-log-files-2074>.
- Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: WWW'10
- Shabbir Tayabali (December 11, 2020). [Electronic resource]: A simple guide to building a confusion matrix. URL: <https://blogs.oracle.com/ai-and-datascience/post/a-simple-guide-to-building-a-confusion-matrix>.



ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР
ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ
SMART SYSTEMS

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 67–81

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.007>

УДК 004.03; 658.785

МФТИ 20.15.05

**ANALYSIS OF THE NECESSITY OF A LOGISTICS SYSTEM AND SETTING
GOALS TO INCREASE EFFICIENCY**

© *A.C. Abdullayeva**, *A.V. Chan*, *A.K. Aitim*, 2022

Asel S. Abdullayeva — Master of Technical Sciences, Senior Lecturer of the "Information Systems" Department, International Information Technology University

ORCID: 0000-0001-5188-3008. E-mail: a.abdullayeva@iitu.edu.kz;

Alexandra V. Tian — Master of Economical Sciences, Lecturer of the "Information Systems" Department, International Information Technology University

ORCID: 0000-0001-5735-5791;

Aigerim K. Aitim — Master of Technical Sciences, Senior Lecturer of the "Information Systems" Department, International Information Technology University

ORCID: 0000-0003-2982-214X.

© A.C. Abdullayeva, A.V. Chan, A.K. Aitim, 2022

Abstract. This article deals with the analysis of the necessity of the development of the logistics system and analyzes the problems of disorder in the development of logistics enterprises, information platform needs and logistics information standard lag. Today it is impossible to imagine a production or trading company that is not engaged in solving logistics problems. The logistics system is a relatively stable set of links (structural / functional divisions of the company, as well as suppliers, consumers, and logistics intermediaries) interconnected and united by a single management of the corporate strategy of business organization. Logistics systems allow to optimize the functioning of commodity, information, and financial flows, significantly reducing the time interval between the acquisition of raw materials and semi-finished products and the delivery of the finished product to the consumer, contributes to a sharp reduction



in inventories. In the context of globalization, given the vastness of the territory of Kazakhstan, the competitiveness of the economy and the state will largely depend on the efficient operation of the transport and communications complex. The key to the competitiveness of domestic goods, services, and the economy as a whole is a high-tech transport infrastructure corresponding to the state transport and transit policy. Thus, the solution to the problem of improving the efficiency of the transport sector is impossible without the research to determine the future directions for the development of motor vehicles based on a systematic analysis of its state at the present stage.

Keywords: transportation, logistics, system, transport, supply chain, cargo, drivers, scheduling, accounting

For citation: А.А.С. Абдуллаева, А.В. Тянь, А.К. Айтим. Analysis of the necessity of a logistics system and setting goals to increase efficiency // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 67–81 (In Russ.). DOI: 10.54309/IJICT.2022.10.2.007

ЛОГИСТИКАЛЫҚ ЖҮЙЕНІҢ ҚАЖЕТТІЛІГІН ТАЛДАУ ЖӘНЕ ТИІМДІЛІКТІ АРТТЫРУ МАҚСАТТАРЫН БЕЛГІЛЕУ

А.С. Абдуллаева, А.В. Тянь, А.К. Айтим*

Абдуллаева Асель Сейдуллаевна — техника ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0001-5188-3008. E-mail: a.abdullayeva@iitu.edu.kz;

Тянь Александра Владимировна — экономика ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0001-5735-5791;

Әйтім Әйгерім Қайратқызы — техника ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0003-2982-214X.

© А.С. Абдуллаева, А.В. Тянь, А.К. Айтим, 2022

Аннотация. Бұл еңбек логистикалық жүйені дамыту қажеттілігін талдауға негізделген. Логистикалық кәсіпорындардың дамуындағы проблемалар, ақпараттық платформаға сұранысы мен логистикалық ақпараттардың стандартты кешеуілдеуі байқалады. Қазіргі таңда логистикалық мәселелерді шешумен айналысатын өндірісті немесе сауда компаниясын елестету мүмкін емес. Логистикалық жүйе ол бизнестің корпоративтік стратегиясының біртұтас басқаруымен біріктірілген және біріктірілген байланыстардың салыстырмалы тұрақты жиынтығы (компанияның құрылымдық-функционалдық бөлімдері, сондай-ақ жеткізушілер, тұтынушылар және логистикалық делдалдар). Осындай жүйелер тауардың, ақпараттың және қаржы ағындарының жұмысын оңтайландыруға мүмкіндік береді, «шикізат пен жартылай фабрикаттарды сатып алу мен дайын өнімді тұтынушыға жеткізу арасындағы уақыт аралығын едәуір азайтады, материалдық қорлардың күрт қысқаруына ықпал етеді». Жаһандану жағдайында Қазақстан аумағының кең



ауқымын ескере отырып, экономиканың және мемлекеттің бәсекеге қабілеттілігі көліктік-коммуникациялық кешеннің тиімді жұмысына тәуелді болады. Отандық тауарлардың, қызметтердің және тұтастай алғанда экономиканың бәсекеге қабілеттілігінің кілті — мемлекеттік көліктік және транзиттік саясатқа сәйкес жоғары технологиялық көлік инфрақұрылымы. Осылайша, көлік секторының тиімділігін арттыру мәселесін шешу қазіргі кезеңде оның жағдайын жүйелі түрде талдау негізінде автокөлік құралдарын дамытудың болашақ бағыттарын айқындау үшін зерттеулерді нығайту ең өзекті мәселелердің бірі екеніне көз жеткіздік.

Түйін сөздер: тасымалдау, логистика, жүйе, көлік, жеткізу тізбегі, жүк, жүргізушілер, диспетчерлеу, бухгалтерлік есеп

Дәйексөз үшін: А.С. Абдуллаева, А.В. Тян, А.К. Айтим. Логистикалық жүйенің қажеттілігін талдау және тиімділікті арттыру мақсаттарын белгілеу // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 67–81 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.007.

АНАЛИЗ НЕОБХОДИМОСТИ ЛОГИСТИЧЕСКОЙ СИСТЕМЫ И ПОСТАНОВКА ЗАДАЧ ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ

А.С. Абдуллаева, А.В. Тян, А.К. Айтим*

Абдуллаева Асель Сейдуллаевна — магистр технических наук, сениор-лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0001-5188-3008. E-mail: a.abdullayeva@iitu.edu.kz;

Тян Александра Владимировна — магистр экономических наук, лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0001-5735-5791;

Айтим Айгерим Кайратовна — магистр технических наук, сениор-лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0003-2982-214X.

© А.С. Абдуллаева, А.В. Тян, А.К. Айтим, 2022

Аннотация. Эта статья основана на анализе необходимости разработки логистической системы. Анализируются проблемы в развитии логистических предприятий, потребности информационной платформы, стандартное отставание логистической информации. Сегодня невозможно представить производственную или коммерческую компанию, которая не занимается решением логистических задач. Логистическая система представляет собой относительно стабильный набор отношений (бизнес/функциональные подразделения, а также поставщики, потребители и логистические посредники), которые взаимосвязаны и объединены общим управлением бизнес-стратегией компании. Эти системы оптимизируют работу товарных, информационных и финансовых потоков, «значительно сокращают время между закупкой сырья и полуфабрикатов и доставкой готовой продукции потребителю, помимо того, что способствуют резкому сокращению



товарно-материальных запасов». Учитывая обширность территории Казахстана, в условиях глобализации, конкурентоспособность экономики и страны во многом будет зависеть от эффективного функционирования транспортно-коммуникационного комплекса. Залогом конкурентоспособности национальных товаров и услуг и экономики в целом является высокотехнологичная транспортная инфраструктура, увязанная с транспортно-транзитной политикой страны. Таким образом, решение проблемы повышения эффективности транспортного сектора невозможно без усиления исследований по определению будущих направлений развития автотранспортных средств на основе систематического анализа его состояния на современном этапе.

Ключевые слова: транспортировка, логистика, система, транспорт, цепочка поставок, груз, водители, диспетчеризация, учет

Для цитирования: А.С. Абдуллаева, А.В. Тянь, А.К. Айтим. Анализ необходимости логистической системы и постановка задач повышение эффективности // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 67–81 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.007.

Введение

Автомобильный транспорт играет важную роль в транспортном комплексе страны, обслуживая в первую очередь предприятия с различной структурой собственности, а также потребности населения страны. Более 80 % грузовых перевозок ежегодно осуществляется автомобильным транспортом, а более 75 % пассажирских перевозок осуществляется общественным транспортом.

Автомобильный транспорт является крупнейшим потребителем ресурсов, потребляемых транспортным комплексом: 66 % нефтепродуктов, 70 % рабочей силы и около половины всех инвестиций. Одной из основных задач управления цепями поставок является обеспечение бесперебойной и эффективной работы на производственной стороне цепи. По мере усложнения производственных процессов, включающих, например, большее количество продуктов и распределенных производственных площадок, проектирование и управление логистической инфраструктурой становится все более важным для эффективности всей цепи поставок. Лицам, принимающим решения, необходимо оценить и сравнить альтернативные логистические решения, принимая во внимание не только их прямые затраты и выгоды, но и, в частности, их и, прежде всего, их влияние на эффективность бизнеса. Например, основной целью транспортной логистики является обеспечение своевременной доставки сырья и промежуточных товаров на производственные площадки с целью минимизации производственных потерь из-за отсутствия/избытка запасов на различных предприятиях.

Когда транспортные ресурсы запланированы и расписаны с точки зрения количества, требуемой мощности и маршрутов, относительно легко оценить прямые транспортные затраты с точки зрения стоимости ресурсов в час/день и фактического времени транспортировки. Однако гораздо более важным с точки



зрения цепочки поставок является оценка потенциальных производственных потерь, которые определяются стохастической изменчивостью темпов производства на заводе и временем прибытия каждого ресурса на конкретные производственные участки в зависимости от имеющихся на каждом участке складских мощностей. В некоторых случаях стоимость таких производственных потерь может значительно превышать прямые транспортные расходы (Бруccione и др., 2003).

В настоящее время на фоне структурной реформы предложения, развитие традиционной логистической отрасли является неустойчивым. Снижение логистических затрат стало важным вопросом экономического развития. Чтобы обеспечить более здоровое, стабильное и устойчивое развитие логистической отрасли в информационную эпоху и значительно повысить конкурентоспособность логистических предприятий на рынке, необходимо ускорить создание и внедрение передовой техники и технологий, увеличить обновление подвижного состава и других технических средств, укрепить материально-техническую базу, развивать транспортную инфраструктуру и сервисные объекты.

Интенсификация производства, экономия всех видов ресурсов и повышение производительности труда - задачи, напрямую связанные с автомобильным транспортом и его подсистемой — технической эксплуатацией транспортных средств (ТОТ), которая обеспечивает работоспособность автомобильного парка. Его развитие и совершенствование диктуется интенсивностью развития самого автомобильного транспорта и его ролью в транспортном комплексе страны, необходимостью экономии трудовых, материальных, энергетических и других ресурсов при перевозках, техническом обслуживании, ремонте и хранении транспортных средств, необходимостью обеспечения надежности транспортного процесса подвижным составом, защиты населения, персонала и окружающей среды (Беккер и др., 2010).

Необходимость комплексного системного подхода к проблеме формирования транспортно-логистических структур и принятия инвестиционных решений в цепи поставок товаров с участием автомобильного транспорта, а также потребность в новых теоретических и практических обобщениях и оптимизации грузовых перевозок в региональном аспекте обусловили актуальность проведенного поиска. Рекомендации, вытекающие из проведенного научно-теоретического и практического исследования, могут быть использованы государственными и экономическими органами для разработки методов регулирования инвестиционных процессов, регулирования деятельности и совершенствования работы транспортно-логистических структур по управлению цепью поставок товаров, реформирования автомобильного транспорта с учетом региональных особенностей страны, разработки и совершенствования законодательства о транспортно-логистической деятельности.

Материалы и методы

Необходимость логистической информационной системы. В цепи поставок товаров с участием автомобильного транспорта взаимодействуют различные структуры транспортно-логистической системы. Логистическая информационная



платформа—этопредоставляющаяуслугипообменуисовместномуиспользованию информации для всех типов пользователей открытая сетевая информационная система. В процессе развития логистики построение информационной платформы стало необходимым воплощением ключевой компетенции (Йоханссон и др., 2010).

Согласно анализу, в настоящее время некоторые ведущие логистические предприятия в нашей стране изначально создали интеллектуальную логистическую систему посредством инноваций информационных технологий, но логистических предприятий, полностью внедривших информационные технологии, относительно немного. Прежде всего, малым и средним логистическим предприятиям сложнее получить эффективные продукты и техническую поддержку, они остаются только на этапе выпуска и получения информации, роль мониторинга и аудита информации не сыграна, а роль платформы ограничена, в результате чего логистическая информация разбросана, фрагментирована. Затрудненное совместное использование информационных ресурсов, предприятия не могут своевременно получить важную информацию для своего развития, не могут в полной мере использовать преимущества информации в логистической отрасли, что приводит к увеличению общих логистических затрат.

Во-вторых, умная логистика в основном основана на информационной системе, и предприятия не могут получить информацию о собственной благоприятной политике, что приведет к трудностям в реализации связанных связей, и трудно сформировать интеграцию ресурсов. С точки зрения общего развития логистической отрасли, большинству предприятий все еще не хватает общей оптимизации информационной системы компании, и часть функций не полностью реализована, например, по данным, взятым из журнала "Transportation on Systems Engineering and Information Technology" большое количество логистических ресурсов простаивает, в Китае частота пустых грузовых автомобилей превышает 30 %, доля вакантных мест на складе составляет около 15 %. Инкрементные ресурсы ограничены асимметричной информацией, и трудно создать новые требования (Вэй Куи).

Процесс транспортировки - очень сложная и ответственная операция. Так как в этот процесс вкладываются большие деньги и задействовано большое количество людей. Технологическая схема процесса перевозки грузов одним видом транспорта (рис.1).



Рисунок 1- Процесс транспортировки груза

Учитывая значительные объемы перевозимой продукции, перевозки осуществляются в основном на автомобилях: необходимо управлять целым комплексом, определяя количество автомобилей, их индивидуальные мощности, маршруты, планируя их поездки в соответствии с объемами продукции и требованиями к перевозкам (совместимость и т.д.). В этом контексте первой целью транспортной логистики является максимизация использования мощностей автотранспорта путем определения задач автотранспорта, связанных с одновременной транспортировкой нескольких продуктов, где могут иметь место как погрузочные, так и разгрузочные операции.

Второй задачей является минимизация изменчивости времени доставки/забора груза при сохранении транспортных расходов в рамках заданного бюджета. В идеале, по мере увеличения количества автомобилей, задействованных на данном маршруте, при неизменной общей пропускной способности, влияние временных задержек для одного конкретного транспорта становится все менее значительным с точки зрения производительности всей производственной системы. Однако соответствующие транспортные расходы значительно возрастают, поскольку экономия от масштаба обменивается на меньшую вариативность процесса доставки/забора: необходимо помнить, что стоимость аренды автомобилей, как и стоимость топлива, не изменяется линейно в зависимости от мощности. Сложность компромисса между загрузкой мощностей, транспортными расходами и изменчивостью процесса приводит к выявлению множества логистических решений, которые можно оценить с точки зрения общих затрат, включая как фактические логистические расходы, так и ожидаемые производственные потери. Для того чтобы рассчитать весь этот процесс, нам нужна система. Однако существующие решения для планирования и планирования логистики, лишь частично покрывающие современные потребности компаний (например, SAP и Oracle), слишком дороги (для покупки и обслуживания) для малых и средних компаний. Поэтому важно создать отдельные модули, которые будут работать по назначению (Международная конференция по Интернету и электронному бизнесу 2018).

Недостатки логистической системы Казахстана. Эксперт консалтинговой компании Logistics Systems отмечает, что "неразвитость логистической инфраструктуры оказывает значительное влияние на экономическое развитие Казахстана сегодня. В настоящее время сектор логистики сталкивается с очень важными проблемами, такими как: — обеспечение доступности и конкурентоспособности услуг, увеличение мощностей и устранение дисбалансов; — создание интегрированной системы логистических центров и обеспечение информационной поддержки. Логистические затраты в Казахстане очень высоки и зачастую выше, чем в развитых странах. Доля логистических затрат в себестоимости продукции достигает 25 %, что значительно выше среднемирового показателя в 11 % и 10 % в Канаде и США. Динамика экономического роста требует развития транспортной системы для удовлетворения логистических потребностей новой экономики. Казахстан остро нуждается в развитой



транспортно-логистической системе, способной обслуживать экономику страны. В логистике хорошо известно, что эффективность перевозок определяется тремя факторами: Стоимость, скорость и непрерывность. Все эти критерии плохо подходят для железнодорожного транспорта (Городецкий и др., 2012).

Низкая эффективность национального оператора в виде нехватки парка и нерегулярных поставок, а также постоянный рост тарифов являются реальным препятствием для развития производителей и экономики в целом. Если раньше нехватка вагонов была сезонной проблемой, то сегодня это почти ежедневная проблема. За этот период объем казахстанских производителей вырос в несколько раз, как и ожидания потребителей, которые мы теряем из-за нерегулярных поставок в пользу иностранных производителей.

Транзитный потенциал Казахстана следует рассматривать как точку экономического роста страны. Поэтому необходимо добиться того, чтобы СНГ стало более привлекательным и имело самую современную и эффективную транспортно-логистическую систему в СНГ. Необходимо также улучшить транспортную деятельность на всех видах транспорта с целью предоставления широкого спектра услуг по конкурентоспособным ценам; дальнейшее совершенствование коридоров для транзитных потоков товаров на постоянной линейной основе, четко определяя транзитное время, стоимость и систематическое использование этих коридоров; организация и

В целом, современная логистика в Казахстане должна представлять собой интегрированный комплекс, призванный обеспечить необходимый объем количественных и качественных услуг для бесперебойной транспортировки всех товаров и услуг во всех направлениях и на все расстояния. Кроме того, как и другие инфраструктурные системы, логистическая система должна способствовать повышению конкурентоспособности отечественной продукции за счет сокращения времени доставки, снижения затрат, диверсификации направлений и т.д. В то же время логистика должна оставаться прибыльной и самодостаточной, с высоким уровнем реинвестирования в ее развитие.

Прибыль и расходы на перевозку грузов. Ускорение исследований и разработки логистических информационных технологий, стандартов безопасности, управления и обслуживания, а также продвижение стандартизации соответствующих информационных технологий, управления и услуг является важной мерой для продвижения логистики. Стандартизация и стандартизация интерфейса данных делают обмен и сотрудничество между предприятиями более удобным и быстрым. Например, создание таких модулей, как учет работы водителей, даст возможность рассчитать затраты, которые компания должна распределить.

С учетом контекста применения, описанного в предыдущем разделе, тактическая миссия определяется как транспортный маршрут, состоящий из определенной последовательности пунктов и заданных групп совместимых количеств продукции, которые должны быть загружены и разгружены в каждом пункте последовательности. Логистическая эффективность каждой тактической миссии может быть оценена как совокупная эффективность каждой отдельной



авто миссии для всех авто, выделенных на этот маршрут. Фактическая стоимость каждой авто задачи определяется пропускной способностью автомобилей, временем навигации, ожиданием и временем выполнения операций (Киздарбекова, 2013).

Другим релевантным показателем эффективности является фактическая загрузка мощностей, которая может быть выражена в виде затрат как стоимость неиспользованных мощностей. Поскольку конечной целью транспортной логистики является обеспечение множества производственных процессов, географически распределенных по нескольким производственным площадкам, определенным количеством сырья и промежуточных продуктов, третья мера эффективности может быть построена на основе так называемого риска миссии, который определяется как совокупная вероятность того, что определенная миссия приведет к возникновению ситуаций нехватки или избытка запасов на любом из задействованных складов, с соответствующими производственными потерями и/или дополнительными затратами из-за неучтенных задержек и времени ожидания.

Распределение ресурсов — это непрерывный процесс принятия решений в реальном времени, в котором критерии меняются "на лету". Чтобы иметь возможность правильно рассчитать время, прибыль, затраты, а также сформировать правильное расписание для клиентов и водителей.

Эффективный планировщик автомобильных перевозок должен обрабатывать транспортные инструкции (ТИ) от множества различных точек погрузки до множества различных пунктов назначения (например, мест расположения клиентов и кросс-доков, где грузы разгружаются и консолидируются) и множество различных маршрутов, по которым могут быть доставлены заказы. Выбор наилучшего маршрута на основе консолидации или других критериев называется динамической маршрутизацией. Планировщик также должен уметь распределять грузы разных размеров и веса по разным типам грузовиков и прицепов; учитывать предпочтения владельцев, операторов и водителей и вписывать расписание в многочисленные ограничения, накладываемые часами работы склада, правилами работы водителей, правилами безопасности и политикой предприятия, например, по выбору между собственным парком и сторонними перевозчиками. Разные компании имеют разные критические ограничения, например разрешение отменить временные или другие ограничения для достижения более эффективного расписания. Составленное расписание должно быть не только выполнимым, но и эффективным, т.е. должны быть найдены возможности для обратных перевозок и консолидации.

Сложность также определяется количеством и разнообразием заказов (и других событий, влияющих на составление расписания) в день, а также количеством и разнообразием транспортных ресурсов, таких как грузовики.

Кроме того, от планировщика ожидается, что он будет быстро изменять график выполнения заказов и транспортных ресурсов, пострадавших от неожиданных событий, таких как поступление новых заказов, отмены, сбои, плохие погодные условия, дорожные работы и неявка водителей или погрузочных бригад (Coyle и др., 2013).

На этапе планирования заказы назначаются грузовику и строится его маршрут. На этом этапе заказы могут быть добавлены или удалены, а маршрут, запланированный для грузовика, может быть изменен в результате последующих событий.

В определенный момент планировщик должен зафиксировать грузовик. Это инициирует связь со складами, планировщиками для смены водителей, обслуживания грузовиков и т.д., чтобы подготовить грузовик к поездке. На этом этапе изменения в расписании грузовиков нежелательны, так как это может повлиять на работу склада, назначение водителя и т.д.

Этап выполнения начинается с выполнения водителем проверок перед поездкой и продолжается до подведения итогов в конце смены. На этом этапе потребуется высокий уровень сложности для изменения графика движения грузовиков во время перевозки. Стоит помнить, что при перевозке грузов водитель играет большую роль, так как расписание составляется на основе водителей. Соответственно, расчет заработной платы водителей основывается на графике.

После обширного обмена информацией возникает финансовый поток: клиент платит за товар и его доставку. Информационная система должна связать информационные потоки воедино, начиная с заявки на доставку и заканчивая расчетом заработной платы водителей. (рис.2).



Рисунок 2 - Фрагмент схемы взаимосвязи документов информационной системы

В современных условиях возрастает важность обеспечения дистанционного обслуживания клиентов, переход на автоматизированную систему электронной обработки различных видов документов, необходимых для осуществления перевозок (Дж. Химофф и др., 2006).

Международное соглашение "Европейское соглашение, касающееся работы экипажей транспортных средств, осуществляющих международные перевозки" (ЕСТР), к которому Республика Казахстан присоединилась Указом Президента Республики Казахстан № 2272 от 12 мая 1995 года, закрепило широкое распространение тахографов, с помощью которых фиксируется рабочее время

водителя. Для транспортных средств стран, не являющихся членами ЕСТР, въезд в Европу без тахографов был запрещен (Джейн Грей, 2012).

Сравнительный анализ аналогов. Был проведен полный анализ и изучена работа конкурентов. Есть две таблицы. В таблице 1 приведены технические характеристики программного обеспечения, а в таблице 2 — функциональные характеристики системы.

Таблица 1- Технические характеристики программного обеспечения

Название программного обеспечения	Производитель	Дата первой публичной версии	Последний релиз	Стоимость	Лицензия
1С Управление транспортом	Форест	декабрь 2015 г.	8.4	Платно без дополнительной оплаты платил 600\$	GPL
Автодоставка	Экспресс Компания	март 2011 г.	4	Платно 150\$	GPL
Транспортный менеджер	Компания Транс_менеджер	июль 2017 г.	6.1	Платно 50\$-125\$ зависит от класса (облегченный, стандартный, профессиональный)	GPL
Смарт-Логистика	ООО «Смарт Логистикс»	Февраль 2014	7.2	Платно 58\$-169\$	Собственный
Ogacle Джей Ди Эдвардс	Оракл	сентябрь 2010 г.	9.4	Платно 1000\$	Собственный
Прологиста	Компания «Транко Логистик»	Декабрь 2017 г.	1.2	Бесплатно	Условно-бесплатная

Таблица 2 – Функциональные характеристики системы

Название программного обеспечения	Бухгалтерская накладная	Учет водителя	Отслеживание оборудования	Обслуживание	Расчет заработной платы по рабочему времени	Автоматизировать заполнение д по ИНН	Бухгалтерия завода	Создание отчетов
1С Управление транспортом	+	+	+	+	+	+	+	+
Автодоставка	-	+	+	-	-	-	+	-
Транспортный менеджер	-	-	-	+	-	-	+	-
СмартЛогистика	+	-	+	-	-	-	-	-
Ogacle Джей Ди Эдвардс	-	+	-	-	+	+	+	+
Прологиста	-	-	-	+	-	+	-	-
Новый проект	+	+	+	-	+	-	+	+

Учетная накладная означает процесс оформления автомобиля и установок. В путевом листе указан пробег, количество часов, заправка и остаток топлива при возврате. С помощью этой функции расход топлива рассчитывается автоматически на основе пробега и часов работы, а также норм расхода топлива.

Учет водительских часов означает фиксацию рабочего времени водителя, то есть того, сколько времени он потратил на перевозку.

Планирование технического обслуживания — это планирование технического обслуживания каждого транспорта, то есть когда и какой транспорт должен пройти техосмотр.

Расчет заработной платы по отработанному времени - данный функционал основан на учете работы водителей. Он учитывает, сколько часов отработал водитель и сколько работодатель платит за каждый час. И по соответствующим формулам рассчитывается месячная зарплата водителя.

Автоматизировать заполнение д по ИНН — это означает, что информация может быть добавлена автоматически, если в базе данных был ИНН работника. Таким образом, вам не нужно заполнять все вручную

Учет работы установки означает учет агрегатов, запасных частей, выраженный в пробеге или часах. При приближении к завершению срока службы программа сигнализирует о необходимости замены установок, агрегатов, запасных частей.

Создание отчетов означает функцию автоматического формирования отчетов по всем видам учета. В отчетах отражаются остатки на складах, ресурсы и производство запасных частей и комплектующих. Также реализован отчет о расходе топлива по норме и по факту. Реализован журнал накладных. Отчет о наличии автомобилей в парке. Отчет по талонам в разрезе литров и штук.

Сравнивая все программы, можно сделать вывод, что 1С Форес: Учет автотранспорта является наиболее эффективным инструментом для логистической отрасли. Итак, есть отчеты и автоматизированный процесс подсчета работы. Этот софт удобен для крупных логистических компаний с большими данными.

Если смотреть на цену, то бюджетным вариантом будет AUTOsis, но в этом софте нет многих функций, которые нам понадобились бы при большом количестве данных. Поэтому крупные компании не могут с ним работать.

В качестве научной базы, активно развивающейся в последние годы, посвящено большое количество публикаций: от тезисов конференций до учебников и монографий. Известно, что внедрение принципов логистики в управление материальными потоками позволяет снизить уровень запасов на 30–50 % и сократить время движения продукции на 25–45 %.

Основными причинами быстрого роста автомобильных перевозок являются присущая гибкость доставки «от двери до двери» и высокоскоростные междугородние перевозки. Именно поэтому современная логистика в Казахстане должна представлять собой интегрированный комплекс, призванный оказывать необходимый объем количественных и качественных услуг, обеспечивать беспрепятственную транспортировку в любом направлении, на любое расстояние любых товаров и услуг. Кроме того, как и другие инфраструктурные

системы, логистическая система должна способствовать повышению конкурентоспособности отечественной продукции за счет сокращения сроков доставки, снижения себестоимости, диверсификации направлений и т.д. При этом логистика должна оставаться прибыльной, самоокупаемой с высоким уровнем реинвестирования в собственное развитие. Для достижения этой цели мы поставили задачу разработать модуль, который позволяет:

- Сохранить информацию о водителях
- Предварительно рассчитать расстояния
- Рассчитать заработную плату на основе информации о расстоянии и часов.

Это будет теоретически и экспериментально обоснованная система.

Заключение

Обзор практики и теории грузовых автомобильных перевозок показал, что постановка, условия и критерии, методы решения задачи устанавливались собственником — государством, а потому реализовывались в условиях централизованного управления народным хозяйством и в его интересы. Перевозка грузов ведомственным транспортом была организована аналогично работе общественного транспорта.

Анализ исследований показал, что оперативное планирование таких перевозок, основанный на учете большого количества факторов технологического и организационного порядка, также остается малоизученным. Транзитный потенциал Казахстана следует рассматривать как точку экономического роста страны. Для этого необходимо обеспечить повышение привлекательности и создание самой современной эффективной транспортно-логистической системы в СНГ. А также совершенствование транспортных операций на любом виде транспорта с учетом оказания широкого спектра услуг, предлагая конкурентоспособные тарифы. Для достижения этой цели были сформированы такие задачи как;

- формирование эффективной системы анализа, регулирования и мониторинга рынка транспортно-логистических услуг, повышение качества обслуживания клиентов транспортно-логистической сферы;
- достижение мировых стандартов транспортной инфраструктуры и согласование нормативно-правовой базы и систем контроля в сфере логистики;
- укрепление единого экономического пространства и развитие межрегиональных связей;
- реализация единой инвестиционно-экономической политики в сфере транспортно-логистических услуг, развитие инфраструктуры транспортной системы страны и отдельного региона.

Повышение эффективности транспортной системы сделает ее конкурентоспособным элементом казахстанской экономики. Значительно возрастут пассажирские и грузовые перевозки.

Кроме того, как и другие инфраструктурные системы, логистическая система должна способствовать повышению конкурентоспособности отечественной продукции за счет сокращения сроков и стоимости доставки, диверсификации направлений и т. д. При этом логистика должна оставаться прибыльной, самоокупаемой с высоким уровнем реинвестирования в собственное развитие.



Сегодня правительства многих стран делают ставку на развитие логистики, которая является основным фактором стимулирования экономического развития. Поэтому развитие логистических систем является важнейшей задачей для всех стран мира.

ЛИТЕРАТУРЫ

Р.С. Айтеленов (2013). Развитие транспортно-логистических услуг в Казахстане. Журнал "Вестник КазНПУ". — Алматы, 2013.

Аникин Б.А. (2017). Коммерческая логистика: учебник / Б.А. Аникин, А.П. Тяпукхин. — Москва: Проспект, —2017. — 426 с.

Бруццоне А., Орсоли А. (2003). "Методы ИИ и моделирования для оценки логистической эффективности цепи поставок". 36-й ежегодный симпозиум по моделированию. Орландо, штат Флорида, — США 2003.

Беккер М. Веннинг Б.-Л. Гёрг К. Йедерманн Р. Тимм-Гил А. (2010). "Логистические приложения с беспроводными сенсорными сетями". In: Proc. of HotEmNets 2010.

Б. Йоханссон, С. Джайн, Дж. Монтойя-Торрес, Дж. Хуган и Э. Ючесан (2010). "Расширенная среда возможностей анализа логистики". Зимняя конференция по моделированию 2010.

Вэй Куи (2018). "Исследование проблем и контрмер развития умной логистики в Китае". Международная конференция по Интернету и электронному бизнесу 2018.

В. Городецкий, О. Карсаев, В. Конюший, В. Самойлов (2012). "Транспортные логистические сервисы из облака". IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology 2012.

Киздарбекова М. (2013). Проблемы и перспективы развития логистики в Казахстане. Журнал "Молодой ученый". — Алматы, 2013.

J.J. Coyle, C.J. Langley Jr., B.J. Gibson, R.A. Novack, and E.J. Bardi (2013). Supply Chain Management: A Logistics Perspective, 9th ed. Мейсон, Огайо: South-Western Cengage Learning, — 2013.

Дж. Химофф, П. Скобелев, Г. Ржевски (2006). "Технология MAGENTA мультиагентной логистики i-Scheduler для автомобильных перевозок". Конференция: 5-я Международная объединенная конференция по автономным агентам и мультиагентным системам, Хакодате. — Япония, 2006.

Джейн Грей (2012). "Перевозка удовольствия: Снижение затрат в производственной цепочке поставок", The Manufacturer. — 16 марта 2012 г.

Международное соглашение "Европейское соглашение, касающееся работы экипажей транспортных средств, осуществляющих международные перевозки", — 1970.

Логистика: интегрированная цепь поставок / Доналд Дж. Бауэрсокс, Дэйвид Дж. Клосс. — Москва: Олимп-Бизнес, 2017. — 635 с.

Логистика: учебник для академического бакалавриата / Ю.М. Неруш, А.Ю. Неруш. — Москва: Юрайт, 2017. — 558 с.

Логистика: учебно-методическое пособие: [для студентов, аспирантов и слушателей экономических специальностей] / О.М. Овечкина. — Минск: Амалфея, 2020. — 214 с.

Логистика: учебное пособие / Р.Б. Ивуть. — Минск: БНТУ, 2021. — 461 с.

Логистика во внешнеэкономической деятельности: учебное пособие / Е.С. Аكوпова, Т.Е. Евтодиева. — Москва: Русайнс, 2020. — 189 с.

Общий курс транспортной логистики: учебное пособие / Л.С. Фёдоров, В.А. Персианов, И.Б. Мухаметдинов. — 2-е изд., стереотипное. — Москва: КноРус, 2020. — 309 с.

REFERENCES

Aitelenov R.S. (2013). Development of transport and logistics services in Kazakhstan. Journal "Bulletin KazNPU". — Almaty, 2013.

Anikin B.A. (2017). Commercial logistics: textbook / Anikin B.A., Tyapukhin A.P. — Moscow: Prospect, 2017. — 426 p.

Bruzzzone A., Orsoni A. (2003). "AI and simulation-based techniques for the assessment of supply chain logistic performance." 36th Annual Simulation Symposium. Orlando, FL, — USA 2003.



- Becker M. Wenning B., Görg L., Jedermann C., Timm-Giel R. (2010). "Logistic applications with Wireless Sensor Networks". In: Proc. of HotEmNets 2010.
- Coyle J.J., Langley Jr C.J., Gibson B.J., Novack R.A., Bardi E.J. (2013). *Supply Chain Management: A Logistics Perspective*, 9th ed. Mason, Ohio: South-Western Cengage Learning, 2013.
- Johansson B., Jain S., Montoya-Torres J., Hagan J., Yücesan E. (2010). "Advanced logistics analysis capabilities environment". Winter Simulation Conference 2010.
- Jane Gray (2012). "A Transport of Delight: Reducing Costs in the Manufacturing Supply Chain," *The Manufacturer*. — March 16, 2012.
- Wei Cui. "Study on Problems and Countermeasures of Smart Logistics Development in China". International Conference on Internet and e-Business 2018.
- Gorodetsky V., Karsaev O., Konyushiy V., Samoylov V. (2012). "Transportation Logistics Services from Cloud". IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology 2012.
- Kizdarbekova M. Problems and Prospects for the Development of Logistics in Kazakhstan. *Journal "Young Scientist"*. — Almaty, 2013.
- Himoff J., Skobelev P., Rzevski G. (2006). "MAGENTA technology multi-agent logistics i-Scheduler for road transportation". Conference: 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate. — Japan, 2006.
- The international agreement "European Agreement Concerning the Work of the Crew of Vehicles Carrying International Transportation", 1970.
- Logistics: an integrated supply chain / Donald J. Bowersox, David J. Kloss. — Moscow: Olymp-Business, 2017. — 635 p.
- Logistics: a textbook for academic undergraduate studies / Yu.M. Nerush, A.Yu. Nerush. — Moscow: Yurayt, 2017. — 558 p.
- Logistics: teaching aid: [for students, graduate students and students of economic specialties] / O.M. Ovechkina. — Minsk: Amalfeya, 2020. — 214 p.
- Logistics: textbook / R.B. Ivut. — Minsk: BNTU, 2021. — 461 p.
- Logistics in foreign economic activity: textbook / E.S. Akopova, T.E. Evtodieva. — Moscow: Rusajns, 2020. — 189 p.
- General course of transport logistics: textbook / L.S. Fedorov, V.A. Persianov, I.B. Mukhametdinov. — 2nd ed., stereotypical. - Moscow: KnoRus, 2020. — 309 p.



INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 82–91

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.008>

UDK 004.91

MODELS OF NATURAL LANGUAGE PROCESSING FOR IMPROVING SEMANTIC SEARCH RESULTS

A.K. Aitim

Aigerim K. Aitim — Master of Technical Sciences, Senior Lecturer of the "Information Systems" Department, International Information Technology University
ORCID: 0000-0003-2982-214X. E-mail: a.aitim@iitu.edu.kz.

© A.K. Aitim, 2022

Abstract. Most people rely on search engines to get and share information from all sorts of resources. All totals returned by search engines are not always important because they are drawn from heterogeneous data sources. Moreover, it is not easy for a user to prove that the acquired results are relevant for the request. As a result, the semantic network plays a significant role in interpreting the relevance of search results. This paper proposes a new method for searching for appropriate documents using the semantic web, based on the concept of natural language processing (NLP). In the proposed system, NLP is used to analyze a user request in terms of parts of speech. The extracted definitions are compared with a domain dictionary to identify the appropriate domain of the user's interest. The retrieved user request papers are examined with natural language processing support to identify the respective domain.

Keywords: semantic search engine, natural language processing, context analysis, information retrieval systems, graphematic analysis, ontology

For citation: A.K. Aitim. Models of natural language processing for improving semantic search results // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 82–91 (In Russ.). DOI: [10.54309/IJICT.2022.10.2.008](https://doi.org/10.54309/IJICT.2022.10.2.008).



СЕМАНТИКАЛЫҚ ІЗДЕУ НӘТИЖЕЛЕРІН ЖЕТІЛДІРУ ҮШІН ТАБИҒИ ТІЛДЕРДІ ӨНДЕУ МОДЕЛДЕРІ

Ә.Қ. Әйтiм

Әйтiм Әйгерiм Қайратқызы — техника ғылымдарының магистрі, «Ақпараттық жүйелер» кафедрасының аға оқытушысы, Халықаралық ақпараттық технологиялар университеті
ORCID: 0000-0003-2982-214X. E-mail: a.aitim@iitu.edu.kz.

© Ә.Қ. Әйтiм, 2022

Аннотация. Көптеген адамдар әртүрлі ресурстардан ақпаратты алу және бөлісу үшін іздеу жүйелеріне сүйенеді. Іздеу жүйелері қайтарған барлық нәтижелер жаңартылмайды, өйткені олар әртүрлі деректер көздерінен алынған. Сонымен қатар, алынған нәтижелердің пайдаланушы сұранысы үшін маңызды екенін растау қиын. Сондықтан семантикалық веб іздеу нәтижелерінің өзектілігін түсіндіруде маңызды рөл атқарады. Бұл мақалада табиғи тілді өңдеу тұжырымдамасына негізделген семантикалық вебті пайдалана отырып, сәйкес құжаттарды іздеу үлгісі ұсынылады. Бұл жүйеде қолданушы сұранысын сөйлеу бөліктері бойынша талдау үшін қолданылады. Шығарылған терминдер пайдаланушыны қызықтыратын сәйкес доменді анықтау үшін домен сөздігімен салыстырылады. Екінші жағынан, алынған пайдаланушы сұрауының құжаттары сәйкес доменді анықтау үшін табиғи тілді өңдеу арқылы зерттеледі.

Түйін сөздер: семантикалық іздеу жүйесі, табиғи тілді өңдеу, контекстік талдау, ақпаратты іздеу жүйелері, графематикалық талдау, онтология

Дәйексөз үшін: Ә.Қ. Әйтiм. Семантикалық іздеу нәтижелерін жетілдіру үшін табиғи тілдерді өңдеу моделдері // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 82–91 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.008.

МОДЕЛИ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ УЛУЧШЕНИЯ СЕМАНТИЧЕСКИХ РЕЗУЛЬТАТОВ ПОИСКА

А.К. Айтим

Айтим Айгерим Кайратовна — магистр технических наук, сениор-лектор кафедры «Информационных систем», Международный университет информационных технологий
ORCID: 0000-0003-2982-214X. E-mail: a.aitim@iitu.edu.kz.

© А.К. Айтим, 2022

Аннотация. Большинство людей полагаются на поисковые системы для получения и обмена информацией из различных ресурсов. Все результаты, возвращаемые поисковыми системами, не всегда актуальны, поскольку они извлекаются из разнородных источников данных. Более того, обычному



пользователю трудно подтвердить, что полученные результаты важны для пользовательского запроса. Поэтому семантическая сеть играет важную роль в интерпретации релевантности результатов поиска. В этой работе предлагается модель для поиска соответствующих документов с использованием семантической сети, основанный на концепции обработки естественного языка (ЕЯ). В этой предлагаемой системе ЕЯ используется для анализа пользовательского запроса с точки зрения частей речи. Извлеченные термины сравниваются со словарем домена, чтобы идентифицировать соответствующий домен интереса пользователя. С другой стороны, извлеченные документы пользовательского запроса исследуются с помощью обработки естественного языка для идентификации соответствующего домена.

Ключевые слова: семантическая поисковая система, обработка естественного языка, контекстный анализ, информационно-поисковые системы, графематический анализ, онтология

Для цитирования: А.К. Айтим. Модели обработки естественного языка для улучшения семантических результатов поиска // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 82–91 (на русском языке). DOI: 10.54309/IJICT.2022.10.2. 008.

Introduction

One of the leading functions of advanced information systems (IS) is the exploration of data components that satisfy some parameters. The specificity of the methodological and technological tasks in a search arises due to several reasons. First, it is due to the nature of the content of the information resources included in the system. Textual content still dominates in advanced information systems, however, multi-format resources that have multimedia content and use all kinds of forms of content structuring to increase performance are becoming more widespread. The result of structurization is the division of information into precisely structured data, metadata that describe their structure, including "metadata" that define all kinds of data structures.

These features of the content, in an obvious or implicit form, determine the layout for the organization of an effective search for information within the framework of an appropriate set of resources. Another necessary circumstance that has a significant impact on the effectiveness of information retrieval devices is the distributed and, therefore, heterogeneous nature of advanced information resources and systems. Search mechanisms aimed at implementation in the criteria of homogeneous information systems and resources quickly lose their effectiveness when used in distributed heterogeneous systems, where data presentation formats and, in accordance with this, metadata are allocated from resource to resource or from system to system. This event forces prospectors and creators to find ways to create universal methods and technologies for information search, adequate to the claims of advanced information systems (Aitim et al., 2020: 118–124).

The word is considered one of the leading forms of information exchange in society.



Textual information in various formats is an important part of the information resources of information systems. As a result, the creation and development of word processing technologies attracted great concern at all stages of the formation of information systems. The most common systems of this category are text search systems, the task of which is to search for a given collection of documents in natural language that meet the information needs of users. This work presents the main foundations of text search, natural language processing methods and their implementation in search models.

Materials and methods

Information retrieval methods and natural language text processing. An important place in the development of text search is occupied by the processing of NL. Under the processing of NLP (Natural Language Processing) is understood the conclusion of tasks related to awareness, analysis, performance of various operations on words, and for example, their generation. Examples of similar tasks: classification, clustering of stored collections of documents, deep word test, translation of documents from one language to another, etc.

All the variety of information search methods is based on the processing and analysis of the words of indexed documents. The bulk of information retrieval systems (IRS) are the systems with preprocessing - preparatory processing (indexing) of all documents available in the system. Exceptions are made for metasearch engines. Here is the list of the main problems that arise when processing words in NL (Aitim et al., 2020: 1–4):

- the problem of synonymy;
- the problem of homonymy;
- stable combinations of words;
- morphological variations.

The problem of synonyms. One concept can be expressed by different texts. As a result, relevant papers that use synonyms for the opinions indicated by the user in the request have every chance of not being detected by the system.

The problem of homonymy and phenomena "adjacent to the homonymy". Grammatical homonyms are the words different in meaning of the text, but coinciding in spelling in separate grammatical forms. It has every chance of being the text of one or different parts of speech. Lexical homonyms are texts of the same part of speech, similar in sound and spelling, but different in lexical meaning.

Stable combinations of texts. Phrases have every chance of owning a remarkable meaning from the meaning that the texts have separately. Morphological options in many natural languages, texts have several morphological forms that differ in spelling.

The IRS that uses all kinds of ways to process the words of the NL. In the advanced text search, not only the installation of linguistics for the analysis of words is used, but also statistical methods, mathematical logic and the doctrine of possibilities, a cluster test, methods of artificial origin of the mind, and for example, control technologies data. Let us consider two leading scenarios for the processing and analysis of NL words — statistical and linguistic.

An effective layout based on statistical analysis is latent semantic indexing. Latent Semantic Testing is a doctrine and method for extracting the context-dependent

meanings of texts while supporting statistical processing of large text datasets. The latent semantic test is based on the idea that the totality of all contexts in which a given text collides and does not collide sets many mutual restrictions, which to a large extent make it possible to qualify the similarity of the semantic meanings of texts and sets of texts among themselves.

The key defect of statistical methods is the impossibility of considering the coherence of a word, and the presentation of a word as an unpretentious large number of texts is not enough to reflect its content. The word gives a set of texts arranged in a specific given sequence. To overcome this defect allows the introduction of linguistic methods of word analysis.

In figure 1 proper meanings of linguistic analysis such as graphematic, morphological, syntactic, semantic are given. The output of each value is used by the next level of analysis as input. The purpose of graphematic analysis is to highlight the components of the word structure: paragraphs, services, individual texts, etc (Karpagam et al., 2015: 5679–5688).

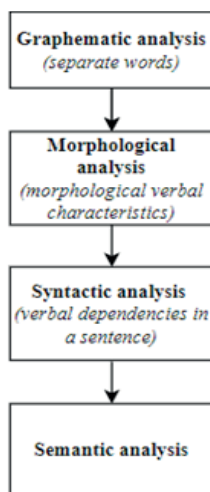


Figure 1 - Levels of linguistic analysis

The purpose of morphological analysis is to determine the morphological data of the text and its leading word form. The features of analysis strongly depend on the chosen natural language.

The purpose of syntactic analysis is to determine the syntactic dependence of texts in a sentence. Due to the presence in the Kazakh language of many syntactically homonymous systems, the presence of a close connection between semantics and syntax, the procedure for automatic parsing of a word is considered labor-intensive. The complexity of the method increases exponentially as the number of texts in a sentence and the number of applied rules increase (Madhu et al., 2016: 34–42).

Developments in the field of semantic analysis of the word are associated with the field of artificial origin of the mind, reprimanding the semantic awareness of the word. In real time, success in this direction is rather limited. The created semantic analyzers

have the highest computational complexity and ambiguity of the output results.

Information retrieval models

During the formation of information retrieval systems, many models of information search were proposed, then we will consider the main ones.

The model is a cobweb of proper elements:

1. Format of submission of documents.
2. Format of submission of requests. A request is a formalized method for expressing the information needs of an IRS user. For this, a search query language is used, the syntax of which varies from system to system.

3. The function of the ratio of the document to the request. The level of correlation between the query and the found document (relevance) is a personal concept because the results of the search, suitable for 1 user, have every chance of being unsuitable for another.

In all possible IRS models, the picture of the relevance aspect of documents depends on the appearance of the information search model, for example, in semantic search models, a clear entry of query texts into the document is not considered the main aspect, as, for example, in set-theoretic models (Dou et al., 2018: 59–68).

Variants of these elements define many implementations of search systems. Let's look at the more well-known search models. These models of conventional information search given in Figure 2 are usually divided into 3 types: set-theoretic (Boolean, fuzzy sets, extended Boolean), algebraic (vector, generalized vector, latent-semantic, neural network), probabilistic.

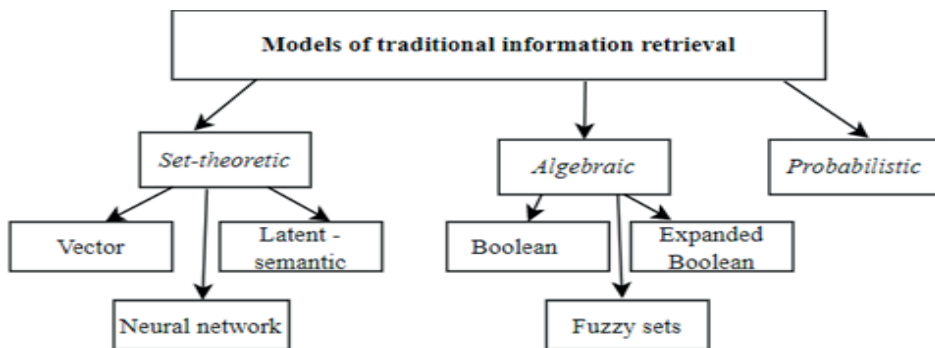


Figure 2 - Models of traditional information retrieval

Boolean model is a search model based on the operations of crossing, joining, and subtracting sets. Requirements are presented in the form of Boolean expressions from texts and regular operators. Relevant securities are those that satisfy the Boolean expression in the query. The leading defect of the Boolean model is its unsuitability for ranking search results.

A vector model is a representation of a collection of documents by vectors from the 1st common vector place for the entire collection. Papers and claims are represented as vectors in n-dimensional Euclidean space. The authority of a term in a document

can be qualified in different ways. For example, it is possible to count the number of occurrences of a term in a document, for example, the named frequency of the term - the more often the text collides in the document, the greater its weight will become. If the term does not collide in the document, then its authority in this document is equal to zero. All definitions that are seen in the documents of the collection being processed can be ordered. If now for some documents write out by covering those that are not in this document, a vector will come out, which will become a representation of the provided document in the vector place (Deng et al., 2018: 75–96).

Results and discussion

Semantic search. One of the promising directions in the formation of information retrieval systems is the construction of models of "semantic" search. Semantic intelligence is a picture of automatic full-text information search, considering the semantic content of texts and phrases of the user's request and services of words of indexed information resources. Semantic intelligence, for example, allows you to find papers that do not have texts from the search query, but are related to it.

Ontologies have personal methods of processing (logical inference), corresponding to the tasks of semantic processing of information. As a result, ontologies have become widespread in the conclusion of problems of knowledge representation and knowledge engineering, semantic integration of information resources, information search, etc.

Ontology defines a vocabulary for scientists who need to apply information together in the subject area. It incorporates machine-interpretable formulations of the domain's leading opinions and the dealings between them.

The main problem in the implementation of the use of the ontological layout is the inaccessibility of quite large and high-quality ontologies of subject areas, especially in the Kazakh language.

The main idea of communicative grammar is that, in fact, syntax must investigate just reasonable speech, and syntactic criteria must consider the categorical meanings of texts to be able to predetermine the generalized meanings of any syntactic system - from a text to a phrase and an unpretentious sentence. There is no doubt that 1 morphological data is not enough for the text to become a constructive unit of syntax. The word-lexeme is not yet considered a syntactic unit, the text is not a collection, and in its various forms, various aspects of its joint meaning have every chance to be realized or actualized. Thus, the decisive role here is played by the generalized meaning, that is, the categorical-semantic class of the text. The generalized meaning determines the syntactic abilities of the text and the methods of its functioning. Creating and exploring coherent speech, syntax deals with reasonable units that carry personal, not individually lexical, but generalized, categorical meaning in systems of varying degrees of difficulty. These units are characterized each time by the interaction of morphological, semantic, and active symptoms. These units are called syntaxes. It is important to emphasize that the semantic meaning is formed because of the combination of the categorical meaning and the morphological form and is realized in a specific syntactic position. Discussing the text in isolation, in isolation from the word, does not allow the introduction of a syntactic meaning, and therefore - to bring to life semantic intelligence.



The need for ontologization is associated with the inability of existing methods to process texts adequately automatically in the native language. For high-quality text processing, it is necessary to have a detailed description of the problem area with many logical connections that show the connections between terms. The use of ontologies makes it possible to process text in the native language in such a way that it becomes available for automatic processing (Deng et al., 2018: 75–96).

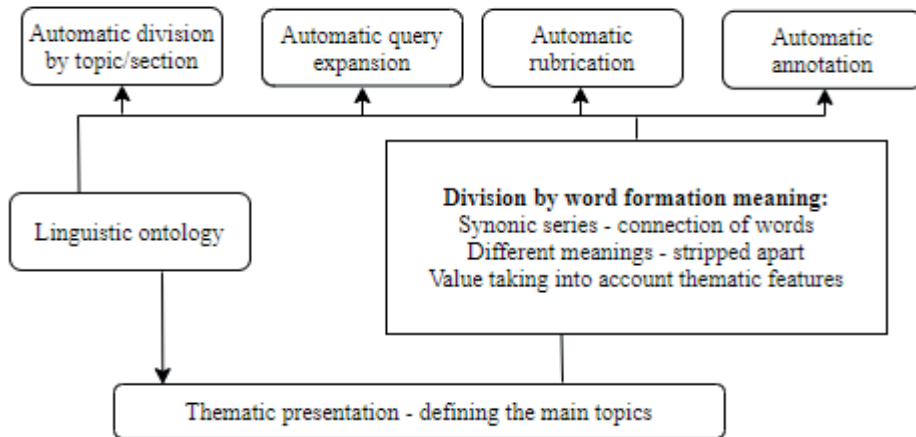


Figure 3 - Block diagram of automatic text processing based on the linguistic method

Based on the results of the study, the following stages of automatic text analysis are performed:

- automatic resolution of lexical ambiguity,
- building a thematic representation of the text,
- division of words according to the structure of the structure,
- formation of a block of words,
- based on the built thematic presentation
- automatic division of words into sections / topics
- automatic categorization,
- automatic annotation (Fig.3)

Many words processing models in the field of information retrieval assume the independent use of words in a connected text. Meanwhile it is known that the text contains many related words and has an internal hierarchical structure. There are quite a few different applications of automatic text processing that could produce better results if it were possible to automatically identify the content structure of a coherent text. Among them are applications such as automatic segmentation of texts, resolution of polysemy, proper information search, better definition of term weights in a document, text categorization, automatic annotation of texts, etc. The concept of text connectivity can be considered in several aspects.

Structural coherence and coherence of the text are distinguished. In fact, we are talking about internal (structural) and external (pragmatic) connectivity. Cohesion is the

relationship of text elements, in which the interpretation of some text elements depends on others. Coherence is the connectedness introduced by something external to the text, first by the knowledge of its addressee. Based on this knowledge, the addressee can construct a certain list and complete the links that are not explicitly present in the text. From another point of view, the global and local connectedness of the text are distinguished. The global coherence of the text is ensured by the fact that the text has a single theme. The local connectedness of discourse manifests itself in the relationship between adjacent minimal units of the text.

Determining the main topic of a text is an important step for many information retrieval applications. The concept of the main (or global) theme of the text is associated with such properties of the text as thematic coherence and thematic structure. A text can be formally connected through various types of connectivity, but if it does not have a single theme, then it cannot be considered as a text. The topic of a whole text can be characterized in terms of subtopics, and subtopics in terms of even more local subtopics. Each sentence of the text corresponds to one or another subtopic of the hierarchical structure of the text. The macrostructure of a text determines its global connectivity. "Without such global connectivity, it would be impossible to control local connections. sentences can be well connected according to local connectivity criteria, but they might deviate if there were no global restrictions on their content".

Semantic networks are one of the artificial intelligence methods for representing knowledge in natural language. The semantic network is an information model of the subject area, which has the form of a directed graph, the vertices of which correspond to the objects of the subject area, and the edges determine the relationship between them. Objects can be concepts, events, properties, processes. Thus, semantic networks are one of the ways to represent knowledge in natural language. In recent years, in the works of many scientists, it is proposed to use graphodynamic systems as a formal basis for projected intelligent systems, abstract logical-semantic models of intelligent systems.

Conclusion

The described classical search engine models at the beginning hoped to discuss documents as many separate texts that do not depend on each other. The probabilistic model is characterized by low computational scalability, the need to constantly study the system. Algebraic set-theoretic models are more widespread, since their practical effectiveness is usually higher. It follows in the footsteps to indicate that the recent implementations of information search plans proposed recently are often considered hybrid models and have the qualities of models of various classes. One of the promising directions in the formation of information retrieval systems is the construction of semantic search models, the main task of which is to analyze the word, i.e., extracting meaning from a word and reflecting it into a formal model that allows you to find the semantic proximity of two words. It is worth accepting that the potential of these systems is indeed enormous, however, not all possible semantic technologies have been



sold far away in real time. In essence, now they can only help to mark the main texts from tirades built in natural language and select auxiliary word forms for making a correct search query. This purpose of search methods urgently asks for formation.

REFERENCES

- Aitim A.K., Satybaldiyeva R.Zh. (2020). Analysis of methods and models for automatic processing systems of speech synthesis, IJICT. — 2020. — Pp.118–124.
- Aitim A.K., Satybaldiyeva R.Zh., Wojcik W. (2020). The construction of the Kazakh language thesauri in automatic word processing system, Proceedings of the 6th International Conference on Engineering & MIS — 2020. — Pp.1–4.
- Karpagam G.R., Uma Maheswari J. (2015). A conceptual framework for ontology information retrieval. Int. J. Eng. Sci. Technol. 2(10). — 2015. — Pp. 5679–5688.
- Madhu G., Govardhan A., Rajinikanth T.V. (2016). Intelligent semantic web search engines: a brief survey. Int. J. Web Semant. Technol. 2(1). — 2016. — Pp. 34–42.
- Dou T., Rockett P. (2018). Comparison of semantic-based local search methods for multiobjective genetic programming. In Web of Science. Retrieved October 20. — 2018. — Pp.59–68.
- Deng Z., Zheng X., Zhang L.W., Zhang J.W. (2018). A learning framework for semantic reach-to-grasp tasks integrating machine learning and optimization. In Web of Science. — Retrieved October 19. — 2018. — Pp.75–96.



МАТЕМАТИКАЛЫҚ ЖӘНЕ КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ
МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ
MATHEMATICAL AND COMPUTER MODELING

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 3. Is. 2. Number 10 (2022). Pp. 92–102
Journal homepage: <https://journal.iitu.edu.kz>
<https://doi.org/10.54309/IJICT.2022.10.2.009>

УДК 517.927.6; 519.622

**ON THE NUMERICAL SOLUTION OF SYSTEMS OF ESSENTIALLY
LOADED DIFFERENTIAL EQUATIONS WITH A THREE-POINT
CONDITION**

Zh.M. Kadirbayeva, D.S. Abilkair, B.S. Massalimov*

Zhazira M. Kadirbayeva — Cand.Ph.-M.S, Associate Professor of the «Mathematical and computer modeling» Department, International Information Technology University

ORCID: 0000-0001-8861-4100. E-mail: zhkadirbayeva@gmail.com;

Abilkair Din-Mukhammed Serzhanuly — Master's student, «Mathematical and computer modeling» Department, International Information Technology University;

Massalimov S. Bekzhan — Master's student, «Mathematical and computer modeling» Department, International Information Technology University.

© Zh.M. Kadirbayeva, D.S. Abilkair, B.S. Massalimov, 2022

Abstract. A linear boundary value problem for systems of essentially loaded ordinary differential equations with a three-point condition is considered. The considering problem is reduced to a boundary value problem for loaded ordinary differential equations with a three-point condition. Based on D.S. Dzhumabaev's parameterization method, a numerical method for solving a boundary value problem for loaded ordinary differential equations with a three-point condition is developed and an algorithm for its implementation is proposed. By partitioning the interval and introducing additional parameters, the boundary value problem for loaded ordinary differential equations is reduced to an equivalent boundary value problem with a parameter. An equivalent boundary value problem with parameters consists of the Cauchy problem for a system of ordinary differential equations with parameters, a three-point condition, and a gluing condition. The solution of the Cauchy problem for a system of ordinary differential equations with parameters is constructed using the fundamental matrix of the differential



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

equation. Substituting the values at the corresponding points of the constructed solution into the three-point condition and the gluing condition, a system of linear algebraic equations with respect to the parameters is compiled. A numerical method for solving a boundary value problem for essentially ordinary differential equations with a three-point condition based on the solving of the constructed system is proposed. The proposed numerical method is illustrated by an example.

Keywords: boundary value problem, loaded differential equation, multipoint condition, numerical method, parametrization method

For citation: Zh.M. Kadirbayeva, D.S. Abilkair, B.S. Massalimov. On the numerical solution of systems of essentially loaded differential equations with a three-point condition // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 92–102 (In Russ.). DOI: 10.54309/IJICT.2022.10.2.009.

ҮШ НҮКТЕЛІ ШАРТЫ БАР ЕЛЕУЛІ ТҮРДЕ ЖҮКТЕЛГЕН ДИФФЕРЕНЦИАЛДЫҚ ТЕНДЕУЛЕР ЖҮЙЕСІНІҢ САНДЫҚ ШЕШІМІ ТУРАЛЫ

Ж.М. Қадырбаева, Д.С. Абилкаир, Б.С. Масалимов*

Қадырбаева Жазира Мұратбекқызы — ф.-м.ғ.к., «Математика және компьютерлік моделдеу» кафедрасының қауымдастырылған профессоры, Халықаралық ақпараттық технологиялар университеті

ORCID: 0000-0001-8861-4100. E-mail: zhkadirbayeva@gmail.com;

Абилкаир Дін-Мұхаммед Сержанұлы — «Математика және компьютерлік моделдеу» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті;

Масалимов Бекжан Серікбақытұлы — «Математика және компьютерлік моделдеу» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті.

© Ж.М. Қадырбаева, Д.С. Абилкаир, Б.С. Масалимов, 2022

Аннотация. Үш нүктелі шарты бар елеулі түрде жүктелген жәй дифференциалдық тендеулер жүйесі үшін сызықтық шеттік есеп қарастырылады. Қарастырылып отырған есеп үш нүктелі шарты бар жүктелген жәй дифференциалдық тендеулер үшін сызықтық шеттік есепке келтіріледі. Д.С. Жұмабаевтың параметрлеу әдісі негізінде үш нүктелі шарты бар жүктелген жәй дифференциалдық тендеулер үшін шекті есепті шешудің сандық әдісі әзірленді және оны жүзеге асыру алгоритмі ұсынылды. Аралықты бөлу және қосымша параметрлерді енгізу арқылы жүктелген жәй дифференциалдық тендеулер үшін шеттік есеп параметрі бар эквивалентті шеттік есепке келтіріледі. Параметрлері бар эквивалентті шеттік есеп параметрлері бар жәй дифференциалдық тендеулер жүйесі үшін Коши есебінен, үш нүктелі шартынан және жалғау шартынан тұрады. Параметрлері бар жәй дифференциалдық тендеулер жүйесі үшін Коши есебінің шешімі дифференциалдық тендеудің фундаменталдық матрицасын қолдана отырып құрылады. Құрылған шешімнің тиісті нүктелеріндегі мәндерді үш нүктелі шартқа



және жалғау шарттарына қойып, параметрлерге қатысты сызықтық алгебралық теңдеулер жүйесі құрылады. Құрылған жүйені шешуге негізделген үш нүктелі шарты бар елеулі түрде жүктелген жәй дифференциалдық теңдеулер үшін шеттік есепті шешудің сандық әдісі ұсынылған. Ұсынылған сандық әдіс мысалмен көркемделген.

Түйін сөздер: шеттік есеп, жүктелген дифференциалдық теңдеу, көпнүктелі шарт, сандық әдіс, параметрлеу әдісі

Дәйексөз үшін: Ж.М. Қадырбаева, Д.С. Абилкаир, Б.С. Масалимов. Үш нүктелі шарты бар елеулі түрде жүктелген дифференциалдық теңдеулер жүйесінің сандық шешімі туралы // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 92–102 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.009.

О ЧИСЛЕННОМ РЕШЕНИИ СИСТЕМ СУЩЕСТВЕННО НАГРУЖЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ТРЕХТОЧЕЧНЫМ УСЛОВИЕМ

Ж.М. Кадирбаева, Д.С. Абилкаир, Б.С. Масалимов*

Кадирбаева Жазира Муратбековна — к.ф.-м.н., ассоциированный профессор кафедры «Математическое и компьютерное моделирование», Международный университет информационных технологий

ORCID: 0000-0001-8861-4100. E-mail: zhkadirbayeva@gmail.com;

Абилкаир Дин-Мухаммед Сержанулы — магистрант кафедры «Математическое и компьютерное моделирование», Международный университет информационных технологий;

Масалимов Бекжан Серикбакирович — магистрант кафедры «Математическое и компьютерное моделирование», Международный университет информационных технологий.

© Ж.М. Кадирбаева, Д.С. Абилкаир, Б.С. Масалимов, 2022

Аннотация. Рассматривается линейная краевая задача для систем существенно нагруженных обыкновенных дифференциальных уравнений с трехточечным условием. Рассматриваемая задача сводится к краевой задаче для нагруженных обыкновенных дифференциальных уравнений с трехточечным условием. На основе метода параметризации Д.С. Джумабаева разработан численный метод решения краевой задачи для нагруженных обыкновенных дифференциальных уравнений с трехточечным условием и предложен алгоритм его реализации. Разбиением интервала и введением дополнительных параметров краевая задача для нагруженных обыкновенных дифференциальных уравнений сводится к эквивалентной краевой задаче с параметром. Эквивалентная краевая задача с параметрами состоит из задачи Коши для системы обыкновенных дифференциальных уравнений с параметрами, трехточечного условия и условия склеивания. Решение задачи Коши для системы обыкновенных дифференциальных уравнений с параметрами строится с помощью фундаментальной матрицы дифференциального уравнения. Подставляя значения в соответствующих точках построенного решения в



трехточечное условие и условие склеивания, составляется система линейных алгебраических уравнений относительно параметров. Предложен численный метод решения краевой задачи для существенно нагруженных обыкновенных дифференциальных уравнений с трехточечным условием, основанный на решении построенной системы. Предлагаемый численный метод иллюстрируется примером.

Ключевые слова: краевая задача, нагруженное дифференциальное уравнение, многоточечное условие, численный метод, метод параметризации

Для цитирования: Ж.М. Кадирбаева, Д.С. Абилкаир, Б.С. Масалимов. О численном решении систем существенно нагруженных дифференциальных уравнений с трехточечным условием // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 92–102 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.009.

Введение

Исследование нагруженных дифференциальных уравнений представляет интерес как с точки зрения построения общей теории дифференциальных уравнений, так и с точки зрения приложений, причем приложений как в математическом моделировании, так и в собственно математике. Разные краевые задачи для нагруженных дифференциальных уравнений и методы нахождения их решений рассмотрены в (Нахушев, 2012: 232; Нахушев, 1995: 301; Джениалиев и др., 2010: 334; Abdullaev и др., 2014: 1096–1109; Assanova и др., 2018: 508–516; Assanova и др., 2018: 4966–4976; Bakirova и др., 2020: 77–86).

Термин существенно нагруженное дифференциальное уравнение означает, что правая часть дифференциального уравнения зависит от значения искомого решения и его производных в заданных точках. Краевые задачи для существенно нагруженных дифференциальных уравнений различного типа исследованы в (Джениалиев и др., 2010: 334; Krall, 1975: 493–542; Kadirbayeva и др., 2020: 47–57; Kadirbayeva и др., 2021: 6–14; Kadirbayeva, 2021: 551–559). Так как наличие производной решения в нагруженной точке сильно влияет на свойства уравнений, в настоящей работе исследуется линейная краевая задача для существенно нагруженных дифференциальных уравнений с трехточечным условием.

На $[0, T]$ рассматривается линейная трехточечная краевая задача для существенно нагруженных обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = A_0(t)x + A_1(t)x(\theta_0) + A_2(t)x(\theta_1) + A_3(t)\dot{x}(\theta_0) + f(t), \quad t \in (0, T), \quad (1)$$

$$B_0x(\theta_0) + B_1x(\theta_1) + B_2x(\theta_2) = d, \quad d \in R^n, \quad x \in R^n, \quad (2)$$

где $(n \times n)$ -матрицы $A_i(t)$, $(i = \overline{0,3})$, и n -вектор-функция $f(t)$ непрерывны на $[0, T]$, B_j , $(j = \overline{0,2})$, постоянные $(n \times n)$ -матрицы, d – постоянный n -вектор, и $0 = \theta_0 < \theta_1 < \theta_2 = T$, $\|x\| = \max_i |x_i|$.



Через $C([0, T], R^n)$ обозначим пространство непрерывных функций $x: [0, T] \rightarrow R^n$ с нормой $\|x\|_1 = \max_{t \in [0, T]} \|x(t)\|$.

Непрерывно дифференцируема на $(0, T)$ функция $x(t) \in C([0, T], R^n)$ называется решением задачи (1), (2), если она удовлетворяет системе существенно нагруженных дифференциальных уравнений (1) и трехточечному условию (2).

В уравнении (1) при $t = \theta_0$ и получим

$$[I - A_3(\theta_0)]\dot{x}(\theta_0) = A_0(\theta_0)x(\theta_0) + A_1(\theta_0)x(\theta_0) + A_2(\theta_0)x(\theta_1) + f(\theta_0), \quad (3)$$

где I – единичная матрица размерности $(n \times n)$. Предположим матрица $[I - A_3(\theta_0)]$ обратима, тогда из уравнения (3) получим

$$\dot{x}(\theta_0) = [I - A_3(\theta_0)]^{-1}\{A_0(\theta_0)x(\theta_0) + A_1(\theta_0)x(\theta_0) + A_2(\theta_0)x(\theta_1) + f(\theta_0)\}.$$

Подставляя значение $\dot{x}(\theta_0)$ в уравнение (1) получим следующую трехточечную краевую задачу для систем нагруженных обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = A_0(t)x + K_1(t)x(\theta_0) + K_2(t)x(\theta_1) + F(t), \quad t \in (0, T), \quad (4)$$

$$B_0x(\theta_0) + B_1x(\theta_1) + B_2x(\theta_2) = d, \quad d \in R^n, \quad x \in R^n, \quad (5)$$

где $K_1(t) = A_1(t) + A_3(t)[I - A_3(\theta_0)]^{-1}\{A_0(\theta_0) + A_1(\theta_0)\}$,

$$K_2(t) = A_2(t) + A_3(t)[I - A_3(\theta_0)]^{-1}A_2(\theta_0),$$

$$F(t) = A_3(t)[I - A_3(\theta_0)]^{-1}f(\theta_0) + f(t).$$

Материалы и методы исследования

Краевая задача (4), (5) исследуется методом параметризации Д.С. Джумабаева (Dzhumabayev, 1989: 34–46). Интервал $[0, T]$ разбиваем на две части:

$$[0, T] = [\theta_0, \theta_1] \cup [\theta_1, \theta_2].$$

Введем пространство $C([0, T], \theta, R^{2n})$ систем функций $x[t] = (x_1(t), x_2(t))$, где $x_r: [\theta_{r-1}, \theta_r] \rightarrow R^n$ непрерывны на $[\theta_{r-1}, \theta_r]$, $r = 1, 2$, имеют конечный левосторонний предел $\lim_{t \rightarrow \theta_1-0} x_1(t)$ и $\lim_{t \rightarrow \theta_2-0} x_2(t)$, с нормой $\|x[\cdot]\|_2 = \max \left\{ \sup_{t \in [\theta_0, \theta_1]} \|x_1(t)\|, \sup_{t \in [\theta_1, \theta_2]} \|x_2(t)\| \right\}$.

Сужение функции $x(t)$ на $[\theta_{r-1}, \theta_r]$ $r = 1, 2$, обозначим через $x_r(t)$, т.е. $x_r(t) = x(t)$ для всех $t \in [\theta_{r-1}, \theta_r]$, $r = 1, 2$. Введем дополнительные параметры $\lambda_1 = x_1(\theta_0)$, $\lambda_2 = x_2(\theta_1)$. И на каждом интервале $[\theta_{r-1}, \theta_r]$, $r = 1, 2$, произведем замену функции $x_r(t) = u_r(t) + \lambda_r$, $r = 1, 2$. Тогда задача (4), (5) перейдет к следующей эквивалентной задаче:

$$\frac{du_r}{dt} = A_0(t)[u_r + \lambda_r] + K_1(t)\lambda_1 + K_2(t)\lambda_2 + F(t), \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2, \quad (6)$$

$$u_r(\theta_{r-1}) = 0, \quad r = 1, 2 \quad (7)$$



$$B_0\lambda_1 + B_1\lambda_2 + B_2\lambda_2 + B_2 \lim_{t \rightarrow \theta_2-0} u_2(t) = d, \quad (8)$$

$$\lambda_1 + \lim_{t \rightarrow \theta_1-0} u_1(t) = \lambda_2, \quad (9)$$

Решением задачи (6)-(9) является пара $(u^*[t], \lambda^*)$ с элементами $(u_1^*(t), u_2^*(t)) \in C([0, T], \theta, R^{2n})$, $\lambda^* = (\lambda_1^*, \lambda_2^*) \in R^{2n}$ где функции $u_r^*(t)$, $r = 1, 2$, непрерывно дифференцируемы на $[\theta_{r-1}, \theta_r]$ и при $\lambda_j = \lambda_j^*$, $j = 1, 2$ удовлетворяют системе дифференциальных уравнений (6) и условиям (7)–(9).

Задачи (4), (5) и (6)–(9) эквивалентны. Если $x^*(t)$ является решением задачи (4), (5), тогда пара $(u^*[t], \lambda^*)$, где $u^*[t] = (x^*(t) - x^*(\theta_0), x^*(t) - x^*(\theta_1))$, and $\lambda^* = (x^*(\theta_0), x^*(\theta_1))$, будет решением задачи (6)-(9). И наоборот, если пара $(\tilde{u}[t], \tilde{\lambda})$ с элементами $\tilde{u}[t] = (\tilde{u}_1(t), \tilde{u}_2(t)) \in C([0, T], \theta, R^{2n})$, $\tilde{\lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2) \in R^{2n}$ - решение задачи (6)-(9), тогда функция пара $\tilde{x}(t)$ определяемая равенствами $\tilde{x}(t) = \tilde{u}(t) + \tilde{\lambda}_r$, $t \in [\theta_{r-1}, \theta_r]$, $r = 1, 2$, будет решением исходной задачи (4), (5).

Пусть $\Phi_r(t)$ – фундаментальная матрица дифференциального уравнения $\frac{dx}{dt} = A_0(t)x$ на $[\theta_{r-1}, \theta_r]$, $r = 1, 2$. Тогда решение задачи Коши (6), (7) можно записать в виде

$$u_r(t) = \Phi_r(t) \int_{\theta_{r-1}}^t \Phi_r^{-1}(\tau) A_0(\tau) d\tau \lambda_r + \Phi_r(t) \int_{\theta_{r-1}}^t \Phi_r^{-1}(\tau) [K_1(\tau) \lambda_1 + K_2(\tau) \lambda_2] d\tau + \\ + \Phi_r(t) \int_{\theta_{r-1}}^t \Phi_r^{-1}(\tau) F(\tau) d\tau, \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2. \quad (10)$$

Подставив правую часть (10) в условия (8), (9) при соответствующих предельных значениях, получим следующую систему алгебраических уравнений относительно параметров λ_1 и λ_2 :

$$B_0\lambda_1 + B_1\lambda_2 + B_2\lambda_2 + B_2\Phi_2(\theta_2) \int_{\theta_1}^{\theta_2} \Phi_2^{-1}(\tau) [A_0(\tau) \lambda_2 + K_1(\tau) \lambda_1 + K_2(\tau) \lambda_2] d\tau = \\ = d - B_2\Phi_2(\theta_2) \int_{\theta_1}^{\theta_2} \Phi_2^{-1}(\tau) F(\tau) d\tau, \quad (11)$$

$$\lambda_1 + \Phi_1(\theta_1) \int_{\theta_0}^{\theta_1} \Phi_1^{-1}(\tau) [A_0(\tau) \lambda_1 + K_1(\tau) \lambda_1 + K_2(\tau) \lambda_2] d\tau - \\ - \lambda_2 = -\Phi_1(\theta_1) \int_{\theta_0}^{\theta_1} \Phi_1^{-1}(\tau) F(\tau) d\tau. \quad (12)$$

Далее рассмотрим следующие задачи Коши для обыкновенных дифференциальных уравнений на подинтервалах

$$\frac{dy}{dt} = A(t)y + P(t), \quad y(\theta_{r-1}) = 0, \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2, \quad (13)$$

где $P(t)$ — непрерывные на $[\theta_{r-1}, \theta_r]$, $r = 1, 2$, $(n \times n)$ матрица, либо n вектор. Следовательно, решение задачи (13) представляет собой квадратную матрицу или вектор. Обозначим через $\alpha_r(P, t)$ решение задачи Коши (13). Очевидно, что

$$a_r(P, t) = \Phi_r(t) \int_{\theta_{r-1}}^t \Phi_r^{-1}(\tau) P(\tau) d\tau, \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2, \quad (14)$$

Перезапишем систему алгебраических уравнений (11), (12) используя (14):

$$B_0 \lambda_1 + B_1 \lambda_2 + B_2 \lambda_2 + B_2 a_2(A_0, \theta_2) \lambda_2 + B_2 a_2(K_1, \theta_2) \lambda_1 + B_2 a_2(K_2, \theta_2) \lambda_2 = d - B_2 a_2(F, \theta_2), \quad (15)$$

$$\lambda_1 + a_1(A_0, \theta_1) \lambda_1 + a_1(K_1, \theta_1) \lambda_1 + a_1(K_2, \theta_1) \lambda_2 - \lambda_2 = -a_1(F, \theta_1), \quad (16)$$

Матрицу, соответствующую левой части системы уравнений (15), (16), обозначим через $Q_*(\theta)$ и запишем систему в виде

$$Q_*(\theta) \lambda = F_*(\theta), \quad \lambda \in R^{2n}, \quad (17)$$

где

$$Q_*(\theta) = \begin{bmatrix} B_0 + B_2 a_2(K_1, \theta_2) & B_1 + B_2 [I + a_2(A_0, \theta_2) + a_2(K_2, \theta_2)] \\ I + a_1(A_0, \theta_1) + a_1(K_1, \theta_1) & a_1(K_2, \theta_1) - I \end{bmatrix},$$

$$F_*(\theta) = \begin{bmatrix} d - B_2 a_2(F, \theta_2) \\ -a_1(F, \theta_1) \end{bmatrix}.$$

Нетрудно установить, что разрешимость краевой задачи (4), (5) эквивалентна разрешимости системы (17). Решением системы (17) является вектор $\lambda^* = (\lambda_1^*, \lambda_2^*) \in R^{2n}$ состоит из значений решений исходной задачи (4), (5) в начальных точках подинтервалов, т.е. $\lambda_1^* = x^*(\theta_0)$ и $\lambda_2^* = x^*(\theta_1)$.

Мы предлагаем следующую численную реализацию метода параметризации Д.С.Джумабаева.

1. Берем разбиение $0 = \theta_0 < \theta_1 < \theta_2 = T$. Разделим первый интервал $[\theta_0, \theta_1]$ на N_1 частей с шагом $h_1 = (\theta_1 - \theta_0)/N_1$, а второй интервал $[\theta_1, \theta_2]$ на N_2 частей с шагом $h_2 = (\theta_2 - \theta_1)/N_2$. Предположим, что на каждом интервале $[\theta_{r-1}, \theta_r]$, $r = 1, 2$, переменная $\hat{t}$ принимает дискретные значения: $\hat{\theta} = \theta_{r-1}$, $\hat{\theta} = \theta_{r-1} + h_r, \dots, \hat{\theta} = \theta_{r-1} + (N_r - 1)h_r$, $\hat{\theta} = \theta_r$, и обозначим через $\{\theta_{r-1}, \theta_r\}$ множество таких точек.

2. Решая следующие задачи Коши для обыкновенных дифференциальных уравнений

$$\begin{aligned} \frac{dy}{dt} &= A_0(t)y + A_0(t), \quad y(\theta_{r-1}) = 0, \quad t \in [\theta_{r-1}, \theta_r], \\ \frac{dy}{dt} &= A_0(t)y + K_1(t), \quad y(\theta_{r-1}) = 0, \quad t \in [\theta_{r-1}, \theta_r], \\ \frac{dy}{dt} &= A_0(t)y + K_2(t), \quad y(\theta_{r-1}) = 0, \quad t \in [\theta_{r-1}, \theta_r], \\ \frac{dy}{dt} &= A_0(t)y + F(t), \quad y(\theta_{r-1}) = 0, \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2, \end{aligned}$$

находим значения $(n \times n)$ матриц $a_r(A_0, \hat{t})$, $a_r(K_1, \hat{t})$, $a_r(K_2, \hat{t})$ и n вектор $a_r(F, \hat{t})$ на $\{\theta_{r-1}, \theta_r\}$, $r = 1, 2$.

3. Построим систему линейных алгебраических уравнений относительно параметров



$$Q_{*}^{\tilde{h}}(\theta)\lambda = -F_{*}^{\tilde{h}}(\theta), \quad \lambda \in R^{2n}, \quad (18)$$

Решая систему (18), находим $\lambda^{\tilde{h}}$: $x^{\tilde{h}_1}(\theta_0) = \lambda_1^{\tilde{h}}$ и $x^{\tilde{h}_2}(\theta_1) = \lambda_2^{\tilde{h}}$.

4. Чтобы определить значения приближенного решения в оставшихся точках множества $\{\theta_{r-1}, \theta_r\}$, нужно решить задачи Коши

$$\begin{aligned} \frac{dx}{dt} &= A_0(t)x + K_1(t)\lambda_1^{\tilde{h}} + K_2(t)\lambda_2^{\tilde{h}} + F(t), \\ x(\theta_{r-1}) &= \lambda_r^{\tilde{h}}, \quad t \in [\theta_{r-1}, \theta_r], \quad r = 1, 2. \end{aligned}$$

Таким образом, предложенный алгоритм позволяет найти численное решение задачи (4), (5). Мы видим, что решение краевой задачи (4), (5) также является решением краевой задачи (1), (2), когда матрица $[I - A_3(\theta_0)]$ обратима.

Чтобы проиллюстрировать предложенный алгоритм для численного решения линейной краевой задачи для существенно нагруженных обыкновенных дифференциальных уравнений с трехточечным условием (1), (2) на основе метода параметризации Д.С. Джумабаева, рассмотрим следующий пример.

Результат исследования

На отрезке $[0, 1]$ рассматривается линейная краевая задача для существенно нагруженных обыкновенных дифференциальных уравнений с с трехточечным условием

$$\frac{dx}{dt} = A_0(t)x + A_1(t)x(0) + A_2(t)x\left(\frac{1}{2}\right) + A_3(t)\dot{x}(0) + f(t), \quad t \in (0, 1), \quad (19)$$

$$B_0x(0) + B_1x\left(\frac{1}{2}\right) + B_2x(1) = d, \quad d \in R^2, \quad x \in R^2. \quad (20)$$

$$\text{Здесь } \theta_0 = 0, \quad \theta_1 = \frac{1}{2}, \quad \theta_2 = T = 1, \quad A_0(t) = \begin{pmatrix} t & t^2 \\ 2 & t-4 \end{pmatrix},$$

$$A_1(t) = \begin{pmatrix} t^2 & 5 \\ 2t & t^3 \end{pmatrix}, \quad A_2(t) = \begin{pmatrix} 3 & t \\ 5t & t^2 \end{pmatrix}, \quad A_3(t) = \begin{pmatrix} t^2 + 6 & 4t \\ -2 & t-5 \end{pmatrix},$$

$$B_0 = \begin{pmatrix} 2 & 5 \\ 1 & 7 \end{pmatrix}, \quad B_1 = \begin{pmatrix} -3 & 5 \\ 0 & 2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 3 & 6 \\ -5 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} 55 \\ -44 \end{pmatrix},$$

$$f(t) = \begin{pmatrix} -12t^3 - 3t^2 - 31t - 8 \\ 2t^3 - 18t^2 - 14t + 30 \end{pmatrix}.$$

Значение $\dot{x}(0)$ находим из (19):

$$\dot{x}(0) = \begin{pmatrix} -\frac{1}{5} & 0 \\ 1 & 1 \\ \frac{1}{15} & \frac{1}{6} \end{pmatrix} \left\{ A_0(0)x(0) + A_1(0)x(0) + A_2(0)x\left(\frac{1}{2}\right) + f(0) \right\}.$$

Подставляя значение $\dot{x}(0)$ в уравнение (19) получим следующую краевую задачу для систем нагруженных обыкновенных дифференциальных уравнений с трехточечным условием

$$\frac{dx}{dt} = A_0(t)x + K_1(t)x(0) + K_2(t)x\left(\frac{1}{2}\right) + F(t), \quad t \in (0,1), \quad (21)$$

$$B_0x(0) + B_1x\left(\frac{1}{2}\right) + B_2x(1) = d, \quad d \in R^2, \quad x \in R^2. \quad (22)$$

где

$$K_1(t) = A_1(t) + A_3(t)[I - A_3(0)]^{-1}\{A_0(0) + A_1(0)\} = \begin{pmatrix} t^2 + \frac{4}{3}t & -t^2 - \frac{4}{3}t - 1 \\ \frac{7}{3}t - \frac{5}{3} & t^3 - \frac{t}{3} + \frac{11}{3} \end{pmatrix},$$

$$K_2(t) = A_2(t) + A_3(t)[I - A_3(0)]^{-1}A_2(0) = \begin{pmatrix} \frac{4}{5}t - \frac{3}{2}t^2 - \frac{3}{5} & t \\ \frac{26}{5}t + \frac{1}{5} & t^2 \end{pmatrix},$$

$$F(t) = A_3(t)[I - A_3(0)]^{-1}f(0) + f(t) = \begin{pmatrix} \frac{8}{5} - \frac{7}{5}t^2 - \frac{197}{15}t - 12t^3 \\ 2t^3 - 18t^2 - \frac{143}{15}t + \frac{67}{15} \end{pmatrix}.$$

Для решения краевой задачи (21), (22) используем численную реализацию метода параметризации Д.С. Джумабаева. Мы предоставляем результаты численной реализации алгоритма путем разбиения подинтервалов $[0, 0.5]$, $[0.5, 1]$ с шагом $h = 0.05$.

Точным решением задачи (19), (20) является функция $x^*(t) = \begin{pmatrix} 4t^2 + 5 \\ 8t - 2 \end{pmatrix}$.

Результаты расчетов численных решений в точках разбиения представлены в таблице 1. Эти результаты получены с помощью системы MathCad.

Таблица 1 – Численное решение задачи (19), (20)

t	$\tilde{x}_1(t)$	$\tilde{x}_2(t)$	t	$\tilde{x}_1(t)$	$\tilde{x}_2(t)$
0	5.000001235	-1.999998855	0.5	6.000000851	1.999998545
0.05	5.010001158	-1.599999601	0.55	6.210000854	2.399998592
0.1	5.04000109	-1.200000173	0.6	6.440000865	2.79999866
0.15	5.090001031	-0.800000606	0.65	6.690000882	3.199998746
0.2	5.160000981	-0.400000927	0.7	6.960000905	3.599998847
0.25	5.25000094	-0.000001158	0.75	7.250000936	3.999998961
0.3	5.360000907	0.399998685	0.8	7.560000973	4.399999086
0.35	5.490000882	0.799998586	0.85	7.890001018	4.79999922
0.4	5.640000864	1.199998536	0.9	8.240001071	5.199999363
0.45	5.810000854	1.599998524	0.95	8.610001132	5.599999514
0.5	6.000000851	1.999998545	1	9.000001202	5.999999671

Для разности соответствующих значений точных и построенных решений задачи справедлива следующая оценка: $\max_{j=0,20} \|x^*(t_j) - \tilde{x}(t_j)\| < 0.000002$.

Заключение

В данной работе предложена численная реализация метода параметризации Д.С. Джумабаева для решения линейной трехточечной краевой задачи для системы существенно нагруженных обыкновенных дифференциальных уравнений. Приведен пример, иллюстрирующий численные алгоритмы метода параметризации Д.С. Джумабаева.

ЛИТЕРАТУРЫ

Abdullaev V.M., Aida-zade K.R. (2014). Numerical method of solution to loaded nonlocal boundary value problems for ordinary differential equations. *Computational Mathematics and Mathematical Physics*. — 2014. — Vol. 54. — No. 7. — Pp. 1096–1109.

Assanova A.T., Imanchiyev A.E., Kadirbayeva Zh.M. (2018). Numerical solution of systems of loaded ordinary differential equations with multipoint conditions. *Computational Mathematics and Mathematical Physics*. — 2018. — Vol. 58. — No. 4. — Pp. 508–516.

Assanova A.T., Kadirbayeva Zh.M. (2018). On the numerical algorithms of parametrization method for solving a two-point boundary-value problem for impulsive systems of loaded differential equations. *Computational and Applied Mathematics*. — 2018. — Vol. 37. — No. 4. — Pp. 4966–4976.

Bakirova E.A., Kadirbayeva Zh.M. (2020). Numerical solution of the boundary value problems for the loaded differential and Fredholm integro-differential equations // *International Journal of Information and Communication Technologies*. — 2020. — Vol. 1. — No 2. — Pp. 77–86.

Дженалиев М.Т., Рамазанов М.И. (2010). Нагруженные уравнения как возмущения дифференциальных уравнений. -Алматы: Ғылым, 2010. — 334 с.

Dzhumabayev D.S. (1989). Criteria for the unique solvability of a linear boundary-value problem for an ordinary differential equation // *U.S.S.R. Comput. Maths. Math. Phys.* — 1989. — Vol. 29. — No 1. — Pp. 34–46.

Krall A.M. (1975). The development of general differential and general differential-boundary systems, *Rocky Mountain Journal of Mathematics*. — 1975. — Vol. 5. — Pp. 493–542.

Kadirbayeva Zh.M., Karakenova S.G. (2020). Numerical solution of the multipoint boundary value problems for essentially loaded ordinary differential equations // *Kazakh Mathematical Journal*. — 2020. — Vol. 20. — No. 4. — Pp. 47–57.

Kadirbayeva Zh.M., Bakirova E.A., Dauletbayeva A.Sh., Kassymgali A.A. (2021). An algorithm for solving a boundary value problem for essentially loaded differential equations // *News of the NAS RK. Phys.-Math. Series*. — 2021. — Vol. 2. — No. 336. — Pp. 6–14.

Kadirbayeva Zh.M. (2021). A numerical method for solving boundary value problem for essentially loaded differential equations // *Lobachevskii Journal of Mathematics*. — 2021. — Vol. 42. — No. 3. — Pp. 551–559.

Нахушев А.М. (2012). Нагруженные уравнения и их применение. — М.: Наука, 2012. — 232 с.

Нахушев А.М. (1995). Уравнения математической биологии. — М.: Высшая школа, 1995. — 301 с.

REFERENCES

Abdullaev V.M., Aida-zade K.R. (2014). Numerical method of solution to loaded nonlocal boundary value problems for ordinary differential equations. *Computational Mathematics and Mathematical Physics*. — 2014. — Vol. 54. — No. 7. — Pp. 1096–1109.

Assanova A.T., Imanchiyev A.E., Kadirbayeva Zh.M. (2018). Numerical solution of systems of



loaded ordinary differential equations with multipoint conditions. *Computational Mathematics and Mathematical Physics*. — 2018. — Vol. 58. — No. 4. — Pp. 508–516.

Assanova A.T., Kadirbayeva Zh.M. (2018). On the numerical algorithms of parametrization method for solving a two-point boundary-value problem for impulsive systems of loaded differential equations. *Computational and Applied Mathematics*. — 2018. — Vol. 37. — No. 4. — Pp. 4966–4976.

Bakirova E.A., Kadirbayeva Zh.M. (2020). Numerical solution of the boundary value problems for the loaded differential and Fredholm integro-differential equations // *International Journal of Information and Communication Technologies*. — 2020. — Vol. 1. — No 2. — Pp. 77–86.

Dzhumabayev D.S. (1989). Criteria for the unique solvability of a linear boundary-value problem for an ordinary differential equation // *U.S.S.R. Comput. Maths. Math. Phys.* — 1989. — Vol. 29. — No 1. — Pp. 34–46.

Krall A.M. (1975). The development of general differential and general differential-boundary systems, *Rocky Mountain Journal of Mathematics*. — 1975. — Vol. 5. — Pp. 493–542.

Dzhenaliev M.T., Ramazanov M.I. (2010). *Nagruzhennye uravneniya kak vozmushheniya differentsial'nyh uravnenij*. - Almaty: Gylym, 2010. — 334 p. [in Russ.]

Kadirbayeva Zh.M., Karakenova S.G. (2020). Numerical solution of the multipoint boundary value problems for essentially loaded ordinary differential equations // *Kazakh Mathematical Journal*. — 2020. — Vol. 20. — No. 4. — Pp. 47–57.

Kadirbayeva Zh.M., Bakirova E.A., Dauletbayeva A.Sh., Kassymgali A.A. (2021). An algorithm for solving a boundary value problem for essentially loaded differential equations // *News of the NAS RK. Phys.-Math. Series*. — 2021. — Vol. 2. — No. 336. — P. 6–14.

Kadirbayeva Zh.M. (2021). A numerical method for solving boundary value problem for essentially loaded differential equations // *Lobachevskii Journal of Mathematics*. — 2021. — Vol. 42. — No. 3. — Pp. 551–559.

Nakhushev A.M. (2012). *Nagruzhennye uravneniya i ih primeneniye*. — M.: Nauka, 2012. — 232 p. [in Russ.]

Nakhushev A.M. (1995). *Uraveniya matematicheskoy biologii*. — M.: Vyshaiya shkola, 1995. — 301 p. [in Russ.]

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 2. Number 10 (2022). Pp. 103–113

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2022.10.2.010>

UDC 519.257, 005

RESEARCH ON RISK ANALYSIS METHODS USING MODELS OF DEFAULT PROBABILITY IN THE FINANCIAL INDUSTRY

A.R. Suleimenova^{1}, A.Zh. Sayabayeva¹, A.N. Moldagulova²*

¹International Information Technology University, Almaty, Kazakhstan;

²Kazakh National Research Technical University, Almaty, Kazakhstan.

Modagulova Aiman Nikolayevna — Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department "Software Engineering", Kazakh National Research Technical University named after K.I. Satpayev

ORCID:0000-0002-1596-561X. E-mail: aigerimsuleimenova98@gmail.com;

Suleimenova Aigerim Raikhanovna — Master's student, specialty "Business Analytics", the International University of Information Technologies

ORCID: 0000-0002-5754-651X;

Sayabayeva Arida Zhomartkyzy — Master's student, specialty "Business Analytics", the International University of Information Technologies

ORCID:0000-0003-4403-7625. E-mail: arida0304z@gmail.com.

© A.R. Suleimenova, A.Zh. Sayabayeva, A.N. Moldagulova, 2022

Abstract. The article discusses the issues of methods for analyzing the probability of default (bankruptcy) of enterprises in the context of the financial industry. Based on the research, it proposes approaches to assess the probability of default. Many models help analyze credit risks, for example, the probability of default, migration risk, and default on losses. Each of these models is vital for assessing credit risk, however, one of the basic and essential models is the probability of the default model, i.e., used in this article. To solve this problem, a standard mathematical apparatus is used: linear regression, matrix theory, and nonlinear programming.

Keywords: the probability of default; loan quality category; reserves for possible loan losses; binary regression

For citation: A.R. Suleimenova, A.Zh. Sayabayeva, A.N. Moldagulova. Research on risk analysis methods using models of default probability in the financial industry // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 2. Number 10. Pp. 103–113 (In Russ.). DOI: [10.54309/IJICT.2022.10.2.010](https://doi.org/10.54309/IJICT.2022.10.2.010).



ҚАРЖЫ САЛАСЫНДАҒЫ ТӘУЕКЕЛДЕРДІ ТАЛДАУ ӘДІСТЕРІН ҚАРЖЫ САЛАСЫНДАҒЫ ҚАРЖЫЛЫҚ ЫҚТИМАЛДЫҚ ҮЛГІЛЕРІН ПАЙДАЛАНҒАН ЗЕРТТЕУ

А.Н. Молдагулова^{1}, А.Р. Сулейменова¹, А.Ж. Саябаева²*

Молдагулова Айман Николаевна — физика-математика ғылымдарының кандидаты, доцент, Қ.И. Сәтбаев атындағы Қазақ ұлттық ғылыми-зерттеу техникалық университеті «Программалық инженерия» кафедрасының меңгерушісі

ORCID:0000-0002-1596-561X. E-mail: aigerimsuleimenova98@gmail.com;

Сулейменова Айгерим Райхановна — Халықаралық ақпараттық технологиялар университетінің «Бизнес-аналитика» мамандығының екінші курс магистранты

ORCID: 0000-0002-5754-651X.

Саябаева Арида Жомартқызы — Халықаралық ақпараттық технологиялар университетінің «Бизнес-аналитика» мамандығының екінші курс магистранты

ORCID:0000-0003-4403-7625. E-mail: arida0304z@gmail.com.

© Молдагулова А.Н., Сулейменова А.Р., Саябаева А.Ж., 2022

Аннотация. Мақалада қаржылық индустрия контекстінде кәсіпорындардың дефолт (банкрот) ықтималдығын талдау әдістері қарастырылады. Зерттеу негізінде ол дефолт ықтималдығын бағалау тәсілдерін ұсынады. Көптеген модельдер несиелік тәуекелдерді талдауға көмектеседі, мысалы, дефолт ықтималдығы, көші-қон тәуекелі және шығындар бойынша дефолт. Осы үлгілердің әрқайсысы несиелік тәуекелді бағалау үшін өте маңызды, дегенмен негізгі және маңызды үлгілердің бірі әдепкі үлгінің ықтималдығы болып табылады, яғни осы мақалада пайдаланылады. Бұл мәселені шешу үшін стандартты математикалық аппарат қолданылады: сызықтық регрессия, матрицалық теория және сызықты емес бағдарламалау.

Түйін сөздер: дефолт ықтималдығы, несие сапасы санаты, несие бойынша ықтимал шығындарға арналған резервтер, екілік регрессия

Дәйексөз үшін: Молдагулова А.Н., Сулейменова А.Р., Саябаева А.Ж. Қаржы саласындағы тәуекелдерді талдау әдістерін қаржы саласындағы қаржылық ықтималдық үлгілерін пайдаланған зерттеу // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 2. Нөмірі 10. 103–113 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.10.2.010.

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА РИСКА С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ ВЕРОЯТНОСТИ ДЕФОЛТА В ФИНАНСОВОЙ ОТРАСЛИ

А.Н. Молдагулова^{1}, А.Р. Сулейменова¹, А.Ж. Саябаева²*

Молдагулова Айман Николаевна — кандидат физико-математических наук, ассоциированный профессор, заведующая кафедрой «Программная инженерия», Казахского национального исследовательского технического университета имени К.И. Сатпаева



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ORCID:0000-0002-1596-561X. E-mail: aigerimsuleimenova98@gmail.com;

Сулейменова Айгерим Райхановна — магистрант второго курса специальности «Бизнес аналитика» Международного университета информационных технологий

ORCID:0000-0002-5754-651X;

Саябаева Арида Жомарткызы — магистрант второго курса специальности «Бизнес аналитика» Международного университета информационных технологий

ORCID:0000-0003-4403-7625. E-mail: arida0304z@gmail.com.

© Молдагулова А.Н., Сулейменова А.Р., Саябаева А.Ж., 2022

Аннотация. В статье рассматриваются вопросы методов анализа вероятности дефолта (банкротства) предприятий в разрезе финансовой отрасли. На основе исследования предлагаются подходы к оценке вероятности дефолта. Многие модели помогают анализировать кредитные риски, например вероятность дефолта, миграционный риск и дефолт по убыткам. Каждая из этих моделей жизненно необходима для оценки кредитного риска, однако одной из основных и существенных моделей является модель вероятности дефолта, т.е. используемая в данной статье. Для решения этой задачи используется стандартный математический аппарат: линейная регрессия, теория матриц и нелинейное программирование.

Ключевые слова: вероятность дефолта, категория качества кредита, резервы на возможные потери по ссудам, бинарная регрессия

Для цитирования: Молдагулова А.Н., Сулейменова А.Р., Саябаева А.Ж. Исследование методов анализа риска с использованием моделей вероятности дефолта в финансовой отрасли // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 2. Номер 10. Стр. 103–113 (на русском языке). DOI: 10.54309/IJICT.2022.10.2.010.

Introduction

To date, commercial banks in the developed countries have designed and tested many mathematical models for assessing borrowers' credit risk. Many of them have stood the test of time and are now used in world practice. Thus, the Basel Committee on Banking Supervision proposes a standardized approach for assessing credit risk, based on the assessment of external rating agencies, and an approach based on internal ratings (Internal Rating Based approach, IRB) (<http://www.bis.org/publ/bcbasc111.htm>).

The problem of solving issues related to default in the financial sphere is considered by us based on constructing a mathematical integral model that evaluates the individual risk of the borrower and acts as a basis for assessing the aggregate risk. The main disadvantage of this approach is the unaccounted fact that over the life of the loan there is a change in the individual probability of going into default. The actual assessment of the loan is investigated not only for the first year of the loan's life but also for the entire term of its life. According to the IFRS standard, the volume of reserves formed should depend on the dynamics of the risk level, which is determined by assessing the risk of default throughout the life of the loan (lifetime PD). A model is proposed for obtaining a dynamic estimate of the probability of default based on its data on the quality of



fulfillment of obligations under the loan agreement (future engineering) and the phase of the macroeconomic cycle. The model is based on the method of logistic regression with regularization, the quality of which is determined through the calculation of the Gini coefficient. Building a model for estimating the probability of going into default.

Methods

An essential class is market models that are based on information available to stock market players. Firstly, it is the market data of the borrower's listed securities. Such models can be divided into structural and abbreviated models.

At the center of the structure is the idea that the value of the company's shares is an option on the company's assets with a transaction price equal to its liabilities. The founders of this class of models are BLACK and Shoals, Merton. Merton considered the company's accounts payable as a claim that could be changed in its value and used the BLACK — S choles option pricing formula to assess the probability of the company's default (Black et al., 1973: 637–654). Within the framework of this model, the provision of a loan is interpreted as the purchase of company assets from shareholders and the transfer of an option on these assets with an exercise price equal to the cost of the loan and an execution time equal to the repayment of the loan. V_t represents the value of the company's assets at time t , and after the payment expires at time T , creditors demand the nominal amount of debt D . the model uses a simplified description of the debt structure: it is assumed that the company has one debt in the amount of D in the form of a zero-coupon bond, that is, a zero-coupon debt. the total amount of debt must be paid in a lump sum at the time of T . Suppose that the amount of debt is fully secured by the assets of the company, that is, $V_T \geq D$, then the shareholders receive the difference $V_T - D$. If the debt is not secured by the assets of the enterprise, that is, the creditors receive what is available to them, the shareholders do not receive anything (this follows from the priority of fulfilling creditors' claims to shareholders in the event of bankruptcy of the enterprise).

With increasing uncertainty in financial markets, credit risk management is a priority in banking risk management. By the Basel Agreement, the basis of credit risk management is the formation of a reserve capitalization corresponding to the expected losses calculated for the portfolio (Karminsky et al., 2005). Therefore, the correctness of the assessment of expected losses becomes more relevant.

Models that are based on financial and accounting data of corporate borrowers of the bank are quite common. Depending on the statistical method used, they can be divided into scoring models, linear discriminant analysis models (among which, in turn, discriminant analysis models with one and several variables differ), and binary choice models. A credit score is a statistical method proposed by the American scientist David Durand in 1941, which was originally used to rank retail borrowers. When using the scoring model, each borrower is assigned a rating that characterizes his financial condition and ability to repay his obligations to the lender promptly. In the future, the entire range of possible values is divided into intervals, rating groups. By calibrating the model based on historical data, each rating score is adjusted to take into account the probability of default, which for the most part is a determination of



the proportion of companies in this group that have defaulted during the year. Now, credit-scoring models are quite common in banking practice, especially for assessing the probability of default on single retail loan portfolios. However, Durand's original scoring model was quite simplistic. The author took into account certain characteristics of each retail borrower (such as gender, age, living in the place, profession, place and work experience, financial condition), and depending on their cost, he assigned each client a certain point. If the total score for all characteristics exceeded 1.25 points, the client was considered solvent. Modern credit rating systems of Moderna have become much more complex and universal, but their disadvantage is discretion, as well as a prerequisite for having a wide base of credit histories. Since valuation models need a lot of background information about loans and borrowers' credit histories to determine dependencies, it is necessary to regularly update the reference data and, consequently, the identified dependencies. In the group of discriminant analysis models, discriminant analysis models with one variable, the ancestor of which is Beaver, are the simplest for assessing the borrower's default.

According to Basel II and IFRS, the assessment of the expected losses of the loan portfolio is based on the IRB approach to credit risk, which, if the bank complies with certain minimum conditions and disclosure requirements, allows it to rely on its internal assessments of risk components when calculating capital coverage of a certain risk [4]. According to the Basel II agreement, the calculation of expected losses EL (Expected Loss) on the loan is made according to the following formula:

$$EL \square PD \square EAD \square LGD, \quad (1)$$

Where PD (Probability of Default) — the probability of default.

EAD (Exposure at Default) — exposure to credit risk, which is an economic assessment of the value of assets at risk at the time of default exposure to credit risk, which is an economic assessment of the value of assets at risk at the time of default;

LGD (Loss Given Default) — losses in case of default, reflecting the share of irretrievable loss in case of default. losses in case of default, reflecting the share of irretrievable loss in case of default.

At the same time, most often, like PD , an individual assessment of the probability of default is used, which is calculated based on a scoring model. According to the recommendations of the Basel II Agreement, overdue principal debt or interest for 90 days or more during the first year of the loan term is considered unpaid, and the probability of overdue principal debt or interest for more than 90 days during the first year of the loan term is defined as overdue (in the future, the the year will be equal to 365 calendar days). The main problem when using the current definition is the unregistered fact that over the life of the loan there is a change in the individual probability of default. In this regard, the issue of credit assessment not only for the first year of the loan term but also for its entire life is becoming more and more relevant. This provision is enshrined in the standard IFRS 9, according to which the number of reserves created should depend on the dynamics of the risk level, which is determined by the assessment of the risk of default on the loan during the entire loan term (lifetime estimation).

Loan portfolio data for the last month for a certain period (for example, 2012–01–

31, 2012–02–29, 2012–03–31, ...). For each cutting date t , a loan portfolio of assets is formed: loans opened for which there was no violation event and for which the actual values above the output violation indicator in the interval are known. Loan portfolios for each cut-off point t in one sample. The division into a training sample and a validation sample is based on a random selection of credits relative to 70/30 so that each credit falls only in the training sample or validation sample. It is worth noting that when using a simple random selection of observations, we will get an obvious retrain of the model: the coefficient in the test sample may be greater than in the training sample. The correct construction of training and validation samples will allow obtaining more stable estimates of models, which after the implementation of the model will exclude the occurrence of unpredictable situations when assessing the probability of default of the loan portfolio in any segment (provided that the portfolio structure will not change, and the macroeconomic situation will be stable).

The variables that will be used to build a behavior assessment model are based on the credit history of each loan. The raised issue of choosing variables that will correctly reflect the quality of fulfillment of obligations under the loan agreement (future engineering) is relevant. Let's introduce the following notation: let PR be a set of values for the number of days of delay that the client allows, and M is a set of values for the number of months during which the presence of a delay is estimated.

Then it is proposed to use the following quantitative estimates as variables:

- the number of exits for overdue PR in the first m months of the loan term -these variables allow you to assess the quality of loan servicing in the first year of the loan term;
- the number of exits on overdue PR in the last m months of the loan term (the countdown of months begins with the court date, timer) — allows you to assess the quality of credit service in the last year of the loan life before the court date, thus, the credit history of the last loan is evaluated;
- the ratio of the total number of outflows of overdue public relations for the entire loan term to the loan term in months;
- the ratio of the total number of exits from PR delinquencies for the entire term of the loan to the number of months that the loan has lived on the cut-off date;
- the proportion of the maximum number of months during which the loan was in permanent public relations debt, from the number of months during which the loan lived until the cut-off date.

Similarly, variables are calculated based on the overdue amount of loans:

- the ratio of the overdue PR amount in the first m months of the loan's life to the total loan amount;
- the ratio of the overdue amount of PR for the last m months of the life of the loan, to the total amount of the loan, the months are counted from the countdown cut-off date;
- the ratio between the total amount of overdue public relations debt during the entire term of the loan, divided into months, and the total amount of the loan;
- the ratio of the total amount of overdue PR for the entire term of the loan, divided

by the number of months during which the loan exists at the cut-off date, to the total amount of the loan;

- the ratio between the average amount overdue when the loan was in permanent public relations debt and the total amount of the loan.

The described variables can be attributed to a block of variables that change over the life of the loan. In addition, it is worth taking into account the fact that the historical sample of the loan portfolio from some segment T contains loans with different useful lives. Using the credit history of loans as variables, we get that there is incomplete and time-limited data for building the model, i.e. censored data on the right, i.e. there is some history of the customer's behavior on the loan before the cutoff, but there is no information about how the customer's behavior will look in the next period $[t, T]$, where T is the cutoff time, T is the loan term. Because of this, an additional variable is introduced: the fact that the loan is valid until a certain point in time.

In addition to the block of variables that change over the life of the loan, it is possible to allocate a block of variables that are fixed at the time of loan issuance, namely:

- part of the initial payment of the amount of the loan issued;
- advance payment of the cost of the goods;
- PD0 score, which determines the probability of default during the first year of the loan's life and is calculated at the time of loan issuance based on the characteristics of the borrower's questionnaire. At the same time, most often, like PD, an individual assessment of the probability of default is used, which is calculated based on a scoring model. By the recommendations of the Basel II Agreement, overdue principal debt or interest for 90 days or more during the first year of the loan term is considered unpaid, and the probability of overdue principal debt or interest for more than 90 days during the first year of the loan term is defined as overdue (in the future, the year will be equal to 365 calendar days). The main problem when using the current definition is the unregistered fact that over the life of the loan there is a change in the individual probability of default. In this regard, the issue of credit assessment not only for the first year of the loan term but also for its entire life is becoming more and more relevant. This provision is enshrined in the standard IFRS 9, according to which the number of reserves created should depend on the dynamics of the risk level, which is determined by the assessment of the risk of default on the loan during the entire loan term (lifetime estimation).

Loan portfolio data for the last month for a certain period (for example, 2012–01–31, 2012–02–29, 2012–03–31, ...). For each cutting date t, a loan portfolio of assets is formed: loans opened for which there was no violation event and for which the actual values above the output violation indicator in the interval are known. Loan portfolios for each cut-off point t in one sample. The division into a training sample and a validation sample is based on a random selection of credits relative to 70/30 so that each credit falls only in the training sample or validation sample. It is worth noting that when using a simple random selection of observations, we will get an obvious retrain of the model: the coefficient in the test sample may be greater than in the training sample (Altman,

2003: 201–214). The correct construction of training and validation samples will allow obtaining more stable estimates of models, which after the implementation of the model will exclude the occurrence of unpredictable situations when assessing the probability of default of the loan portfolio in any segment (provided that the portfolio structure will not change and the macroeconomic situation will be stable).

Result

The variables that will be used to build a behavior assessment model are based on the credit history of each loan. The raised issue of choosing variables that will correctly reflect the quality of fulfillment of obligations under the loan agreement (future engineering) is relevant. Let us introduce the following notation: let PR be a set of values for the number of days of delay that the client allows, and M is a set of values for the number of months during which the presence of a delay is estimated.

Then it is proposed to use the following quantitative estimates as variables:

- the number of exits for overdue PR in the first m months of the loan term -these variables allow you to assess the quality of loan servicing in the first year of the loan term;

- the number of exits on overdue PR in the last m months of the loan term (the countdown of months begins with the court date, timer) - allows you to assess the quality of credit service in the last year of the loan life before the court date, thus, the credit history of the last loan is evaluated;

- the ratio of the total number of outflows of overdue public relations for the entire loan term to the loan term in months;

- the ratio of the total number of exits from PR delinquencies for the entire term of the loan to the number of months that the loan has lived on the cut-off date;

- the proportion of the maximum number of months during which the loan was in permanent public relations debt, from the number of months during which the loan lived until the cut-off date.

Similarly, variables are calculated based on the overdue amount of loans:

- the ratio of the overdue PR amount in the first m months of the loan's life to the total loan amount;

- the ratio of the overdue amount of PR for the last m months of the life of the loan, to the total amount of the loan, the months are counted from the countdown cut-off date;

- the ratio between the total amount of overdue public relations debt during the entire term of the loan, divided into months, and the total amount of the loan;

- the ratio of the total amount of overdue PR for the entire term of the loan, divided by the number of months during which the loan exists at the cut-off date, to the total amount of the loan;

- the ratio between the average amount overdue when the loan was in permanent public relations debt and the total amount of the loan.

The described variables can be attributed to a block of variables that change over the life of the loan. In addition, it is worth taking into account the fact that the historical sample of the loan portfolio from some segment T contains loans with different useful

lives. Using the credit history of loans as variables, we get that there is incomplete and time-limited data for building the model, i.e. censored data on the right, i.e. there is some history of the customer's behavior on the loan before the cutoff, but there is no information about how the customer's behavior will look in the next period $[t, T]$, where T is the cutoff time, T is the loan term. Because of this, an additional variable is introduced: the fact that the loan is valid until a certain point in time.

In addition to the block of variables that change over the life of the loan, it is possible to allocate a block of variables that are fixed at the time of loan issuance, namely:

- part of the initial payment of the amount of the loan issued;
- advance payment of the cost of the goods;
- PD0 score, which determines the probability of default during the first year of the loan's life and is calculated at the time of loan issuance based on the characteristics of the borrower's questionnaire.

Analysis

Thus, variables are calculated for each contract, based on which a behavioral scoring model for assessing the default is built. From the available variety of model construction methods in the current example, the method of logistic regression with regularization was used as a method of model construction. The formula for calculating the scoring score PD:

$$PD = \frac{w}{1 + e^{-z}} \quad (2)$$

where z is calculated as the sum of the constant and the points corresponding to the characteristics of the borrower and their combinations:

$$z = b_1 x_1 + b_2 x_2 + \dots + b_n x_n + b_0,$$

$x_1, x_2, \dots, x_n$ – values of variables

$b_1, b_2, \dots, b_n$ – coefficients for variables,

b_0 – some constant

w – correction factor for the macroeconomic cycle.

The z value is determined by the calculated variables, which are the characteristics of the borrower. Additional constraints are applied to the vector of weights in the model by regularizing the model. This approach consists in choosing small weights in the absolute value on average, which leads to less instability of the model, that is, dependence on training data. The elastic network method is used as a regularization scheme, which combines loop and ridge, regression models (Altman, 1968).

Discussion

Let us consider the construction of a model for estimating the probability of PDT default at the moment (slice) t on the example of a regional retail bank. The data of the loan portfolio for the last month, taking into account the fact of the "survival" of the loan to each segment under study, serve as reference data for constructing a model for estimating the PDT score. The model is built in the recent segment of the portfolio. For

each cut-off date, a portfolio of current loans was formed-loans opened from the cut-off date and with a current default on principal or interest of no more than 15 days. The portfolios obtained for each cut-off point were combined into one training sample by the described method. The volume of the study sample amounted to 151586 loans, the share of overdue loans in which 0.04146 by 31-10-2017. The volume of the validation sample is 66310, the share of overdue loans is 0.04340. The sample size is 22038, the share of overdue loans is 0.03825. The division of the sample into formation and validation is based on the rule described above. For each of the samples, the above variables were calculated, assuming PR 1, 15, 30, 90, and M {1, 3, 6, 9, 12}, and based on it, a logistic regression model with regularization was built. The coefficient in the training sample was 0.7163, in the validation sample - 0.7024, in the test sample - the actual score according to the state (Figure 1). The z value is determined by the calculated variables, which are the characteristics of the borrower. Additional constraints are applied to the vector of weights in the model by regularizing the model. This approach consists in choosing small weights in the absolute value on average, which leads to less instability of the model, that is, dependence on training data. The elastic network method is used as a regularization scheme, which combines loop and ridge, regression models.

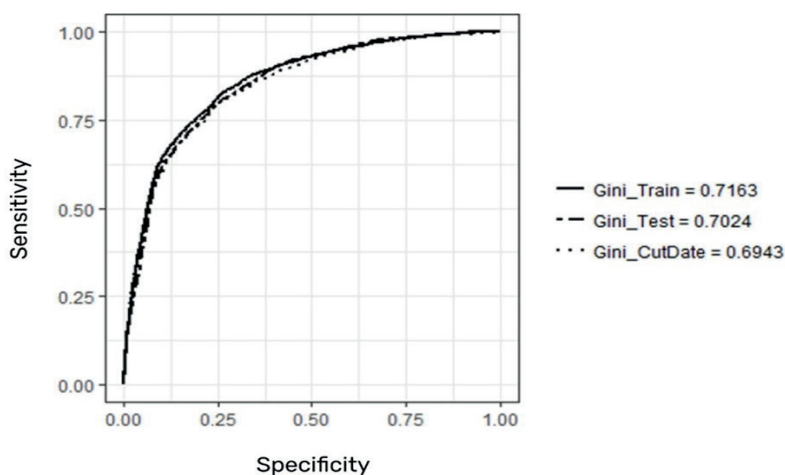


Figure 1 - ROC curve graph

Conclusion

In this paper, a solution to the problem was presented not only for the first year of the loan "life", but also for the entire lifetime (life, assessment). Based on the behavioral model of architecture evaluation. For instance, the regional retail bank used a car as an example of how to build a PDt default probability model to estimate at some point (slice) t , taking into account the terms of the loan agreement of fulfillment of obligations, the quality estimated by the described variable sizes (future engineering) and the macroeconomic phase of the cycle. The obtained quality indicators from the model indicate its high-power prediction. This credit risk assessment of the credit risk

during the action complies with the requirements of the standard and allows to create a reserve volume of the portfolio based on the level of risk in dynamics.

REFERENCES

- Altman E.I. (2003). «Managing credit risk: a challenge for the new millennium». *Economic Notes*. — Vol. 31. — Issue 2 (December), — Pp. 201–214.
- Altman E.I. (1968). «Financial ratios. Discriminant analysis, and the prediction of corporate bankruptcy». *Journal of Finance*, September.
- Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards, Comprehensive Version. — <http://www.bis.org/publ/bcbasc111.htm>.
- Black F., Scholes M. (1973). «The pricing of options and corporate liabilities». *The Journal of Political Economy*. — Vol. 81. — No. 3 (May — Jun.). — Pp. 637–654.
- Karminsky A.M., Peresetsky A.A., Petrov A.E. (2005). *Ratings in economics: methodology and practice: Monograph / Edited by A.M. Karminsky. - M.: Finance and Statistics, — 2005.*
- Bobyshev A., Galperin F., Mishchenko Ya. (2005). *The practice of applying the VaR methodology for assessing and managing credit risk in Alfa-Bank // Financial Risk Management. — 2005. — № 2.*



**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

Правила оформления статьи для публикации в журнале на сайте:

<https://journal.iitu.edu.kz>

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных
технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР

Ералы Диана Русланқызы

КОМПЬЮТЕРНАЯ ВЕРСТКА

Жадыранова Гульнур Даутбековна

Подписано в печать 15.06.2022.

Формат 60x881/8. Бумага офсетная. Печать - ризограф. 7,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.