

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



**INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION
TECHNOLOGIES**

Published since 2020.
Volume 7. 1 (25). 2026
January–March

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

2020 жылдан бері шығарылады
Том 7. 1 (25). 2026
Қаңтар-Наурыз

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Издается с 2020 г.
Том 7. 1 (25). 2026
Январь-Март

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

EDITOR-IN-CHIEF:

Kateryna Kolesnikova — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

DEPUTY EDITOR-IN-CHIEF:

Madina Ipalakova — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

EDITORIAL BOARD:

Abdul Razak — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Lucio Tommaso De Paolis — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)

Liz Bacon — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)

Michele Pagano — PhD, Professor, University of Pisa (Italy)

Mukhtarbay Otelbayev — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Bolatbek Rysbauly — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

Yevgeniya Daineko — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurzhan Duzbayev — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)

Bakhtgerci Sinchev — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)

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

Ardak Mukhamediyeva — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)

Zamira Abdikalikova — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

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

Damilya Yeskendirova — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)

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

Altai Aitmagambetov — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Yelena Bakhtiyarova — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Kanibek Sansyrbay — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Sakhybay Tynymbayev — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)

Ali Abd Almisreb — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Yang Im Chu — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, Vice-Rector, Adam Mickiewicz University (Poland)

Orken Mamyrbayev — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)

Sergey Bushuyev — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association "UKRNET," Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)

Svetlana Beloshitskaya — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

MANAGING EDITOR

Raushan Mrzabayeva — Master of Science, editor, International Information Technology University (Kazakhstan)

International Journal of Information and Communication Technologies

Periodicity: 4 times a year.

Languages: Kazakh, Russian, English

DOI prefix: 10.54309

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Distribution: Materials are distributed under the Creative Commons Attribution 4.0

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

Owner: International Information Technology University JSC (Almaty).

Copyright: © International Journal of Information and Communication Technologies, 2026

РЕДАКЦИЯ

БАС РЕДАКТОР:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

- Разак Абдул** — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)
Луччо Томмазо де Паолис — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры
Лиз Бэкон — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары
Микеле Пагано — PhD, Пиза Университетінің (Италия) профессоры
Өтелбаев Мухтарбай Өтелбайұлы — физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)
Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Қазақстан)
Дайнеко Евгения Александровна — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)
Дузаев Нуржан Тоқсулжанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)
Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)
Сейлова Нургуль Абдуллаевна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)
Мухамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнес медиа және басқару факультетінің деканы (Қазақстан)
Абдикаликова Замира Турсынбаевна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)
Шильдибеков Ерлан Жаржанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)
Дамелия Максустовна Ескендрова — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)
Ниязгулова Айгуль Аскарбековна — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)
Айтмағамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)
Бахтиярова Елена Ажибековна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)
Канибек Сансызбай — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)
Тынымбаев Сахибай — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)
Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)
Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)
Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)
Талеуш Валлас — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры
Мамырбаев Оркен Жумажанович — PhD, ҚР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)
Бушув Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сулет университеті жобаларды басқару кафедрасының меңгерушісі (Украина)
Белюшицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Мерзімділігі: жылына 4 рет.

Басылым тілі: қазақ, орыс, ағылшын.

Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".

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

Тарату: материалдар Creative Commons Attribution 4.0 лицензиясы бойынша таратылады

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

Авторлық құқық: © Халықаралық ақпараттық және коммуникациялық технологиялар журналы, 2026

РЕДАКЦИЯ

ГЛАВНЫЙ РЕДАКТОР:

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

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

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)

Лучио 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 Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токсуажевич — 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 бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

Абдикаликова Замира Турсынбаевна — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

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

Дамеля Максютнова Ескендрова — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

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

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

Бахтиярова Елена Ажибековна — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Канибек Сансызбай — PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

Тынымбаев Сахпай — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алимураб 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)

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

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Периодичность: 4 выпусков в год.

Язык издания: казахский, русский, английский.

Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социально-экономических систем".

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

Распространение: материалы распространяются по лицензии Creative Commons Attribution 4.0

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

Авторские права: © Международный журнал информационных и коммуникационных технологий, 2026

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

A.B. Zhalgas, Y.N. Kalpakov, B.Ye. Amirgaliyev
MACHINE LEARNING-DRIVEN OPTIMIZATION OF LOGISTICS IN SMART CITIES: A CASE STUDY OF ASTANA9

L. Kurmangaziyeva, Sh. Kodanova, M. Urazgaliyeva, O. Findik, S. Iskakova
INTEGRATING FUZZY LOGIC AND ARTIFICIAL INTELLIGENCE IN OPTIMIZING BUSINESS PROCESS AUTOMATION DECISIONS24

Y. Mailybayev, U. Adilbayeva, R. Amanova
ORGANIZATION OF AN ONLINE SURVEY OF PARTICIPANTS IN THE EDUCATIONAL PROCESS AND ANALYSIS OF THE RESULTS BASED ON THE MODIFIED DELPHI METHOD46

V.A. Takizhanov, A.Z. Ibragimov, A. Shalakhmetov
SIMULATION-BASED ROBUSTNESS ASSESSMENT OF ASTANA'S BUS NETWORK UNDER RANDOM AND TARGETED FAILURES61

INFORMATION TECHNOLOGY

M. Zh. Aitimov, G. K. Muratova, Zh. K. Bissenbayeva, I.M. Bapiyev, M. Kassim
SEMANTIC COMPLETENESS IN KAZAKH-LANGUAGE EXTRACTIVE QA THROUGH ONTOLOGY AND RETRIEVAL MECHANISMS76

O.N. Akylbekov, Y.T. Dauletbek, A.N. Moldagulova, G.S. Zakariya, D.A. Gura
MACHINE LEARNING METHODS FOR ANALYSING THREE-DIMENSIONAL SPATIAL DATA IN KAZAKHSTAN'S LAND USE PLANNING.....89

S.Zh. Aliaskarov, R.K. Uskenbayeva, A. Razaque, A.B. Kassymova, A.M. Anartayeva
TOWARDS EFFICIENT BIG DATA ANALYTICS IN REGIONAL SYSTEMS: PRACTICAL INSIGHTS FROM HYBRID ARCHITECTURE DEPLOYMENT.....109

A. Ismailova, G. Yessenbayeva, K. Kadyrkulov, R. Moldasheva, A. Amangeldi
DEVELOPMENT OF A HYBRID DEEP LEARNING MODEL FOR MULTICLASS CLASSIFICATION OF MICROSCOPIC IMAGES OF BACTERIA128

G. Kalman, J. Kultan, A.N. Ismukamova, N.M. Ausilova, Y.V. Makhatova
A DOMAIN-KNOWLEDGE-BASED MODEL FOR REFERENCE RESOLUTION IN LOW-RESOURCE LANGUAGES141

Y. Kamen, Zh. Yessendauletova, L. Fazylova, M. Rakhimzhanova, A.M. Nedzved
USING NEURAL NETWORKS FOR OBJECTIVE ASSESSMENT OF ATTENTION IN CHILDREN BASED ON EEG DATA158

A.Ye. Kulakayeva, Ye.A. Bakhtiyarova, G.T. Jakanova, Sh. Nursultan
COMPARATIVE ANALYSIS OF VARIOUS RADIO WAVE PROPAGATION MODELS FOR MOBILE NETWORK COVERAGE PREDICTION173

M.B. Nurpeissova, Sh.K. Aitkazinova, A.M. Abenov, N.S. Donenbayeva
METHODOLOGY FOR TRANSFORMING SATELLITE COORDINATES INTO A TOPOCENTRIC RECTANGULAR COORDINATE SYSTEM189

A. Ospanov, P. Alonso-Jordá, A. Zhumadillayeva
BLOCKCHAIN-ENABLED ERP WAREHOUSE INTEGRATION WITH IOT DIMENSIONERS AND MACHINE LEARNING-OPTIMIZED DIMENSIONAL WEIGHT RECONCILIATION202

A.A. Sakhipov, R.B. Seitbek
EVENT-DRIVEN MICROSERVICES FOR INCIDENT DETECTION AND RESPONSE IN INTELLIGENT TRAFFIC SYSTEM218

G. Yusupova, K.S. Shadinova, D. Ussipbekova, Zh.Zh. Azhibekova, P. Schmidt
DETERMINATION OF SOIL PROFILE STRATIFICATION AT 0–200 CM DEPTH USING A MULTILEVEL STACKING MODEL231

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

S.A. Adilzhanova, M.Zh. Sakypbekova, L.Sh. Cherikbaeva, G.A. Tyulepberdinova, G.T. Zhubanysheva SYSTEMATIC ANALYSIS OF RISK ASSESSMENT METHODS AND MODELS IN INFORMATION SECURITY.....	244
T. K. Zhukabayeva, D.B. Baumuratova, E. Benkhelifa, N.A. Niyetbayeva EDGE COMPUTING-BASED TECHNIQUE FOR CONSTRUCTION OF ATTACK DETECTION MEANS IN CYBER-PHYSICAL SYSTEMS OF INDUSTRIAL INTERNET-OF-THINGS	270
N.E. Karabayev, S.K. Serikbayeva, Y.M. Mardenov, B. Tassuov, M. Fajkus DETECTION OF CYBER ATTACKS IN TRANSPORT NETWORKS BASED ON MACHINE LEARNING METHODS	292
V.A. Kumalakov, A.O. Dargulova A HYBRID FRAMEWORK FOR RESUME-JOB MATCHING SYSTEM	311
V. Makhatova, B. Dzhugembayeva, A. Gabdulova, L. Nurgaliyeva, A. Abdigaliyeva MATHEMATICAL MODEL FOR OPTIMAL SENSOR SELECTION IN SIEM SYSTEMS USING THE ANALYTIC HIERARCHY PROCESS	326

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев АҚЫЛДЫ ҚАЛАЛАРДАҒЫ ЛОГИСТИКАНЫ МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН ОҢТАЙЛАНДЫРУ: АСТАНАНЫҢ ЖАҒДАЙЫН ЗЕРТТЕУ.....	9
Л.Курманғазиева, Ш. Қоданова, М. Уразғалиева, О. Findik, С. Искакова ЖАСАНДЫ ИНТЕЛЛЕКТ ПЕН АЙҚЫН ЕМЕС ЛОГИКАНЫ БІРІКТІРУ АРҚЫЛЫ БИЗНЕС-ПРОЦЕСТЕРДІ АВТОМАТТАНДЫРУ ШЕШІМДЕРІН ОҢТАЙЛАНДЫРУ	24
Е. Майлыбаев, У. Адилбаева, Р. Аманова ҰЙЫМДАСТЫРЫЛҒАН ОНЛАЙН САУАЛНАМА АРҚЫЛЫ БІЛІМ БЕРУ ПРОЦЕСІНЕ ҚАТЫСУШЫЛАРДЫҢ ПІКІРЛЕРІН ЖИНАУ ЖӘНЕ НӘТИЖЕЛЕРІН МОДИФИКАЦИЯЛАНҒАН ДЕЛЬФИ ӘДІСІ НЕГІЗІНДЕ ТАЛДАУ	46
В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов МОДЕЛЬДЕУ НЕГІЗІНДЕ АСТАНАНЫҢ АВТОБУС ЖЕЛІСІНІҢ ТҰРАҚТЫЛЫҒЫН БАҒАЛАУ: КЕЗДЕЙСОҚ ЖӘНЕ МАҚСАТТЫ ІСТЕН ШЫҒУЛАР ЖАҒДАЙЫНДА	61

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим ОНТОЛОГИЯ ЖӘНЕ ІЗДЕУ МЕХАНИЗМДЕРІ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРАКЦИЯЛЫҚ ҚАДАҒЫ СЕМАНТИКАЛЫҚ ТОЛЫҚТЫҚ	76
О.Н. Ақылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура ҚАЗАҚСТАННЫҢ АУМАҚТЫҚ ЖОСПАРЛАУЫНДАҒЫ ҮШ ӨЛШЕМДІ КЕҢІСТІКТІК МӨЛІМЕТТЕРДІ ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ	89
С.Ж. Алиасқаров, Р.К. Ускенбаева, А. Разак, А.Б. Қасымова, А.М. Анартаева АЙМАҚТЫҚ ЖҮЙЕЛЕРДЕГІ ҮЛКЕН ДЕРЕКТЕРДІ ТИІМДІ ТАЛДАУҒА ҚАРАЙ: ГИБРИДТІ АРХИТЕКТУРАНЫ ЕНГІЗУДІҢ ПРАКТИКАЛЫҚ ТҮСІНІКТЕР.....	109
А.А. Исмаилова, Г.Р. Есенбаева, Қ.К. Кадиркулов, Р.Н. Молдашева, А. Амангелді РОСКОПИЯЛЫҚ БЕЙНЕЛЕРІН КӨПКЛАССТЫ ЖІКТЕУГЕ АРНАЛҒАН ГИБРИДТІ ТЕРЕҢ ОҚЫТУ МОДЕЛІН ӘЗІРЛЕУ	128
Г. Қалман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова ПӘНДІК САЛА БІЛІМ НЕГІЗІНДЕ РЕУСРСТАРЫ АЗ ТІЛДЕРДЕГІ РЕФЕРЕНЦИЯНЫ ШЕШУДІҢ МОДЕЛІ.....	141
Е.Г. Кәмен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзьведь ЭЭГ ДЕРЕКТЕРІ БОЙЫНША БАЛАЛАРДЫҢ ЗЕЙІНІН ОБЪЕКТИВТІ БАҒАЛАУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРДІ ҚОЛДАНУ	158
А.Е. Кулакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нурсултан ҰЯЛЫ БАЙЛАНЫС ЖЕЛІЛЕРІНІҢ ҚАМТУ АЙМАҒЫН БОЛЖАУҒА АРНАЛҒАН ӨРТҮРЛІ РАДИОТОЛҚЫН ТАРАЛУ МОДЕЛЬДЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ	173

М.Б. Нұрпейісова, Ш.Қ. Айтқазынова, А.М. Абенов, Н.С. Дөненбаева
СПУТНИКТИК КООРДИНАТТАРДЫ ТОПОЦЕНТРЛІК ТІК БҰРЫШТЫ КООРДИНАТТАР ЖҮЙЕСІНЕ ТҮРЛЕНДІРУДІҢ ӘДІСТЕМЕСІ189

А. Оспанов, П. Алонсо-Хорда, А. Жұмаділлаева
БЛОКЧЕЙН-ТЕХНОЛОГИЯСЫМЕН ЫҚПАЛДАС ERP ҚОЙМА ЖҮЙЕСІН ІОТ ДИМЕНСИОНЕРЛЕР ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ АРҚЫЛЫ ОПТИМИЗАЦИЯЛАНҒАН ӨЛШЕМДІ САЛМАҚ ЕСЕПТЕУМЕН ИНТЕГРАЦИЯЛАУ202

А.А. Сахипов, Р.Б. Сейітбек
ОҚИҒАҒА БАҒДАРЛАНҒАН МИКРОҚЫЗМЕТТЕР ЖҮЙЕСІ АРҚЫЛЫ АҚЫЛДЫ ТРАФИК ЖҮЙЕЛЕРІНДЕ ОҚИҒАЛАРДЫ АНЫҚТАУ ЖӘНЕ ШАРАЛАР ҚОЛДАНУ218

Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, Р. Schmidt
ТОПЫРАҚ ПРОФИЛІНІҢ 0–200 СМ ТЕРЕҢДІКТЕГІ СТРАТИФИКАЦИЯСЫН КӨПДЕҢГЕЙЛІ СТЕКИНГ-МОДЕЛІМЕН АНЫҚТАУ.....231

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

С.А. Адилжанова, М.Ж. Сақыпбекова, Л.Ш. Черикбаева, Г.А. Тюлепбердинова, Г.Т. Жубанышева
АҚПАРАТТЫҚ ҚАУІПСІЗДІКТЕ ТӘУЕКЕЛДЕРДІ БАҒАЛАУ ӘДІСТЕРІ МЕН МОДЕЛЬДЕРІН ЖҮЙЕЛІ ТАЛДАУ.....244

Т.К. Жукабаева, Д. Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниегбаева
ШЕКАРАЛЫҚ ЕСЕПТЕУЛЕРДІ ҚОЛДАНА ОТЫРЫП, ЗАТТАРДЫҢ ӨНЕРКӘСІПТІК ИНТЕРНЕТІНІҢ КИБЕРФИЗИКАЛЫҚ ЖҮЙЕЛЕРІНДЕГІ ШАБУЫЛДАРДЫ АНЫҚТАУ ҚҰРАЛДАРЫН ҚҰРУ ӘДІСТЕМЕСІ.....270

Н.Е. Қарабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН КӨЛІК ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ292

Б.А. Кумалаков, А.О. Даргулова
ТҮЙІНДЕМЕЛЕР МЕН ВАКАНСИЯЛАРДЫ АВТОМАТТАНДЫРЫЛҒАН СӘЙКЕСТЕНДІРУГЕ НЕГІЗДЕЛГЕН ГИБРИДТІ ҮМІТКЕРЛЕРДІ ІРІКТЕУ ЖҮЙЕСІ311

В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нурғалиева, А. Абдигалиева
ИЕРАРХИЯЛАРДЫ ТАЛДАУ ӘДІСІ НЕГІЗІНДЕ SIEM ЖҮЙЕЛЕРІНДЕ ОҢТАЙЛЫ СЕНСОРДЫ ТАҢДАУДЫҢ МАТЕМАТИКАЛЫҚ МОДЕЛІ326

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев
ОПТИМИЗАЦИЯ ЛОГИСТИКИ В УМНЫХ ГОРОДАХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ: НА ПРИМЕРЕ АСТАНЫ9

Л. Курмангазиева, Ш. Коданова, М. Уразғалиева, О. Финдик, С. Исакова
ИНТЕГРАЦИЯ НЕЧЕТКОЙ ЛОГИКИ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОПТИМИЗАЦИИ РЕШЕНИЙ ПО АВТОМАТИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ24

Е. Майлыбаев, У. Адилбаева, Р. Аманова
СБОР МНЕНИЙ УЧАСТНИКОВ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПОСРЕДСТВОМ ОРГАНИЗОВАННОГО ОНЛАЙН-АНКЕТИРОВАНИЯ И АНАЛИЗ РЕЗУЛЬТАТОВ НА ОСНОВЕ МОДИФИЦИРОВАННОГО МЕТОДА ДЕЛЬФИ46

В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов
ОЦЕНКА УСТОЙЧИВОСТИ АВТОБУСНОЙ СЕТИ АСТАНЫ НА ОСНОВЕ МОДЕЛИРОВАНИЯ ПРИ СЛУЧАЙНЫХ И ЦЕЛЕНАПРАВЛЕННЫХ ОТКАЗАХ61

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим
СЕМАНТИЧЕСКАЯ ПОЛНОТА В КАЗАХСКОЯЗЫЧНОМ EXTRACTIVE QA ЧЕРЕЗ ОНТОЛОГИЮ И RETRIEVAL-МЕХАНИЗМЫ76

О.Н. Акылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ТРЁХМЕРНЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ В ТЕРРИТОРИАЛЬНОМ ПЛАНИРОВАНИИ КАЗАХСТАНА	89
С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева НА ПУТИ К ЭФФЕКТИВНОЙ АНАЛИТИКЕ БОЛЬШИХ ДАННЫХ В РЕГИОНАЛЬНЫХ СИСТЕМАХ: ПРАКТИЧЕСКИЕ ВЫВОДЫ ИЗ ВНЕДРЕНИЯ ГИБРИДНОЙ АРХИТЕКТУРЫ	109
А.А. Исмаилова, Г.Р. Есенбаева, К.К. Кадиркулов, Р.Н. Молдашева, А. Амангелды РАЗРАБОТКА ГИБРИДНОЙ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ БАКТЕРИЙ	128
Г. Калман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова МОДЕЛЬ НА ОСНОВЕ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ РАЗРЕШЕНИЯ КОРЕФЕРЕНЦИИ В МАЛОРЕСУРСНЫХ ЯЗЫКАХ	141
Е.Г. Камен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзьведь ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБЪЕКТИВНОЙ ОЦЕНКИ ВНИМАНИЯ У ДЕТЕЙ ПО ДАНЫМ ЭЭГ	158
А.Е. Кулакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нурсултан СРАВНИТЕЛЬНЫЙ АНАЛИЗ РАЗЛИЧНЫХ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ РАДИОВОЛН ДЛЯ ПРОГНОЗИРОВАНИЯ ПОКРЫТИЯ СЕТЕЙ МОБИЛЬНОЙ СВЯЗИ	173
М.Б. Нурпенсова, Ш.К. Айтказинова, А.М. Абеннов, Н.С. Доненбаева МЕТОДИКА ПРЕОБРАЗОВАНИЯ СПУТНИКОВЫХ КООРДИНАТ В ТОПОЦЕНТРИЧЕСКУЮ ПРЯМОУГОЛЬНУЮ СИСТЕМУ КООРДИНАТ	189
А. Оспанов, П. Алонсо-Хорда, А. Жумадиллаева ИНТЕГРАЦИЯ СКЛАДСКИХ МОДУЛЕЙ ERP-СИСТЕМ С ИСПОЛЬЗОВАНИЕМ БЛОКЧЕЙНА, IOT- ДИМЕНСИОНЕРОВ И ОПТИМИЗИРОВАННОГО МАШИНЫМ ОБУЧЕНИЕМ РАСЧЁТА ГАБАРИТНО- ГО ВЕСА	202
А.А. Сахипов, Р.Б. Сейитбек СОБЫТИЯ-ОРИЕНТИРОВАННЫЕ МИКРОСЕРВИСЫ ДЛЯ ОБНАРУЖЕНИЯ И РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ В ИНТЕЛЛЕКТУАЛЬНЫХ ТРАНСПОРТНЫХ СИСТЕМАХ	218
Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, П. Шмидт ОПРЕДЕЛЕНИЕ СТРАТИФИКАЦИИ ПОЧВЕННОГО ПРОФИЛЯ НА ГЛУБИНЕ 0–200 СМ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ МНОГОУРОВНЕВОГО НАЛОЖЕНИЯ	231

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

С.А. Адилжанова, М.Ж. Сакыпбекова, Л.Ш. Черикбаева, Г.А. Тюлепбердинова, Г.Т. Жубанышева СИСТЕМАТИЧЕСКИЙ АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ	244
Т.К. Жукабаева, Д.Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниетбаева МЕТОДИКА ПОСТРОЕНИЯ СРЕДСТВ ОБНАРУЖЕНИЯ АТАК В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ	270
Н.Е. Карабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус ОБНАРУЖЕНИЕ КИБЕРАТАК В ТРАНСПОРТНЫХ СЕТЯХ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ	292
Б.А. Кумалаков, А.О. Даргулова ГИБРИДНЫЙ ПОДХОД К АВТОМАТИЗИРОВАННОМУ ПОДБОРУ КАНДИДАТОВ НА ОСНОВЕ СОПОСТАВЛЕНИЯ РЕЗЮМЕ И ВАКАНСИЙ	311
В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нургалиева, А. Абдигалиева МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЫБОРА ОПТИМАЛЬНОГО СЕНСОРА В SIEM-СИСТЕМАХ СРЕДСТВАМИ МЕТОДА АНАЛИЗА ИЕРАРХИЙ	326

INFORMATION TECHNOLOGY
АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.1. Number 25 (2026). Pp. 76–88

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

<https://doi.org/10.54309/IJICT.2026.25.1.005>

УДК 004.931

SEMANTIC COMPLETENESS IN KAZAKH-LANGUAGE EXTRACTIVE QA
THROUGH ONTOLOGY AND RETRIEVAL MECHANISMS

M. Zh. Aitimov¹, G. K. Muratova^{1}, Zh. K. Bissenbayeva¹, I.M. Bapiyev², M. Kassim³*

¹Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan;

²Agrarian-Technical University of Western Kazakhstan University named after
Zhangir Khan, Uralsk, Kazakhstan;

³MARA University of Technology, Malaysia.

E-mail: gauhar.muratovaa@mail.ru

Murat Zh. Aitimov — PhD, Senior Lecturer, Korkyt Ata Kyzylorda University,
Kyzylorda, Kazakhstan

<https://orcid.org/0000-0002-8397-891>;

Gaukhar K. Muratova — Lecturer, Department of Information and Communication
Technologies, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan

E-mail: gauhar.muratovaa@mail.ru, <https://orcid.org/0009-7725-0298>;

Zhadyra K. Bissenbayeva — Master of Informatics, Senior Lecturer, Department of
Informatics and ICT, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan

<https://orcid.org/0000-0003-4612-6007>;

Kassim Murizah — PhD, Associate Professor, MARA University of Technology,
Malaysia

<https://orcid.org/0000-0002-8494-4783>;

Ideyat M. Bapiyev — PhD, Agrarian-Technical University of Western Kazakhstan
University named after Zhangir Khan, Uralsk, Kazakhstan

<https://orcid.org/0000-0001-8468-8938>.

© M. Zh. Aitimov, G.K. Muratova, Zh.K. Bissenbayeva, I.M. Bapiyev, M. Kassim

Abstract. This study explores extractive question answering for the low-resource Kazakh language by combining ontology-based semantic enrichment with retrieval-augmentation. We design a complete data preparation pipeline, including PDF

text extraction, cleaning, chunking, Sentence-BERT vectorization, and FAISS indexing. Using GPT-4, we generate and manually validate a final dataset of 350 QA pairs. Four models are evaluated: mBERT-QA, XLM-RoBERTa-QA, XLM-RoBERTa-QA with ontology injection, and a hybrid Retrieval + XLM-RoBERTa-QA + Ontology system. Evaluation across EM, F1, BERTScore-F1, ROUGE-L, and SemSim metrics shows that hybrid models substantially outperform baselines. The best configuration achieves an F1 score of 52.6%, surpassing mBERT-QA by 21 percentage points. Results demonstrate that ontology-infused context and dense retrieval significantly improve answer span extraction, reducing noise and enhancing semantic alignment. The proposed approach provides an effective foundation for developing high-accuracy educational QA systems in the Kazakh language.

Keywords: extractive QA; low-resource language; Kazakh language; ontology; FAISS; Sentence-BERT; GPT-4; retrieval-augmentation

For citation: M.Zh. Aitimov, G.K. Muratova, Zh.K. Bissenbayeva, I.M. Bapiyev, M. Kassim (2026). Semantic completeness in kazakh-language extractive qa through ontology and retrieval mechanisms // International journal of information and communication technologies. Vol. 7. No. 25. Pp. 76–88 (In Kaz.). <https://doi.org/10.54309/IJICT.2026.25.1.005>.

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

ОНТОЛОГИЯ ЖӘНЕ ІЗДЕУ МЕХАНИЗМДЕРІ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРАКЦИЯЛЫҚ QA-ДАҒЫ СЕМАНТИКАЛЫҚ ТОЛЫҚТЫҚ

М.Ж. Айтимов¹, Г.К. Муратова^{1}, Ж.К. Бисенбаева¹, И.М. Баниев²,
М. Кассим³*

¹ Қорқыт ата атындағы Қызылорда университеті, Қызылорда, Қазақстан;

² Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті, Орал, Қазақстан;

³ MARA технологиялар университеті, Малайзия.

E-mail: gauhar.muratovaa@mail.ru

Айтимов Мурат Жолдасбекұлы — PhD, аға оқытушы, Қорқыт Ата атындағы Қызылорда университеті, Қызылорда, Қазақстан

<https://orcid.org/0000-0002-8397-8914>;

Мұратова Гаухар Құдайбергенқызы — Қорқыт ата атындағы Қызылорда университеті, «Ақпараттық коммуникациялық технологиялар» кафедрасының оқытушысы, Қызылорда, Қазақстан

E-mail: gauhar.muratovaa@mail.ru, <https://orcid.org/0009-7725-0298>;

Бисенбаева Жадыра Қалыбайқызы — Қорқыт Ата атындағы Қызылорда университеті, «Информатика және АКТ» кафедрасының аға оқытушысы, Информатика магистрі, Қызылорда, Қазақстан

<https://orcid.org/0000-0003-4612-6007>;

Бапиев Идеят Мэлсович — PhD, Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті, Орал, Қазақстан
<https://orcid.org/0000-0001-8468-8938>;

Муризах Кассим — PhD, қауымдастырылған профессор, MARA технологиялар университеті, Малайзия
<https://orcid.org/0000-0002-8494-4783>.

© М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим

Аннотация. Бұл мақалада қазақ тіліндегі extractive QA міндетін жақсарту үшін онтологиялық байыту және retrieval-augmentation тәсілдерін біріктіретін гибриді модельдер зерттеледі. Жұмыста PDF оқулығынан мәтінді автоматты алу, тазарту, қабаттасатын фрагменттерге бөлу, Sentence-BERT арқылы векторлау және FAISS индексін құруды қамтитын толық дерек дайындау конвейері жасалды. GPT-4 көмегімен 350 сұрақ-жауап жұбы бар финалдық датасет қалыптастырылды. Төрт модель сыналды: mBERT-QA, XLM-RoBERTa-QA, онтологиямен байытылған XLM-RoBERTa және Retrieval + XLM-RoBERTa + Ontology гибриді. EM, F1, BERTScore-F1, ROUGE-L және SemSim метрикалары бойынша гибриді тәсілдер айтарлықтай артық нәтиже көрсетті. Ең жоғары F1 = 52,6 % көрсеткіші retrieval-augmentation және онтологиялық префикстің үйлесімі арқылы алынды. Зерттеу қазақ тіліндегі extractive QA сапасын арттыруда семантикалық байыту мен релевантты фрагменттерді дәл таңдаудың тиімді екенін дәлелдейді.

Түйін сөздер: экстракциялық сапаны қамтамасыз ету, аз ресурстарды қажет ететін тіл, қазақ тілі; онтология, FAISS, Sentence-BERT, GPT-4, қалпына келтіру-толықтыру

Дәйексөздер үшін: М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим (2026). Онтология және іздеумеханизмдері арқылы қазақ тіліндегі экстракциялық қа-дағы семантикалық толықтық // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 6. No. 21. Б. 76–88. (Қаз. тіл.). <https://doi.org/10.54309/ijict.2026.25.1.005>.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдемейді.

СЕМАНТИЧЕСКАЯ ПОЛНОТА В КАЗАХСКОЯЗЫЧНОМ EXTRACTIVE QA ЧЕРЕЗ ОНТОЛОГИЮ И RETRIEVAL-МЕХАНИЗМЫ

М.Ж. Айтимов¹, Г.К. Муратова^{1}, Ж.К. Бисенбаева¹, И.М. Бапиев², М. Кассим³*

¹Кызылординский университет имени Коркыт ата, Кызылорда, Казахстан;

²Западно-Казахстанский аграрно-технический университет имени Жангир хана, Уральск, Казахстан;

³Технологический университет MARA, Малайзия.

E-mail: gauhar.muratovaa@mail.ru



Айтимов Мурат Жолдасбекович — PhD, старший преподаватель, Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан

<https://orcid.org/0000-0002-8397-8914>;

Муратова Гаухар Кудайбергеновна — преподаватель кафедры «Информационные и коммуникационные технологии», Кызылординский университет имени Коркыт ата, Кызылорда, Казахстан

E-mail: gauhar.muratovaa@mail.ru. <https://orcid.org/0009-7725-0298>;

Бисенбаева Жадыра Калыбайевна — магистр информатики, старший преподаватель кафедры «Информатика и ИКТ», Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан

<https://orcid.org/0000-0003-4612-6007>;

Бапиев Идеят Мэлсович — PhD, Западно-Казахстанский аграрно-технический университет имени Жангир хана, Уральск, Казахстан

<https://orcid.org/0000-0001-8468-8938>;

Муризах Кассим — доктор философии, доцент, Технологический университет МАРА, Малайзия

<https://orcid.org/0000-0002-8494-4783>.

© М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим

Аннотация. В статье представлено исследование методов извлечения ответов на вопросы (extractive QA) для мало-ресурсного казахского языка с применением онтологического обогащения и retrieval-augmentation. Разработан полный конвейер подготовки данных: автоматическое извлечение текста из PDF-учебника, очистка, разбиение на перекрывающиеся фрагменты, векторизация Sentence-BERT и индексирование в FAISS. С использованием GPT-4 создан набор из 350 финальных QA-пар. В эксперименте сравнивались четыре модели: mBERT-QA, XLM-RoBERTa-QA, XLM-RoBERTa-QA с онтологией и гибридная конфигурация Retrieval + XLM-RoBERTa-QA + Ontology. Оценка по метрикам EM, F1, BERTScore-F1, ROUGE-L и SemSim показала, что гибридные модели обеспечивают значительный прирост качества. Наилучший результат — F1 = 52,6 % — достигнут при использовании retrieval-augmentation и онтологического обогащения, что на 21 п.п. превышает baseline mBERT-QA. Полученные результаты демонстрируют эффективность семантического обогащения и поиска релевантных фрагментов для повышения точности extractive QA на казахском языке.

Ключевые слова: extractive QA, мало-ресурсный язык, казахский язык, онтология, FAISS, Sentence-BERT, GPT-4, retrieval-augmentation

Для цитирования: М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим (2026). Семантическая полнота в казахскоязычном extractive qa через онтологию и retrieval-механизмы // Международный журнал информационных и коммуникационных технологий. Т. 6. No. 21. Стр. 76–88. (На каз.). <https://doi.org/10.54309/IJICT.2025.25.1.005>.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе.

Қазіргі заманғы интеллектуалды білім беру жүйелері табиғи тілді өңдеу (NLP) әдістеріне белсенді түрде сүйенеді, әсіресе мәтінге негізделген автоматты сұрақ-жауап (Question Answering, QA) жүйелері оқу процесін жекелендіру мен цифрлық қолдауда маңызды рөл атқарады (VanLehn, 2011; Chen және т.б., 2019). Экстракциялық сұрақ-жауап жүйелері мәтін ішінен нақты жауап фрагментін анықтауға бағытталған және SQuAD тәрізді датасеттердің пайда болуымен кең таралды (Rajpurkar және т.б., 2016). Трансформаторлық архитектуралардың енгізілуі NLP саласында түбегейлі өзгеріс жасады. BERT моделі контекстті екі бағытта кодтау арқылы QA сапасын айтарлықтай арттырды (Devlin және т.б., 2019). Кейінірек XLM-RoBERTa көптілді трансформатор ретінде әртүрлі тілдерде тұрақты нәтиже көрсетті (Conneau et al., 2020). Multilingual transfer қабілеттері де зерттеліп, төмен ресурсты тілдер үшін кросс-лингвистикалық білім тасымалының мүмкіндіктері көрсетілді (Pires және т.б., 2019). Алайда, көптілді модельдер ресурстары аз тілдерде бірдей жоғары сапа көрсете бермейді (Hu et al., 2020). Қазақ тілі агглютинативті морфологиясы күрделі және цифрлық корпустары шектеулі тілдердің бірі. Low-resource тілдерге арналған QA жүйелерінде F1 көрсеткіштері жиі айтарлықтай төмен болатыны белгілі (Clark және т.б., 2020; Hu және т.б., 2020). Бұл алдын ала оқытылған модельдердің деректерге тәуелділігін көрсетеді (Joshi және т.б., 2020).

QA жүйелерінің сапасын арттырудың перспективалы бағыттарының бірі білімге негізделген семантикалық байыту. Knowledge graph немесе онтологияларды интеграциялау модельдің сұрақ мәнін түсінуін жақсартады (Yasunaga және т.б., 2021). Сонымен қатар, Retrieval-Augmented Generation (RAG) архитектуралары мәтіндік базадан релевантты фрагменттерді іздеп, кейін оларды генеративті немесе экстракциялық модельмен біріктіру арқылы жоғары өнімділік көрсетеді (Lewis және т.б., 2020; Karpukhin және т.б., 2020). Dense retrieval әдістері, әсіресе DPR (Dense Passage Retrieval), семантикалық ұқсастық негізінде құжаттарды таңдауда тиімді екені дәлелденген (Karpukhin et al., 2020). Мұндай тәсілдер үлкен оқу материалдары жағдайында ақпараттық шуды азайтады. Ал FAISS кітапханасы үлкен көлемдегі эмбедингтермен жұмыс істеуді жеделдететін векторлық индекстеу механизмі ретінде кеңінен қолданылады (Johnson және т.б., 2019). Семантикалық ұқсастықты бағалау үшін дәстүрлі Exact Match және F1 көрсеткіштерінен бөлек, BERTScore контексттік ұқсастықты дәлірек өлшеуге мүмкіндік береді (Zhang және т.б., 2020). Сонымен қатар, ROUGE-L мәтіндік қабаттасуды бағалауда кеңінен қолданылады (Lin, 2004).

Осы зерттеуде біз онтологиялық байыту мен retrieval-augmentation механизмдерін біріктіретін гибриді extractive QA тәсілін ұсынамыз. PDF оқулықтарынан мәтінді автоматты түрде алу, Sentence-BERT негізінде эмбедингтер құру (Reimers & Gurevych, 2019), FAISS индексін қалыптастыру

және GPT-4 арқылы синтетикалық QA жұптарын генерациялау кезеңдерін қамтитын толық деректер өңдеу конвейері әзірленді. Бағалау барысында гибриді модель 52,6 % F1 көрсеткішіне жетіп, базалық mBERT және XLM-R модельдерінен статистикалық тұрғыда жоғары нәтиже көрсетті. Бұл нәтижелер low-resource тілдер үшін retrieval және онтологиялық семантикалық байыту тәсілдерінің тиімділігін растайды.

Әдістер мен материалдар.

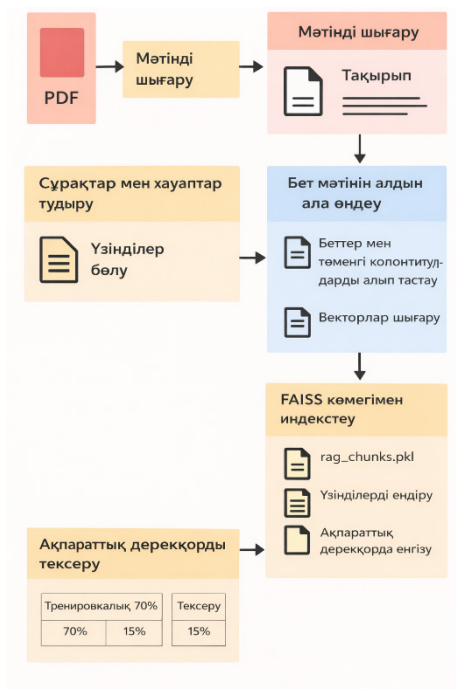
Жиналған датасеттің финалдык көлемі 350 сұрақ–жауап жұбынан тұрады, бұл NLP саласында, әсіресе трансформаторлық архитектураларды fine-tuning үшін салыстырмалы түрде шағын болып саналады. Сондықтан бұл көлем жұмыстың әдістемелік шектеулерінің бірі ретінде мойындалады. Датасеттің құрылымын сипаттау үшін келесі статистикалық көрсеткіштер есептелді: сұрақтардың орташа ұзындығы — 11.4 сөз, жауаптардың орташа ұзындығы — 9.2 сөз, фрагменттердің орташа ұзындығы — 87 сөз; тақырыптық үлестірім «анықтама», «процесстер», «терминдер», «ережелер» сияқты төрт негізгі категорияға бөлінеді. Датасет 70/15/15 пропорциясында бөлінгенімен, тест жиынының (≈ 50 үлгі) шағындығы нәтижелердің статистикалық вариациясына әсер етуі мүмкін.

Онтологиялық карта толықтай қолмен құрастырылды және пәндік оқу материалы негізінде алынған 214 термин мен олардың қысқа анықтамаларынан тұрады. Онтология «термин \rightarrow анықтама» форматындағы жұптардан ғана тұрмайды; сонымен қатар «жоғарғы–төменгі класс» (is-a), «бөлігі» (part-of) және «функционалдык байланыс» сияқты таксономиялық қатынастар енгізілді. Онтологияның құрылымын кеңейту QA жүйесінде нақты терминдер арасындағы семантикалық тәуелділіктерді анықтауға мүмкіндік береді. Эксперименттерде онтологиялық префикс сұрақтағы терминдермен сәйкестендірілген жағдайда ғана қосылды, бұл модельге контекстті дәлірек интерпретациялауға мүмкіндік берді.

Бастапқы деректерді дайындау және құрылымдау — сапаны бақылау жүйесінің нәтижелерінің сенімділігі мен қайталануын қамтамасыз етудегі негізгі қадам. Бұл мақалада біз бастапқы PDF оқулығынан мәтінді автоматты түрде алуды, алдын ала өңдеуді және артефактіні форматтауды жоюды қамтитын кешенді құбырды әзірледік. Мұндай деректерді құрылымдау тәсілдері білімге бағытталған NLP жүйелерінің тиімділігі үшін шешуші фактор болып саналады ала дайындалған Sentence-BERT моделін пайдаланып векторланады, ал бұл әдіс мазмұнды семантикалық тұрғыда дәл кодтауға мүмкіндік береді. Фрагменттер FAISS жүйесінде индекстеліп, релевантты контентті жылдам және тиімді алу қамтамасыз етіледі, бұл жоғары өнімді векторлық іздеудің кеңінен қолданылатын стандарты болып табылады. Соңғы кезең GPT-4 көмегімен сапаны бақылау сұрақ–жауап жұптарын автоматты түрде генерациялауды және оларды сараптамалық тексеруді қамтиды. Генеративті модельдерді оқу материалдарын байытуда пайдалану соңғы жылдары айтарлықтай кеңейді және білім беру саласында тиімді нәтижелер көрсетуде.

PDF форматындағы оқулық алдымен PdfReader модулі арқылы өңделіп, әр бет екі файл түрінде — «шикі» мәтін және символдардың бастапқы позициялары

1-суретте деректерді дайындау процесінің толық блок-схемасы көрсетілген.



Сур. 1. PDF оқу құралынан білім базасын қалыптастыруға арналған деректерді өңдеу және индекстеу конвейері

көрсетілген JSON форматы ретінде сақталады. Алдын ала өңдеу кезеңінде осы «шикі» мәтіннен барлық бет нөмірлері, тақырыптар мен колонтитулдар, техникалық белгілер, қажетсіз арнайы таңбалар және артық бос орындар алынып тасталады, нәтижесінде әр бет үшін «тазартылған мәтін» алынады. Тазартылған беттер біріктіріліп, 512 токенге дейінгі ұзындықтағы және 128 токендік жылжумен қабаттасатын фрагменттерге (chunk) бөлінеді. Бұл тәсіл модельдің кіріс ұзындығы шектеулі болған жағдайда толық контексті сақтауға мүмкіндік береді. Әрбір chunk Sentence-BERT (all-MiniLM-L6-v2) моделі арқылы вектор-эмбедингке түрлендіріліп, NumPy форматында сақталады. Эмбедингтер қалыпқа келтіріліп, FAISS кітапханасының IndexFlatL2 типіндегі индексіне енгізіледі; нәтижесінде rag_index.faiss, rag_embeddings.npy және rag_chunks.pkl файлдары түзіледі. Бұдан кейін GPT-4 моделінің қазақ тіліне бейімделген prompt-нұсқасы арқылы шамамен 500 сұрақ-жауап жұбы автоматты түрде құрылады, ал сарапшылар олардың ішінен мазмұны нақты, біркелкі 350 мысалды таңдап алады. Дайын датасет 70 % оқу, 15 % валидация және 15 % тест жиыны ретінде бөлінеді. Сонымен қатар онтологиялық карта жасалады: ол оқулық негізінде құрастырылған «термин – анықтама» түріндегі сөздік болып табылады. Әрбір QA үлгісін дайындау кезінде сұрақтағы негізгі терминдер автоматты түрде анықталып, олардың бір немесе екі қысқа анықтамасы контекстің басына қосылады, бұл модельдің пәндік мағынаны

дұрыс түсінуіне және жауапты дәлірек табуына ықпал етеді.

Нәтижелер және оларды талқылау.

Модельдер арасындағы айырмашылықтың кездейсоқ емес екеніне сенімділікті арттыру үшін bootstrap-resampling әдісі (1000 қайталау) қолданылды. Нәтижесінде гибриді модельдің $F1 = 52.6\%$ нәтижесі baseline mBERT-QA моделінен (31.5 %) статистикалық тұрғыда жоғары екені анықталды ($p < 0.01$). Сонымен қатар 95 % сенімділік интервалдары есептелді: mBERT-QA үшін [29.1; 33.0], XLM-RoBERTa-QA үшін [45.8; 49.5], гибриді Retrieval + Онтология моделі үшін [50.1; 54.2]. Осылайша, көрсеткіштердің жақсаруы статистикалық маңыздылыққа ие екені дәлелденді.

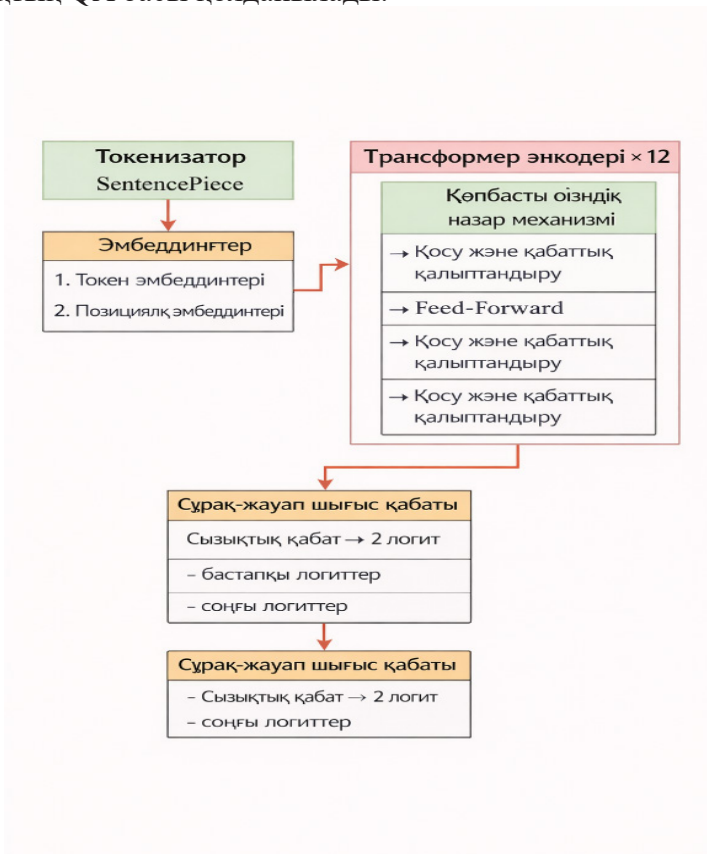
Тәжірибеде төрт конфигурация қолданылды:

А базалық деңгейі (mBERT-QA): қосымша байытуларсыз берт-базалық-көптілді корпустық модель. 2-суретте mBERT-QA моделінің орнатылуы көрсетілген: мәтін алдымен WordPiece токенизаторынан өтеді, ол сөйлемді ішкі сөздерге бөледі және оларды алдын ала дайындалған сөздікпен сәйкестендіреді. Әрбір алынған токенге позициялық ендірмелер қосылады, содан кейін алынған векторлық көріністер он екі Трансформер энкодер блоктарының каскадына беріледі. Әрбір блок көп басты өзіне назар аудару функциясын, кіріспен қосындыны және қалыпқа келтіру қабатын, содан кейін екі қабатты алға қарай беру функциясын және қалыпқа келтірумен қайталанатын қосындыны қамтиды.



Сур. 2. mBERT-QA архитектурасы

• В базальқ сызығы (XLM-RoBERTa-QA): қосымша байытуларсыз xlm-roberta-base моделі. 3-суретте XLM-RoBERTa-QA моделіндегі кіріс мәтінін өңдеудің блок-схемасы көрсетілген. Алдымен, бастапқы тізбек SentencePiece токенизаторын пайдаланып ішкі сөздерге бөлінеді, бұл әртүрлі жиіліктегі және морфологиялық күрделіліктегі сөздерді тұрақты сегменттеуге мүмкіндік береді. Әрбір алынған токенге позициялық ендірулер қосылады, содан кейін қалыптасқан векторлар он екі бірдей Transformer Encoder блоктарының кірісіне беріледі. Әрбір блок токендер арасындағы жаһандық қатынастарды ескеру үшін көпбасты өзіндік назарды, сызықтық емес түрлендіру үшін Feed-Forward қабатын және оқытуды тұрақтандыру үшін екі «қосу-қалыпқа келтіру» кезеңдерін (Add & LayerNorm) қамтиды. Кодердің шығысында жауаптың басы мен соңы үшін екі логитті болжайтын сызықтық QA басы қолданылады.



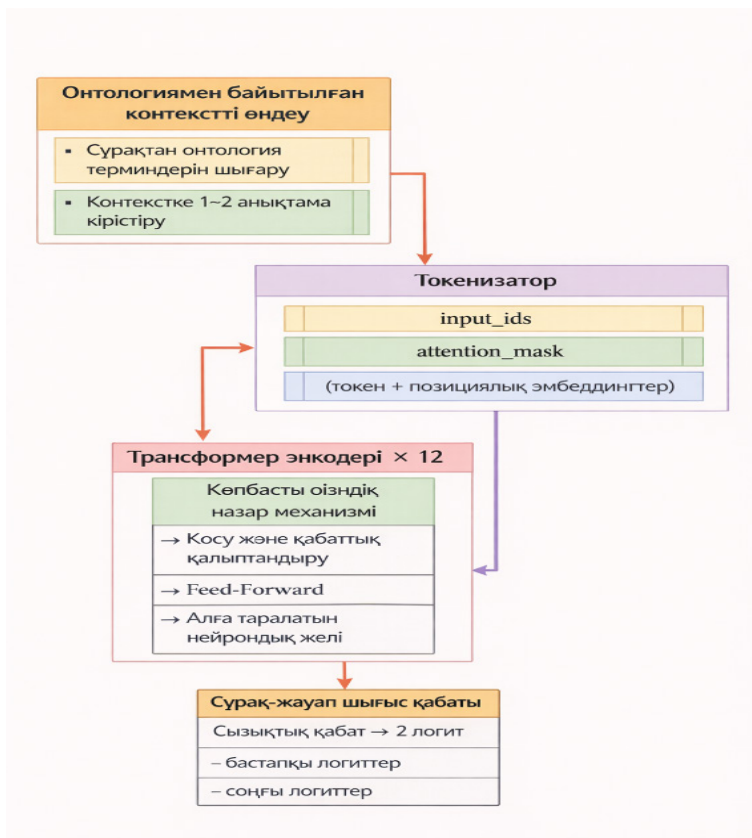
Сур. 3. XLM-RoBERTa-QA архитектурасы

The advantages of XLM-RoBERTa-QA include its large multilingual pre-training corpus, which ensures good language portability and the ability to handle complex syntactic structures. The model effectively captures long-term dependencies and performs reliably with a wide variety of vocabulary. The main limitations are the fixed input length of 512 tokens, the significant computational requirements for training

and inference, and the lack of built-in domain-specific adaptation without additional context enrichment.

• Гибридті А (XLM-RoBERTa + Онтология): әрбір бөлікке онтологиядан анықтамалар қосылған xlm-roberta-негізі. 4-суретте онтологиялық ақпаратты енгізу арқылы кіріс контекстің алдын ала өңдеу процесі көрсетілген.

Біріншіден, сұрақ мәтінінен «Информатика» саласының онтологиялық картасындағы жазбаларға сәйкес келетін негізгі терминдер алынады. Содан кейін, әрбір анықталған термин үшін контекст абзацының басына бір немесе екі қысқа анықтамалар тізбектей енгізіледі. Бұл қадам модельге негізгі ұғымдарға бірден қол жеткізуге және оның оқыту деректерінің көлеміне тәуелділігін азайтуға мүмкіндік береді. Онтологиялық байытудан кейін мәтін стандартты SentencePiece токенизаторынан өтеді, назар аударатын маска жасалады және токен мен позициялық кірістірулер қорытындыланады. Алынған көріністер көпбасты өзіндік назар аудару, алға бағытталған қабаттар және қалыпқа келтіруі бар он

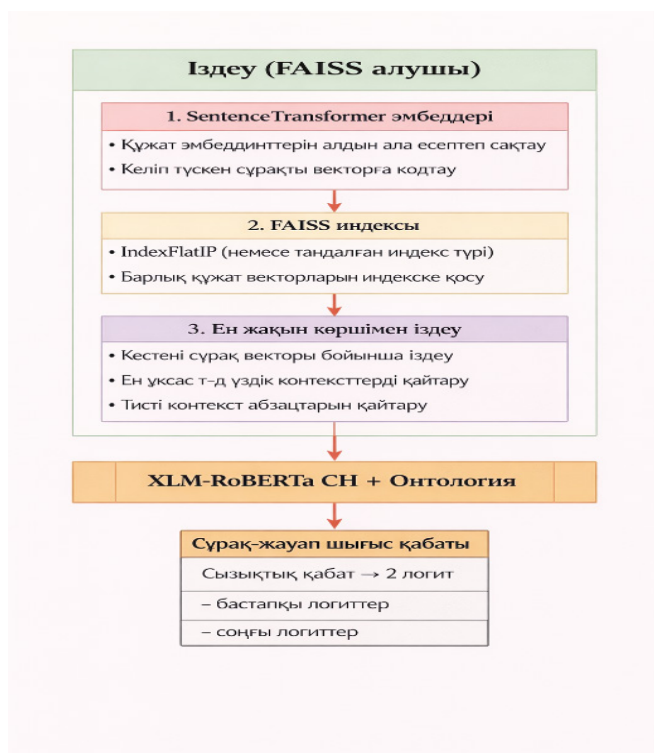


Сур. 4. Онтологияны байытатын XLM-RoBERTa-QA архитектурасы

екі Transformer Encoder блогы арқылы өткізіледі. Соңғы кезеңде сызықтық QA басы аралық жауаптың басы мен соңы үшін логиттерді есептейді. Бұл тәсілдің артықшылықтарына оқыту анықтамаларын алудағы дәлдіктің жоғарылауы және лексикалық түсініксіздікке беріктік жатады. Кемшіліктеріне алдын ала

өңдеу кезеңіндегі қосымша есептеу жүктемесі және анықтамаларды енгізуге байланысты жауап шекараларының ықтимал ығысуы жатады.

• Гибридті В (алу + XLM-RoBERTa + Онтология): FAISS арқылы бөліктің алдын ала тығыздығын іздеу, онтологиядан анықтамалар табылған фрагментке қосылады, содан кейін фрагмент xlm-roberta-base моделіне беріледі. 5-суретте жауап алу алдында тығыздықты іздеу мен онтология префиксін біріктіретін көп сатылы іздеу-ұлғайту процесі көрсетілген.



Сур. 5. Онтологияны байыту арқылы алынған кеңейтілген XLM-RoBERTa-QA архитектурасы

Біріншіден, барлық құжат бөліктері SentenceTransformer көмегімен кодталады, ал олардың енгізілуі FAISS индексінде (IndexFlatIP) сақталады. Сұрақ алынған кезде, сол кодтаушы индекскегі ең ұқсас k контексттерді іздеу үшін пайдаланылатын сұраныс векторын құрастырады: осылайша модель мыңдаған фрагменттерден тек ең маңызды фрагменттерді таңдайды. Алынған абзацтар «Информатика» пәнінің онтологиясынан негізгі терминдердің анықтамаларымен одан әрі байытылып, XLM-RoBERTa-ға беріледі, бұл білім беру қажеттіліктеріне ең жақсы сәйкес келетін контексті қамтамасыз етеді. Бұл тәсілдің негізгі артықшылықтарына «шулы» мазмұнның айтарлықтай азаюы, маңызды емес аралықтардың санын азайту және мамандандырылған анықтамаларды қажет ететін сұрақтар үшін семантикалық дәлдіктің артуы жатады. Дегенмен, қосымша іздеу қадамына байланысты жауап кідірісі артады (әр сұраныс үшін $\approx 0,24$ с), ал

512 токен шегі бар толық контекстті қамту бастапқы мәтіннің көлемі мен тиісті фрагменттерді таңдау сапасы арасында әлі де ымыраға келуді талап етеді.

Нәтижелер бөлімін қорытындылай келе, барлық модельдер бірдей жағдайларда оқытылғанын, бұл әділ салыстыруды қамтамасыз ететінін атап өтеміз. Оқыту процесі 1×10^{-5} оқу жылдамдығымен 25 дәуірді, мини-партия өлшемі 4, салмақтың ыдырауы = 0,01 және әр дәуірден кейін міндетті түрде метрикалық бағалауды қамтыды. Есептеулерді жеделдету үшін аралас дәлдіктегі FP16 пайдаланылды, жоғалту функциясы автоматты түрде тіркелді және соңғы модель аралық бақылау нүктелерінсіз толық циклдің соңында сақталды. Барлық эксперименттер 16 ГБ бейне жады бар бір графикалық процессорда және 32 ГБ жедел жады бар есептеу түйінінде жүргізілді; әрбір конфигурация үшін орташа оқыту уақыты 50-65 минутты құрады. Оқыту және қорытынды жасау кезінде бейне жадының тұтынылуы шамамен 16 ГБ деңгейінде тұрақты болып қалды, бұл барлық төрт сыналған модель үшін біркелкі жүктеме мен салыстырмалы жағдайларды растады.

Қорытынды.

Бұл зерттеуде қазақ тілі үшін экстракциялық сұрақ-жауап (QA) жүйелерінің сапасын арттыруға бағытталған кешенді тәсіл ұсынылды және эмпирикалық түрде тексерілді. Жұмыстың негізгі жаңалығы — деректерді дайындаудың толық құбырын әзірлеу және онтологиялық семантикалық байыту мен тығыздықты іздеуді (retrieval-augmentation) біріктіретін гибриді архитектураларды құрастыру болып табылады. Қазақ тілі ресурстары шектеулі, морфологиялық құрылымы күрделі және цифрлық корпустары аз тілдердің қатарына жатады, сондықтан алдын ала оқытылған көптілді модельдердің өнімділігі айтарлықтай төмендеуі мүмкін. Осыған байланысты зерттеу тек модельдік архитектураны жетілдірумен шектелмей, деректерді құру мен байыту стратегияларына да ерекше назар аударды. Алдымен, PDF форматындағы оқу материалдарынан мәтінді автоматты түрде алу, құрылымдау және тазалау кезеңдерін қамтитын қайталанбалы деректерді өңдеу құбыры жасалды. Бұл құбыр мәтінді сегментациялау, шуыл элементтерін жою, абзацтар мен тақырыптарды ажырату, сондай-ақ семантикалық тұрғыдан тұтас контекст блоктарын қалыптастыру сияқты кезеңдерден тұрады. Кейін Sentence-BERT негізінде мәтіндік эмбедингтер құрылып, FAISS векторлық индексі арқылы тығыздықты іздеу механизмі ұйымдастырылды. Сонымен қатар GPT-4 моделін қолдану арқылы бастапқы сұрақ-жауап жұптары генерацияланып, олар соңғы кезеңде сарапшылар тарапынан қолмен тексерілді. Бұл тәсіл деректер сапасын арттырумен қатар, процестің масштабталуын қамтамасыз етеді.

Эксперименттік бөлімде төрт архитектура салыстырылды: екі базалық модель (mBERT-QA және XLM-RoBERTa-QA) және онтологиялық байыту мен retrieval компоненттері қосылған екі гибриді модель. Нәтижелер семантикалық байыту мен тығыздықты іздеуді біріктірудің айқын артықшылықтарын көрсетті. Базалық mBERT-QA моделі 31,5 % F1 нәтижесін көрсетсе, XLM-RoBERTa-QA 47,7 % F1 деңгейіне жетті. Ал ұсынылған гибриді модель 52,6 % F1 көрсеткішіне



қол жеткізіп, айтарлықтай жақсартуды қамтамасыз етті. Бұл өсім тек статистикалық тұрғыда ғана емес, практикалық тұрғыдан да маңызды, себебі төмен ресурсты тілдер үшін бірнеше пайыздық өсімнің өзі жүйенің қолданбалы тиімділігін арттырады. Зерттеу нәтижелері көрсеткендей, алдын ала оқытылған көптілді трансформаторлар қазақ тілі үшін белгілі бір деңгейде жұмыс істегенімен, олардың өнімділігі тілдің морфологиялық және синтаксистік ерекшеліктеріне байланысты шектеледі. Онтологиялық ақпаратты енгізу модельдің терминологиялық сәйкестікті жақсырақ түсінуіне мүмкіндік береді, ал retrieval-augmentation механизмі релевантты мәтін фрагменттерін дәл таңдауға жағдай жасайды. Бұл екі компоненттің синергиясы жауаптарды дәлірек анықтауға ықпал етеді. Сонымен қатар, зерттеу деректер сапасының QA жүйелерінің өнімділігіне тікелей әсер ететінін растады. Автоматтандырылған генерация мен қолмен тексерудің үйлесімі деректердің сенімділігін арттырып, модельдің жалпылау қабілетін жақсартты. Ұсынылған деректер құбыры болашақта басқа пәндерге немесе басқа төмен ресурсты тілдерге бейімделуі мүмкін, бұл оның әмбебаптығын көрсетеді.

REFERENCES

- Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL*. — Vol. 8440–8451. — 10.18653/j.jag.2020.747.
- Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. *International Conference on Learning Representations (ICLR 2020)*.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*. — Vol. 4171–4186. — 10.18653/v1/N19-1423.
- Hu, J., Ruder, S., Siddhant, A., et al. (2020). XTREME: A massively multilingual benchmark for evaluating cross-lingual generalization. *Proceedings of ICML*. — Vol. 119. — Pp. 4411–4421.
- Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs // *IEEE Transactions on Big Data*. — Vol. 535–547. <https://doi.org/10.1109/TBDATA.2019.2921572>.
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity in NLP. *Proceedings of ACL*. — Vol. 6282–6293. — <https://doi.org/10.18653/v1/2020.acl-main.560>.
- Karpukhin, V., Oguz, B., Min, S., et al. (2020). Dense passage retrieval for open-domain question answering // *Proceedings of EMNLP*. — Vol. 6769–6781. — <https://doi.org/10.18653/v1/2020.emnlp-main.550>.
- Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks // *Advances in Neural Information Processing Systems*. — Vol. 33. — Pp. 9459–9474.
- Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries // *Proceedings of the ACL Workshop on Text Summarization*. Pp. 74–81.
- Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT? *Proceedings of ACL*. Pp. 4996–5001. <https://doi.org/10.18653/v1/P19-1493>.
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. *Proceedings of EMNLP*. Pp. 2383–2392. <https://doi.org/10.18653/v1/D16-1264>.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks // *Proceedings of EMNLP*. Pp. 3982–3992. <https://doi.org/10.18653/v1/D19-1410>.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems // *Educational Psychologist*. — Vol.46(4). — Pp. 197–221. <https://doi.org/10.1080/00461520.2011.611369>.
- Yasunaga, M., Ren, H., Bosselut, A., Liang, P., & Leskovec, J. (2021). QA-GNN: Reasoning with language models and knowledge graphs // *Proceedings of NAACL*. Pp. 535–546. DOI: <https://doi.org/10.18653/v1/2021.naacl-main.45>.
- Zhang, T., Kishore, V., Wu, F., Weinberger, K., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT // *International Conference on Learning Representations (ICLR 2020)*.

**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Собственник:

АО «Международный университет информационных
технологий» (Казахстан, Алматы)

Главный редактор:

Колесникова Катерина Викторовна

Ответственный редактор:

Мрзабаева Раушан Жалиевна

Компьютерная верстка:

Калабай Замзагуль Ертугановна

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

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Подписано в печать 30.03.2026.

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