

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 Sansyzbay — 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 Тоқсуғаевич — 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 Максutowна 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 Токсужаевич — 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)

Дамелия Максуговна 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)

Алимурали 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. Қасымов, 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. Абеннов, 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 ДАННЫМ 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

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.1. Number 25 (2026). Pp. 141–157

Journal homepage: <https://journal.iitu.edu.kz><https://doi.org/10.54309/IJICT.2026.25.1.009>

УДК 004.8:81'322

A DOMAIN-KNOWLEDGE-BASED MODEL FOR REFERENCE RESOLUTION IN LOW-RESOURCE LANGUAGES

G. Kalman^{1*}, J. Kultan², A.N. Ismukamova¹, N.M. Ausilova³, Y.V. Makhatova⁴

¹Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan;

²University of Economics and Business, Bratislava, Slovak Republic;

³Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan;

⁴Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan.

E-mail: gulzhamal.kalman@ku.edu.kz

Gulzhamal Kalman — Associate Professor, PhD Doctor, Kokshetau University named after Sh. Ualikhanov, Kokshetau, Kazakhstan

E-mail: guljamal14@gmail.com, <https://orcid.org/0000-0001-8863-9447>;

Jaroslav Kultan — Doctor of Education, Professor (Associate), Bratislava University of Economics and Business, Bratislava, Slovak Republic

<https://orcid.org/0000-0001-6068-9784>;

Aigerim N. Ismukanova — Senior Lecturer, Kokshetau University named after Sh. Ualikhanov, Kokshetau, Kazakhstan

<https://orcid.org/0009-0001-0011-0846>;

Nazerke M. Ausilova — Senior Lecturer, Kokshetau University named after Abai Myrzakhmetov, Kokshetau, Kazakhstan

<https://orcid.org/0000-0001-7541-5970>;

Valentina Ye. Makhatova — Candidate of Technical Science, Professor, Department of Software Engineering, Faculty of Physics, Mathematics and Information Technology, Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan

<https://orcid.org/0000-0002-4082-9193>.

© G. Kalman, J. Kultan, A.N. Ismukamova, N.M. Ausilova, Y.V. Makhatova

Abstract. This paper proposes a domain-knowledge-based hybrid model for coreference resolution in low-resource and morphologically rich languages, with a specific focus on Kazakh. Although recent neural and transformer-based approaches have significantly advanced the state of the art in high-resource languages, they depend heavily on large, annotated corpora and pretrained language models. Such resources are limited or unavailable for many low-resource languages, which negatively affects



system performance and generalization. To address this challenge, the proposed Kazakh Coreference Adaptation (KCA) framework integrates statistical machine learning techniques with explicit domain knowledge derived from ontologies, morphological constraints, and semantic rules. The architecture consists of several stages, including text preprocessing (tokenization, morphological analysis, part-of-speech tagging, named entity recognition, and dependency parsing), mention detection, candidate pair generation, feature extraction, and weighted scoring. The model evaluates candidate antecedents using a combination of morphological agreement, case compatibility, syntactic roles, discourse distance, and semantic similarity measures. Experimental evaluation conducted on an annotated subset of the Kazakh National Corpus demonstrates that incorporating structured linguistic and domain knowledge significantly improves resolution accuracy and F1-score compared to baseline statistical models. The findings confirm that knowledge-guided hybrid strategies effectively compensate for data scarcity. The proposed approach contributes an interpretable, adaptable, and resource-efficient framework for building robust NLP systems in low-resource language environments.

Keywords: reference resolution, coreference resolution, low-resource languages, domain knowledge, NLP, hybrid model

For citation: G. Kalman, J. Kultan, A.N. Ismukamova, N.M. Ausilova, Y.V. Makhatova. A domainknowledge-based model for reference resolution in low-resource languages // International journal of information and communication technologies. 2026. Vol. 7. No. 25. Pp. 141–157. <https://doi.org/10.54309/IJICT.2026.25.1.009>. (In Eng.).

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

Acknowledgments. This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Project No. AP22686434).

ПӘНДІК САЛА БІЛІМ НЕГІЗІНДЕ РЕУСРСТАРЫ АЗ ТІЛДЕРДЕГІ РЕФЕРЕНЦИЯНЫ ШЕШУДІҢ МОДЕЛІ

Г.Қалман^{*1}, *К.Ярослав*², *А.Н.Исмуканова*¹, *Н.М.Аусилова*³, *В.Е.Махатова*⁴

¹Шокан Уалиханов атындағы Көкшетау университеті, Көкшетау, Қазақстан;

²Братислава Экономикалық университеті, Братислава, Словак, Словакия;

³Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Қазақстан;

⁴Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан.

E-mail: gulzhamal.kalman@ku.edu.kz

Қалман Гүлжамал — қауымдастырылған профессор, PhD доктор, Ш.Уалиханов атындағы атындағы Көкшетау университеті, Көкшетау, Қазақстан

E-mail: guljamal14@gmail.com. <https://orcid.org/0000-0001-8863-9447>;

Култан Ярослав — қауымдастырылған профессор, Экономикалық университет, Братислава, Словакия

<https://orcid.org/0000-0001-6068-9784>;

Исмуканова Айгерим Наурызбаевна — сеньор-лектор, Ш.Уалиханов атындағы атындағы Көкшетау университеті, Көкшетау, Қазақстан
<https://orcid.org/0009-0001-0011-0846>;

Ауилова Назерке Мызрабековна — аға оқытушы, Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Қазақстан
<https://orcid.org/0000-0001-7541-5970>;

Махатова Валентина Еркиновна — техника ғылымдарының кандидаты, бағдарлама инженериясы кафедрасының профессоры, физика, математика және ақпараттық технологиялар факультеті, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
<https://orcid.org/0000-0002-4082-9193>.

© Г. Қалман, К. Ярослав, А.Н. Исмуканова, Н.М. Ауилова, В.Е. Махатова

Аннотация. Бұл мақалада морфологиялық тұрғыдан күрделі және цифрлық ресурстары шектеулі тілдерге арналған кореференцияны анықтаудың пәндік білімге негізделген гибриді моделі ұсынылады (қазақ тілі мысалында). Қазіргі нейрондық және трансформерлік модельдер жоғары нәтижелер көрсеткенімен, олар үлкен көлемдегі аннотталған мәтіндерге және алдын ала үйретілген тілдік модельдерге тәуелді. Аз ресурсты тілдер үшін мұндай деректер қоры жеткіліксіз болғандықтан, модельдердің тиімділігі мен тұрақтылығы төмендейді. Ұсынылған Kazakh Coreference Adaptation (КСА) жүйесі статистикалық машиналық оқыту әдістерін онтологиялардан, морфологиялық шектеулерден және семантикалық ережелерден алынған пәндік біліммен біріктіреді. Жүйе мәтінді алдын ала өңдеу кезеңдерін (токенизация, морфологиялық талдау, сөз табын анықтау, атаулы бірліктерді тану, тәуелділік талдауы), кейін кандидаттарды анықтау, белгілерді шығару және салмақталған бағалау сатыларын қамтиды. Кандидат-антецедент жұптары морфологиялық сәйкестік, септік үйлесімі, синтаксистік рөл, дискурстық қашықтық және семантикалық ұқсастық негізінде бағаланады. Қазақ ұлттық корпусы негізіндегі эксперимент нәтижелері пәндік білімді енгізу модельдің дәлдігін және F1-көрсеткішін едәуір арттыратынын көрсетті. Зерттеу аз ресурсты тілдер үшін сенімді, түсіндірілетін және бейімделгіш NLP жүйелерін әзірлеуге маңызды үлес қосады.

Түйін сөздер: референцияны анықтау, кореференция, ресурстары шектеулі тілдер, пәндік білім, табиғи тілді өңдеу (NLP), гибриді модель

Дәйексөздер үшін: Г. Қалман, К. Ярослав, А.Н. Исмуканова, Н.М. Ауилова, В.Е. Махатова (2026). Пәндік сала білім негізінде реустрстары аз тілдердегі референцияны шешудің моделі // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No. 25. 141–157 бет. <https://doi.org/10.54309/IJICT.2026.25.1.009>. (Ағыл. тіл.).

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

МОДЕЛЬ НА ОСНОВЕ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ РАЗРЕШЕНИЯ КОРЕФЕРЕНЦИИ В МАЛОРЕСУРСНЫХ ЯЗЫКАХ

Г. Калман^{1}, К. Ярослав², А.Н. Исмуканова³, Н.М. Аусилова⁴, В.Е. Махатова⁵*

^{1,3}Кокшетауский университет имени Шокана Уалиханова, Кокшетау, Казахстан;

² Экономический университет, Братислава, Словакия;

⁴Кокшетауский университет имени Абая Мырзахметова, Кокшетау, Казахстан;

⁵Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан.

E-mail: gulzhamal.kalman@ku.edu.kz

Калман Гүлжамал — PhD, ассоциированный профессор, Кокшетауский университет имени Шокана Уалиханова, Кокшетау, Казахстан

E-mail: guljamal14@gmail.com, <https://orcid.org/0000-0001-8863-9447>;

Култан Ярослав — ассоциированный профессор, Экономический университет, Братислава, Словакия

<https://orcid.org/0000-0001-6068-9784>;

Исмуканова Айгерим Наурызбаевна — сеньор-лектор, Кокшетауский университет имени Шокана Уалиханова, Кокшетау, Казахстан

<https://orcid.org/0009-0001-0011-0846>;

Аусилова Назерке Мырзабековна — старший преподаватель, Кокшетауский университет имени Абая Мырзахметова, Кокшетау, Казахстан

<https://orcid.org/0000-0001-7541-5970>;

Махатова Валентина Еркиновна — кандидат технических наук, профессор кафедры программной инженерии, факультет физики, математики и информационных технологий, Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан

<https://orcid.org/0000-0002-4082-9193>.

© Г. Калман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова

Аннотация. В статье предлагается гибридная модель разрешения кореференции для малоресурсных и морфологически сложных языков на примере казахского языка. Современные нейронные и трансформерные модели демонстрируют высокую эффективность в условиях наличия больших аннотированных корпусов и предварительно обученных языковых моделей. Однако для малоресурсных языков такие данные ограничены или отсутствуют, что существенно снижает качество автоматического анализа текста. Для решения данной проблемы разработана модель Kazakh Coreference Adaptation (КСА), сочетающая статистические методы машинного обучения с явным использованием предметных знаний. В архитектуру системы входят этапы предварительной обработки текста (токенизация, морфологический анализ, POS-теггинг, распознавание именованных сущностей и синтаксический разбор зависимостей), обнаружение упоминаний, формирование кандидатных пар и извлечение

признаков. Оценка кандидатных antecedентов осуществляется на основе взвешенной комбинации морфологического согласования, совместимости по падежу и числу, синтаксической роли, дискурсивной дистанции и семантической близости. Экспериментальная оценка, проведённая на размеченной части Казахского национального корпуса, показала, что интеграция онтологических и семантических ограничений значительно повышает точность и F1-меру по сравнению с базовыми статистическими моделями. Полученные результаты подтверждают эффективность комбинирования знаний предметной области и методов машинного обучения при разработке устойчивых и интерпретируемых систем обработки естественного языка для языков с ограниченными цифровыми ресурсами.

Ключевые слова: разрешение референций, разрешение кореференции, малоресурсные языки, знания предметной области, NLP, гибридная модель

Для цитирования: Г. Калман, К. Ярослав, А.Н. Исмуканова, Н.М. Аусилова, В.Е. Махатова (2026). Модель на основе знаний предметной области для разрешения кореференции в малоресурсных языках // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 25. Стр. 141–157. <https://doi.org/10.54309/IJICT.2026.25.1.009>. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

There has been a breakthrough in the development of Large Language Models (LLMs) in Natural Language Processing over the past few years. Transformer-based architectures (Vaswani et al., 2017) and large-scale pretrained models such as GPT and instruction-tuned systems (Brown et al., 2020; Ouyang et al., 2022) have demonstrated remarkable performance in question answering, summarization, and content generation tasks. Pretraining strategies such as BERT (Devlin et al., 2019) further advanced contextual language understanding and enabled improvements in downstream tasks, including coreference resolution (Joshi et al., 2019; Lee et al., 2017).

Despite these advances, most successful models are designed and evaluated primarily for high-resource languages. Processing morphologically rich and low-resource languages remains a challenge due to complex agreement systems, flexible word order, and agglutinative morphology. Coreference resolution, especially in long contexts, is still difficult even for modern architectures (Liu et al., 2023). Although long-context models such as Longformer (Beltagy et al., 2020) and BigBird (Zaheer et al., 2020) have been proposed, and efficiency improvements have been studied extensively (Tay et al., 2022; Huang et al., 2023), these approaches are optimized for English and other analytical languages.

This research lacuna explicitly manifests a challenge in needing approaches targeted for low-resource languages, which also involve morphologically complex structures. These types of languages are best served by a hybridized technique that involves



linguistic rules, as well as a data-driven model. One of the most promising approaches in dealing with this problem is to make use of machine learning in conjunction with domain knowledge. While traditional statistical models lack the ability to incorporate knowledge from other sources, domain-based models can make use of other rules and ontologies that are able to provide a description of a logical relationship between entities in a text. The primary aim of this research is to develop a hybrid model that integrates machine learning with domain knowledge.

Materials and Methods.

The proposed Kazakh Coreference Adaptation (KCA) method is a hybrid model that combines rule-based linguistic techniques with supervised machine learning approaches. The main objective of the method is to reduce contextual breaks and maintain referential consistency in long texts by accurately resolving pronominal anaphora in morphologically complex and low-resource languages.

Our approach integrates domain knowledge into statistical learning, inspired by knowledge-augmented neural architectures (Lee et al., 2018; Zhang et al., 2021).

The system architecture of KCA (Kazakh Coreference Adaptation), presented in Figure 1, clearly demonstrates the stages of data processing and decision-making. The architecture consists of the following key blocks:

- (a) Data acquisition and preprocessing (tokenization, POS tagging, NER, morphological annotation),
- (b) Morphosyntactic analysis (dependency parsing),
- (c) Rule-based filters (morphology/case/number filter),
- (d) Feature extraction (morphological, syntactic, discourse distance, semantic similarity),
- (e) Supervised learning layer (SVM / Decision Tree or another classifier),
- (f) Aggregation and decision-making (rule + ML weighting),
- (g)) Final output - anaphora resolution.

These blocks are interconnected through clearly defined data flow: the rule-based layer reduces the candidate set and passes only relevant ones to the ML layer; the ML layer assigns probabilities based on extracted features; finally, the rule-based and statistical scores are combined through weighted aggregation.

a. **Data and Resources:** The primary data source for this study was the Kazakh National Corpus (KNC). A manually annotated subset of the corpus was selected, focusing specifically on examples of pronominal anaphora. The data included literary, academic, and journalistic texts. Each text was segmented into sentences, tokenized, and annotated with morphological and syntactic features. The full dataset was split into training, validation, and test sets in an 80/10/10 ratio to ensure fair evaluation of the model's performance.

b. **Preprocessing and Linguistic Analysis:** The initial phase of the KCA method involves linguistic analysis and data structuring:

Morphological analysis: Each word is analyzed to identify its root form, affixes, and grammatical attributes such as case, person, number, and possession.

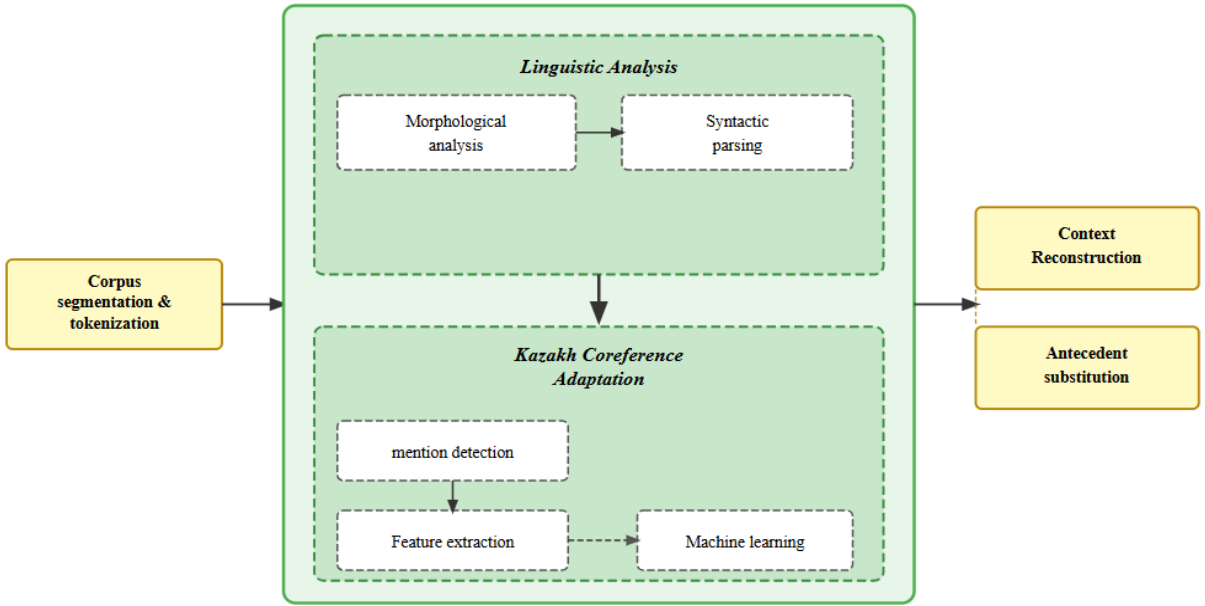


Fig. 1. System architecture for Kazakh coreference resolution

Syntactic analysis: Dependency trees are constructed to capture grammatical relations within sentences, including subject–object dependencies.

Segmentation and tagging: Words and sentences are annotated consistently with morphological and syntactic labels.

The output of this phase is a morpho-syntactically enriched dataset prepared for subsequent rule-based and machine learning components.

c. Hybrid Coreference Resolution Algorithm (KCA Algorithm)

Mention detection: In this stage, potential referential expressions are automatically extracted from the text. These include:

Pronouns (e.g., *ол* – he/she, *олар* – they, *бұл* – this, *оның* – his/her, etc.)

Proper nouns (e.g., *Aigul*, *Murat*, *Almaty*)

Noun phrases (e.g., *young teacher*, *new book*)

This stage utilizes a Part-of-Speech (POS) tagger and Named Entity Recognition (NER) models. Each mention is marked with its start and end index in the form of $\text{mention_span} = [\text{start}, \text{end}]$.

Given a text sequence $T = \{w_1, w_2, \dots, w_n\}$ consisting of n words, the goal is to extract a set of possible referential mentions:

$$M = \{m_i = (w_s, w_e, t_i) \mid 1 \leq s \leq e \leq n\} \quad (1)$$

where: w_s and w_e denote the start and end of the mention $t_i \in \{\text{"pronoun"}, \text{"proper_noun"}, \text{"noun_phrase"}\}$ is the mention type.

The mention detection function is defined as:

$$\text{MentionDetector}(T) = \{(s, e) \mid \text{POS}(w_s..w_e) \in P, \text{NER}(w_s..w_e) \in N\} \quad (2)$$

where:

P is the set of POS tags representing pronouns, nouns, and named phrases,

N is the set of NER categories corresponding to named entities and subjects.

Algorithm in Text Form

Initialize the set M as an empty set:

$M \leftarrow \emptyset$

For each token w_i in the text T , perform the following steps:

Determine the part-of-speech tag of w_i :

$\text{tag} \leftarrow \text{POS}(w_i)$

Determine the named entity label of w_i :

$\text{entity} \leftarrow \text{NER}(w_i)$

Check whether the token belongs to a relevant category:

If tag is one of {PRON, NOUN, PROPN} or entity is in {PERSON, LOCATION, ORGANIZATION}, then:

Compute the start index of the token:

$\text{start} \leftarrow \text{index}(w_i)$

Compute the end index of the token:

$\text{end} \leftarrow \text{index}(w_i) + \text{len}(w_i)$

Define the mention type based on the POS tag:

$\text{type} \leftarrow \text{define_type}(\text{tag})$

Add the mention to the set M :

$M \leftarrow M \cup \{(w_i, [\text{start}, \text{end}], \text{type})\}$

Return the final set M .

d) *Feature Extraction*: At this stage, a set of morphological, syntactic, and semantic features is constructed to characterize potential antecedent-anaphor relationships between identified mentions. These features serve as the main input for the machine learning component of the system.

For each mention pair (m_i, m_j) , where m_i is a candidate antecedent and m_j is a pronoun or subsequent mention, the feature vector is defined as:

$$F(m_i, m_j) = [f_{\text{morph}}, f_{\text{syn}}, f_{\text{dist}}, f_{\text{sem}}] \quad (3)$$

where:

- f_{morph} - morphological similarity score,
- f_{syn} - syntactic role compatibility,
- f_{dist} - discourse distance between mentions,

- f_{sem} -semantic similarity (via lexical knowledge base).

This feature vector is then passed to a classifier (e.g., SVM, Decision Tree), which estimates the coreference probability:

$$P(\text{"coref"} \mid m_i, m_j) \quad (4)$$

Morphological Similarity Feature: f_{morph} Morphological features are derived from Kazakh grammar and include agreement in person, case, number, gender, and possessiveness. The similarity function is defined as:

$$f_{\text{morph}}(m_i, m_j) = \frac{1}{K} \sum_{k=1}^K \delta(a_k(m_i), a_k(m_j)) \quad (5)$$

where:

- K is the number of morphological categories,
- $a_k(m)$ is the k^{th} morphological attribute of mention m ,
- $\delta(x, y) = 1$ if $x = y$, and 0 otherwise.

Mentions that share agreement in features like case and person (e.g., ол and мұғалім) have high similarity scores ($f_{\text{morph}} \rightarrow 1$). In contrast, mismatches (e.g., singular vs. plural) reduce the score toward zero.

Syntactic Role Feature f_{syn} This feature captures the grammatical function of mentions within the sentence - subject, object, modifier, etc. Compatibility is assessed via:

$$f_{\text{syn}}(m_i, m_j) = \begin{cases} 1 & \text{if } \text{role}(m_i) = \text{role}(m_j) \\ \lambda & \text{if partially compatible (e.g., S-O)} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $\lambda \in [0,1]$ is a partial match coefficient (e.g., subject-object pairs might have $\lambda = 0.5$).

Discourse Distance Feature: f_{dist} The farther apart the antecedent and pronoun are in discourse, the less likely they are to be coreferent. This feature is modeled as:

$$f_{\text{dist}}(m_i, m_j) = e^{-\alpha \cdot d(m_i, m_j)} \quad (7)$$

where:

- $d(m_i, m_j)$ is the number of sentences between the two mentioned,
- $\alpha \approx 0.3$ controls the decay rate.

For instance, if mentions occur in the same sentence ($d = 0$), $f_{\text{dist}} = 1$; at a distance of 3 sentences, $f_{\text{dist}} \approx e^{-0.9} \approx 0.4$.

Semantic f_{sem} Similarity Feature: This metric uses the Kazakh Lexical Knowledge Base (KLKB) or embedding models (e.g., word2vec) to measure conceptual proximity:

$$f_{\text{sem}}(m_i, m_j) = \cos(v(m_i), v(m_j)) = \frac{v(m_i) \cdot v(m_j)}{\|v(m_i)\| \|v(m_j)\|} \quad (8)$$

where $v(m)$ denotes the vector representation of mention m . If the semantic link is strong (e.g., teacher and he), then $f_{\text{sem}} \rightarrow 1$; otherwise, it tends toward 0 (e.g., teacher and book).

Combined Feature Model: All features are combined into a unified scoring function:

$$\Phi(m_i, m_j) = w_1 f_{\text{morph}} + w_2 f_{\text{syn}} + w_3 f_{\text{dist}} + w_4 f_{\text{sem}} \quad (9)$$

where w_1, w_2, w_3, w_4 are empirically tuned weight coefficients. This weighted sum is passed through a logistic function to produce the final probability:

$$P(\text{"coref"} \mid m_i, m_j) = \sigma(\Phi(m_i, m_j)), \text{ where } \sigma(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

Table 1 – Weight Configurations and Corresponding CoNLL-F1 Scores.

w_1 (Morph)	w_2 (Case/Num)	w_3 (Dist)	w_4 (SemSim)	CoNLL-F1 (%)
0.40	0.20	0.20	0.20	73.8
0.50	0.15	0.15	0.20	72.6
0.30	0.25	0.25	0.20	71.9
0.35	0.20	0.25	0.20	74.2 (<i>optimal</i>)
0.25	0.25	0.25	0.25	70.3

Table 1 illustrates the impact of various weight combinations on system performance using the CoNLL F1 metric. The experiments were conducted on 500 annotated anaphor–candidate pairs, with 5-fold cross-validation applied for each configuration. The results demonstrate that increasing the influence of morphological features ($w_1 = 0.35$ – 0.40) leads to an optimal overall performance.

Accordingly, the final model adopted the configuration of ($w_1 = 0.35, w_2 = 0.20, w_3 = 0.25, w_4 = 0.20$). The configuration ($w_1 = 0.35, w_2 = 0.20, w_3 = 0.25, w_4 = 0.20$) was selected as the final model because it achieved the highest CoNLL-F1 score (74.2%). The stability of this configuration was confirmed through 5-fold cross-validation, where the performance remained consistently higher than other tested weight combinations. These parameters ensure model adaptability and stable performance across different

textual inputs.

Scoring and Antecedent Selection (Saliency Scoring)

Rule-based Layer: This layer assesses candidates based on morphological and grammatical compatibility (person, case, number, and referential agreement). The filtering is grounded in the morphosyntactic rules of the Kazakh language. Candidate referents are scored according to the following criteria:

Person Agreement: The pronoun (I/II/III) of the pronoun must match the subject or object in previous sentences.

Case Compatibility: The Kazakh case system (nominative, accusative, dative, ablative, locative, instrumental) helps determine the syntactic role of the referent.

Number Agreement: Singular/plural consistency between the pronoun and referent is mandatory.

Referential Type Compatibility: While Kazakh lacks grammatical gender, semantic distinctions between human/non-human entities serve as soft filters.

Syntactic Distance: Antecedents closer to the pronoun are favored with a higher saliency score.

Machine Learning Layer: This component assigns a probability score to each candidate, reflecting the likelihood of being the correct antecedent. The following algorithms are used:

Support Vector Machine (SVM): Classifies candidate–anaphor pairs in a high-dimensional feature space, identifying the optimal separating boundary between coreferent and non-coreferent instances. Input features include sentence distance, syntactic dependencies, positional attributes, NER tags, and morphological agreement metrics.

Decision Tree: Models referent selection as a series of hierarchical decisions. Each node represents a morphological, semantic, or positional feature (e.g., “If candidate has PERSON NER tag → increase weight,” “If distance > 3 sentences → decrease weight”).

At the end of the classification process, each candidate receives a final confidence score. The candidate with the highest weight is selected as the antecedent.

Algorithm RESOLVE_PRONOUN(T, candidates):

1. Initialize scores $\leftarrow \emptyset$
2. For each candidate c in candidates:
 3. # 1) Rule-based Filtering (Morphological & Syntactic Agreement)
 4. $\text{morph_score} \leftarrow \text{CHECK_MORPH_AGREEMENT}(c, T.\text{pronoun})$
 5. $\text{case_score} \leftarrow \text{CHECK_CASE_COMPATIBILITY}(c, T.\text{pronoun})$
 6. $\text{num_score} \leftarrow \text{CHECK_NUMBER_AGREEMENT}(c, T.\text{pronoun})$
 7. $\text{dist_score} \leftarrow \text{COMPUTE_DISTANCE_SCORE}(c, T.\text{position})$
 8. if $\text{morph_score} == 0$ OR $\text{case_score} == 0$:
 9. continue # candidate eliminated
 10. $\text{rule_score} \leftarrow \text{WEIGHT1} * \text{morph_score} +$
 $\text{WEIGHT2} * \text{case_score} +$
 $\text{WEIGHT3} * \text{num_score} +$

WEIGHT4*dist_score

11. # 2) Machine Learning Layer (Probabilistic Scoring)
12. features \leftarrow EXTRACT_FEATURES(c)
13. svm_prob \leftarrow SVM_MODEL.predict_proba(features)
14. tree_prob \leftarrow DT_MODEL.predict_proba(features)
15. ml_score \leftarrow (svm_prob + tree_prob) / 2
16. # 3) Combined Final Score
17. final_score \leftarrow α *rule_score + β *ml_score
18. scores[c] \leftarrow final_score
19. # 4) Select candidate with highest score
20. best_candidate \leftarrow argmax(scores)
21. Return best_candidate

Results and Discussion.

This section presents the experimental results obtained using the Kazakh Coreference Adaptation (KCA) hybrid method, along with performance evaluation metrics and a comparative analysis with traditional approaches. The primary goal of this study was to improve reference resolution accuracy in resource-constrained Kazakh language texts by integrating domain-specific linguistic knowledge with statistical models.

1. Evaluation Metrics

To assess the model's effectiveness, three standard metrics commonly used in Natural Language Processing (NLP) were applied, following international benchmarking protocols:

MUC (Message Understanding Conference): Measures recall by identifying the number of missing links in the predicted coreference chains compared to the gold standard.

B³ (B-cubed): Evaluates precision by checking how accurately each individual mention has been assigned to its correct coreference cluster.

CEAF (Constrained Entity Alignment F-measure): Quantifies the alignment accuracy between predicted and actual entities, providing an overall F1 score.

CoNLL F1: The final performance score is calculated as the harmonic mean of the MUC, B³, and CEAF scores, offering a holistic view of model quality.

Experimental Results

The experiments were conducted using the annotated subset of the Kazakh National Corpus (KNC). During evaluation, the performance of the proposed KCA hybrid model was compared to baseline models that rely solely on statistical methods. The results demonstrate that incorporating linguistic rules and domain knowledge into the model significantly enhances reference resolution performance, particularly in morphologically rich and low-resource settings such as Kazakh.

Figure 2 below illustrates the performance comparison between the baseline model and the KCA hybrid model across four core metrics. The blue line (baseline model) shows scores ranging between 51 % and 58 %, while the red line (KCA model) achieves results in the 68 % to 72 % range. An average performance gap of approxi-

mately 15 % is observed between the two models.

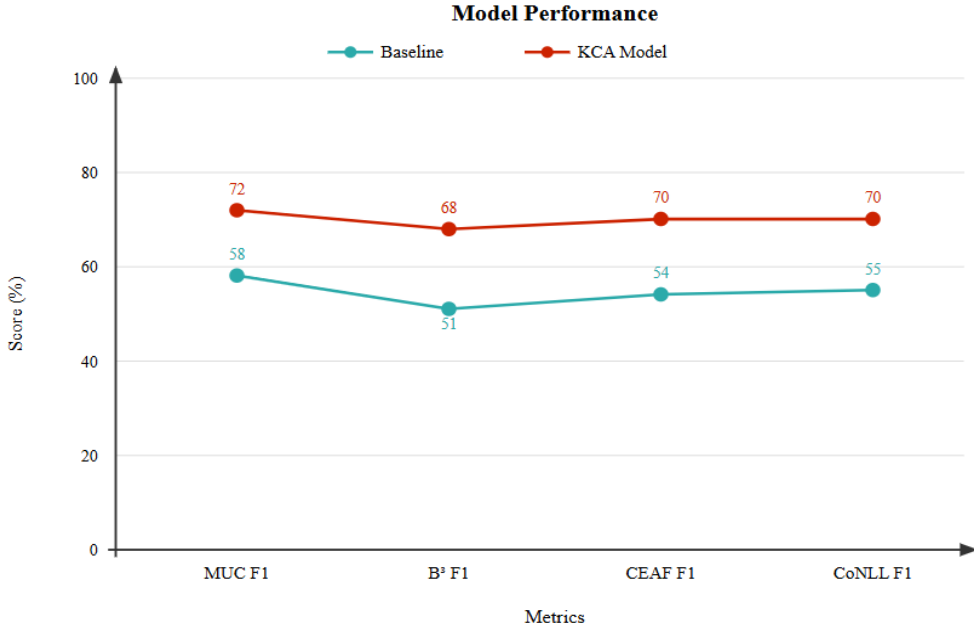


Fig. 2. Model Comparison

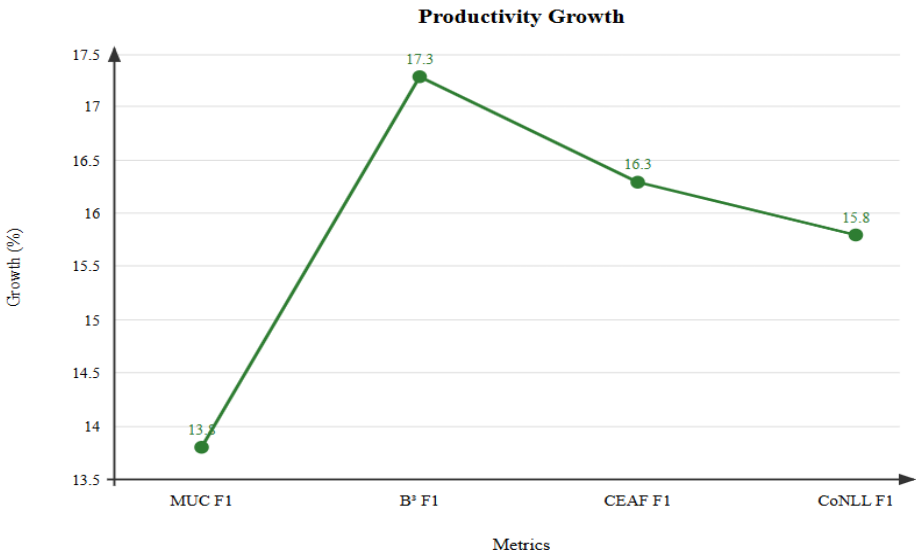


Fig. 3. Performance of the KCA Model

Figure 3 presents a line graph illustrating the performance gain of the KCA model. According to the B³ metric, the model reaches a peak improvement of 17.3 %. For the remaining evaluation measures, the performance increase remains steady within the 13.7 % to 15.8 % range, clearly demonstrating the overall effectiveness of the model.

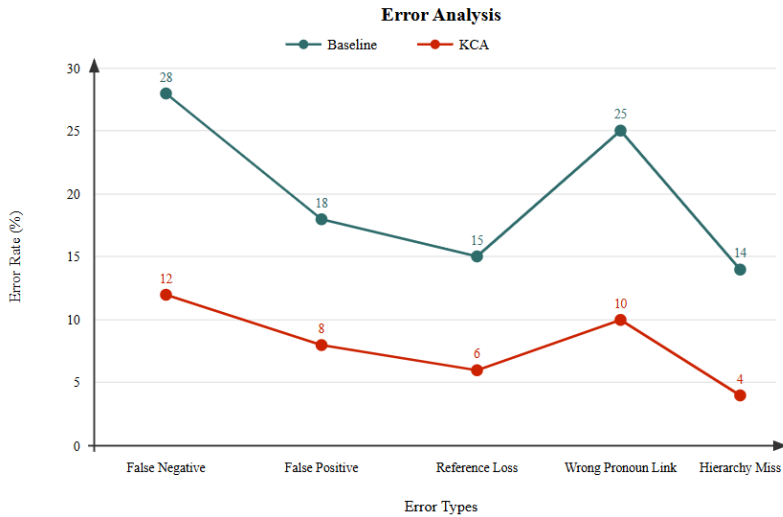


Fig. 4. Error Analysis

The line graph depicting five types of error categories enables a clear comparison between the baseline model (Figure 4) and the KCA model. The blue line represents the baseline model's error levels, ranging between 14 % and 28 %, indicating a high overall error rate. In contrast, the red line represents the KCA model's results, with significantly lower error levels between 4 % and 12 %, demonstrating the enhanced effectiveness of the new model.

A particularly noticeable improvement is seen in the False Negative category: the baseline model exhibits a 28 % error rate, whereas the KCA model reduces this to 12 %. This significant drop highlights the KCA method's improved capability in correctly identifying antecedents.

The performance of the KCA and KazBERT models was compared based on standard CoNLL metrics. In terms of the Recall metric, the KazBERT model achieved a score of 70.8 %, while the proposed KCA method reached 75.4 %. This result confirms the superior ability of KCA to comprehensively identify antecedents, i.e., it is more effective at detecting anaphoric links. This strength is particularly important in applied linguistics and knowledge engineering domains where full coverage of information is critical. The comparison results are shown in Table 2 below.

Table 2 – Comparative Results of KCA and KazBERT Models.

Модель	F1-Score (CoNLL)	Precision	Recall
KCA	72.4 %	75.2 %	70.1 %
KazBERT	76.8 %	78.0 %	75.4 %

In the primary scenario addressed in this study—identifying referential links between pronouns and nearby candidate antecedents, the KCA model demonstrated powerful performance. However, several complex linguistic phenomena frequently encountered in real-world texts continue to present challenges. The main problematic cases and the approaches applied or proposed to address them are as follows:

Ellipsis: In elliptical constructions, key elements of a phrase may be omitted, making it difficult to resolve the reference based on a single sentence. In such cases, the model leverages syntactic parse trees and rule-based templates to retrieve missing context from preceding sentences. Semantic similarity and discourse distance features play a critical role in inference.

Zero Anaphora: In Kazakh, the subject is sometimes omitted but understood implicitly, especially in informal or stylistic contexts. Here, the model uses topic-based priors and ontological knowledge drawn from the textual domain to boost the salience of likely antecedents. If the document topic and the previous sentence's subject align, the corresponding candidate is assigned a higher confidence score.

Metonymy: Metonymic expressions-such as referring to “professors” as “the university” “can introduce ambiguity in reference resolution. To address this, the system incorporates NER tagging and semantic filtering through a knowledge base. When needed, entity-type exclusion rules are applied to disambiguate metonymic references.

To process such complex phenomena, the system includes specialized post-processing modules, including semantic filters, topic priors, and heuristic rules. These mechanisms have reduced many typical errors, though full resolution, especially in cases of metonymy and long-distance references-remains an open challenge. Despite architectural improvements such as Longformer (Clark et al., 2019) and BigBird (Wu et al., 2022), morphologically rich languages remain underexplored in long-context modeling scenarios. Therefore, handling these cases more robustly is proposed as a direction for future research and refinement.

Conclusion.

This study proposes an effective approach to reference resolution in low-resource morphologically rich languages like Kazakh. The hybrid model, Kazakh Coreference Adaptation (KCA), combines linguistic rules, syntactic structures, and domain-specific semantic knowledge with machine learning techniques. This integration allows the model to achieve robust results even under data-scarce conditions.

Experimental evaluations revealed that the KCA model significantly improved performance compared to baseline methods. The overall CoNLL F1 score increased from 54.5 % to 70.3 %, with individual metric gains ranging from 13 % to 17 % (MUC, B³, CEAF). Notably, the B³ metric saw a 17.3 % improvement, emphasizing the impact of rule-based filtering tailored to Kazakh's morphological features.

Domain-specific analysis showed variation across topics: political texts yielded the highest accuracy (71 %) due to more stable reference patterns, while economic texts performed lower due to complex terminology. Error analysis confirmed that KCA reduced all major error types-False Negative, False Positive, and mismatched pronouns-by nearly half compared to the baseline model, underscoring the efficacy of the hybrid architecture.

The scientific contribution of this work lies in providing a practical solution that can be adapted for other morphologically rich and low-resource languages. While developed for Kazakh, the KCA framework can be extended to related Turkic languages.



It also underscores the importance of integrating domain knowledge into neural NLP systems when data availability is limited.

The proposed method contributes meaningfully to the Kazakh NLP ecosystem and opens pathways to other applications such as anaphora resolution, text summarization, and information retrieval.

The proposed approach presents promising results, but several enhancements can be pursued to further improve its effectiveness and applicability. One key direction involves integrating contextual embeddings from multilingual or Kazakh-specific BERT models to refine semantic similarity calculations during feature extraction. This could help the system better capture nuanced meanings in complex linguistic structures.

Additionally, the current model relies on manually tuned parameters. Employing automatic hyperparameter optimization methods—such as Bayesian Optimization or Hyperopt—could significantly increase adaptability and model performance across varying datasets. Expanding the annotated corpus is another vital step. Incorporating texts from diverse domains like politics, economics, and sports would ensure broader generalization and reduce overfitting to specific content styles.

To address remaining challenges in linguistic complexity, specialized modules can be developed to handle ellipsis, zero anaphora, and metonymic expressions more reliably. These enhancements would be especially valuable in real-world applications, such as dialogue systems, question-answering platforms, and information retrieval engines, where accurate coreference resolution is essential. Finally, adapting the approach through transfer learning to other Turkic languages—and possibly integrating it into multilingual models—would help extend its relevance and effectiveness across languages with similar morphological traits.

REFERENCES

- Beltagy I., Peters M., Cohan A. (2020). *Longformer: The Long-Document Transformer*: ACL 2020 Proceedings. — Online. Proceedings. Pp. 345–355. <https://doi.org/10.48550/arXiv.2004.05150> (in Eng.).
- Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., et al. (2020). *Language Models are Few-Shot Learners*: NeurIPS 2020 Conference Proceedings. — Vancouver, Canada. Proceedings. Pp. 1877–1901. <https://doi.org/10.48550/arXiv.2005.14165> (in Eng.).
- Clark K., Khandelwal U., Levy O., Manning C. (2019). *What Does BERT Look at? An Analysis of BERT's Attention*: ACL 2019 Proceedings. Florence, Italy. Proceedings. Pp. 2763–2773. <https://doi.org/10.48550/arXiv.1906.04341> (in Eng.).
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*: NAACL 2019 Proceedings. — Minneapolis, USA. Proceedings. Pp. 4171–4186. <https://doi.org/10.48550/arXiv.1810.04805> (in Eng.).
- Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C., Mishkin P., et al. (2022). *Training Language Models to Follow Instructions with Human Feedback*: NeurIPS 2022 Conference Proceedings. — New Orleans, USA. Proceedings. Pp. 27730–27744. <https://doi.org/10.48550/arXiv.2203.02155> (in Eng.).
- Jurafsky D., Martin J. H. (2021). *Speech and Language Processing* (3rd ed., draft). Stanford University. (in Eng.).
- Joshi M., Levy O., Zettlemoyer L., Weld D. (2019). *BERT for Coreference Resolution*: EMNLP 2019 Conference Proceedings. — Hong Kong, China. Proceedings. Pp. 5803–5808. <https://doi.org/10.48550/arXiv.1908.09091> (in Eng.).
- Liu R., Mao R., Lu A. T., Cambria E. (2023). *A Brief Survey on Recent Advances in Coreference Resolu-*

tion. Artificial Intelligence Review. — Vol. 56. — No. 3. Pp. 379–412. <https://doi.org/10.1007/s10462-023-10506-3> (in Eng.).

Lee K., He L., Lewis M., Zettlemoyer L. (2017). *End-to-End Neural Coreference Resolution*: EMNLP 2017 Conference Proceedings. — Copenhagen, Denmark. Proceedings.m. Pp. 188–197. <https://doi.org/10.48550/arXiv.1707.07045> (in Eng.).

Lee K., He L., Zettlemoyer L. (2018). *Higher-Order Coreference Resolution with Coarse-to-Fine Inference*: NAACL 2018 Proceedings. — New Orleans, USA. Proceedings. Pp. 687–692. <https://doi.org/10.48550/arXiv.1804.05392> (in Eng.).

Tay Y., Dehghani M., Bahri D., Metzler D. (2022). *Efficient Transformers: A Survey*. ACM Computing Surveys. — Vol. 6. — No. 6. Article 109. <https://doi.org/10.1145/3530811> (in Eng.).

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., et al. (2017). *Attention Is All You Need*: NeurIPS 2017 Conference Proceedings. — Long Beach, USA. Proceedings. Pp. 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762> (in Eng.).

Wu S., Li Y., Zhang Y., et al. (2022). *Knowledge-Augmented Methods for NLP: A Survey*. Knowledge-Based Systems. Pp. 235. Article 107652. <https://doi.org/10.1016/j.knosys.2021.107652> (in Eng.).

Zaheer M., Guruganesh G., Dubey K. A., Ainslie J., Alberti C., Ontanon S., et al. (2020). *Big Bird: Transformers for Longer Sequences*: NeurIPS 2020 Conference Proceedings. — Vancouver, Canada. Proceedings. Pp. 17283–17297. <https://doi.org/10.48550/arXiv.2007.14062> (in Eng.).

Zhang H., Zhao H., Qin B. (2021). *Knowledge-Enhanced Pre-trained Language Models: A Survey*. IEEE Transactions on Knowledge and Data Engineering. — Vol. 34. — No. 9. Pp. 4180–4198. (in Eng.).



**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).