

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



**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION
AND COMMUNICATION TECHNOLOGIES**

2022 (3) 3
Маусым-қыркүйек

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

БАС РЕДАКТОР:

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

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

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

ҒАЛЫМ ХАТШЫ:

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

РЕДАКЦИЯЛЫҚ АЛҚА:

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

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

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

Микеле Пагано — PhD, Пиза университетінің профессоры (Италия)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ЖАУАПТЫ РЕДАКТОР:

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

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

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

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

№ KZ82VPY00020475 мерзімдік басылым тіркеуіне қойылу туралы куәлік.

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

Мерзімділігі: жылына 4 рет.

Тиражы: 100 дана

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

E-mail: ijiet@iitu.edu.kz

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

© Халықаралық ақпараттық технологиялар университеті АҚ, 2022

© Авторлар ұжымы, 2022

ГЛАВНЫЙ РЕДАКТОР:

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

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

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

УЧЕНЫЙ СЕКРЕТАРЬ:

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

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

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

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

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

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Тадеш 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)

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

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

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

Международный журнал информационных и коммуникационных технологий

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

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

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

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

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

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

E-mail: ijict@iitu.edu.kz

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

© АО Международный университет информационных технологий, 2022

© Коллектив авторов, 2022

EDITOR-IN-CHIEF:

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

DEPUTY CHIEF DIRECTOR:

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

SCIENTIFIC SECRETARY:

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

EDITORIAL BOARD:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

EXECUTIVE EDITOR

Eraly Diana Ruslankyzy — International Information Technology University (Kazakhstan)

«International Journal of Information and Communication Technologies»

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Owner: International Information Technology University JSC (Almaty).

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

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

Periodicity: 4 times a year.

Circulation: 100 copies.

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

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

© International Information Technology University JSC, 2022

© Group of authors, 2022

МАЗМҰНЫ

БАҒДАРЛАМАЛЫҚ ҚАМТАМАНЫ ӨЗІРЛЕУ ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ

Чинибаева Т.Т., Таймас Н., Жексенкадыр Е. СТУДЕНТТЕРДІҢ ҮЛГЕРІМІН ЕСЕПКЕ АЛУДЫ АВТОМАТТАНДЫРУ ЖӘНЕ СЫНАУ.....	8
Төлегенова А. МӘТІНДІ НОРМАЛАУ ҮШІН NAIVE BAYES КЛАССИФИКАТОРЫ: ҚАЗАҚ ТІЛІНДЕ ЗЕРТТЕУ.....	17

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУІПСІЗДІК

Шаповаленко О.Д., Бедрий Д.И. КИБЕРҚАУІПСІЗДІКТІҢ ҚАЗІРГІ ЖАҒДАЙЫНА ШОЛУ.....	24
Ахметова Д. ТҮРЛІ СТЕГАНОГРАФИЯЛЫҚ ӘДІСТЕРДІ ШІФРЛЕУ ТИІМДІЛІГІ.....	36

ЭКОНОМИКАДАҒЫ ЖӘНЕ МЕНЕДЖМЕНТТЕГІ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

Бердіқұлова Ғ.М., Омарова А.Ш., Сағандықова С.Ш., Абдинова М.Х., Батай М.А. УНИВЕРСИТЕТТЕРДІ ЦИФРЛАНДЫРУДЫҢ ҚАЗАҚСТАНДЫҚ ЖӘНЕ ШЕТЕЛДІК ТӘЖІРИБЕСІ.....	48
Гогунский В.Д., Лукьянов Д.В., Колесников А.Е. ҚЫЗМЕТКЕРЛЕРДІҢ БІЛІКТІЛІГІН ДАМУ ЖӘНЕ ҚАЙТА ДАЙЫНДАУ БОЙЫНША ҚЫЗМЕТ МОДЕЛІН ӨЗІРЛЕУ.....	58

БҰҚАРАЛЫҚ АҚПАРАТ ҚҰРАЛДАРЫНДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

Алхабаев Ш.Е. TENGRINEWS.KZ САЙТЫ МЫСАЛЫНДА ОНЛАЙН ЖУРНАЛИСТИКАДАҒЫ ЖОБАНЫ БАСҚАРУ.....	69
Асылбек А. ФЕЙК АҚПАРАТТЫҢ ҚОҒАМДЫҚ ШІКІР ҚАЛЫПТАСТЫРУҒА БЫҚПАЛЫ.....	78

СОДЕРЖАНИЕ

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

Чинибаева Т.Т., Таймас Н., Жексенкадыр Е. АВТОМАТИЗАЦИЯ И ТЕСТИРОВАНИЕ УСПЕВАЕМОСТИ СТУДЕНТОВ.....	8
Толегенова А. НАИВНЫЙ БАЙЕСОВСКИЙ КЛАССИФИКАТОР ДЛЯ НОРМАЛИЗАЦИИ ТЕКСТА: ПРИМЕР ДЛЯ КАЗАХСКОГО ЯЗЫКА.....	17

ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

Шаповаленко О.Д., Бедрий Д.И. ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ КИБЕРБЕЗОПАСНОСТИ.....	24
Ахметова Д. ЭФФЕКТИВНОСТЬ ШИФРОВАНИЯ РАЗЛИЧНЫХ СТЕГАНОГРАФИЧЕСКИХ МЕТОДОВ.....	36

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ

Бердыкулова Г.М., Омарова А.Ш., Сагандыкова С.Ш., Абднова М.Х., Багай М.А. КАЗАХСТАНСКИЙ И ЗАРУБЕЖНЫЙ ОПЫТ ЦИФРОВИЗАЦИИ УНИВЕРСИТЕТОВ.....	48
Гогунский В.Д., Лукьянов Д.В., Колесников А.Е. РАЗРАБОТКА ДЕЯТЕЛЬНОСТНОЙ МОДЕЛИ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ И ПЕРЕПОДГОТОВКИ КАДРОВ.....	58

ЦИФРОВЫЕ ТЕХНОЛОГИИ В МАСС-МЕДИА

Алхабаев Ш.Е. УПРАВЛЕНИЕ ПРОЕКТАМИ В ОНЛАЙН ЖУРНАЛИСТИКЕ НА ПРИМЕРЕ САЙТА TENGRINEWS.KZ.....	69
Асылбек А. СИЛА ФЕЙКОВОЙ ИНФОРМАЦИИ В ФОРМИРОВАНИИ ОБЩЕСТВЕННОГО МНЕНИЯ.....	78

CONTENTS

SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

Chinibayeva T.T., Taimas N., Zhexenkadyr Y. AUTOMATION AND TESTING OF STUDENT ACHIEVEMENT.....	8
Tolegenova A. A NAIVE BAYESIAN CLASSIFIER FOR NORMALIZATION OF TEXT: A CASE STUDY FOR KAZAKH LANGUAGE.....	17

INFORMATION AND COMMUNICATION NETWORKS AND CYBERSECURITY

Shapovalenko O.D., Bedrii D.I. OVERVIEW OF THE PRESENT STATE OF CYBER SECURITY.....	24
Akhmetova D. ENCRYPTION EFFICIENCY OF VARIOUS STEGANOGRAPHIC METHODS.....	36

DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

Berdykulova G.M., Omarova A.Sh., Sagandykova S.Sh., Abdinova M.Kh., Batai M.A. KAZAKHSTANI AND FOREIGN EXPERIENCE OF UNIVERSITY DIGITALIZATION.....	48
Gogunskii V.D., Lukianov D.V., Kolesnikov O.Ye. DEVELOPMENT OF AN ACTIVITY MODEL FOR PROFESSIONAL DEVELOPMENT AND RETRAINING OF STAFF.....	58

DIGITAL TECHNOLOGIES IN THE MASS MEDIA

Alkhabayev Sh.Ye. PROJECT MANAGEMENT IN ONLINE JOURNALISM ON THE EXAMPLE OF THE SITE TENGRINEWS.KZ.....	69
Asylbek A. THE POWER OF FAKE INFORMATION IN FORMING PUBLIC OPINION.....	78

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 3. Is. 3. Number 11 (2022). Pp. 17–23

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

<https://doi.org/10.54309/IJICT.2022.11.3.002>

A NAIVE BAYESIAN CLASSIFIER FOR NORMALIZATION OF TEXT: A CASE STUDY FOR KAZAKH LANGUAGE

A. Tolegenova

Tolegenova Assylay — master student of the department of "Computer Science" of the Suleiman Demirel University, Kaskelen, Kazakhstan

© A. Tolegenova, 2022

Abstract. The amount of complicated documents and texts has increased exponentially in recent years, necessitating a deeper understanding of machine learning technologies in order to effectively identify texts in numerous applications. Text normalization is one of the best decisions. It is the reduction of all words of the text to the original form. This paper investigates a layered strategy for fixing mistakes in Kazakh language literature downloaded from the Internet. Because of the widespread use of social media as a source for linguistic study, error correction is a critical issue. The goal of this research was to look at the current Naive Bayes algorithm in English, as well as the normalization of words and sentences in natural languages, in order to create a similar algorithm for the Kazakh language. The method of morphology of Kazakh words and their difference from English was considered suitable for processing words in a dictionary. As a result of the normalization system, the efficiency of this method for the Kazakh language was proved.

Keywords: text normalization, Naïve Bayes algorithm, natural language, processing of text, classifier

For citation: A. Tolegenova. A naive bayesian classifier for normalization of text: a case study for kazakh language // INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2022. Vol. 3. Is. 3. Number 11. Pp.17–23 (In Russ.). DOI: *10.54309/IJICT.2022.11.3.002*.

МӘТІНДІ НОРМАЛАУ ҮШІН NAIVE BAYES КЛАССИФИКАТОРЫ: ҚАЗАҚ ТІЛІНДЕЗЕРТТЕУ

A. Толегенова

Төлегенова Асыллай — Сүлеймен Демирел атындағы университетінің «Компьютерлік ғылымдар» кафедрасының магистранты.

© A. Толегенова, 2022



Аннотация. Соңғы жылдары күрделі құжаттар мен мәтіндердің саны экспоненциалды түрде өсті, бұл көптеген қолданбаларда мәтіндерді тиімді анықтау үшін машиналық оқытутехнологияларын тереңірек түсінуді талап етеді. Мәтінді қалыпқа келтіру — ең жақсы шешімдердің бірі. Бұл мәтіндегі барлық сөздерді бастапқы түріне келтіру. Бұл мақалада интернеттен жүктеп алынған қазақ тіліндегі әдебиеттердегі қателерді түзетудің көп деңгейлі стратегиясы қарастырылған. Әлеуметтік желілерді лингвистикалық зерттеулердің көзі ретінде кеңінен пайдалануына байланысты қателерді түзету өзекті мәселе болып табылады. Бұл зерттеудің мақсаты ағылшын тіліндегі қазіргі аңғал Байес алгоритмін, сонымен қатар қазақ тіліне ұқсас алгоритм құру үшін табиғи тілдегі сөздер мен сөйлемдерді нормалауды зерттеу болды. Бұл жұмыстың мақсаты ағылшын тіліндегі қолданыстағы Naive Bayes алгоритмін және табиғи тілдердегі сөздер мен сөйлемдерді нормалауды зерттеу, сондай-ақ қазақ тіліне арналған ұқсас алгоритмді жасау болды. Сөздің түбірін алудың қолданыстағы алгоритмдері және қалыпты форманы синтездеудің мүмкін жолдары қарастырылады. Сөздікте сөздік өңдеуге қолайлы қазақ сөздерінің морфологиялық әдістемесі және олардың ағылшын тілінен айырмашылығы қарастырылады. Нормалау жүйесінің нәтижесінде бұл әдістің қазақ тілі үшін тиімділігі дәлелденді.

Түйін сөздер: мәтінді қалыпқа келтіру, Naive Bayes алгоритмі, табиғи тіл, мәтінді өңдеу, классификатор

Дәйексөз үшін: А. Толегенова. Мәтінді нормалау үшін naive bayes классификаторы: қазақ тілінде зерттеу // ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2022. Том. 3. Is. 3. Нөмірі 11. 17–23 бет (орыс тілінде). DOI: 10.54309/IJICT.2022.11.3.002.

КЛАССИФИКАТОР NAIVE BAYES ДЛЯ НОРМАЛИЗАЦИИ ТЕКСТА: ПРИМЕР ДЛЯ КАЗАХСКОГО ЯЗЫКА

А. Толегенова

Толегенова Асылрай — магистрант кафедры «Компьютерные науки» университета им. Сулеймана Демиреля

© А. Толегенова, 2022

Аннотация. Количество сложных документов и текстов в последние годы увеличилось в геометрической прогрессии, что требует более глубокого понимания технологий машинного обучения для эффективной идентификации текстов в многочисленных приложениях. Нормализация текста — одно из лучших решений. Это приведение всех слов текста к исходному виду. В данной статье исследуется многоуровневая стратегия исправления ошибок в казахскоязычной литературе, скачанной из Интернета. Из-за широкого использования социальных сетей в качестве источника для лингвистических исследований исправление ошибок является критической проблемой. Целью данного исследования было изучить текущий



алгоритм наивного Байеса на английском языке, а также нормализацию слов и предложений на естественных языках, чтобы создать аналогичный алгоритм для казахского языка. Целью данной работы было изучение существующего алгоритма Наивного Байеса на английском языке и нормализации слов и предложений на естественных языках, а также разработка аналогичного алгоритма для казахского языка. Рассмотрены существующие алгоритмы извлечения основы слова и возможные способы синтеза нормальной формы. Рассмотрен метод морфологии казахских слов и их отличие от английских, пригодных для обработки слов в словаре. В результате системы нормализации доказана эффективность данного метода для казахского языка.

Ключевые слова: нормализация текста, алгоритм наивного Байеса, естественный язык, обработка текста, классификатор

Для цитирования: А. Толегенова. Классификатор naive bayes для нормализации текста: пример для казахского языка // МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2022. Том. 3. Is. 3. Номер 11. Стр. 17–23 (на русском языке). DOI: 10.54309/IJICT.2022.11.3.002.

Introduction

Currently, the main information in the form of a text format is constantly growing, so there is a variety of solutions for automated analysis, spelling processing, and analysis of the generation of social networks. In order to show the user information in an understandable language, word normalization is used. For text normalization, there are many automatic text processing systems designed to solve individual or a small set of tasks. In connection with such tasks, the most urgent request is the introduction of a suitable algorithm for the language to be used, since each language has its own peculiarities.

Different languages have different grammatical and orthographic features, and in this regard, algorithms that have been successfully performed to normalize one language may show certain flaws and low performance in another language. This work is devoted to checking, processing, and solving problems with Bayer's theory for Kazakh language texts. Since the processing of all words in the text to normal form greatly simplifies the work not only for search engines, but also for projects with automatic texting. The object of this study is the process of automatic processing of texts of social networks and algorithms for their processing.

Currently, social networks are the main source of people's information and their opinions. The relevance of this study is mainly related to the development of social networks, as they can become a source of additional data. Social networks and portals contain a lot of valuable material, and correctly written words will be useful in further work. To use sources such as social networks, special tools are required. Many people have reduced or modified the words that are used in social networks, in connection with this, the database of Kazakh words has been replenished. For the study 50 words were entered which were collected during the survey.

The Bayes theorem was chosen to solve problems because the Naive Bayes

classifier is one of the easiest classification methods. However, it frequently performs more complicated algorithms. This classification model is based on the concept of the conditional probability that document d belongs to class c . In connection with the above, this paper considers the Naive Bayes algorithm that has been used in many projects related to natural language processing, based on supervised machine learning methods.

The main goal of the study is to create and implement a Bayesian theorem-based algorithm for text normalization in social networks for the Kazakh language, as well as to analyze the use of this method for an existing project in English that was created for Kazakh, in order to demonstrate the differences in text normalization between the two languages.

During the study, research works related to agglutinative languages were considered, such as "Text normalization and spelling correction in Kazakh language" by Gaukhar Slamova and Meruyert Mukhanova [1]. In which the morphological analysis and aspects of the Kazakh language were viewed, using the division of words into several groups and the division of the text into modules of words. Abdalhkim A.M. Damegi in his work has developed an application that gives advice on correcting texts while typing texts [2]. This paper presents the benefits of using Naive Bayes theorems for the English language. Nazarbayev University students Zhanibek Kozhimbayev, Zhandos Yessenbayev presented the databases collected from the Kazakhstan segment of the Internet [3]. Here the methods of statistical machine translation, in which a parallel corpus was created with a set of aligned sentences in canonical and non-canonical forms, were applied. And for more comparison of English with Kazakh, the work of Ahmed Yunus and Md Masum was studied, in which the main task was to correct spelling. In this work, the author created a System that used a Naive Bayes classifier with a Python NLTK implementation to find out which one was most likely to be a misword correction [4].

Experimentally, it was revealed what results the algorithm with naive bayes gives for both languages.

Methodology

A Naive Bayesian Classifier (Bayesian Naive, TIS) is an example of how vector analysis techniques can be used. Because of its comparative simplicity in implementation and testing, NB C is one of the most widely used classifiers. While naive Bayesian classifiers show no worse results than more complex classifiers, they are not in any way superior to each other [5].

Bayes' theorem (or formula) governs how the naive Bayesian classifier operates:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

A and B are referred to as events. This means that $P(A \text{ and } B)$ indicates whether event A occurs if event B occurs. Evidence is also considered event B. In probability theory, the probability that something will occur before it becomes evident that is $P(A)$, i.e., its prior independent probability. If B comes about due to evidence A, then $P(A|B)$ is



the probability of B following from evidence A, i.e., probability B will happen due to evidence A [6].

With a Naive Bayesian classification algorithm, it is possible to predict whether a variable consists of more than two classes [7].

Finding out which classification the text belongs to. A Naive Bayesian algorithm for classifying text can be extremely effective due to multi-level prediction [8]. It can also be used for applications such as spam detection and sentiment analysis [9].

Combining Naive Bayes with algorithms for filtering unseen information, such as collaborative filtering, can help predict a user's need for a resource.

Result

For this model, a document is a vector: $d = \{w_1, w_2, \dots, w_n\}$, where w_i is the weight of the i -th term, and n is the size of the sample dictionary [10]. Thus, according to Bayes' theorem, the probability of class c for document d will be:

$$P(c | d) = \frac{P(c)P(d | c)}{P(d)}$$

Thus, probability under a given condition for all classes is calculated.

The most probable class c^* to which document d belongs is the one in which the conditional probability of document d belonging to class c is maximal:

$$c^* = \arg \max_c P(c | d).$$

There are two types of suffixes in the Kazakh language:

- 1) word-forming;
- 2) formative

Word-forming suffixes are used to form new words (the meaning of the word changes). For example: ән — song, ән-ші — singer; ақыл — mind, ақыл-ды — smart. Modified suffixes include: derived noun suffixes; suffixes of the degree of comparison of adjectives; suffixes of ordinal names of numerals; suffixes of verbs forming the voice, the amplifying and negative types of the verb.

Table 1: Testing results of Naive Bayesian Classifier model for normalization for Kazakh and English Language

Language	English	Kazakh
The average number of correctly corrected words	6,2	0,35
The average number of incorrectly corrected words	5,92	0,2
The average number of unnecessary words corrected	1,3	0,05
Accuracy	95,78	99,19
Recall	61,98	81,33
Result of F1	82,4	89,38



Using Naive Bayesian classifications has some advantages of using classifiers:

- A Naïve Bayesian algorithm can be implemented quickly and easily.
- Convergence rates will be faster than those of discriminative models.
- Less data is required for training.
- These algorithms tend to scale very well, or they scale linearly as the number of predictors and variables increases.
- It can analyze continuous and discrete data as well as make probabilistic predictions.
- Both binary and multiclass classification problems can be solved with the naive Bayesian algorithm.

These algorithms tend to scale very well, or they increase exponentially when the amount of input increases.

- It can analyze discrete and continuous data and make probabilistic predictions.
- In both binary and multiclass classification tasks can be solved with the Naive Bayesian algorithm. An approach based on Naive Bayesian classification has several disadvantages.

- A naive Bayesian classification suffers from the drawback that features tend to be strongly independent of one another in reality, since it is extremely unlikely to find in real life a collection of characteristics that is 100 % independent.

- Moreover, Naïve Bayesian classification cannot deal with categorical variables with zero frequencies. Suppose the Naïve Bayesian model is limited to assigning zero probability to a categorical variable that does not appear in the test data set.

- The obvious drawback of Naive Bayesian classification is that in real life it consists of traits that are highly independent of one another, since it is impossible to have a set of traits that are absolutely independent.

Conclusion

The outcomes of this investigation were analyzed utilizing the algorithm for both the English and Kazakh languages. Finally, methodologies for normalizing texts in social networks based on Naive Bayes were demonstrated, and a comparative analysis was performed. The method was also demonstrated for text checking on the Django framework.

Concepts like text spelling and morphology were theoretically understood while creating this thesis. As a consequence of the software, we can see that when utilizing Naive Bayes theorems for the English language, we get better results than when using them for the Kazakh language.

The program was created by applying Naive Bayes system methods to various text analysis possibilities, and a comparison study was performed. The effects of normalization on the Naive Bayes's algorithm quality for analyzing text sentiment are demonstrated. The implementation can be used for correcting texts with pre-normalized content for future work.

REFERENCES

Abdalkhim A., Damegi M. (2011). Implementation of Naïve Bayes Algorithm for Spell Correction in Web-Based Learning for written English, Seminar on electrical, informatics and its education, 2011



- Bird S., Klein E., Loper E. (2009). Natural language processing with python. Sebastopol, CA: O'Reilly Media, — 2009. — Pp. 245–250.
- Chen J., Huang H., Tian S., Qu Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36 (3) (April 2009). — Pp. 5432–5435
- Eyheramendy S., Lewis D.D., Madigan D. (2003). On the Naive Bayes model for text categorization. Bishop C.M., Frey B.J. (Eds.), *AI & Statistics 2003: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics (2003)*. — Pp. 332–339
- John G.H., Langley P. (1995). Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the 11th international conference on uncertainty in artificial intelligence*, San Francisco, CA, — 18–22. — August 1995. — Pp. 338–345. San Mateo, CA: Morgan Kaufmann.
- Kozhirbayev Zh., Yessenbayev Zh., Makazhanov A. (2020). KazNLP: A Pipeline for Automated Processing of Texts Written in Kazakh Language, *International Conference on Speech and Computer (SPECOM 2020)*, — 2020
- Kaviani P., Dhotre S. (2017). Short Survey on Naive Bayes Algorithm, *International Journal of Advance Engineering and Research Development (IJAERD)*. — Volume 4. — Issue 11. — 2017
- Yunus A., Masum M. (2020). A Context Free Spell Correction Method using Supervised Machine Learning Algorithms, *International Journal of Computer Applications (0975 – 8887)*. — Volume. — 176. — No. 27. — 2020
- Rish I. (2001). An empirical study of the naive Bayes classifier. In: *IJCAI workshop on empirical methods in artificial intelligence*, — 2001.
- Slamova G., Mukhanova M. (2018). Text normalization and spelling correction in Kazakh language, *AIST*, 2018.



**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

Правила оформления статьи для публикации в журнале на сайте:

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

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

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

ОТВЕТСТВЕННЫЙ РЕДАКТОР

Ералы Диана Русланқызы

КОМПЬЮТЕРНАЯ ВЕРСТКА

Жадыранова Гульнур Даутбековна

Подписано в печать 15.09.2022.

Формат 60x881/8. Бумага офсетная. Печать - ризограф. 7,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.