

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Қазақстан Республикасының
Ғылым Академиясының
Әл-Фараби атындағы
Қазақ ұлттық университеті

NEWS

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

PHYSICO-MATHEMATICAL SERIES

4 (344)

OCTOBER – DECEMBER 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас директорының м.а. (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), **Н=7**

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), **Н=5**

БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), **Н=3**

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), **Н=17**

ӘМІРҒАЛИЕВ Еділхан Несіпханұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Жасанды интеллект және робототехника зертханасының меңгерушісі (Алматы, Қазақстан), **Н=12**

КИЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония), ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=6**

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=4**

ОТМАН Мохаммед, PhD, Информатика, коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті (Селангор, Малайзия), **Н=23**

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының аға ғылыми қызметкері (Алматы, Қазақстан), **Н=3**

БИЯШЕВ Рустам Гакашевич, техника ғылымдарының докторы, профессор, Информатика және басқару мәселелері институты директорының орынбасары, Ақпараттық қауіпсіздік зертханасының меңгерушісі (Қазақстан), **Н=3**

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), **Н=3**

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), **Н=2**

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РКБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика-математикалық сериясы*».

Қазіргі уақытта: «*ақпараттық технологиялар*» бағыты бойынша *ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді*.

Мерзімділігі: *жылына 4 рет*.

Тиражы: *300 дана*.

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022
Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

Главный редактор:

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

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), **Н=7**

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), **Н=5**

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Саптаева (Алматы, Казахстан), **Н=3**

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **Н=23**

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), **Н=17**

АМИРГАЛИЕВ Едилхан Несипханович, доктор технических наук, профессор, академик Национальной инженерной академии РК, заведующий лабораторией «Искусственного интеллекта и робототехники» (Алматы, Казахстан), **Н=12**

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=6**

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=4**

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), **Н=23**

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

БИЯШЕВ Рустам Гакашевич, доктор технических наук, профессор, заместитель директора Института проблем информатики и управления, заведующий лабораторией информационной безопасности (Казахстан), **Н=3**

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

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

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), **Н=2**

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

«Известия НАН РК. Серия физико-математическая».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика-математическая.*

В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

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

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, acting General Director of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), **H = 7**

Mamyrbayev Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H = 5**

BAIGUNCHEKOV Zhumadil Zhanabaevich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), **H=23**

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), **H= 17**

AMIRGALIEV Edilkhan Nesipkhanovich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Head of the Laboratory of Artificial Intelligence and Robotics (Almaty, Kazakhstan), **H= 12**

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 6**

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 4**

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), **H= 23**

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H= 3**

BIYASHEV Rustam Gakashevich, doctor of technical sciences, professor, Deputy Director of the Institute for Informatics and Management Problems, Head of the Information Security Laboratory (Kazakhstan), **H= 3**

KAPALOVA Nursulu Aldazarhovna, Candidate of Technical Sciences, Head of the Laboratory cyber-security, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=3**

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), **H=5**

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), **H=2**

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

News of the National Academy of Sciences of the Republic of Kazakhstan.

Physico-matematical series.

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж**, issued 14.02.2018

Thematic scope: *physical-mathematical series.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

© National Academy of Sciences of the Republic of Kazakhstan, 2022

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

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

Volume 4, Number 344 (2022), 56-67
<https://doi.org/10.32014/2022.2518-1726.156>

ӘОЖ 004.9

Г. Қалман¹, М.А. Самбетбаева^{1*}, Д.А. Ақтаева², А.С. Илюбаев^{2*}

¹Л.Н. Гумилев атындағы Еуразиялық ұлттық университеті,
Астана, Қазақстан;

²Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Қазақстан.
E-mail: guljamal14@gmail.com

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН АНАФОРАНЫ ШЕШУ МОДЕЛІ

Аннотация. Қазіргі кезде референцияны шешу табиғи тілдегі ең күрделі және әлі күнге дейін толық шешімін таппаған мәселелердің бірі, осыған қарамастан, көптеген лингвистердің және зерттеушілердің еңбектері талқыланып зерттеу жүргізіліп референцияны толық шешу модельдерін жобалауды талап етіп отыр.

Атап айтсақ, компьютерлік лингвистиканың алдында тұрған өзекті міндеттердің бірі мәтіндік құжаттарда әртүрлі нысандар туралы: адамдар, ұйымдар, оқиғалар, орындар және т.б., сондай-ақ олардың арасындағы байланыстар туралы мәліметтерді бөліп көрсету болып табылады. Әрбір Ақпараттық нысан белгілі бір ұғымға/пәндік аймақтың қатынасына сәйкес келетін белгілі бір құрылымға ие. Болашақта мәтінді өңдеу белгілі бір ресми тілде шектеулі және нақты сипатталған ақпараттық жүйе аясында жүзеге асырылады деп болжауға болады. Мәтінді автоматты өңдеудің маңызды элементтері анафориялық байланыстарды орнату және бір нысанның әртүрлі атауларын анықтау болғандықтан, белгілі бір контексте кез-келген объектіге, ұйымға, адамға және т.б., бірнеше рет айтылады.

Бұл зерттеу жұмысында машиналық оқыту әдістеріне негізделген қазақ тіліндегі анафоралық қатынасты шешу моделі ұсынылды. Қазақ тілінің негізгі тілдік ерекшеліктері ескеріле отырып жіктеу алгоритмдері пайдаланылды. Машиналық оқытуда тірек векторлық машиналарды(SVM) және Libsvm кітапханасы қолданылды. SVM-тің негізгі идеясы – шешім қабылдау жазықтығына гиперплант орнату, ең үлкен маргинализацияға қол жеткізу үшін оң даналар мен теріс даналарды оқшаулау.

Зерттеу нәтижелерін салыстыру мақсатында бірнеше эксперименттер жасалды. Сынақ деректері жинағы ретінде TengriNews.kz-тен жаңалықтар топтамасы және Ғ.Мұстафин әңгімелерінен үзінділер қолданылды. Эксперименттер нәтижелерін салыстыру мақсатында 2-кестеде көрсетілген нәтижелер алынды. Алынған нәтижелер қойылған міндеттерді шешуге ұсынылған модельдің тиімділігін растайды.

Түйін сөздер: анафора, антецедент, референциялық қатынас, машиналық оқыту, тірек векторлық машиналар (SVM).

Ғ. Қалман¹, М.А. Самбетбаева^{1*}, Д.А. Ақтаева², А.С. Илюбаев^{2*}

¹Евразийского национального университета им. Л.Н. Гумилева,
Астана, Казахстан;

²Кокшетауский университет имени Абая Мырзахметова,
Кокшетау, Казахстан.

E-mail: guljamal14@gmail.com

МОДЕЛЬ РАЗРЕШЕНИЯ АНАФОРЫ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В настоящее время решение референции является одной из самых сложных и до сих пор не полностью решенных проблем на естественном языке несмотря на то, что работы многих лингвистов и исследователей обсуждаются и требуют разработки моделей полного решения референции.

В частности, одной из актуальных задач, стоящих перед компьютерной лингвистикой, является выделение информации о различных объектах: людях, организациях, событиях, местах и т.д., а также о связях между ними в текстовых документах. Каждый информационный объект соответствует определенному отношению области понятия/дисциплины и имеет определенную структуру. Можно предположить, что в будущем обработка текстов будет осуществляться в рамках определенной информационной системы, ограниченной и четко описанной на определенном официальном языке. Важными элементами автоматической обработки текста являются установление анафорических связей и отождествление разных названий одного и того же объекта, например, какого-либо объекта, организации, лица и т.п., упоминаемых более одного раза в определенном контексте.

В данной исследовательской работе была предложена модель решения анафорических отношений в казахском языке на основе методов машинного обучения. Мы использовали алгоритмы классификации, учитывающие основные лингвистические особенности казахского языка. Машинное обучение использовало машины опорных векторов (SVM) и библиотеку Libsvm.

Основная идея SVM состоит в том, чтобы разместить гиперплоскость

на плоскости принятия решений, изолируя положительные экземпляры и отрицательные экземпляры для достижения максимальной маргинализации.

Было проведено несколько экспериментов для сравнения результатов исследований. В качестве набора тестовых данных мы использовали подборку новостей Tengrinews.kz и выдержки из рассказов Г. Мустафина. Для сравнения результатов экспериментов были получены результаты, представленные в таблице 2. Полученные результаты подтвердили эффективность предложенной модели для решения поставленных задач.

Ключевые слова: анафора, антецедент, референтные отношения, машинное обучение, машины опорных векторов (SVM).

G. Kalman¹, M.A. Sambetbayeva^{1*}, A.C. Ilyubayev², D.A. Aktaeva^{2*}

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

²Abay Myrzakmetov Kokshetau University, Kokshetau, Kazakhstan.

E-mail: guljamal14@gmail.com

ANAPHORA RESOLUTION MODEL BASED ON MACHINE LEARNING METHODS

Abstract. At present, solving referencing is one of the most difficult and still not completely solved problems in natural language, despite the fact that the works of many linguists and researchers are discussed and require the development of models for the complete solution of referencing.

In particular, one of the urgent tasks facing computer linguistics is to isolate in textual documents information about various objects: people, organizations, events, places, etc., as well as the relationships between them. Each information object corresponds to a certain concept/relationship of the subject area and has a certain structure. It can be assumed that in the future text processing will be carried out within a certain information system, limited and clearly described in a certain official language. Important elements of automatic text processing are the establishment of anaphoric relationships and the definition of different names for the same object, for example, for any object, organization, person, etc. In a particular context, repeated pronunciation.

This research paper proposes a model for solving anaphoric relations in the Kazakh language, based on machine learning methods. We used classification algorithms taking into account the main linguistic features of the Kazakh language. Machine learning used support vector machines (SVM) and Libsvm library.

The basic idea of SVM-establish a hyperplane on the decision plane, isolate positive and negative instances to achieve the greatest marginalization.

Several experiments were conducted in order to compare the results of the study. As a set of test data Tengrinews.kz-we used a selection of news and excerpts from

the stories of G. Mustafin. In order to compare the results of the experiments we obtained the results shown in Table 2. The results confirmed the effectiveness of the proposed model for solving the tasks.

Key words: anaphora, antecedent, reference relations, machine learning, support vector machines (SVMs).

Кіріспе. Табиғи тілдердің көпшілігі үшін мәтінді талдау кезінде ең үлкен қиындық морфологиялық талдаудан сәтті өнделетін жеке сөздерді талдау емес, сөздер арасындағы қарым-қатынастарды немесе, басқаша айтқанда, байланыстарды талдау болып табылады. Бұл мақалада референциялық қатынастар егжей-тегжейлі қарастырылады. Екі сөз бір нақты объектіні білдірсе, сілтеме бойынша байланысқан болып саналады. Бұл құбылыс әдетте анафора деп аталады. Атап айтқанда, есімдік анафора – сілтеулік байланысқа қатысушылардың бірі есімдік болатын анафора. Бұл құбылысты келесі сөйлемнің мысалында қарастырамыз.

Ерасыл мектептен келді, **ол** бірден ағасына көмектесуге цехқа кетті.

Мұндағы «ол» есімдігі «Ерасыл» деген жалқы есімді білдіреді. Есімдік сілтеп тұрған сөзді АНТЕЦЕДЕНТ деп атайды (бұл жағдайда антецедент «Ерасыл»), сілтеме жасау құралы анафора деп аталады (бұл жағдайда анафора «ол»). Есімдік анафорасының шешілуі оның төл сөзімен сәйкес келуі болып табылады.

Анафора мен корференцияны шешу міндеті 1960 жылдардан бастап белсенді зерттеліп келеді, дегенмен, шешілмеген мәселелері әлі де бар. Бұл мәселені шешудің негізгі тәсілдерін Р.Митковтың (Митков, 1999:281) және Elango P (Elango, 2006:8) зерттеулерімен Prokofyev R. (Prokofyev et al., 2015:458) жұмыстарынан көреміз. Ағылшын тіліне арналған анафораларды автоматты түрде шешу саласындағы зерттеулер 70 жылдары басталды. Виноградов, Уилкс, Хоббстың алғашқы әдістері мен жүйелері негізінен синтаксистік ақпаратқа негізделген ережелермен жұмыс істеді, сонымен қатар энциклопедиялық білім де кеңінен қолданылды. 80 жылдары бұрын бөлек қолданылған әртүрлі белгілерді біріктіру тенденциясы пайда болды. Э. Рич пен С. Луперфойдың еңбектерінде Дж. Карбонелла, Р. Митков гендер мен сандарды, синтаксистік және семантикалық қатынастарды үйлестіретін алгоритмдерді сипаттады.

Орыс тіліне арналған анафораның шешу эксперименталды түрде аз зерттелген. Кибрик А. еңбектерінде (Кибрик, 2013: 118) орыс тіліндегі анафора құбылысының теориялық аспектілерін талқылайды және анафораның табиғатын көрсететін тілдік белгілер қатарын сипаттайды. Толпегин П (Толпегин, 2006: 88) еңбектерінде машиналық оқыту әдістерін қолдана отырып, орыс мәтіндеріндегі есімдік анафораны шешудің статистикалық моделін құру алгоритмін ұсынады. Абрамов В еңбектерінде (Абрамов, 2011:127) әлеуметтік-саяси мәтіндердің үйлесімділік ережелерін талдау үшін пайдаланатын әртүрлі сөйлемдер мен жағдайларда анафориялық қатынастарды анықтау принциптерін егжей-тегжейлі сипаттайды.

Қазіргі тәсілдер аннотацияланған корпустарды қолдана отырып, автоматты оқытуға негізделген. Олар дәстүрлі лингвистикалық әдістерді статистикалық әдістермен біріктіреді және морфологиялық, синтаксистік, семантикалық және тезаурус жиынтығы сияқты әр түрлі оқыту түрлерін қолданады. Бұл жұмыстың мақсаты қазақ тіліндегі анафоралық қатынасты тірек векторлық машиналар көмегімен шешу жолдарын зерттеу, қазақ тілінің тілдік ерекшеліктерін жіктеу, үлкен көлемдегі деректерді масштабтау мен деректерді оңтайландыру жолдарын ұсыну болып табылады.

Зерттеу әдістемесі мен материалдары. Жұмыс мақсатына жету үшін келесі зерттеу есептері қойылып оларды шешу жолдары қарастырылады:

Үлкен деректерді масштабтау, анафораның оң және теріс мәндерін алу, жіктеу алгоритмін құру, анафоралық шешу моделін жасау.

Машиналық оқытуға негізделген анафоралық ажыратымдылықтың жіктелу алгоритмі негізінен келесі төрт бөлікті қамтиды: корпусты алдын ала өңдеу, функция векторының құрылысы, оқыту классификациясының моделі және белгісіз қатынастар классификациясы. Қолдау векторлық машиналарының машиналық оқыту алгоритмінің сипаттамаларына сәйкес біз екі қадамды арттырамыз: оқыту деректерінің масштабтау процесі және оқыту параметрлерін оңтайландыру, 1-суретте көрсетілген жалпы процесс.



Сурет 1. Анафоралық қатынастарды шешу процесі

Тірек вектор моделі (svm) - тірек векторлық машиналары (Rong-en et al., 2005:1889) максималды маржа классификаторы болып табылады. $y \in \{1, -1\}$ i белгілері бар X_i $i = 1, 2, \dots$ даналарының жиынын ескере отырып, SVM-ді оқытудағы негізгі міндет келесі квадраттық оңтайландыру есебін шешу болып табылады:

$$\min_a f(a) = \frac{1}{2} a^T Q a - e^T a \quad (1)$$

$$0 \leq a_i \leq C, i = 1, \dots, i, y^T a = 0$$

Мұндағы e -барлық бірліктердің векторы, C - барлық айнымалылардың жоғарғы шегі;

Q-симметриялық матрица L-ге $Q_Y = y_i y_j k(X_i X_j)$ және $k(X_i X_j)$ ядро функциясы болып табылады.

Тірек векторлық машиналар негізінен жіктеу алгоритмінің идеясын қолданады, ол QR(квадраттық бағдарламалау мәселесі) (Zhang et all., 2003:32) мәселесін бірнеше кіші ішкі тапсырмаларға ыдыратады, бұл ішкі есептерді итеративті түрде шешуге және QR туралы оңтайлы шешім береді. Осы жіктеу алгоритмінің идеясының арқасында жіктеу стратегиясы, конвергенция шарттары және конвергенция сияқты басқа да мәселелерді шешуде, уақытты үнемдеу және жылдамдықты арттыру үшін жіктеу алгоритм орындалады.

SVMs-тің негізгі идеясы-шешім қабылдау жазықтығына гиперплант орнату, ең үлкен маргинализацияға қол жеткізу үшін оң даналар мен теріс даналарды окшаулау.

Табиғи тілді өңдеудің көптеген мәселелерін жіктелген мәселе ретінде қарастыруға болады; Жұмыстың негізгі мәселесін шешу идеясы оны классификацияға айналдыру болып табылады. Негізгі анафоралық қатынасының оң даналары мен теріс даналарына үміткерді болжаған кезде, SVM алгоритмі оң даналарды және теріс даналарды дұрыс ажырататын жіктеу сызығын табу және жіктелген интервалды максимум ету, бұл оңтайлы жіктеу сызығы деп аталады. Көп өлшемді кеңістіктің оқу деректерінде классификация қатесін минималды етіп жасай алатын категория бетін табу керек, оны оптималды гипер жазықтық деп атайды.

Векторлық Функцияның құрылысы – бұл мақаланың негізгі идеясы – негізгі анафоралық қатынасты түрлендіру классификацияға қатысты, содан кейін машиналық оқытуды пайдалану. Табиғи тілді өңдеуде - бұл үлкен көлемдегі «шектеулерді» қамтамасыз ету арқылы ғана сапалы талдау мен мәселені оңтайлы шешуге мүмкіндік бар. Бұл жұмыста келесі категорияларды қамтитын белгілерді қарастырады: сөздік белгілер, сөздер арасындағы қашықтық, жүйелілік белгілер және т.б.

Сөздік белгілер – екі үміткердің жолдарының сәйкестік дәрежесін білдіреді; жалпы алғанда, кандидаттардың жолдарының сәйкестік дәрежесі неғұрлым жоғары болса, кандидат сызықтарының анафоралық байланысы соғұрлым жоғары болады. Ол әдетте сөздіктің нобайы туралы ақпаратты, сөздіктің ішкі жолы туралы қысқартылған ақпаратты шығарып, сөздіктің сәйкестігі туралы толық ақпаратты қамтиды. Мысалы, "Елбасы" – бұл "президент", яғни бұл екі сөз бір бірімен анафоралық қатынастарды қалыптастырады.

Сөздер арасындағы қашықтық – екі үміткердің қашықтығын қарастырады; екі үміткердің арасындағы қашықтық неғұрлым жақын болса, кандидаттар арасындағы ықтималдық соғұрлым жоғары болады. Тұтастай алғанда, екі кандидат үштен көп сөйлеммен бөлінген, бұл негізгі анафоралық қатынас мүмкіндігі өте аз болады.

Жүйелілік белгілер – көптелу, гендерлік категория, семантикалық категорияға және т.б. бөлуге болады. «Біз» және «студенттер» сияқты барлық көпше түрдегі зат есімдер, олар жеке адамға сілтеме жасай алмайды.

Қазақ тілінде басқа тілдер сияқты гендерлік категорияға бөлінбейді, сөйлемдегі сөздің жалпы мағынасына және адам аттарының ерекшеліктеріне қарай біз сілтеменің ер немесе әйел не басқада объектілерге қатысты айтылғанын біле аламыз, осындай жүйелік белгілерге сәйкес. Топтаманың алдыңғы жағындағы антецедент А үшін, ал топтаманың артқы жағындағы анафора R үшін. 1-кестеде көрсетілгендей мүмкіндік векторын ұсынамыз.

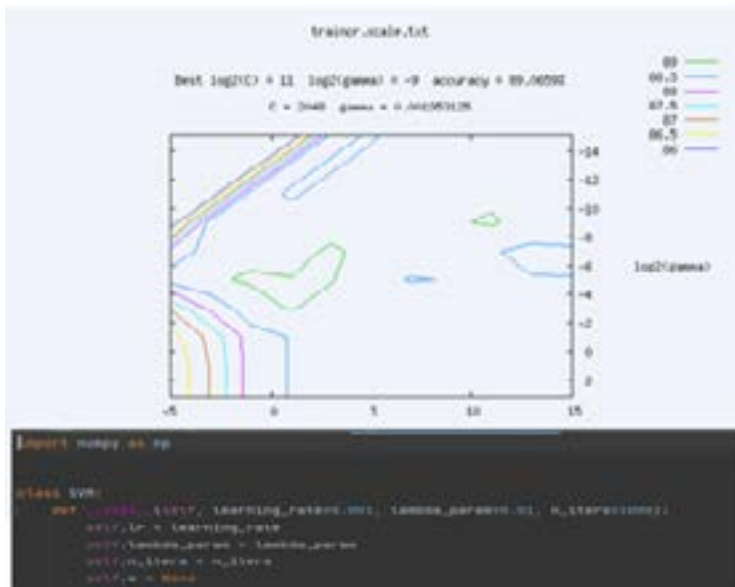
Кесте 1-анафоралық қатынасты шешудегі мүмкіндік векторлары

№	мүмкіндік векторы	Үйлесімділік функциясы
1	Синтаксистік құрлым	А және R Синтаксистік құрылымы бірдей немесе жоқ, Y иә, N жоқ
2	Өзгертілетін шектеу түрі	Өзгертілген шекті түрі А және R бірдей немесе жоқ, Y иә, N жоқ
3	Қашықтық	А және R бір сөйлемде, 1 иә, 2 интервал, 3 жоқ
4	Дара немесе көпше	А және R бірдей жекеше немесе көпше, Y иә, N әртүрлі, U жоқ
5	Гендерлік категория	А және R гендерлік категориясы бірдей Y бірдей, N әртүрлі ретінде, U жоқ
6	Түр	А және R бірдей нысан түрі .Y бірдей, N әртүрлі ретінде, U жоқ
7	Сәйкестік туралы ақпарат	А және R толық сәйкестігі. Y иә, N әртүрлі
8	Сөздердің ұқсастығы туралы ақпарат	А және R нобайы сәйкес .Y иә, N әртүрлі
9	Қысқартылған сөздер	А және R қысқартылған сөздер.Y иә, N әртүрлі

Тірек векторлары машиналардың моделіне негізделген іске асыру

1) Оқыту деректерін масштабтау. LibSVM - бұл векторлық машиналарды қолдауға арналған кітапхана, әрбір деректер данасы нақты сандар векторы ретінде ұсынылуын талап етеді. Демек, категориялық атрибуттар болса, біз оларды біріншісінде сандық деректерге түрлендіруіміз керек. m -санатының төл сипатын көрсету үшін m санын пайдалануды ұсынамыз. m санының біреуін бір, ал басқаларын нөл деп есептейміз. Мысалы, қызыл, жасыл, көк сияқты үш санатты атрибутты $(0,0,1)$, $(0,1,0)$ және $(1,0,0)$ ретінде көрсетуге болады. SVM қолданбас бұрын оларды масштабтау өте маңызды. Негізгі артықшылығы - үлкен сандық диапазондағы атрибуттардың үстемділігін болдырмау. Тағы бір артықшылығы - есептеу кезінде сандық қиындықтарды болдырмау. Себебі, ядро мәндері әдетте сызықтық ядро және көпмүшелік ядро сияқты сипаттамалық векторлардың ішкі көбейтіндісіне тәуелді болғандықтан, атрибуттардың үлкен мәндері сандық мәселелерді тудыруы мүмкін. Әр атрибутты $[-1; +1]$ немесе $[0; 1]$ ауқымына сызықтық түрде масштабтауды ұсынамыз.

2) оқыту параметрін оңтайландыру. Іздеу параметрлерінде біз Libsvm, Python, Gnuplot үш құралын қолдандық. Python gnuplot бағдарламасын орнатқаннан және орнатқаннан кейін сіз параметрлерді автоматты түрде іздеуді жүзеге асыра аласыз. 2-суретте оңтайлы параметрлерді табу үшін Python Пайдалану процесі көрсетілген.



Сурет 2. Python көмегімен параметрлерді іздеу процесі

3) модельді іске асыру. Модельді жүзеге асыру кезінде осы мақалада Тайвань ұлттық университетінің жирен Лин жасаған support vector machines toolkit құралдар жиынтығы қолданылды, содан кейін, анафоралық қатынасы деп белгіленген корпусты оқыту және анафоралық қатынасының жіктеу моделін алу үшін, біз анафоралық қатынастар жіктеушітерін құру, алдын-ала өңделген анафоралық қатынасын болжау және анафоралық қатынасын тану нәтижелерін алуға арналған оқыту моделін қолдандық.

Эксперименттік нәтижелер. Бұл зерттеуде тірек-векторлық машиналарды анафоралық қатынасты шешу үшін жіктеу ретінде пайдаланудың тиімділігін зерттеу үшін біз сәйкесінше максималды энтропия мен SVM қолданылған салыстырмалы эксперименттер жинағын жасадық. Сынақ деректері жинағы ретінде Tengrinews.kz-тен жаңалықтар топтамасы және F. Мұстафин әңгімелерінен үзінділерді қолдандық.

Біз анафоралық қатынастың шешім нәтижесін алу үшін бірнеше эксперимент жасадық:

Салыстырмалы эксперимент -мұнда максималды энтропия мен SVM ды қолдана отырып екі әдістің нәтижесі салыстырылды. Нәтижесінде SVM әдісі максималды энтропия әдісіне қарағанда жақсы нәтиже берді. Оқу деректерін масштабты өңдеу және оқыту параметрлерін оңтайландыру эксперименті- бұл эксперимент деректерді өңдеу анафоралық қатынасты алуда өте тиімді екенін көрсетті. Эксперименттік нәтижелер мен салыстырмалы деректер 2-кестеде көрсетілген нәтиже айтарлықтай жақсарғанын көрсетеді.

Кесте 2- Эксперименттік нәтижелерін салыстыру

Эксперимент	Жіктеу даналары	Дәлдік нәтижелері
максималды энтропия әдісі	Оң даналары	74%
	Теріс даналары	79%
SVM	Оң даналары	76%
	Теріс даналары	82%
Оқу деректерін масштабты өңдеу және оқыту параметрлерін оңтайландыру эксперименті	Оң даналары	80%
	Теріс даналары	83%

Кестеден эксперименттік деректерді салыстыра отырып, жіктеу үшін максималды энтропиямен жіктеу алгоритмі және тірек векторларының машиналары қолданылғанын көреміз, екі әдістің нәтижесінде айтарлықтай айырмашылықтар бар, бірақ үшінші экспериментте деректерді масштабты өңдеуді және оқыту параметрлерін оңтайландыру нәтижелерінде айқын жақсартуларға ие. бұл сонымен қатар деректерді өңдеудің оң нәтиже беретінін және өте тиімді екенін түсіндіреді. Жоғарыда келтірілген эксперименттік деректерден оң даналардың нәтижелері теріс даналардың нәтижелерінен жақсырақ екенін анық көруге болады, сонымен қатар біз статистикалық талдау арқылы теріс даналардың саны оң даналардың санынан көп екенін оңай табамыз.

Талқылау. Зерттеу жұмысы көрсеткендей, қазақ тіліндегі анафоралық қатынасты шешуде тірек векторлық машина арқылы шешу тиімділігі жоғары екендігін, машиналық оқытуда тірек векторлық машиналарды(SVM) және Libsvm кітапханасын қолданылуы табиғи тілді өңдеуде нәтижесін жақсартудағы тиімділігін көрсетті. Бұл зерттеу жұмысымызда ең басты артықшылықтар тірек векторлық машинаны(SVM) қолданылған кезінде сынақ-оқу деректерін масштабтау, оқыту параметрін оңтайландыру, сөздік белгілер, сөздер арасындағы қашықтық, жүйелілік белгілер және т.б. пайдаланылды. Оқу деректерін масштабты өңдеу және оқыту параметрлерін оңтайландыру эксперименті жасалынып максималды энтропия әдісі мен SVM әдісімен салыстырғанда әлде қайда жақсы нәтиже алынды(2-кестеде көрсетілген).

Қазақ тіліндегі есімдік анафорасын шешу бойынша жасалынған зерттеу жұмыстармен (Қалман т.б., 2022:125) салыстырғанда әлде қайда жақсы нәтижелер алынды.

Қортынды. Бұл мақалада тірек векторлық машиналар моделіне негізделген анафораны шешу модельі ұсынылды. модельді құру және эксперименттер жүргізу үшін сөздік белгілер, сөздер арасындағы қашықтықты, дәйектілікті және басқа сипаттамаларын жан-жақты қарастырады.

Модель жұмысын тестілеу мақсатында деректерді оқу жүргізілді. Сынақ деректері жинағы ретінде Tengrinews.kz- тен жаңалықтар топтамасы және F. Мұстафин әңгімелерінен үзінділерді қолдандық. оқыту параметрін оңтайландыруда Libsvm, Python, Gnuplot үш құралын қолдандық және модельді

жүзеге асыру кезінде support vector machines toolkit құралдар жиынтығы қолданылды.

Анафоралық қатынастың шешілу дәлдігін бағалау мақсатында эксперименттер жасалды және нәтижесін талдау мақсатында кесте түрінде алынды (кесте 2). Таңдалған белгілер екілік жіктеу құрылымына шешуші әсер ететіндіктен, біз зерттеу барысында екілік жіктеуге байланысты көптеген мәселелердің бар екенін ескердік.

Эксперименттер нәтижесінде анафоралық қатынасқа деректерді өңдеу және масштабтау тиімді болғанын және тірек векторлық машиналар арқылы анафоралық қатынасты шешу жақсы нәтиже бергенін көреміз.

Information about the authors:

Kalman Gulzhamal – graduate student at the Department of Information Systems, L.N. Gumilyov Eurasian National University, E-mail: guljamal14@gmail.com, DOI:<https://orcid.org/0000-0001-8863-9447>;

Sambetbayeva Madina Aralbayevna – Ph.D., Associate Professor, Nur-Sultan, Kazakhstan, Department of Information Systems, Eurasian National University named after L.N. Gumilyov. Senior Researcher, Institute of Information and Computing Technologies, Almaty, Kazakhstan, E-mail: madina_jgtu@mail.ru, <http://orcid.org/0000-0001-9358-1614>;

Aktaeva Dilara Aidosovna – Master of technical sciences, Department of Information Systems, Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan, E-mail: dilaraabilmazinova@gmail.com, <https://orcid.org/0000-0002-7307-2481>;

Ilyubayev Adelzhan – Master of technical sciences, Kokshetau, Kazakhstan department of information systems and computer science, A. Myrzakhmetov Kokshetau university. Tel. +7776104747. E-mail: adik_better@mail.ru , DOI: <https://orcid.org/0000-0001-6333-0718>.

ӘДЕБИЕТТЕР:

Абрамов В.Е., Абрамова Н.Н., Некрасова Е.В., Росс Г.Н. (2011) Статистический анализ связности текстов по общественно-политической тематике // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: тр. 13 Всерос. науч. конф. Воронеж, с 127–133. (орыс тілінде).

Қалман Г., Самбетбаева М.А., Жұмабай Е.С. (2022) Қазақ тіліндегі есімдік анафорасын шешу алгоритімі, Инновациялық Еуразия университетінің Хабаршысы. № 2 ISSN 2709-3077, 126 б. (қазақ тілінде).

Толпегин П.В. (2006) Новые методы и алгоритмы автоматического разрешения референции местоимений третьего лица русскоязычных текстов. Ком-Книга, – 88 с. (орыс тілінде).

Толпегин П.В., Ветров Д.П., Кропотов Д.А. (2006) Алгоритм автоматизированного разрешения анафоры местоимений третьего лица на основе методов машинного обучения. Международной конференции «Диалог 2006» – Компьютерная лингвистика и интеллектуальные технологии – Бекасово, с 504–507. (орыс тілінде).

Chang C.C., Lin C.J. (2014). LIBSVM-ALibraryforSupport Vector Machines, availableat [Электронныйресурс]. – Режимдоступа: www.csie.ntu.edu.tw/~cjlin/libsvm.

Elango P. (2006) Coreference Resolution: A Survey. Technical Report. UW-Madison. (ағылшын тілінде).

Jun Lang, Bing Qin, Ting Liu, Sheng Li (2007). Intra-document Coreference Resolution: The state of the art. *Journal of Chinese Language and Computing*, pp. 227-253. (ағылшын тілінде).

Kibrik A.A. (1996), Anaphora in Russian narrative discourse: A cognitive calculative account in B, Fox (ed.) *Studies in anaphora*, Amsterdam, pp. 255–304. (ағылшын тілінде).

Kibrik A.A., Dobrov G.B., Khudyakova M.V., Loukachevitch N.V., Pechenyj A. (2013), A corpus-based study of referential choice: Multiplicity of factors and machine learning techniques, *Text processing and cognitive technologies. Cognitive modeling in linguistics: Proceedings of the 13th International Conference, Corfu*, pp. 118–126. (ағылшын тілінде).

Mitkov R. (1999) Anaphora resolution: the state of the art. *Citeseer*. (ағылшын тілінде).

Mitkov R. (2003) Anaphora resolution. *The Oxford handbook of computational linguistics*, ch.14, N.Y.: Oxford university. (ағылшын тілінде).

Prokofyev R., Tonon A., Luggen M., Vouilloz L., Difallah D.E., Cudr'e-Mauroux P. (2015) SANAPHOR: Ontology-Based Coreference Resolution. *14th International Semantic Web Conference, part I, LNCS, vol. 9366*, pp. 458-473. (ағылшын тілінде).

Qiu Rong sheng, Dong Yunjie. (2003) Research developments of decomposition algorithm for SVM QP problem[J]. *Pattern Recognition and Artificial Intelligence*. vol.16, pp 63-69. (ағылшын тілінде).

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin. (2005) Working Set Selection Using Second Order Information for Training Support Vector Machines[J]. *Journal of Machine Learning Research* pp 889-1918. (ағылшын тілінде).

Zhou, H., Li, Y., Huang, D., Zhang, Y., Wu, C., Yang, Y. (2011). *Combining syntactic. Weka 3: Data Mining Software in Java*, available at University of Waikato. Retrieved from: www.cs.waikato.ac.nz/ml/weka/ 20. (ағылшын тілінде).

REFERENCES:

Abramova N.N., Abramov V.E., Nekrasova E.V., Ross G.N. (2011), Statistic analysis of social and political texts coherence [Statisticheskij analiz svjaznosti tekstov po obshchestvenno-politicheskoy tematike], *Proceedings of the 13th All-Russian Scientific Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections” [Trudy 13j Vserossijskoj nauchnoj konferentsii “Elektronnyye biblioteki: perspektivnye metody i tehnologii, elektronnyye kollekcii”]*, Voronezh, pp. 127–133. [in Russian].

G. Kalman, M.A. Sambetbayeva, Y.S. Zhumabay (2022) Algorithm for solving the anaphora of a pronoun in the Kazakh language, *bulletin of the innovative university of Eurasia*, № 2 ISSN 2709-3077, pp 126 (in Kazakh).

Tolpegin P.V. (2006), The new methods and algorithms of automated third person pronominal reference resolution of Russian texts, [Novye metody i algoritmy avtomaticheskogo razresheniya referentsii mestoimenij tret'ego litsa russkojazychnyh tekstov], *Komkniga*, p-88. [in Russian].

Tolpegin P.V., Vetrov D.P., Kropotov D.A. (2006), Automated third person anaphora resolution algorithm based on machine learning methods [Algoritm avtomatizirovannogo razresheniya anafory mestoimenij tret'ego litsa na osnove metodov mashinnogo obuchenija], *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2006”*, Bekasovo, pp. 504–507. [in Russian].

Elango P. (2006) Coreference Resolution: A Survey. Technical Report. UW-Madison. (in Eng.).

Jun Lang, Bing Qin, Ting Liu, Sheng Li. Intra-document Coreference Resolution: The state of the art [J]. *Journal of Chinese Language and Computing*, 2007, 17(4):227-253. (in Eng.).

Kibrik A.A. (1996), Anaphora in Russian narrative discourse: A cognitive calculative account in B, Fox (ed.) *Studies in anaphora*, Amsterdam, pp. 255–304. (in Eng.).

Kibrik A.A., Dobrov G.B., Khudyakova M.V., Loukachevitch N.V., Pechenyj A. (2013), A corpus-based study of referential choice: Multiplicity of factors and machine learning techniques, Text processing and cognitive technologies. Cognitive modeling in linguistics: Proceedings of the 13th International Conference, Corfu, pp. 118–126. (in Eng.).

Mitkov R. (2003) Anaphora resolution. The Oxford handbook of computational linguistics, ch.14, N.Y.: Oxford university. (in Eng.).

Prokofyev R., Tonon A., Luggen M., Vouilloz L., Difallah D.E., Cudr'e-Mauroux P.(2015) SANAPHOR: Ontology-Based Coreference Resolution.14th International Semantic Web Conference, part I, LNCS, vol. 9366, pp. 458-473. (in Eng.).

Qiu Rong sheng, Dong Yunjie. Research developments of decomposition algorithm for SVM QP problem[J]. Pattern Recognition and Artificial Intelligence.vol.16, March.2003.63-69. (in Eng.).

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin. Working Set Selection Using Second Order Information for Training Support Vector Machines[J]. Journal of Machine Learning Research 6(2005)1889- 1918. (in Eng.)

Zhou, H., Li, Y., Huang, D., Zhang, Y., Wu, C., Yang, Y. (2011). Combining syntactic. Weka 3: Data Mining Software in Java, available at University of Waikato. Retrieved from: www.cs.waikato.ac.nz/ml/weka/ 20.

МАЗМҰНЫ

А.С. Баймаханова, А.Ж. Сейтмуратов DEEP LEARNING АЛГОРИТМІН ҚОЛДАНУ НЕГІЗІНДЕ ЦИФРЛЫҚ ҚҰЖАТТАРДЫ ЖІКТЕУ.....	5
М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайұлы ВЕБ-РЕСУРСТАРДАҒЫ ФИШИНГТІК ХАБАРЛАМАЛАР ЖӘНЕ ОЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ.....	16
М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова ЭВАКУАЦИЯ ДИНАМИКАСЫНА АРНАЛҒАН FLOOR FIELD МОДЕЛІ...30	30
А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько ДЕРЕКТЕРДІ ИНТЕЛЕКТУАЛДЫ ТАЛДАУ АРҚЫЛЫ ЭПИДЕМИОЛОГИЯЛЫҚ ЖАҒДАЙДЫ БАҚЫЛАУ ЖӘНЕ МОДЕЛЬДЕУ.....	43
Г. Қалман, М.А. Самбетбаева, Д.А. Ақтаева, А.С. Илюбаев МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН АНАФОРАНЫ ШЕШУ МОДЕЛІ.....	56
С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова АҚПАРАТТЫҚ ЖҮЙЕНІҢ ҚАУІПТЕРІ МЕН ОСАЛ ТҰСТАРЫНЫҢ МОНИТОРИНГІ ТУРАЛЫ.....	68
У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков БҰЛДЫР ЖИЫНДАР ТЕОРИЯСЫ НЕГІЗІНДЕ БИЗНЕСТІҢ НЕСИЕ ҚАБІЛЕТІЛІГІН БАҒАЛАУ АЛГОРИТМІ.....	81
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова СУ ЭКО ЖҮЙЕЛЕРІН ЗЕРТТЕУДЕ АТЖ ӨЗІРЛЕУГЕ ҚОЙЫЛАТЫН ТАЛАПТАР.....	93
А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ТАЛДАУДА БҰЛДЫР ЛОГИКАНЫ ҚОЛДАНУ.....	106

Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов ШУ КЕДЕЛДЕРІН БҮРКЕУДІҢ САПАСЫН БАҒАЛАУ ӘДІСТЕМЕСІ.....	120
А.Ш. Хасенов, М.А. Кантурсева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова АГЕНТТІК МОДЕЛЬДЕУ ЖҮЙЕСІНДЕ ЭВАКУАЦИЯ МОДЕЛІН ЖҮЗЕГЕ АСЫРУ ТӘСІЛІ.....	134
А. Шаушенова, А. Нурпейсова, Д. Досалянов, Г. Мауина ПРОКТОРИНГ ЖҮЙЕСІНДЕ ЖАСАНДЫ НЕЙРОНДЫҚ ЖЕЛІЛЕРГЕ НЕГІЗДЕЛГЕН СӨЙЛЕУДІ ТАҢУ МӘСЕЛЕЛЕРІ.....	146
А.Ә. Шекербек, Г.Б. Абдикеримова, Ә.М. Сабыр, Ж.С. Әбілқайыр КЕУДЕ КЛЕТКАСЫНЫҢ ПАТОЛОГИЯСЫН АНЫҚТАУ ҮШІН ӘДІС ПЕН АЛГОРИТМДІ ҚОЛДАНУ.....	159

СОДЕРЖАНИЕ

А.С. Баймаханова, А.Ж. Сейтмуратов КЛАССИФИКАЦИЯ ЦИФРОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА DEEP LEARNING.....	5
М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайулы ФИШИНГОВЫЕ СООБЩЕНИЯ НА ВЕБ-РЕСУРСАХ И ИХ ОПРЕДЕЛЕНИЕ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ.....	16
М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова FLOOR FIELD МОДЕЛЬ ДЛЯ ДИНАМИКИ ЭВАКУАЦИИ.....	30
А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько МОНИТОРИНГ И МОДЕЛИРОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ С ПОМОЩЬЮ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ.....	43
Г. Қалман, М.А. Самбетбаева, Д.А. Актаева, А.С. Илюбаев МОДЕЛЬ РАЗРЕШЕНИЯ АНАФОРЫ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	56
С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова О МОНИТОРИНГЕ УГРОЗ И УЯЗВИМОСТЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ.....	68
У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков АЛГОРИТМ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ БИЗНЕСА НА ОСНОВЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ.....	81
Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова ТРЕБОВАНИЯ К РАЗРАБОТКЕ ИАС-ИССЛЕДОВАНИЙ ВОДНЫХ ЭКОСИСТЕМ.....	93
А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ПРИ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ.....	106

Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов МЕТОДИКА ОЦЕНКИ КАЧЕСТВА МАСКИРУЮЩИХ ШУМОВЫХ ПОМЕХ.....	120
А.Ш. Хасенов, М.А. Кантуреева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова ПОДХОД К РЕАЛИЗАЦИИ МОДЕЛИ ЭВАКУАЦИИ В СИСТЕМЕ АГЕНТНОГО МОДЕЛИРОВАНИЯ.....	134
А.Г. Шаушенова, А.А. Нурпейсова, Д.Б. Досалянов, Г.М. Мауина ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМЕ ПРОКТОРИНГА.....	146
А.А. Шекербек, Г.Б. Абдикеримова, А.М. Сабыр, Ж.С. Абулхаир ПРИМЕНЕНИЕ МЕТОДА И АЛГОРИТМА ДЛЯ ВЫЯВЛЕНИЯ ПАТОЛОГИИ ГРУДНОЙ КЛЕТКИ.....	159

CONTENTS

A. Baimakhanova, A. Seitmuratov CLASSIFICATION OF DIGITAL DOCUMENTS USING DEEP LEARNING ALGORITHM.....	5
M. Bolatbek, Sh. Musiralieva, K Bagitova, A. Нюсупов, E. Abaiuly PHISHING MESSAGES ON WEB RESOURCES AND THEIR DETECTION BY MACHINE LEARNING METHODS.....	16
M. Kantureyeva, A. Khassenov, D. Tussupov, A. Zakirova, A. Alimagambetova FLOOR FIELD MODEL FOR EVACUATION DYNAMICS.....	30
A.D. Kubegenova, K.T. Iskakov, E.S. Kubegenov, O.I. Krivorotko MONITORING AND MODELING OF THE EPIDEMIOLOGICAL SITUATION USING DATA MINING.....	43
G. Kalman, M.A. Sambetbayeva, A.C. Ilyubayev, D.A. Aktaeva ANAPHORA RESOLUTION MODEL BASED ON MACHINE LEARNING METHODS.....	56
S.T. Mambetov, Ye.Ye. Begimbayeva, S. Joldasbayev, B.O. Kulambayev, G.N. Kazbekova ABOUT MONITORING THREATS AND VULNERABILITIES OF THE INFORMATION SYSTEM.....	68
U. Makhazhanova, B. Tassuov, A. Mukhanova, A. Mukhiyadin, R. Zetkinshekov AN ALGORITHM FOR ASSESSING THE CREDITWORTHINESS OF A BUSINESS BASED ON THE THEORY OF FUZZY SETS.....	81
R.M. Moldasheva, A.A. Ismailova, A.K. Zhamangara, A.M. Zadagali, G.B. Turmukhanova REQUIREMENTS TO DEVELOPMENT OF IAS FOR RESEARCH OF AQUEOUS ECOSYSTEMS.....	93
A. Mukhanova, U. Makhazhanova, N. Markhabatov, B. Tassuov, Zh. Lamasheva APPLICATION OF FUZZY LOGIC IN THE ANALYSIS OF ECONOMIC SYSTEMS N.....	106

N.A. Seilova, A. Batyrgaliyev, Zh. Dzhangozin, D. Baibatchayeva, N. Nurgabylov METHOD FOR ASSESSING THE QUALITY OF MASKING NOISE INTERFERENCES.....	120
A. Khassenov, M. Kantureyeva, D. Tussupov, A. Omarbekova, G. Abdikerimova APPROACH TO THE IMPLEMENTATION OF EVACUATION MODEL IN THE AGENT-BASED MODELING SYSTEM.....	134
A.G. Shaushenova, A.A. Nurpeisova, D.B. Dosalyanov, G.M. Mauina PROBLEMS OF SPEECH RECOGNITION BASED ON ARTIFICIAL NEURAL NETWORKS IN THE PROCTORING SYSTEM.....	146
A. Shekerbek, G. Abdikerimova, A. Sabyr, Zh. Abilkaiyr APPLICATION OF THE METHOD AND ALGORITHM FOR THE DETECTION OF CHEST PATHOLOGY.....	159

**Publication Ethics and Publication Malpractice
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Заместитель директор отдела издания научных журналов НАН РК *Р. Жәліқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 15.09.2022.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

10,5 п.л. Тираж 300. Заказ 4.