

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



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ
«ХАЛЫҚ» ЖҚ

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ
АКАДЕМИИ НАУК РЕСПУБЛИКИ
КАЗАХСТАН»
ЧФ «Халық»

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
«Halyk» Private Foundation

SERIES
PHYSICS AND INFORMATION TECHNOLOGY

4 (348)

OCTOBER – DECEMBER 2023

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK



ЧФ «ХАЛЫҚ»

В 2016 году для развития и улучшения качества жизни казахстанцев был создан частный Благотворительный фонд «Халық». За годы своей деятельности на реализацию благотворительных проектов в областях образования и науки, социальной защиты, культуры, здравоохранения и спорта, Фонд выделил более 45 миллиардов тенге.

Особое внимание Благотворительный фонд «Халық» уделяет образовательным программам, считая это направление одним из ключевых в своей деятельности. Оказывая поддержку отечественному образованию, Фонд вносит свой посильный вклад в развитие качественного образования в Казахстане. Тем самым способствуя росту числа людей, способных менять жизнь в стране к лучшему – профессионалов в различных сферах, потенциальных лидеров и «великих умов». Одной из значимых инициатив фонда «Халық» в образовательной сфере стал проект *Ozgeris powered by Halyk Fund* – первый в стране бизнес-инкубатор для учащихся 9-11 классов, который помогает развивать необходимые в современном мире предпринимательские навыки. Так, на содействие малому бизнесу школьников было выделено более 200 грантов. Для поддержки талантливых и мотивированных детей Фонд неоднократно выделял гранты на обучение в Международной школе «Мирас» и в *Astana IT University*, а также помог казахстанским школьникам принять участие в престижном конкурсе «*USTEM Robotics*» в США. Авторские работы в рамках проекта «Тәлімгер», которому Фонд оказал поддержку, легли в основу учебной программы, учебников и учебно-методических книг по предмету «Основы предпринимательства и бизнеса», преподаваемого в 10-11 классах казахстанских школ и колледжей.

Помимо помощи школьникам, учащимся колледжей и студентам Фонд считает важным внести свой вклад в повышение квалификации педагогов, совершенствование их знаний и навыков, поскольку именно они являются проводниками знаний будущих поколений казахстанцев. При поддержке Фонда «Халық» в южной столице был организован ежегодный городской конкурс педагогов «*Almaty Digital Ustaz*».

Важной инициативой стал реализуемый проект по обучению основам финансовой грамотности преподавателей из восьми областей Казахстана, что должно оказать существенное влияние на воспитание финансовой грамотности и предпринимательского мышления у нового поколения граждан страны.

Необходимую помощь Фонд «Халык» оказывает и тем, кто особенно остро в ней нуждается. В рамках социальной защиты населения активно проводится работа по поддержке детей, оставшихся без родителей, детей и взрослых из социально уязвимых слоев населения, людей с ограниченными возможностями, а также обеспечению нуждающихся социальным жильем, строительству социально важных объектов, таких как детские сады, детские площадки и физкультурно-оздоровительные комплексы.

В копилку добрых дел Фонда «Халык» можно добавить оказание помощи детскому спорту, куда относится поддержка в развитии детского футбола и карате в нашей стране. Жизненно важную помощь Благотворительный фонд «Халык» оказал нашим соотечественникам во время недавней пандемии COVID-19. Тогда, в разгар тяжелой борьбы с коронавирусной инфекцией Фонд выделил свыше 11 миллиардов тенге на приобретение необходимого медицинского оборудования и дорогостоящих медицинских препаратов, автомобилей скорой медицинской помощи и средств защиты, адресную материальную помощь социально уязвимым слоям населения и денежные выплаты медицинским работникам.

В 2023 году наряду с другими проектами, нацеленными на повышение благосостояния казахстанских граждан Фонд решил уделить особое внимание науке, поскольку она является частью общественной культуры, а уровень ее развития определяет уровень развития государства.

Поддержка Фондом выпуска журналов Национальной Академии наук Республики Казахстан, которые входят в международные фонды Scopus и Wos и в которых публикуются статьи отечественных ученых, докторантов и магистрантов, а также научных сотрудников высших учебных заведений и научно-исследовательских институтов нашей страны является не менее значимым вкладом Фонда в развитие казахстанского общества.

**С уважением,
Благотворительный Фонд «Халык»!**

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н=5**

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

МАМЫРБАЕВ Өркен Жұмажанұлы, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

РЕДАКЦИЯ АЛҚАСЫ:

ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

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

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

БОШКАЕВ Қуантай Авғазыұлы, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

QUEVEDO Nemando, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

ЖҮСІПОВ Марат Абжанұлы, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

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

РАМАЗАНОВ Тілекқабұл Сәбитұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

ТАКИБАЕВ Нұрғали Жабағаұлы, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

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

ХАРИН Станислав Николаевич, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

ДАВЛЕТОВ Асқар Ербуланович, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

КАЛАНДРА Пьетро, Ph.D (физика), Нанокұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

«ҚР ҰҒА Хабарлары. Физика және информатика сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (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 және ақпараттық коммуникациялық технологиялар 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/>

ГЛАВНЫЙ РЕДАКТОР:

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

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

МАМЫРБАЕВ Оркен Жумажанович, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **Н=5**

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

КАЛИМОЛДАЕВ Максат Нурадилович, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **Н=7**

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

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

БОШКАЕВ Куантай Авгазыевич, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=10**

QUEVEDO Hemando, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **Н=28**

ЖУСУПОВ Марат Абжанович, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=7**

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

РАМАЗАНОВ Тлексабул Сабитович, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=26**

ТАКИБАЕВ Нурғали Жабағевич, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=5**

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

ХАРИН Станислав Николаевич, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **Н=10**

ДАВЛЕТОВ Аскар Ербуланович, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=12**

КАЛАНДРА Пьетро, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **Н=26**

«Известия НАН РК. Серия физика и информатики».

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 и информационные коммуникационные технологии.* 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/>

EDITOR IN CHIEF:

MUTANOV Galimkair Mutanovich, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

DEPUTY EDITOR-IN-CHIEF

MAMYRBAYEV Orken Zhumazhanovich, Ph.D. in the specialty "Information systems, executive secretary of the RSE "Institute of Information and Computational Technologies", Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

BAYGUNCHEKOV Zhumadil Zhanabayevich, 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 Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

BOSHKAYEV Kuantai Avgazievich, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

QUEVEDO Hemando, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

ZHUSSUPOV Marat Abzhanovich, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

KOVALEV Alexander Mikhailovich, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

RAMAZANOV Tlekkabul Sabitovich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

TAKIBAYEV Nurgali Zhabagaevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

TIGHINEANU Ion Mikhailovich, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

KHARIN Stanislav Nikolayevich, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

DAVLETOV Askar Erbulanovich, Doctor in Physics and Mathematics, Professor, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

CALANDRA Pietro, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

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

Series of physics and informatics.

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-ЖК**, issued 14.02.2018
Thematic scope: *series physics and information technology.*

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/>

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 4. Number 348 (2023). 253–267

<https://doi.org/10.32014/2023.2518-1726.233>

UDC 004.9

© **A. Tanirbergenov¹, Zh. Taszhurekova², S.Serikbayeva^{1*}, A. Shorayev¹,
A. Abduvalova², 2023**

¹ Eurasian National University named after L.N. Gumilyov, Department of
Information Systems, Astana, Kazakhstan;

² Taraz Regional University named after M.KH. Dulaty, Taraz, Kazakhstan.
E-mail: inf_8585@mail.ru

METHODS OF CONSTRUCTING A MODEL AND AN INFORMATION SYSTEM FOR SEARCHING FULL-TEXT DOCUMENTS

Tanirbergenov Adilbek — associate professor of the Department of algebra and geometry, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

E-mail: t.adilbek@mail.ru, <https://orcid.org/0009-0000-8401-5434>;

Taszhurekova Zhazira Kudaibergenovna — acting associate professor of the Department «applied informatics and programming» Taraz regional university named after M.KH. Dulaty, Taraz, Kazakhstan

E-mail: taszhurekova@mail.ru, <https://orcid.org/0000-0002-8307-9417>;

Serikbayeva Sandugash — PhD, Senior Lecturer of the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

E-mail: inf_8585@mail.ru, <https://orcid.org/0000-0003-1729-6875>;

Shorayev Abylaikhan — Master's degree in Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

E-mail: a_khan02@mail.ru, <https://orcid.org/0009-0001-7190-6016>;

Abduvalova Ainur — acting associate professor of the Department «Information Systems», Taraz regional university named after M.KH. Dulaty, Taraz, Kazakhstan.

E-mail: abduvalova_ad@mail.ru, <https://orcid.org/0000-0002-4683-7821>.

Abstract. The article discusses the methods of constructing a model and an information system for searching full-text documents. The article presents various approaches to building a search model for full-text documents. One approach is based on the use of a document index, which is a data structure that stores information about documents and their contents. Another approach is based on the use of probabilistic models that allow you to evaluate the relevance of documents to a query. The search for full-text documents is an urgent task, because it allows users to quickly and efficiently find the necessary information. The focus is on developing systems capable of processing large amounts of data and providing accurate search results. The article also discusses the problems associated with current search systems and offers practical solutions to overcome them. This

research represents an important contribution to the field of information retrieval and may be useful for specialists in this field. It discusses methods of information retrieval and text processing in natural language, as well as models of information retrieval. Particular attention is paid to the use of natural language text processing in the context of building an information system for searching full-text documents. A brief overview of natural language text processing methods and their use in search models is given. The main models of information search, existing semantic search systems are considered.

Keywords: Full-text search, information system, search model, text analysis, document search methods, information search

Financing: There is no source of funding for this study.

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

© **А.Ж. Танирбергенов¹, Ж.К. Тасжурекова², С.К. Серикбаева^{1*},
А.А. Шораев¹, А.Д. Абдувалова², 2023**

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

²М.Х. Дулати атындағы Тараз өңірлік университеті, Тараз, Қазақстан.

E-mail: inf_8585@mail.ru

ТОЛЫҚ МӘТІНДІ ҚҰЖАТТАРДЫ ІЗДЕУДІҢ МОДЕЛІ МЕН АҚПАРАТТЫҚ ЖҮЙЕСІН ҚҰРУ ӘДІСТЕРІ

Танирбергенов Адилбек Жұматаевич – Л.Н. Гумилев атындағы Еуразия ұлттық университетінің алгебра және геометрия кафедрасының доцент м.а., Астана, Қазақстан

E-mail: t.adilbek@mail.ru, <https://orcid.org/0009-0000-8401-5434>;

Тасжурекова Жазира Қудайбергеновна — М.Х. Дулати атындағы Тараз өңірлік университеті, «Қолданбалы информатика және бағдарламалау» кафедрасының доцент м.а., т.ғ.к., Тараз, Қазақстан

E-mail: tashjurekova@mail.ru, <https://orcid.org/0000-0002-8307-9417>;

Серикбаева Сандугаш Курманбековна — Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Ақпараттық жүйелер кафедрасының аға оқытушысы, PhD, Астана, Қазақстан

E-mail: inf_8585@mail.ru, <https://orcid.org/0000-0003-1729-6875>;

Шораев Абылайхан Ануарбекұлы — Л.Н. Гумилев атындағы Еуразия ұлттық университетінің «Ақпараттық жүйелер» мамандығының магистранты, Астана, Қазақстан

E-mail: a_khan02@mail.ru, <https://orcid.org/0009-0001-7190-6016>;

Абдувалова Айнур Джумабевна — М.Х. Дулати атындағы Тараз өңірлік университеті, «Ақпараттық жүйелер» кафедрасының доцент м.а., т.ғ.к., Тараз, Қазақстан.

E-mail: abduvalova_ad@mail.ru, <https://orcid.org/0000-0002-4683-7821>.

Аннотация. Мақалада толық мәтінді құжаттарды іздеудің моделі мен ақпараттық жүйесін құру әдістері қарастырылады. Мақалада толық мәтінді құжаттарды іздеу моделін құрудың әртүрлі тәсілдері келтірілген. Бір тәсіл құжаттар индексін қолдануға негізделген, бұл құжаттар мен олардың мазмұны туралы ақпаратты сақтайтын мәліметтер құрылымы. Тағы бір тәсіл құжаттардың сұранысқа сәйкестігін бағалауға мүмкіндік беретін ықтималдық модельдерін қолдануға негізделген. Толық мәтінді құжаттарды іздеу өзекті

мәселе болып табылады, өйткені ол пайдаланушыларға қажетті ақпаратты тез және тиімді табуға мүмкіндік береді. Үлкен көлемдегі деректерді өңдеуге және нақты іздеу нәтижелерін қамтамасыз етуге қабілетті жүйелерді дамытуға баса назар аударылады. Мақала сонымен қатар қазіргі іздеу жүйелеріне қатысты мәселелерді талқылайды және оларды жеңудің практикалық шешімдерін ұсынады. Бұл зерттеу ақпараттық іздеу саласына маңызды үлес қосады және осы саладағы мамандар үшін пайдалы болуы мүмкін. Онда табиғи тілдегі мәтіндерді ақпараттық іздеу және өңдеу әдістері, сондай-ақ ақпараттық іздеу модельдері қарастырылады. Толық мәтінді құжаттарды іздеу үшін ақпараттық жүйені құру контекстінде табиғи тілде мәтінді өңдеуді қолдануға ерекше назар аударылады. Табиғи тілдегі мәтіндерді өңдеу әдістеріне және оларды іздеу модельдерінде қолдануға қысқаша шолу берілген. Ақпаратты іздеудің негізгі модельдері, қолданыстағы семантикалық іздеу жүйелері қарастырылады.

Түйін сөздер: Толық мәтінді іздеу, ақпараттық жүйе, іздеу моделі, мәтіндік талдау, құжаттарды іздеу әдістері, ақпараттық іздеу

Қаржыландыру: Бұл зерттеу жұмысын қаржыландыру көзі жоқ.

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

© А.Ж. Танирбергенов¹, Ж.К. Тасжурекова², С.К. Серикбаева^{1*},
А.А. Шораев¹, А.Д. Абдувалова², 2023

¹Евразийский национальный университет имени Л.Н. Гумилева,
Астана, Казахстан;

²Таразский региональный университет им. М.Х. Дулати, Тараз, Казахстан.
E-mail: inf_8585@mail.ru

МЕТОДЫ ПОСТРОЕНИЯ МОДЕЛИ И ИНФОРМАЦИОННОЙ СИСТЕМЫ ПОИСКА ПОЛНОТЕКСТОВЫХ ДОКУМЕНТОВ

Танирбергенов Адилбек Жуматаевич — и.о. доцента кафедры алгебры и геометрии, Евразийского национального университета им. Л.Н. Гумилева, Астана, Казахстан
E-mail: t.adilbek@mail.ru, <https://orcid.org/0009-0000-8401-5434>;

Тасжурекова Жазира Кудайбергеновна — к.т.н., и.о. доцента кафедры «Прикладная информатика и программирование» Таразского регионального университета им. М.Х. Дулати, Тараз, Казахстан
E-mail: tashjurekova@mail.ru, <https://orcid.org/0000-0002-8307-9417>;

Серикбаева Сандугаш Курманбековна — PhD, старший преподаватель кафедры информационных систем Евразийского национального университета им. Л.Н. Гумилева, Астана, Казахстан
E-mail: inf_8585@mail.ru, <https://orcid.org/0000-0003-1729-6875>;

Шораев Абылайхан Ануарбекович — магистрант специальности «Информационные системы», Евразийского национального университета им. Л.Н. Гумилева, Астана, Казахстан
E-mail: a_khan02@mail.ru, <https://orcid.org/0009-0001-7190-6016>;

Абдувалова Айнура Джумабевна — к.т.н., и.о. доцента кафедры «Информационные системы», Таразский региональный университет им. М.Х. Дулати, Тараз, Казахстан
E-mail: abduvalova_ad@mail.ru, <https://orcid.org/0000-0002-4683-7821>.

Аннотация. В статье рассматриваются методы построения модели и информационной системы поиска полнотекстовых документов. Авторы представили различные подходы к построению модели поиска полнотекстовых документов. Один подход основан на использовании индекса документов, который представляет собой структуру данных, хранящую информацию о документах и их содержании. Другой подход основан на использовании вероятностных моделей, которые позволяют оценивать релевантность документов запросу. Поиск полнотекстовых документов является актуальной задачей, поскольку позволяет пользователям быстро и эффективно находить необходимую информацию. Основное внимание уделяется разработке систем, способных обрабатывать большие объемы данных и предоставлять точные результаты поиска. В статье также обсуждаются проблемы, связанные с текущими системами поиска, и предлагаются практические решения их преодоления. Это исследование представляет собой важный вклад в область информационного поиска и может быть полезно для специалистов в этой области. В нем рассматриваются методы информационного поиска и обработки текстов на естественном языке, а также модели информационного поиска. Особое внимание уделяется использованию обработки текста на естественном языке в контексте построения информационной системы для поиска полнотекстовых документов. Приведен краткий обзор методов обработки текстов естественного языка и их использование в моделях поиска. Рассмотрены основные модели поиска информации, существующие системы семантического поиска.

Ключевые слова: полнотекстовый поиск, информационная система, модель поиска, текстовый анализ, методы поиска документов, информационный поиск

Финансирование: данная исследовательская работа не имеет источников финансирования.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе

Соңғы онжылдықтар ішінде байқалған электрондық құжаттарды сақтау көлемінің қарқынды өсуіне байланысты ақпаратты іздеудің бағдарламалық құралдарын әзірлеу ерекше маңызға ие. Электрондық құжаттардың бір түрі-табиғи тілдегі мәтіндерді немесе толық мәтінді құжаттарды қамтитын құжаттар. Пайдаланушыға іздеу, алу және жеткізу үшін телекоммуникациялық қол жетімділік құралдары арқылы қол жетімді көптеген толық мәтінді құжаттар толық мәтінді құжаттар жинағы деп аталады. Толық мәтінді құжаттар жинағының ерекше жағдайы толық мәтінді электрондық кітапхана болып табылады, оның құжаттары дұрыс библиографиялық сипаттамамен жабдықталған. Ғылыми және техникалық бағыттағы, сондай-ақ ғаламдық

Интернет желісінде еркін қол жетімді толық мәтінді құжаттар топтамаларының мысалдарын келтірейік:

а) толық мәтінді электрондық кітапханалар «ғылыми электрондық кітапхана «сияқты eLIBRARY.RU», «Ашық орыс электронды кітапханасы» және т.б. электрондық кітапханалар көп жағдайда оқырмандарға қызмет көрсетудің заманауи талаптарын қанағаттандыруға ұмтылатын дәстүрлі кітапханалар қорының бірі болып табылады. Осылайша, «ашық орыс электронды кітапханасы» Ресей мемлекеттік кітапханасының электрондық құжаттар қоры ретінде пайда болды.

б) оқырмандарға жарияланған мақалалардың толық мәтініне қол жеткізуге мүмкіндік беретін ғылыми және техникалық журналдар «ғылым әлемінде» және «ашық жүйелер» сияқты.

в) бір тақырыптық бағытқа сәйкес келетін, орыс тілді мақалалар, кітаптар, citforum Ақпараттық технологиялар жөніндегі нұсқаулықтар жинағы және «ғылым және техника» электрондық кітапханасы сияқты білім беру мақсатымен біріктірілген аналитикалық, шолу немесе жаңалықтар сипатындағы техникалық материалдардың жалпыға қолжетімді жинақтары.

Толық мәтінді құжаттар топтамаларының басты әлеуетті артықшылығы пайдаланушыларға заманауи іздеу мүмкіндіктерін ұсыну болып табылады. Іздеу мүмкіндіктерін іске асырудың негізгі тетіктері:

а) пайдаланушының сұранысы бойынша ақпараттық іздеу (кілт сөздерді іздеу);

б) құжаттар жинағын жіктеу негізінде ақпараттық іздеу.

Пайдаланушының сұранысы бойынша ақпараттық іздеу, пайдаланудың қарапайымдылығына байланысты, көптеген құжаттар жинағында қолданылады. Алайда, бұл механизмнің бірқатар кемшіліктері бар, біріншіден, оқырманның кілт сөздері бойынша құжаттарды іздеуде қиындықтардың туындауымен, қажетті пәндік саламен таныс емес немесе іздеу машиналарын пайдалану мәселелерінде тәжірибесі аз. Екіншіден, сұрауға жауап ретінде іздеу жүйесі тапқан құжаттардың үлкен тізімін қарау арқылы пайдаланушыны қызықтыратын құжаттарды таңдауда қиындықтар туындайды. Бұл мәселе көбінесе пайдаланушының тиімді іздеу сұрауларын жасай алмауынан туындайды (Осипов, 2008).

Құжаттар топтамасының жіктелуіне негізделген ақпараттық іздеуді сұраныс бойынша іздеу мәселелерін шешуде де, тәуелсіз іздеу механизмі ретінде де пайдалануға болады.

Толық мәтінді іздеудің ақпараттық жүйесі-бұл әртүрлі көздерден құжаттардың, мақалалардың, файлдардың және басқа ақпараттың толық мәтіндерін тиімді іздеуге және шығаруға арналған қуатты құрал. Бұл жүйе пайдаланушыларға қажетті ақпаратты жылдам табуға мүмкіндік береді, бұл мәтіндік деректерді іздеу мен талдау процесін жеңілдетеді. Бұл кіріспеде біз толық мәтіндерді іздеудің ақпараттық жүйесінің негізгі сипаттамалары мен мақсаттарымен танысамыз.

Толық мәтіндерді іздеудің ақпараттық жүйесінің мақсаттары мен міндеттері:

Толық мәтіндік деректерге қол жеткізуді қамтамасыз ету: мұндай жүйенің негізгі мақсаты пайдаланушыларға ғылыми мақалаларды, кітаптарды, жаңалықтар мақалаларын, құжаттаманы және басқа мәтіндік ресурстарды қамтуы мүмкін толық мәтіндік материалдарға қол жеткізуді қамтамасыз ету болып табылады.

Іздеу процесін жақсарту: жүйе әр түрлі іздеу параметрлерін, кілт сөздерді және сүзгілерді қолдана отырып, пайдаланушыға қажетті ақпаратты табуды жеңілдету арқылы іздеу тиімділігін арттыру мақсатында жасалады.

Нәтижелердің өзектілігін оңтайландыру: ақпараттық жүйе пайдаланушыларға олардың сұранысына сәйкес келетін бастапқы ақпаратты жылдам табуға мүмкіндік беретін ең маңызды іздеу нәтижелерін ұсынуға тырысады.

Әртүрлі деректер көздерін біріктіру: жүйе әртүрлі көздерден алынған ақпаратты, соның ішінде кітапхана дерекқорларын, мұрағаттарды, веб-сайттарды, электрондық журналдарды және басқа ресурстарды біріктіріп, жан-жақты іздеуді қамтамасыз ете алады.

Мәтінді өңдеу және талдау: кейбір толық мәтінді іздеу ақпараттық жүйелері мәтінді өңдеу және талдау мүмкіндіктерін қамтамасыз ете алады, мысалы, негізгі сөз тіркестерін бөлектеу, семантиканы талдау, түйіндеме жасау және басқа функциялар.

Деректерді басқару және сақтау жүйелері: мәтіндік деректердің үлкен көлемін сақтау және басқару мұндай жүйенің маңызды бөлігі болып табылады. Бұған индекстеу, каталогтау және деректердің қауіпсіздігін қамтамасыз ету кіреді (Соченков, 2013).

Пайдаланушы интерфейсін қолдау: жүйе әдетте пайдаланушыларға іздеу сұрауларын енгізуге, нәтижелерді көруге және табылған мәтіндермен өзара әрекеттесуге мүмкіндік беретін пайдаланушы интерфейсін ұсынады.

Жаңарту және сүйемелдеу: жүйе оның сенімді және өзекті жұмысын қамтамасыз ету үшін үнемі жаңартуды, техникалық қызмет көрсетуді және бақылауды қажет етеді.

Толық мәтінді іздеудің ақпараттық жүйелері зерттеушілер, кәсіпқойлар, студенттер және кең мәтіндік ақпаратқа қол жеткізуді қажет ететін кез келген адам үшін маңызды құрал болып табылады. Олар зерттеулердің, оқу процесінің және адам қызметінің көптеген басқа салаларының өнімділігі мен сапасын жақсартуға ықпал етеді.

Қолданыстағы жүйелік ақпарат пен іздеу әдістеріне шолу

Бүгінгі таңда толық мәтінді құжаттардан ақпаратты талдау мен алудың көптеген ақпараттық жүйелері мен әдістері бар. Бұл жүйелер мен әдістер ғылыми мақалаларды іздеу, жаңалықтарды талдау, медициналық құжаттарды іздеу және т.б. сияқты әртүрлі қолданбаларға арналған. Міне, олардың кейбіреулері:

1. Google және Bing сияқты іздеу жүйелері: бұл жүйелер пайдаланушының кілт сөзге негізделген анықтаманың тиісті нәтижелерін есептеу үшін қуатты индекстеу және рейтинг алгоритмдерін қолданады. Олар Интернеттен ақпарат іздеу үшін кеңінен қолданылады, бірақ олар әрқашан үлкен мәтінмен толық мәтінді құжаттарды өңдей алмайды.

2. PubMed және IEEE Xplore сияқты ғылыми іздеу жүйелері: бұл жүйелер ғылыми мақалалар мен басылымдарды іздеуге маманданған. Олар автор, журнал және кілт сөз сүзгілері сияқты кеңейтілген іздеу мүмкіндіктерін ұсынады. Алайда, олар ғылыми жарияланымдар саласымен шектеледі.

3. Мәтінді талдау және машиналық оқыту жүйелері: кейінгі зерттеулер Машиналық оқыту әдістерін қолдана отырып, мәтінді жүйелік талдауды дамытуға бағытталған. Бұл жүйелер үлгілерді тауып, мәтіндегі барлық ақпаратты анықтай алады, бірақ оқыту көбінесе белгіленген деректердің үлкен көлемін қажет етеді.

4. Бизнеске арналған ақпараттық жүйелер: бизнес саласында Корпоративтік құжаттар мен есептерді талдауға мүмкіндік беретін ақпараттық жүйелер бар. Олар стратегиялық ақпаратты табуға көмектесе алады, бірақ құпиялылық пен деректер қауіпсіздігі мәселелеріне тап болуы мүмкін.

Қолданыстағы тәсілдердің проблемалары мен шектеулерін сипаттау

Көптеген Ақпараттық жүйелер мен тексеру әдістеріне қарамастан, толық мәтінді құжаттардағы тексерудің сапасы мен тиімділігіне теріс әсер етуі мүмкін бірқатар мәселелер мен шектеулер бар:

1. Өзектілік мәселесі: көптеген іздеу жүйелері кілт сөздерге сүйенеді, бұл маңызды емес іздеу нәтижелеріне әкелуі мүмкін, әсіресе мәтіннің үлкен көлемін іздеу кезінде.

2. Қолдану аясын шектеу: жүйенің деректері қолдану саласында шектеулі және олардың тиімділігі олар талдайтын құжаттардың түрлеріне байланысты.

3. Контекстің болмауы: кейбір жүйелер мәтіннің контексті мен семантикасын ескермейді, бұл ақпаратты дұрыс түсінбеуге әкелуі мүмкін.

4. Үлкен көлемдегі деректерді өңдеудің жеткіліксіздігі: үлкен, толық мәтінді құжаттарды өңдеу айтарлықтай есептеу ресурстары мен уақытты қажет етуі мүмкін, бұл жүйелердің қолданылуын шектейді.

5. Құпиялылық және қауіпсіздік мәселелері: деректердің құпиялылығымен жұмыс істейтін бизнес және жүйелер жағдайында ақпараттың құпиялылығы мен қауіпсіздігін қамтамасыз ету қажет.

Қазіргі ақпараттық жүйелердің (АЖ) негізгі функцияларының бірі-кейбір белгілерді қанағаттандыратын деректер элементтерін іздеу (ақпараттық іздеу). Мұндай іздеуді ұйымдастыруда туындайтын әдістемелік және технологиялық проблемалардың ерекшелігі бірнеше факторларға байланысты. Ең алдымен, бұл жүйеге кіретін ақпараттық ресурстардағы мазмұнның сипаты. Қазіргі заманғы АЖ – да мәтіндік мазмұн әлі де басым, бірақ мультимедиялық мазмұнды (графика, аудио және видео ақпарат) қамтитын мультиформатты

ресурстар, сондай - ақ мазмұнды құрылымдаудың әр түрлі формалары жұмыс істеу тиімділігін арттыру үшін қолданылады. Құрылымдаудың нәтижесі ақпаратты нақты деректерге, олардың құрылымын сипаттайтын мета-деректерге, тіпті деректер құрылымдарының әртүрлі нұсқаларын анықтайтын "Мета-деректерге" бөлу болып табылады. Мазмұнның мұндай ерекшеліктері, айқын немесе жасырын түрде, тиісті ресурстар жиынтығы шеңберінде ақпаратты тиімді іздеуді ұйымдастырудың тәсілдерін анықтайды (Смирнов, 2008).

Ақпаратты іздеу механизмдерінің тиімділігіне айтарлықтай әсер ететін тағы бір маңызды жағдай қазіргі заманғы ақпараттық ресурстар мен жүйелердің таратылған және нәтижесінде гетерогенді сипаты болып табылады. Біртекті Ақпараттық жүйелер мен ресурстар жағдайында қолдануға бағытталған іздеу механизмдері (мысалы, қарапайым индекстер негізінде) таратылған гетерогенді жүйелерге қолдануда тиімділігін күрт жоғалтады, мұнда деректерді ұсыну форматтары және сәйкесінше метадеректер ресурстан ресурсқа немесе жүйеден жүйеге ерекшеленеді. Бұл жағдай зерттеушілер мен әзірлеушілерді заманауи ақпараттық жүйелердің талаптарына сәйкес келетін әмбебап ақпараттық іздеу әдістері мен технологияларын құру жолдарын іздеуге мәжбүр етеді.

Мәтін қоғамдағы ақпарат алмасудың негізгі формаларының бірі болып табылады. Әр түрлі форматтағы мәтіндік ақпарат ақпараттық жүйелердің ақпараттық ресурстарының едәуір бөлігін құрайды. Сондықтан өңдеу технологияларын құру және дамыту мәтіндік іздеу жүйелері болып табылады, олардың міндеті пайдаланушылардың ақпараттық қажеттіліктерін қанағаттандыратын құжаттарды табиғи тілде (лерде) берілген құжаттар жинағы бойынша іздеу болып табылады. Бұл жұмыста мәтіндік іздеудің негізгі принциптері, табиғи тілді өңдеу әдістері және оларды іздеу модельдерінде қолдану ұсынылған.

Ақпараттық іздеу әдістері және табиғи тілдегі мәтіндерді өңдеу

Мәтінді іздеу технологиясында табиғи тілді өңдеу маңызды орын алады. Табиғи тіл арқылы (Natural Language Processing, NLP) өңдеу деп мәтіндерді түсінуге, талдауға, әртүрлі операцияларды орындауға, сондай-ақ оларды генерациялауға байланысты мәселелерді шешу түсініледі. Мұндай міндеттердің мысалдары: құжаттардың сақталған коллекцияларын жіктеу, кластерлеу, мәтіндерді терең талдау, құжаттарды бір тілден екінші тілге аудару және т. б.

Ақпараттық іздеудің барлық әдістері индекстелген құжаттардың мәтіндерін өңдеуге және талдауға негізделген (Большакова, 2008). Ақпараттық іздеудің көпшілігі жүйе-бұл процесске дейінгі жүйелер-жүйеде бар барлық құжаттарды алдын-ала өңдеу (индекстеу). Ерекшеліктер-Мета іздеу жүйелері. Мәтіндерді табиғи тілде өңдеу кезінде туындайтын негізгі қиындықтарды тізімдейміз:

* синоним мәселесі;

- * омоним мәселесі;
- * тұрақты сөз тіркестері;
- * морфологиялық вариация.

Синоним мәселесі. Бір ұғымды әртүрлі сөздермен білдіруге болады. Нәтижесінде, пайдаланушы сұраныста көрсеткен ұғымдардың синонимдерін қолданатын тиісті құжаттар омоним проблемалары мен "омонимге іргелес" құбылыстар болуы мүмкін. Грамматикалық омонимдер сөздің мағынасында әр түрлі, бірақ жеке грамматикалық формаларда жазылуымен сәйкес келеді. Бұл сөйлеудің бір немесе әртүрлі бөліктеріндегі сөздер болуы мүмкін. Лексикалық омонимдер-сөйлеудің бір бөлігінің дыбысы мен емлесі бірдей, бірақ лексикалық мағынасы жағынан әртүрлі сөздер.

Тұрақты сөз тіркестері. Сөз тіркестерінің мағынасы болуы мүмкін сөздердің жеке мағынасынан өзгеше.

Морфологиялық вариация. Көптеген табиғи тілдерде сөздердің емлесі бойынша ерекшеленетін бірнеше морфологиялық формалары бар.

Қолданыстағы іздеу жүйелері мәтінді өңдеудің әртүрлі әдістерін қолданады. Мәтінді іздеудің заманауи технологияларында мәтіндерді талдау үшін лингвистика аппараты ғана емес, сонымен қатар статистикалық әдістер, математикалық логика және ықтималдық теориясы, кластерлік талдау, жасанды интеллект әдістері, сонымен қатар деректерді басқару технологиялары қолданылады. Табиғи тіл мәтіндерін өңдеу мен талдаудың екі негізгі тәсілін қарастырыңыз – статистикалық және лингвистикалық (сурет. 1).



Сур.1. Табиғи тілді өңдеу әдістері
(Fig.1. Natural language processing methods)

Статистикалық тәсілдің негізінде мәтіннің мазмұны жиі кездесетін сөздермен көрінеді деген болжам жатыр. Статистикалық талдаудың мәні құжаттағы сөздердің пайда болу санын есептеу болып табылады. Жалпы-құжаттағы әрбір t терминін кейбір теріс емес салмақпен салыстыру. Терминдік салмақтар әртүрлі тәсілдермен есептеледі. Ең қарапайымы- d құжатындағы t терминінің пайда болу санына тең "салмақ" қою, $tf_{t,d}$ (term frequency) деп белгіленеді. Бұл өлшеу әдісі терминнің кемсітушілік Күшін ескермейді. Сондықтан, коллекция бойынша терминдерді қолдану статистикасы қол жетімді болған жағдайда, TF-IDF шкаласын есептеу схемасы келесідей анықталады:

$$tf - idf_{id} = tf_{i,d} \times idf_i$$

мұндағы $idf_i \log - t$ терминінің кері деректі жиілігі (inverse document frequency), DFT – құжаттық жиілігі (document frequency), t, N терминдерін қамтитын коллекциядағы құжаттардың саны – коллекциядағы құжаттардың жалпы саны ретінде анықталады. Tf-IDF схемасы және оның модификациялары іс жүзінде кеңінен қолданылады.

Статистикалық талдауға негізделген тиімді тәсіл жасырын семантикалық индекстеу болып табылады. Жасырын семантикалық талдау бұл мәтіндік деректердің үлкен жиынтығын статистикалық өңдеу арқылы сөздердің контекстке тәуелді мағыналарын алудың теориясы мен әдісі. Жасырын семантикалық талдау берілген сөз кездесетін және кездеспейтін барлық контексттердің жиынтығы сөздердің семантикалық мағыналары мен сөз жиындарының бір-бірімен ұқсастығын анықтауға мүмкіндік беретін көптеген өзара шектеулерді белгілейді деген идеяға негізделген (Браславский, 2006).

Статистикалық әдістердің басты кемшілігі-мәтіннің байланысын есепке алудың мүмкін .стігі, ал мәтінді сөздердің қарапайым жиынтығы ретінде ұсыну оның мазмұнын көрсету үшін жеткіліксіз. Мәтін белгілі бір ретпен құрылған сөздер жиынтығын білдіреді. Бұл кемшілікті жеңу мәтінді талдаудың лингвистикалық әдістерін қолдануға мүмкіндік береді.

Лингвистикалық талдаудың келесі деңгейлері бар: графематикалық, морфологиялық, синтаксистік, семантикалық. Әр деңгейдің нәтижелері келесі талдау деңгейінде кіріс ретінде қолданылады (сурет. 2).



Сур.2. Лингвистикалық талдау деңгейлері
(Fig.2. Levels of linguistic analysis)

Графематикалық талдаудың мақсаты-мәтін құрылымының элементтерін бөліп көрсету: абзацтар, абзацтар, сөйлемдер, жеке сөздер және т.б.

Морфологиялық талдаудың мақсаты-сөздің морфологиялық сипаттамаларын және оның негізгі сөз формасын анықтау. Талдау ерекшеліктері тандалған табиғи тілге байланысты.

Талдаудың мақсаты-сөйлемдегі сөздердің синтаксистік тәуелділігін анықтау. Орыс тілінде көптеген синтаксистік омонимдік конструкциялардың болуына, семантика мен синтаксис арасында тығыз байланыстың болуына байланысты мәтінді автоматтандырылған талдау процедурасы көп уақытты қажет етеді. Алгоритмнің күрделілігі сөйлемдегі сөздер саны мен қолданылатын ережелер саны артқан сайын экспоненциалды түрде артады (Аграновский, 2003).

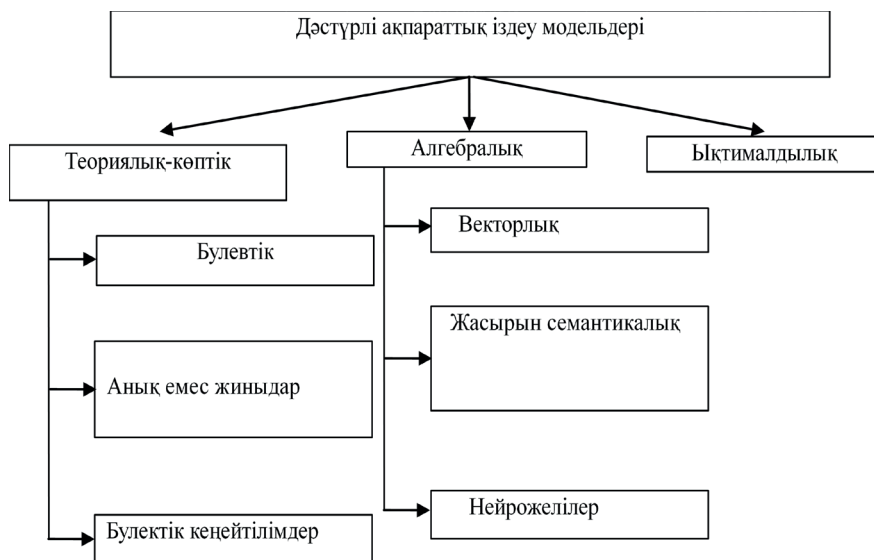
Мәтінді семантикалық талдау саласындағы әзірлемелер мәтінді семантикалық түсінуге баса назар аударатын жасанды интеллект саласымен байланысты. Қазіргі уақытта бұл бағыттағы жетістіктер өте шектеулі. Әзірленген семантикалық анализаторлар есептеудің жоғары күрделілігіне және берілген нәтижелердің түсініксіздігіне ие.

Ақпараттық іздеу модельдері

Ақпараттық-іздеу жүйелерінің әртүрлі модельдерінде құжаттардың өзектілігі критерийінің түрі ақпараттық іздеу моделінің түріне байланысты, мысалы, семантикалық іздеу модельдерінде сұрау сөздерінің құжатқа дәл енуі, мысалы, теориялық және бірнеше модельдердегідей, негізгі критерий емес.

Бұл компоненттердің вариациялары іздеу жүйелерінің көптеген енгізілуін анықтайды. Ең көп таралған іздеу үлгілерін қарастырыңыз.

Дәстүрлі ақпараттық іздеудің модельдері әдетте үш түрге бөлінеді: теориялық - көптік (буль, анық емес жиындықтар, кеңейтілген буль), алгебралық (векторлық, жалпыланған векторлық, жасырын-семантикалық, нейрондық желі), ықтималдық (сурет.3).



Сур.3. Дәстүрлі ақпараттық іздеу модельдері
(Fig.3. Traditional information search models)

Ақпараттық іздеу-Бұл веб-сайттар, дерекқорлар, кітапханалар және басқа ресурстар сияқты әртүрлі көздерден ақпаратты табу және алу процесі. Ақпараттық іздеудің тиімділігін жақсарту үшін әртүрлі модельдер мен әдістер жасалды. Міне, олардың кейбіреулері:

Булевтік іздеу Моделі: бұл қарапайым модельдердің бірі. Пайдаланушы логикалық операторларды (AND, OR, NOT) пайдаланып сұрау салады және жүйе сол операторларға сәйкес нәтижелерді қайтарады.

Векторлық кеңістік моделі (vector Space Model, VSM): бұл модельде әр құжат пен сұрау көп өлшемді кеңістіктегі векторлар түрінде ұсынылған. Іздеу құжат векторлары мен сұрау арасындағы косинус өлшемдерін салыстыру арқылы жүзеге асырылады.

TF-IDF моделі (term Frequency-Inverse Document Frequency): бұл модель сұраудағы сөздің маңыздылығын анықтау үшін құжаттардағы сөздердің пайда болу статистикасын қолданады. Құжаттарда сирек кездесетін, бірақ сұраныста жиі кездесетін сөздер жоғары салмақ алады.

Bm25 Моделі: бұл құжаттардың ұзындығын ескеретін және іздеу нәтижелерін жақсарту үшін параметрлерді қолданатын TF-IDF жетілдірілген нұсқасы.

Ықтималдық ақпараттық іздеу моделі (Probabilistic Information Retrieval): бұл модель құжаттың сұранысқа сәйкес болу ықтималдығын бағалайды. Бұл модельдің белгілі іске асыруларының бірі-bm25 моделі.

LDA моделі (Latent Dirichlet Allocation): бұл модель тақырыптық модельдеу үшін қолданылады. Бұл құжаттардағы жасырын тақырыптарды анықтауға мүмкіндік береді және белгілі бір тақырыпқа қатысты құжаттарды табуға көмектеседі.

Машиналық оқыту және нейрондық желілер: Машиналық оқыту мен терең оқытудың дамуымен ақпараттық іздеуді жақсарту үшін нейрондық желілерге негізделген модельдер дами бастады. Мысал ретінде BERT (Bidirectional Encoder Representations from Transformers) және басқа трансформаторлық модельдер болуы мүмкін.

Іздеу нәтижелерінің рейтингі: іздеу үлгілерінен басқа, іздеу нәтижелерінің рейтингі де маңызды рөл атқарады. Рейтинг өзектілік, танымалдылық және жекелендірілген ұсыныстар сияқты әртүрлі көрсеткіштерге негізделуі мүмкін (Иванов, 2005).

Бұл ақпараттық іздеу модельдері мен әдістерін әртүрлі контексттерде қолдануға болады, соның ішінде интернеттегі іздеу жүйелері, мәліметтер базасын іздеу, кітапханалардағы ақпараттық іздеу және басқа да көптеген салалар. Әрбір модельдің өзіндік артықшылықтары мен шектеулері бар және сәйкес модельді таңдау пайдаланушының нақты тапсырмасы мен талаптарына байланысты.

Мәтінді табиғи тілде өңдеуді қолдану

Күнделікті тапсырмаларды автоматтандыру. NLP технологиясына негізделген чатботтар бүгінде адамдар айналысатын көптеген күнделікті

операцияларды орындай алады және осылайша қызметкерлерге күрделі және қызықты тапсырмаларды орындауға мүмкіндік береді. Мысалы, чатботтар мен сандық көмекшілер пайдаланушының көптеген сұрауларын тани алады, оларды дерекқордағы тиісті жазбалармен сәйкестендіреді және пайдаланушыға сәйкес жауаптарды тұжырымдай алады.

Жақсартылған іздеу. Құжаттар мен жиі қойылатын сұрақтар бойынша іздеу жағдайында NLP контекстке негізделген сөздердің түсініксіздігін жою (мысалы, "тасымалдаушы" сөзі медицина мен өнеркәсіпте әртүрлі мағынаға ие), синонимдерді сәйкестендіру (мысалы, іздеуде "машина" сөзі бар құжаттарды таңдау) арқылы кілт сөздерді іздеуді жақсарта алады. "автомобиль") және морфологиялық нұсқаларды есепке алу (бұл ағылшын тілінен басқа тілдердегі сұраулар үшін маңызды). NLP технологиясына негізделген ғылыми деректерді тиімді іздеу дәрігерлер, заңгерлер және басқа мамандар үшін соңғы зерттеулерге қол жеткізуді айтарлықтай жеңілдетеді.

Іздеу жүйелерін оңтайландыру. NLP-бұл сіздің ұйымыңызға онлайн іздеу нәтижелерінде жоғарылауға мүмкіндік беретін керемет құрал, өйткені ол нәтижелерді талдай алады және мазмұнды оңтайландырады. Іздеу жүйелері іздеу нәтижелерін бағалау үшін NLP пайдаланады. Осы әдістерді қалай тиімді пайдалану керектігін түсіну арқылы Сіздің компанияңыздың бәсекелестерден жоғары іздеу нәтижелерінде көрінуіне оңай қол жеткізуге болады. Бұл Сіздің компанияңыздың көбірек қаралатынын білдіреді.

Құжаттардың үлкен жинақтарын талдау және ретке келтіру. Құжаттарды кластерлеу және тақырыптық модельдеу сияқты NLP әдістері компания есептері, жаңа мақалалар немесе ғылыми құжаттар сияқты үлкен құжаттар жинағындағы әртүрлі материалдарды түсінуді жеңілдетеді. Бұл әдістер көбінесе сұраныс бойынша ақпаратты ашу үшін қолданылады.

Әлеуметтік медианы талдау. NLP ақпараттың үлкен көлемін жақсы түсіну үшін пайдаланушылардың пікірлері мен әлеуметтік медиа пікірлерін талдай алады. Мәтіннің кілтін талдау әлеуметтік желілердегі оң және теріс пікірлерді анықтайды, бұл клиенттердің нақты уақыттағы реакциясын бағалауға мүмкіндік береді. Бұл кейінірек айтарлықтай пайда әкелуі мүмкін, мысалы, сатып алушылардың қанағаттанушылығының жоғарылауы және кірістің өсуі.

Нарықты талдау. Егер сіз NLP-ді компания клиенттерінің тілін талдау үшін қолдансаңыз, олардың не қалайтынын және олармен қалай қарым-қатынас жасау керектігін жақсы түсінуге болады. Мәтіннің кілтін аспектіге бағытталған талдау әлеуметтік желілердегі белгілі бір аспектілерге немесе өнімдерге қатысты мәлімдемелердің кілтін анықтайды (мысалы: "пернетақта керемет және экран тым күңгірт") және өнім дизайнына немесе жарнамалық науқандарға бірден өзгерістер енгізуге болатын ақпаратты береді.

Мазмұнды модерациялау. Егер сіздің компанияңыз көптеген пайдаланушыларды тартса немесе клиенттер көптеген пікірлер қалдырса, NLP олардың сапасы мен сыпайылығын қамтамасыз ету үшін мәлімдемелерді

модерациялауға мүмкіндік береді. Бұған тек сөздерді ғана емес, сонымен қатар түсініктемелердің кілті мен мәнін талдау арқылы қол жеткізуге болады.

Қорытынды

Жоғарыда келтірілген дәстүрлі іздеу жүйесінің модельдері бастапқыда құжаттарды бір-біріне тәуелді емес көптеген жеке сөздер ретінде қарастыруды көздеді. Бқтималдық моделі төмен есептеу масштабталуымен, жүйені үнемі оқыту қажеттілігімен сипатталады. Ең көп тарағандары-алгебралық теориялық-көп модельдер, өйткені олардың практикалық тиімділігі әдетте жоғары. Айта кету керек, жақында ұсынылған ақпараттық іздеу жобаларының жаңа іске асырылуы көбінесе гибриді модельдер болып табылады және әртүрлі сыныптардың модельдерінің қасиеттеріне ие. Ақпараттық-іздеу жүйелерін дамытудың перспективалық бағыттарының бірі-семантикалық іздеу модельдерін құру, оның негізгі міндеті мәтінді талдау, яғни мәтіннен мағынаны алу және оны екі мәтіннің семантикалық жақындығын табуға мүмкіндік беретін формальды модельге айналдыру. Мұндай жүйелердің әлеуеті шынымен үлкен екенін мойындау керек, бірақ қазіргі уақытта барлық мүмкін семантикалық технологиялар жүзеге асырылған жоқ. Шын мәнінде, қазір олар тек табиғи тілде құрылған сөз тіркестерінен кілт сөздерді бөліп көрсетуге және дұрыс іздеу сұрауын жасау үшін қосымша сөз формаларын таңдауға көмектеседі. Іздеу әдістерінің бұл бағыты дамуды қажет етеді.

ӘДЕБИЕТТЕР

Аграновский А.В., Арутюнян Р.Э. (2003). Способы индексации и поиска документов в интернет-порталах // Труды X Всероссийской научно-методической конференция «Телематика-2003». Санкт-Петербург. — 2003. — Т.1. — С. 204–206.

Большакова Е.И., Васильева Н.Э. (2008). Терминологическая вариантность и ее учет при автоматической обработке текстов // Одиннадцатая Национальная конференция по искусственному интеллекту с международным участием КИИ—2008. Труды конференции. —М.: Физматлит, 2008. — Т. 2. — С. 174–182.

Браславский П.И., Соколов Е.А. (2006). Сравнение четырех методов автоматического извлечения двухсловных терминов из текста // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог'2006. — М.: Изд-во РГГУ, — 2006. — С. 88–94.

Иванов В.К., Иванов К.В. (2005). Введение в информационно-поисковые системы: метод. указания по изучению дисциплины "Мировые информ. ресурсы" для студентов спец. "Прикл. информатика (в экономике)" и "Информ. сист. и технол.". — Ч. 2. — Тверь: ТГТУ, — 2005.

Осипов Г.С., Смирнов И.В., Тихомиров И.А. (2008). Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. — М.: ИСА РАН — №2. — 2008. — С. 3–10. 2.

Соченков И.В. (2013). Метод сравнения текстов для решения поисково-аналитических задач // Искусственный интеллект и принятие решений. — М.: ИСА РАН, — № 2. — 2013. — С. 95–106. 3.

Смирнов И.В., Соченков И.В., Муравьев В.В., Тихомиров И.А. (2008). Результаты и перспективы поискового алгоритма Eхactus. // Труды российского семинара по оценке методов информационного поиска РОМИП'2007–2008. — Санкт-Петербург: НУ ЦСИ, — 2008. — С. 66–76.

REFERENCES

Agranovsky A.V., Harutyunyan R.E. (2003). Ways of indexing and searching documents in Internet portals // Proceedings of the X All-Russian Scientific and Methodological Conference "Telematics-2003". St. Petersburg. — 2003. — Vol.1. — Pp. 204–206.

Bolshakova E.I., Vasilyeva N.E. (2008). Terminological variation and its consideration in automatic text processing // Eleventh National Conference on Artificial Intelligence with international participation KII-2008. Proceedings of the conference. -Moscow: Fizmatlit, — 2008. — Vol. 2. — Pp. 174–182.

Braslavsky P.I., Sokolov E.A. (2006). Comparison of four methods of automatic extraction of two-word terms from the text // Computational linguistics and intellectual technologies: Proceedings of the International Conference Dialog'2006. - Moscow: Publishing House of the Russian State University, — 2006. — Pp. 88–94.

Ivanov V.K., Ivanov K.V. (2005). Introduction to information retrieval systems: method. instructions for the study of the discipline "World inform. resources" for students of spec. "Approx. informatics (in economics)" and "Inform. system. and technol.". — Part 2. — Tver: TSTU, — 2005.

Osipov G.S., Smirnov I.V., Tikhomirov I.A. (2008). Relational-situational method of text search and analysis and its applications // Artificial intelligence and decision-making. — Moscow: ISA RAS. — No.2. — 2008. — Pp. 3–10. 2.

Sochenkov I.V. (2013). Method of text comparison for solving search and analytical problems // Artificial Intelligence and decision-making. — Moscow: ISA RAS, — No. 2. — 2013. — Pp. 95–106. 3.

Smirnov I.V., Sochenkov I.V., Muravyev V.V., Tikhomirov I.A. (2008). Results and prospects of the Exactus search algorithm. // Proceedings of the Russian seminar on evaluation of information retrieval methods ROMIP'2007–2008. St. Petersburg: NU CSI, — 2008. — Pp. 66–76.

МАЗМҰНЫ

| | |
|--|-----|
| Г.Б. Абдикеримова, Р.М. Аманов, Г.Т. Азиева, А.М. Заманбекова, Қ. Жеңсқанқызы <i>МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНА ОТЫРЫП, АККОРДТЫ ТАҢУ ТАПСЫРМАСЫНДАҒЫ ДЫБЫСТЫ ӨҢДЕУ ӘДІСТЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ</i> | 7 |
| Л.А. Абдыкеримова, Г.Е. Мырзабекова, Г.С. Омарова, Л. Ақзуллақызы, Г.Ш. Мусагулова ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ АРҚЫЛЫ ЖҮРЕК ПАТОЛОГИЯСЫН АНЫҚТАУ..... | 21 |
| А.Е. Әбжанова, Е.Ә. Әбжанов, А.А. Мырзамуратова, А.Г. Батырханов, А.Б. Бексейтова ҚАШЫҚТАН ЗОНДТАУ АРҚЫЛЫ АЛЫНҒАН ТОПЫРАҚ ЫЛҒАЛДЫЛЫҒЫ..... | 35 |
| У.Ж. Айтимова, М.Ж. Айтимов, Э.Н. Тулегенова, А.У. Есиркепова, Ж.Т. Абилдаева СУРЕТТЕН ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ ӨРТ ОШАҒЫН АНЫҚТАУ..... | 50 |
| К.М. Алдабергенова, М.Ж. Жасұзақова, М.Ж. Айтимов, Н.Т. Мұстафаева, К.К. Дауренбеков АУЫЛ ШАРУАШЫЛЫҒЫН ЦИФРЛАНДЫРУ: ДАМУ МҮМКІНДІКТЕРІ МЕН ПЕРСПЕКТИВАЛАРЫ..... | 64 |
| А.С. Баегизова, Г.И. Мухамедрахимова, Ж.Б. Ламашева, А.З. Абдрахманова, Т.Т. Оспанова ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІМЕН КЕСКІННІҢ САПАСЫН ЖАҚСARTУ..... | 78 |
| Г.Т. Бекманова, А.С. Омарбекова, М.А. Кантуреева, Н.О. Байгабылов, М.М. Құдабеков ӘЛЕУМЕТТАНУЛЫҚ САУАЛНАМАЛЫҚ ЗЕРТТЕУЛЕРДЕГІ АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР..... | 91 |
| М.Ә. Берсүгір, Г.У. Маматова, А.А. Нурпейсова, М.Б. Онгарбаева, Ж.Т. Алтынбекова ТЕКСТУРАЛЫҚ ТИПТЕГІ СУРЕТТЕРДІ ЖАҚСARTУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ..... | 104 |
| М.А. Болатбек, К.Д. Байсылбаева, М. Сағынай, Ш.Ж. Мусиралиева, А.Н. Жумаханова ИНТЕРНЕТ КЕҢІСТІГІНДЕГІ ЖАСТАРҒА БАҒЫТТАЛҒАН ДЕСТРУКТИВТІ МӘТІНДЕРДІ ЖИНАҚТАУҒА ҚАЖЕТТІ ПАРСЕР БАҒДАРЛАМАСЫН ӨЗІРЛЕУ..... | 117 |
| М.Қ. Болсынбек, Г.Б. Абдикеримова, Г.С. Омарова, А.Б. Остаева, А.Г. Батырханов ТОПЫРАҚ ДАЙЫНДАУДЫ БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ..... | 132 |
| Ш.К. Ележанова, А.Г. Батырханов, А.Е. Чукуров, Б.С. Хайржанова, Д.А. Тагиев АҚПАРАТТЫҚ БЕЛГІСІЗДІК ТИПОЛОГИЯСЫ ЖӘНЕ АҚПАРАТТЫ ІЗДЕУ ТҮРЛЕРІ..... | 151 |
| М.М. Есмагамбетова, Т.Т. Оспанова, Л.К. Бобров, Т.Л. Тен, Т.У. Есмагамбетов ҒАРЫШТЫҚ ТӨТЕНШЕ ЖАҒДАЙЛАРДЫ БАҚЫЛАУ КЕСКІНДЕРІН ӨҢДЕУДЕ ТҮСТЕРДІ ӨЛШЕУ БАҒДАРЛАМАЛЫҚ ҚҰРАЛЫН ТАҢДАУ..... | 161 |

| | |
|---|-----|
| Т.К. Жукабаева, А. Адамова, Б.А. Ху Вен-Цен, Е.М. Марденов, Л.З. Жолшиева СЫМСЫЗ СЕНСОР ЖЕЛІСІНДЕГІ SYBIL ЖӘНЕ WORMHOLE ШАБУЫЛДАРДЫ АНЫҚТАУ..... | 171 |
| А.А. Исмаилова, Ж.Т. Бельдеубаева, А.А. Нурпейсова, Г.О. Исакова, Ж.З. Жантасова ӨСІМДІК АУРУЛАРЫН ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ..... | 184 |
| А.Х. Касымова, М.Б. Есенова, М.У. Худойберганов, А.Б. Остаева, М.Г. Байбулова ДАҚЫЛДАРДЫҢ АУРУЛАРЫН ЖІКТЕУ ҮШІН ТЕРЕҢ ОҚЫТУ АЛГОРИТМДЕРІН ҚОЛДАНУ..... | 198 |
| А.Ұ. Мұхиядин, М.У. Мукашева, У.Т. Махажанова, А.А. Муханова, Ж.Б. Ламашева ПРОГРАММАЛЫҚ ҚҰРАЛДАР КӨМЕГІМЕН ЭКСТРЕМАЛДЫ ҚАШЫҚТЫҚТАН ОҚЫТУДЫҢ ОҚУШЫЛАРҒА ӘСЕРІН ЗЕРТТЕУ..... | 209 |
| Б.Б. Оразбаев, Л.Т. Салыбек, К.Н. Оразбаева, Ш.К. Коданова, С.Ш. Исакова МҰНАЙДЫ АЛҒАШҚЫ ӨНДЕУДЕ ЭЛЕКТРОТҰЗСЫЗДАНДЫРЫРУ ЖӘНЕ СУСЫЗДАНДЫРУ ПРОЦЕССТЕРІН ОПТИМИЗАЦИЯЛАУ ҮШІН МОДЕЛЬДЕР ҚҰРУ ТӘСІЛІ..... | 224 |
| С.К. Серикбаева, М.Қ. Болсынбек, А.Д. Абдувалова, А.Т. Абдыхалық, Д.Е. Ануарбек ТОПЫРАҚ САПАСЫН БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНУ: АЛГОРИТМДЕР МЕН ӘДІСТЕР..... | 237 |
| А.Ж. Танирбергенов, Ж.К. Тасжурекова, С.К. Серикбаева, А.А. Шораев, А.Д. Абдувалова ТОЛЫҚ МӘТІНДІ ҚҰЖАТТАРДЫ ІЗДЕУДІҢ МОДЕЛІ МЕН АҚПАРАТТЫҚ ЖҮЙЕСІН ҚҰРУ ӘДІСТЕРІ..... | 253 |
| А.Ә. Таурбекова, Ө.Ж. Мамырбаев, К.Ж. Тұрғанбай СЕЙСМИКАЛЫҚ ҚЫЗМЕТТІ БАҒАЛАУ ҮШІН ГИДРОДИНАМИКАЛЫҚ ТҰРАҚСЫЗДЫҚ ПРОЦЕСІН ЗЕРТТЕУ..... | 268 |
| Н. Т. Тұржанов, Ш. К. Ележанова, С. Н. Идрисов, Ж. К. Дюсембина АҚПАРАТТЫҚ ҮДЕРІСТЕРДІҢ РЕИНЖИНИРИНГІНІҢ ИННОВАЦИЯЛЫҚ КУРСЫН ӘЗІРЛЕУ..... | 290 |
| В. Шевцов, А. Исмаилова, Ж. Белдеубаева, А. Сатыбалдиева, А. Нурпейсова МЛВА ГЕНОТИПТЕУДІҢ ӘДІСІ ЖӘНЕ ОНЫ ЖҮЗЕГЕ АСЫРУ АЛГОРИТМДЕРІ РЕТІНДЕГІ ГЕНОМДЫҚ ДЕРЕКТЕРДІ ПАЙДАЛАНУ..... | 300 |
| А.Ә. Шекербек, А.А. Некесова, Ж.Ж. Молдашева, А.И. Онгарбаева, А.О. Тохаева ФРАКТАЛДЫҚ ӘДІСПЕН ӨКПЕНІҢ ПАТОЛОГИЯЛЫҚ ЖАҒДАЙЫН ТАЛДАУ..... | 313 |

СОДЕРЖАНИЕ

| | |
|---|-----|
| Г.Б. Абдикеримова, Р.М. Аманов, Г.Т. Азиева, А.М. Заманбекова, К. Женсканкызы СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОБРАБОТКИ ЗВУКА В ЗАДАЧЕ РАСПОЗНАВАНИЯ АККОРДОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ..... | 7 |
| Л.А. Абдыкеримова, Г.Е. Мурзабекова, Г.С. Омарова, Л. Акзуллакызы, Г.Ш. Мусагулова ОБНАРУЖЕНИЕ СЕРДЕЧНОЙ ПАТОЛОГИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ..... | 21 |
| А.Е. Абжанова, Е.А. Абжанов, А.А. Мырзамуратова, А.Г. Батырханов, А.Б. Бексейтова ВЛАЖНОСТЬ ПОЧВЫ, ПОЛУЧЕННАЯ ДИСТАНЦИОННЫМ ЗОНДИРОВАНИЕМ..... | 35 |
| У.Ж. Айтимова, М.Ж. Айтимов, Э.Н. Тулегенова, А.У. Есиркепова, Ж.Т. Абилдаева ОБНАРУЖЕНИЕ ОЧАГОВ ПОЖАРА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ПО ИЗОБРАЖЕНИЮ..... | 50 |
| К.М. Алдабергенова, М.Ж. Жасузакова, М.Ж. Айтимов, Н.Т. Мустафаева, К.К. Дауренбеков ЦИФРОВИЗАЦИЯ СЕЛЬСКОГО ХОЗЯЙСТВА: ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ РАЗВИТИЯ..... | 64 |
| А.С. Баегизова, Г.И. Мухамедрахимова, Ж.Б. Ламашева, А.З. Абдрахманова, Т.Т. Оспанова УЛУЧШЕНИЕ КАЧЕСТВА ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ..... | 78 |
| Г.Т. Бекманова, А.С. Омарбекова, М.А. Кантуреева, Н.О. Байгабылов, М.М. Кудабеков ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В СОЦИОЛОГИЧЕСКИХ ОПРОСНЫХ ИССЛЕДОВАНИЯХ..... | 91 |
| М.А. Берсугир, Г.У. Маматова, А.А. Нурпейсова, М.Б. Онгарбаева, Ж.Т. Алтынбекова ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ УЛУЧШЕНИЯ ИЗОБРАЖЕНИЙ ТЕКСТУРНОГО ТИПА..... | 104 |
| М.А. Болатбек, К.Д. Байсылбаева, М. Сагынай, Ш.Ж. Мусиралиева, А.Н. Жумаханова РАЗРАБОТКА ПРОГРАММЫ ПАРСЕРА ДЛЯ СБОРА ДЕСТРУКТИВНЫХ ТЕКСТОВ, ОРИЕНТИРОВАННЫХ НА МОЛОДЕЖЬ В ИНТЕРНЕТ-ПРОСТРАНСТВЕ..... | 117 |
| М.К. Болсынбек, Г.Б. Абдикеримова, Г.С. Омарова, А.Б. Остаева, А.Г. Батырханов ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ПОДГОТОВКИ ПОЧВЫ..... | 132 |
| Ш.К. Ележанова, А.Г. Батырханов, А.Е. Чукуров, Б.С. Хайржанова, Д.А. Тагиев ТИПОЛОГИЯ ИНФОРМАЦИОННОЙ НЕОПРЕДЕЛЕННОСТИ И ТИПЫ ПОИСКА ИНФОРМАЦИИ..... | 151 |

| | |
|---|-----|
| М.М. Есмагамбетова, Т.Т. Оспанова, Л.К. Бобров, Т.Л. Тен, Т.У. Есмагамбетов ВЫБОР ПРОГРАММНЫХ СРЕДСТВ ЦВЕТОМЕТРИИ В ОБРАБОТКЕ ИЗОБРАЖЕНИЙ КОСМИЧЕСКОГО МОНИТОРИНГА ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ..... | 161 |
| Т.К. Жукабаева, А. Адамова, В.А. Ху Вен-Цен, Е.М. Марденов, Л.З. Жолшиева ОБНАРУЖЕНИЕ SYBIL И WORMHOLE АТАК В БЕСПРОВОДНОЙ СЕНСОРНОЙ СЕТИ..... | 171 |
| А.А. Исмаилова, Ж.Т. Бельдеубаева, А.А. Нурпейсова, Г.О. Исакова, Ж.З. Жантасова ОБНАРУЖЕНИЕ БОЛЕЗНЕЙ РАСТЕНИЙ МЕТОДОМ ГЛУБОКОГО ОБУЧЕНИЯ..... | 184 |
| А.Х. Касымова, М.Б. Есенова, М.У. Худойбергенов, А.Б. Остаева, М.Г. Байбулова ПРИМЕНЕНИЕ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ БОЛЕЗНЕЙ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР..... | 198 |
| А.У. Мухиядин, М.У. Мукашева, У.Т. Махажанова, А.А. Муханова, Ж.Б. Ламашева ИЗУЧЕНИЕ ВЛИЯНИЯ ЭКСТРЕМАЛЬНОГО ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ НА УЧАЩИХСЯ С ИСПОЛЬЗОВАНИЕМ ПРОГРАММНЫХ СРЕДСТВ..... | 209 |
| Б.Б. Оразбаев, Л.Т. Салыбек, К.Н. Оразбаева, Ш.К. Коданова, С.Ш. Исакова МЕТОД РАЗРАБОТКИ МОДЕЛЕЙ ДЛЯ ОПТИМИЗАЦИИ ПРОЦЕССОВ ЭЛЕКТРООБЕССОЛИВАНИЯ И ОБЕЗВОЖИВАНИЯ ПРИ ПЕРВИЧНОЙ ПЕРЕРАБОТКЕ НЕФТИ..... | 224 |
| С.К. Серикбаева, М.К. Болсынбек, А.Д. Абдувалова, А.Т. Абдыхалык, Д.Е. Ануарбек ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КАЧЕСТВА ПОЧВЫ: АЛГОРИТМЫ И МЕТОДИКИ..... | 237 |
| А.Ж. Танирбергенов, Ж.К. Тасжурекова, С.К. Серикбаева, А.А. Шораев, А.Д. Абдувалова МЕТОДЫ ПОСТРОЕНИЯ МОДЕЛИ И ИНФОРМАЦИОННОЙ СИСТЕМЫ ПОИСКА ПОЛНОТЕКСТОВЫХ ДОКУМЕНТОВ..... | 253 |
| А.Ә. Taurbekova, O.Zh. Mamyrbayev, K.Zh. Doshtaev, T.K. Eginbaykyzy HYDRODYNAMIC INSTABILITY MECHANISM PROCESS FOR ASSESSMENT SEISMIC ACTIVITY..... | 268 |
| Н.Т. Туржанов, Ш.К. Ележанова, С.Н. Идрисов, Ж.К. Дюсембина РАЗРАБОТКА ИННОВАЦИОННОГО КУРСА ПО РЕИНЖИНИРИНГУ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ..... | 290 |
| В. Шевцов, А. Исмаилова, Ж. Бельдеубаева, А. Сатыбалдиева, А. Нурпейсова MLVA КАК МЕТОД ГЕНОТИПИРОВАНИЯ И АЛГОРИТМЫ ЕГО РЕАЛИЗАЦИИ С ИСПОЛЬЗОВАНИЕМ ПОЛНОГЕНОМНЫХ ДАННЫХ..... | 300 |
| А.А. Шекербек, А.А.Некесова, Ж.Ж.Молдашева, А.И. Онгарбаева, А.О.Тохаева АНАЛИЗ ПАТОЛОГИЧЕСКИХ СОСТОЯНИЙ ЛЕГКИХ С ИСПОЛЬЗОВАНИЕМ ФРАКТАЛЬНОГО МЕТОДА..... | 313 |

CONTENTS

| | |
|--|-----|
| G.B. Abdikerimova, R.M. Amanov, G.T. Azieva, A.M. Zamanbekova, K. Zhengskankyzy COMPARATIVE ANALYSIS OF SOUND PROCESSING METHODS IN THE CHORD RECOGNITION PROBLEM USING MACHINE LEARNING..... | 7 |
| L. Abdykerimova, G. Murzabekova, G. Omarova, L. Akzullakyyzy, G. Mussagulova DETECTION OF CARDIAC PATHOLOGY USING DEEP LEARNING METHODS..... | 21 |
| A.E. Abzhanova, E.A. Abzhanov, A.A. Myrzamuratova, A.G. Batyrkhanov, A.B. Bekseitova SOIL MOISTURE OBTAINED BY REMOTE SENSING..... | 35 |
| U. Zh Aitimova, M.Zh. Aitimov, E.N. Tulegenova, A.U. Yessirkepova, Zh.T. Abildaeva FIRE FOCUS DETECTION USING DEEP LEARNING METHODS FROM IMAGE..... | 50 |
| K.M. Aldabergenova, M.ZH. Zhasuzakova, M.Zh. Aitimov, N.T. Mustafaeva, K.K. Daurenbekov DIGITALIZATION OF AGRICULTURE: OPPORTUNITIES AND PROSPECTS FOR DEVELOPMENT..... | 64 |
| A.S. Baegizova, G.I. Mukhamedrakhimova, Zh.B. Lamasheva, A.Z. Abdrakhmanova, T.T. Ospanova IMPROVE IMAGE QUALITY WITH DEEP LEARNING TECHNIQUES..... | 78 |
| G. Bekmanova, A. Omarbekova, M. Kantureyeva, N. Baigabylov, M. Kudabekov INFORMATION TECHNOLOGIES IN SOCIOLOGICAL SURVEY RESEARCH..... | 91 |
| M.A. Bersugir, G.U. Mamatova, A.A. Nurpeisova, M.B. Ongarbayeva, Zh.T. Altynbekova USING MACHINE LEARNING METHODS TO IMPROVE TEXTURE-TYPE IMAGES..... | 104 |
| M. Bolatbek, K. Baisylbaeva, M. Sagynay, Sh. Mussiraliyeva, A. Zhumakhanova DEVELOPMENT OF A PARSER PROGRAM FOR THE ACCUMULATION OF DESTRUCTIVE TEXTS AIMED AT YOUNG PEOPLE IN THE INTERNET SPACE..... | 117 |
| M. Bolsynbek, G. Abdikerimova, G. Omarova, A. Ostayeva, A. Batyrkhanov APPLICATION OF MACHINE LEARNING TO PREDICT SOIL PREPARATION.... | 132 |
| Sh.K. Yelezhanova, A.G. Batyrkhanov, A.Y. Chukurov, B.S. Khairzhanova, J.A. Taghiyev TYPOLOGY OF INFORMATION UNCERTAINTY AND TYPES OF INFORMATION RETRIEVAL..... | 151 |
| M. Yesmagambetova, T. Ospanova, L. Bobrov, T. Ten, T. Yesmagambetov SELECTION OF COLORIMETRY SOFTWARE TOOLS IN IMAGE PROCESSING OF SPACE MONITORING OF EMERGENCY SITUATIONS..... | 161 |
| T. Zhukabayeva, A. Adamova, B. Khu Ven-Tsen, Y. Mardenov, L. Zholshiyeva DETECTION OF SYBIL AND WORMHOLE ATTACKS IN A WIRELESS SENSOR NETWORK..... | 171 |
| A.A. Ismailova, Zh.T. Beldeubayeva, A.A. Nurpeisova, G.O. Issakova, Zh.Z. Zhantassova | |

| | |
|--|-----|
| DETECTION OF PLANT DISEASES USING DEEP LEARNING METHODS..... | 184 |
| A.K. Kassymova, M.B. Yessenova, M.U. Khudoyberganov, A.B. Ostayeva, M.G. Baibulova | |
| APPLICATION OF DEEP LEARNING ALGORITHMS FOR CLASSIFICATION OF DISEASES OF AGRICULTURAL CROPS..... | 198 |
| A. Mukhiyadin, M. Mukasheva, U. Makhazhanova, A. Mukhanova, Zh. Lamasheva | |
| STUDYING THE EFFECTS OF EXTREME DISTANCE EDUCATION ON STUDENTS USING SOFTWARE TOOLS..... | 209 |
| B. Orazbayev, L. Salybek, K. Orazbayeva, Sn. Kodanova, S. Iskakova | |
| METHOD FOR DEVELOPING MODELS FOR OPTIMIZING PROCESSES OF ELECTRICAL DESALTING AND DEHYDRATION DURING PRIMARY OIL PROCESSING..... | 224 |
| S.Serikbayeva, M.Bolsynbek, A. Abduvalova, A. Abdykhalyk, D. Anuarbek | |
| APPLICATION OF MACHINE LEARNING TO PREDICT SOIL QUALITY: ALGORITHMS AND TECHNIQUES..... | 237 |
| A. Tanirbergenov, Zh. Tashhurekova, S. Serikbayeva, A. Shorayev, A. Abduvalova | |
| METHODS OF CONSTRUCTING A MODEL AND AN INFORMATION SYSTEM FOR SEARCHING FULL-TEXT DOCUMENTS..... | 253 |
| A.Ə. Taurbekova, O.Zh. Mamyrbayev, K.Zh. Doshtaev, T.K. Eginbaykyzy | |
| HYDRODYNAMIC INSTABILITY MECHANISM PROCESS FOR ASSESSMENT SEISMIC ACTIVITY..... | 268 |
| N.T. Turzhanov, Sh.K. Yelezhanova, S.N. Idrissov, Zh.K. Dyusseminina | |
| DEVELOPMENT OF AN INNOVATIVE COURSE REENGINEERING OF INFORMATION PROCESSES..... | 290 |
| V. Shevtsov, A. Ismailova, Zh. Beldeubayeva, A. Satybaldiyeva, A. Nurpeisova | |
| MLVA AS A METHOD OF GENOTYPING AND ALGORITHMS FOR ITS IMPLEMENTATION USING GENOME-WIDE DATA..... | 300 |
| A.A. Shekerbek, A.A. Nekesova, Zh.Zh. Moldasheva, A.I. Ongarbayeva, A. Tokhaeva | |
| ANALYSIS OF PATHOLOGICAL CONDITIONS OF THE LUNG USING THE FRACTAL METHOD..... | 313 |

**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)

Подписано в печать 28.12.2023.

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

21,0 п.л. Тираж 300. Заказ 4.