

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

«Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ» КЕАҚ

**Студенттер мен жас ғалымдардың
«GYLYM JÁNE BILIM - 2023»
XVIII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XVIII Международной научной конференции
студентов и молодых ученых
«GYLYM JÁNE BILIM - 2023»**

**PROCEEDINGS
of the XVIII International Scientific Conference
for students and young scholars
«GYLYM JÁNE BILIM - 2023»**

**2023
Астана**

УДК 001+37
ББК 72+74
G99

«GYLYM JÁNE BILIM – 2023» студенттер мен жас ғалымдардың XVIII Халықаралық ғылыми конференциясы = XVIII Международная научная конференция студентов и молодых ученых «GYLYM JÁNE BILIM – 2023» = The XVIII International Scientific Conference for students and young scholars «GYLYM JÁNE BILIM – 2023». – Астана: – 6865 б. - қазақша, орысша, ағылшынша.

ISBN 978-601-337-871-8

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 001+37
ББК 72+74

ISBN 978-601-337-871-8

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2023

беру. Бұл жағдайда аудио үлгілердің кіші биттері таңдалады және ақпарат биттерімен ауыстырылады.

6. Криптография: LSB алгоритмін ақпаратты шифрлау үшін пайдалануға болады. Мысалы, шифрлау кілті туралы ақпаратты оның анықталуын болдырмау үшін файлдың төменгі биттерінде жасыруға болады.

Жалпы, LSB алгоритмін цифрлық суреттерде ғана емес, бейне файлдарда және басқа цифрлық материалдарда авторлық құқықты қорғау үшін пайдалануға болады.

Қорытындылай келе, LSB – сандық кескіндердегі авторлық құқықты қорғаудың тиімді әдістерінің бірі. Бұл алгоритмді цифрлық кескінге авторлық және пайдалану шектеулері туралы ақпаратты енгізу үшін пайдалануға болады. Дегенмен, бұл алгоритмді қолданған кезде абай болу керек және оны тек заңды мақсаттарда қолдану қажет.

Сандық кескіндерде авторлық құқықты дифференциалды қорғау үшін алгоритмді қолдану қарапайым және тиімді әдіс болып табылады, оны мамандандырылған жабдықты қажет етпей-ақ қолдануға болады. Сондай-ақ, қорғаудың жоғары деңгейіне жету үшін LSB алгоритмімен бірге басқа әдістер мен алгоритмдерді қолдануды қарастырған жөн екендігін ескеру қажет.

Қолданылған әдебиеттер тізімі:

1. Грибунин В.Г. Цифрлық стеганография / И. Н. Оков, И.В. Туринцев. – М.: СОЛОН-ПРЕСС, 2009. – 272 б.

2. Фридрих Дж. Цифрлық медидағы стеганография: принциптер, алгоритмдер және қосымшалар. – Кембридж: Кембридж университетінің баспасы, 2010. – 437 б.

3. Евсютин О.О. Оңтайландыруды қолдана отырып, ауыстыру операциясына негізделген ақпаратты сығылған цифрлық суреттерге енгізу алгоритмі / О.О. Евсютин, А.А. Шелупанов, Р.В. Мещеряков, Д.О. Бондаренко // Компьютерная оптика. – 2017. – Т. 41, № 3. – С. 412–421.

4. Oluwaseun A. A Comparative Analysis of LSB, MSB and PVD Based Image Steganography/ A.E. Adedoyin, A.O. Titilayo, F.O. Deborah // International Journal of Research and Review. – 2021. – Vol. 8. – P. 373–377.

5. Fridrich J. Reliable detection of LSB steganography in grayscale and color images / , Goljan M., Du R.// ACM Workshop on Multimedia and Security. – 2001. – P. 27-30.

ӘОЖ 004.912

ҚАЗАҚ ТІЛІНДЕГІ ҒЫЛЫМИ МӘТІНДЕРДЕН АТАУЛЫ МӘНДЕРДІ ШЫҒАРЫП АЛУ

Ташибаева Раушан Бауржановна

raushan-tashibaeva@mail.ru

Л.Н.Гумилев атындағы ЕҰУ «Ақпараттық жүйелер»

кафедрасының докторанты, Астана, Қазақстан

Ғылыми жетекші – т.ғ.к., PhD А.М. Бакиева

Шетелдік консультант – ф-м.ғ.к., доцент, Т.В. Батура

Интернеттің таралуымен ақпарат саны екпінді дамып келе жатыр. "Nature" журналының мәліметі бойынша, бүкіл әлемде 2008 жылдан 2018 жылға дейінгі ғылыми ақпаратты жариялау саласы қарқынды өсуде. Мәселен, бұл жылдар аралығында биомедициналық тақырыбындағы мақалалар саны 1.8 миллионнан 2.6 миллионға дейін өскен. Дегенмен, мәтіндерден ең маңызды ақпаратты тиімді алу және оны өңдеу көп уақытты қажет етеді. Әр түрлі саладағы мәтіндер құрылымы мен мазмұны бойынша ерекшеленеді. Мысалы, ғылыми есептер мен мақалаларда білімнің әртүрлі салаларындағы озық жетістіктер туралы құнды мәліметтер бар, ал үкіметтік құжаттар (өкімдер, есептер, қаулылар) өңірлерді дамыту бойынша атқарылған жұмыстар мен жоспарланған іс-шараларды сипаттайды. Ақпараттың

үлкен ағынын тиімдірек саралау үшін жаңа автоматты әдістер мен құралдарды жасау қажет екені айқын.

Ғылыми жарияланымдар санының артуымен, олардан негізгі ақпаратты автоматтандырылған түрде алу және оны құрылымдау қажеттілігі артып келеді, мысалы, мақалада қарастырылған тапсырма туралы ақпарат, зерттеу нәтижелері, жұмыста қолданылатын әдістер [1].

Ақпарат алу дегеніміз (ағылш. Information extraction, IE) - мәтіннен қажетті ақпаратты іздеу үдерісі, соның ішінде атаулы мәндерді, қатынастарды және ең қиыны оқиғаларды (оқиғаның қашан және қай жерде болғанын, қатысушылардың кім екенін және т.б. сипаттамалар) алу. Ол кілттік сөздерді іздеуден қарағанда тереңірек талдауды қажет етеді. Тапсырманың нәтижесі құрылымданбаған ақпаратты құрылымдық түрге түрлендіру болып табылады [2].

Ақпаратты шығарып алу бірнеше ішкі мәселелерді қамтиды, олардың бірі-атаулы мәндерді шығару. Атаулы мәндерді шығарып алу міндеті - сөздерден тұратын мәтіндердің үзінділерін табу, содан кейін атаулы мәндердің түрін түгендеу. Атаулы мән дегеніміз - нақтылы салаларға байланысты. Дәстүрлі түрде атаулы мәндер адамдардың аттарын, ұйымдардың атауларын әрі географиялық ерекшеліктерді қамтитын сөз тіркестерін білдіреді [3]. Уақыт өте келе, уақытша өрнектер, валюта, пайыздық өрнектер және т. б. сияқты категориялар атаулы мәндер аумағында қарастырыла бастады.

Мәтіндерді өңдеу технологияларының дамуымен "атаулы мән" ұғымы білім саласына қатысты ажыратыла бастады. Айталық, мәселен, медицина саласында ауру атауларын, дәрі-дәрмектерді, химиялық компоненттерді және т.б. атаулы мәндерін шығарып алу өзекті мәселе болып табылады. [4, 5]. Ақпараттық технологиялар саласындағы ғылыми мәтіндер үшін "тапсырма", "әдіс", "құрал" және т. б. сияқты атаулы мәндер түрлері қарастырылады. Мысалы: Мақала бағыты [[Интернеттен ақпаратты іздеу]_{TASK} әдістерін құруға]_{CONTRIV} арналған. Қазіргі уақытта [әртүрлі дереккөздерден мәліметтерді жылдам автоматтандырылған түрде алу тапсырмасы]_{TASK} өте өзекті болып табылады. Бұл жұмыстың мақсаты – [осы тапсырманы шешуге арналған құралдар жинағын құру]_{GOAL}.

Мәтіндегі сәйкес үзінділерді табу үрдісі мәтіндердің семантикасымен байланысты және атаулы мәндер шекараларының белгісіздігімен қиындайды, демек олардың қай жерде басталатынын және қай жерде аяқталатынын анықтау қажет [6].

Бұл мәселенің классикалық шешімі - тізбекті тәгтеу мәселесінің қорытындысы болып табылады. Яғни кіріс тізбегіндегі әрбір токен үшін классын бағдарлау тиіс (атаулы мәннің құрамына кіре ме, егер ол кірсе, онда қандай типтегі атаулы мәннің құрамына кіреді). Дәстүр бойынша тәсілдер екі топқа бөлінеді: әр түрлі қолмен жасалған ережелерді пайдалану және машиналық оқыту алгоритмдерін қолдану.

Бұл мәселелерді шешудің көптеген тәсілдері баршылық. Соңғы 50 жыл ішінде ғылыми мәтіндерден терминологияны автоматты түрде шығарып алу, компьютерлік тіл білімі бағытында алгоритмдер мен бағдарламалық жасақтаманы дамытудың дербес саласы ретінде қалыптасты және оған айтарлықтай жетістіктер алынды. Алайда, бұл әдістер нақты нәтиже бермейді және айтарлықтай жоғары дамытуды қажет етеді. Мәтінді өңдеудің және әсіресе терминологияны шығарудың қазіргі кездегі ең өзекті тәсілдерінің бірі машиналық оқытуды қолдану деп санауға болады [7].

Сөз деңгейінде жұмыс істейтін машиналық оқыту әдістері word2vec, Elmo (Embeddings from Language Model), BERT (Bidirectional Encoder Representations from Transformers) және т.б. сияқты алдын ала оқытылған тілдік модельдердің векторлық көріністерін әрбір токенге сәйкестендіреді [8]. Бұл жұмыста біз BERT моделін қарастырамыз, өйткені ол мәтіндік ақпаратты өңдеудің өте кең міндеттері үшін өзін жақсы дәлелдеді.

BERT (Bidirectional Encoder Representations from Transformers) – бұл «трансформерлер» ретінде белгілі бір тізбектермен жұмыс істеу үшін жаңа архитектураның машиналық оқытуға негізделен табиғи тілді өңдеу әдісі. Бұл 2018 жылы Google инженерлері жасаған және жариялаған жаңа технология болып табылады [9]. BERT алгоритмінің ерекшелігі – сөйлемдегі

немесе сұрақтағы сөздер жиынтығын жақсырақ түсінуге мүмкіндік беретін екі бағытты «кодерлер» қолданылауында. BERT әдісінің көмегімен тілдік модельдерін жаттықтыру сөйлемдегі сөздердің толық жиынтығы негізінде екі бағытты оқыту болып табылады, ал NLP дәстүрлі модельдері тілдік модельдерді сөздердің реті бойынша (оңнан солға немесе солдан оңға қарай) оқытады [10].

Егер бұл мәселені символдар деңгейінде шешу туралы айтатын болсақ, онда негізгі ой мәтінді символдар деңгейінде кодтау және олардың векторлық көріністерін алу (әдетте, ереже бойынша мұнда үйірткілі немесе рекуррентлі қабаттар қолданылады), содан кейін тізбекті тэгтеудің әдеттегі мәселесін шешу болып табылады. Жұмыста орыс тіліне қатысты архитектурасы сипатталған. Модельде BI-LSTM (Bidirectional long short-term memory) және CRF (Conditional Random Fields) қабаттарын сәтті біріктіреді, бұл атаулы мәндерді тану сапасын айтарлықтай арттыруға мүмкіндік береді [11].

Қазіргі уақытта ғылыми әдебиеттерді автоматты саралау ерекше өзекті мәселе болып отыр. Жарияланымдар санының тез өсуі жағдайында трендтер мен үрдістерді бақылау өте маңызды. Эксперименттердің шарттары, мәліметтер жиынтығы, алынған сапа көрсеткіштері және т.б. туралы ақпарат алудан басқа, шешілетін міндет, авторлардың қосқан үлесі, қолданылатын тәсілдер, алынған тұжырымдар сияқты аспектілерді бөліп көрсету маңызды міндет болып табылады. Мұндай ақпарат мәселені шешу үшін қандай әдістер қолданылатынын, қарастырылып отырған ғылыми жұмыстардың жаңалығын және т. б. анықтауға көмектеседі. Бұл процесті автоматтандыру уақыт пен еңбек шығындарын азайтуға мүмкіндік береді. Сондықтан ғылыми мәтіндерден ақпарат алудың әдістері мен құралдарын әзірлеу бүгінгі таңда ең үлкен сұранысқа ие міндеттердің бірі. Машиналық оқыту әдістерінің көмегімен біз тіл мамандарының жұмысын едәуір қысқарта отырып, сөйлеудің кез-келген бөлігінің сөздерінен (тек зат есімдер, сын есімдер мен етістіктер ғана емес) тұратын атаулы мәндерді ала аламыз.

Осы еңбек шеңберінде атаулы мәндер мен аспектілерді алу міндеттерін шешу үшін жаңа алгоритмдер мен әдістерді ұсыну, сонымен қатар қазақ тіліндегі мәтіндерден ақпарат алу жүйесінің негізгі компоненттерін іске асыру жоспарлануда. Бұл зерттеу белгіленген деректердің өте көп мөлшерін оқыту үшін қажет етпейтін әдістерге бағытталған, білім беру негізіндегі тасымалдау техникасына негізделген (transfer learning) және әртүрлі пәндік салалардың мәтіндеріне өте қарапайым түрде бейімделуі мүмкін. Жобаны іске асыру барысында деректерді дайындау және белгілеу әдістемесі әзірленетін болады. Осы әдістемеге сәйкес ақпаратты шығарып алудың заманауи автоматты әдістерінің сапасын бағалау және оқыту үшін негіз болатын қазақ тіліндегі мәтіндер корпусын әзірлеу жоспарлануда.

Әдебиеттер тізімі

1. . Marshalova A.E., Bruches E.P., Batura T.V. Aspect extraction from scientific paper texts. *Software & Systems*. 2022. Vol. 35. N 4. pp. 698–706. (in Russian)
2. Indurkha N., Damerau F.. *Handbook of Natural Language Processing* (2nd. ed.). Chapman &Hall/CRC. 2010. 704 p.
3. Kim Sang E., Meulder F.. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 2003. pp. 142–147.
4. Krallinger M., Leitner F., Rabal O., Vazquez M., Oyarzabal J., Valencia A. Overview of the chemical compound and drug name recognition (CHEMDNER) task. *Proceedings of the Fourth BioCreative Challenge Evaluation Workshop*. 2013. Vol. 2. pp. 6–37.
5. Jurafsky D., Martin J. *Speech and Language Processing* (2nd Edition). Prentice-Hall, Inc., USA. 2009. 988 p.
6. Dementeva Ya.Yu., Bruches E.P., Batura T.V. Terms extraction from texts of scientific papers. *Software & Systems*. 2022. Vol. 35. N 4. pp. 689–697. (in Russian)

7. Le A., Arkhipov M., Burtsev M. Application of a hybrid Bi-LSTM-CRF model to the task of Russian named entity recognition. *Artificial Intelligence and Natural Language*. Springer International Publishing, 2018. pp. 91–103.
8. Devlin J., Chang M. W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*. ACL, 2019. pp. 4171–4186.
9. Mansurova M.E., Dosanov B.B., Chikibaeva D.Yu. Мәтіндерден атаулы мәндерді алу үшін машиналық оқыту әдістерін қолдану. *Алматы энергетика және байланыс университетінің Хабаршысы*. 2020. № 2(49). pp. 142–147.
10. Ma X., Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 2016. pp. 1064–1074. DOI: 10.18653/v1/P16-1101.

УДК 004.3'2

СОЗДАНИЕ АДАПТИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ОБРАЗОВАНИЕМ С ИСПОЛЬЗОВАНИЕМ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ

Турашев Марат Джумурзаевич
turashev.kz@gmail.com

Магистрант 2 курса факультета информационных технологий ЕНУ им. Л.Н.Гумилева,
Астана, Казахстан

Научный руководитель – Д.А.Тусупов

Аннотация. В последние годы использование интеллектуальных технологий в системах управления образованием становится все более популярным. Целью проводимого исследования является рассмотрение применения технологий биометрического распознавания лица на основе программно-аппаратного обеспечения Hikvision и комплекса программного обеспечения компании Microsoft Power BI в учебных организациях.

Ключевые слова: Hikvision, Microsoft Power BI, интеллектуальная система, терминал, система.

Введение: Главной целью любого учебного заведения является предоставление качественного образования своим студентам. Успех учебного заведения зависит от посещаемости его учащимися занятий, что напрямую связано с качеством получаемого ими образования. Однако традиционные методы учета посещаемости, такие как бумажные реестры посещаемости или ручной ввод данных, отнимают много времени и подвержены ошибкам. Для решения этих проблем использование интеллектуальных технологий, таких как распознавание лиц, визуализация данных и глубокое обучение, становится все более популярным в системах управления образованием.

В этой научной статье предлагается адаптированная система управления образованием, которая использует эти технологии для повышения точности и эффективности учета посещаемости учащихся.

Основная часть:

В сфере образования для интеллектуальной системы управления используются различные технологии, такие как распознавание лиц, RFID и биометрия. Эти системы могут предоставлять данные о посещаемости в режиме реального времени, снижать административную нагрузку и обеспечивать анализ структуры посещаемости. Кроме того, интеллектуальные системы управления образованием могут помочь учебным заведениям выявлять учащихся из группы риска и оказывать адресную поддержку для улучшения результатов обучения учащихся. Как пример, можно рассмотреть различные системы зарубежных учебных заведений: система Техасского университета A&M использует биометрическую систему посещаемости, которая фиксирует лица и отпечатки пальцев