



UBMK'24

**Bildiriler Kitabı
Proceedings**

Editor Eref ADALI

**9. Uluslararası Bilgisayar Bilimleri ve
Mhendislięi Konferansı**

**9th International Conference on
Computer Science and Engineering**

26-27-28 Ekim (October) 2024 Antalya - Trkiye

9. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (UBMK'2024)

9th International Conference on Computer Science and Engineering

26-28 Ekim 2024 Akdeniz Üniversitesi Antalya Türkiye
26-28 October 2024 Akdeniz University Antalya Türkiye

Telif Hakkı

Bu elektronik kitabın içinde yer alan tüm bildirilerin telif hakları IEEE'ye devredilmiştir. Bu kitabın tamamı veya herhangi bir kısmı yayımcının izni olmaksızın yayımlanamaz, basılı veya elektronik biçimde çoğaltılamaz. Tersî davranışta bulunanlara ABD Telif Hakkı Yasalarına göre ceza uygulanır.

Copyright and Reprint Permission

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. Copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyright Manager at pubs-permission@ieee.org

All right reserved. Copyright C 2024

IEEE Catalog Number : CFP24L97-CDR

ISBN : 979-8-3503-6587-0

Additional copies may be ordered from:
Curran Associates, Inc
57 Morehouse Lane Red Hook, NY 12571 USA
Phone: (845) 758 0400
Fax: (845) 758 2633
E-mail: curran@proceeding.com

	ID	Title	Authors	Pages
NLP	1850	HTV-News: A New Dataset with High Novelty Rate for Turkish Text Summarization	Nihal Zuhal Kayalı Sevinç İlhan Omurca	1 - 6
	1853	Sentiment Analysis of Social Network Comments in Uzbek Language	Saboxat Allanazarova	7 - 10
	1861	An Online Platform for Uzbek-Russian and Russian-Uzbek Parallel Corpora	Nigmatova Lolakhon Khamidovna Saidova Mokhira Rasulevna Djuraeva Zulkhumor Radzhabovna Sharipov Sokhib Salimovich Avezov Sukhrob Sobirovich	11 - 16
	1863	Yapım Ekleri ile Özbekçe Sözcük Türetme Deriving Uzbek Words with Derivational Suffixes	Eşref Adalı Khamroeva Shahlo Mirdjonovna	17 - 21
	1864	Database of Grammatical Form and Meaning Correspondence for Uzbek-English Translation Software	Jamoliddinova Odina Rasulovna Muhammedova Saodat Xudoyberdiyevna Khamroeva Shahlo Mirdjanovna Ergasheva Guli Ismoil Qizi Qodirova Madinabonu Murodjon qizi Kholova Muyassar	22 - 26
	1868	Words Speak Louder Than Actions: Decoding Emotions Through NLP	Melda Paksoy Gokhan Bakal	27 - 31
	1872	Özbek Dilinin Pankronik Derleminin Oluşturulması (Kutadgu bilig’ Nüshalarına Göre)	Xamroyeva Şahlo Mircanovna Avezova Nigora Safarovna Toştemirova Sitara Bahodir kızı Qodirova Fazilat Şukurovna Atanazarova Shoxida Baxramovna	32 - 37
	1873	Preserving Semantic Integrity in Paraphrasing Texts: Rule-Based Paraphrasing	Zarnigor M. Khayatova Jurakuziyev Nodirbek Kholova Muyassar	38 - 42
	1975	NLLB-Based Uzbek NMT: Leveraging Multisource Data	Nilufar Abdurakhmonova Adkham Zokhirov Mohirdev Mukhammadali Salokhiddinov Anvar Narzullayev Ayrat Gatiatullin	43 - 47
	1976	The Problem of the Archaic Words’ Semantic Description in the Alisher Navoi Authorship Corpus	Shuhrat Sirojiddinov Manzura Abjalova Sulton Normamatov Nargiza Gulomova	48 - 53
	1979	Multi-Label Emotion Classification with Fine-Tuned BERT and Contrastive Learning	Ammara Naseem Khan Mehmet Nafiz Aydın	54 - 59
	1885	Algorithm for Aligning Paragraphs and Sentences in Aligner Tool	Elov Botir Boltayevich Khamroeva Shahlo Mirdjonovna Dauletov Adilbek Yusupbayevich Matyakubova Noila Shakirjanovna	60 - 63
	1888	Semantic and Grammatical Issues in Translating Idioms with Automatic Translation Systems	Manzura Abjalova Sarvinoz Sharipova	64 - 69
	1889	Improving Methodology for Selecting Fiction Works for Grades 5-9	Khabibulla Madatov Sapura Sattarova	70 - 73
	1897	Innovating SQL Automation: Evaluating Open-Source Large Language Models with a Dual-Stage Approach for Corporate Data Solutions	Erhan Arslan Eugen Harinda	74 - 79
	1898	Development of a Method for Automatic Extraction of Ontology Entity Names from Natural Language Texts	Sadirmekova Zhanna Elmira Daiyrbayeva Karbozova Indira Bekbolatov Samat	80 - 85
	1899	UZBEKCORPORA.UZ: Creation of Author Corpora on the Platform	Karshiev Abduvali Berkinovich Karimov Suyun Amirovich Tursunov Mukhammadsolikh Sadin ugli Kudratov Iskandar Makhmidov Shirinboy B otirovich	86 - 90
	1901	A Technique for Automatic Extraction of Basis Words: a Case Study on “Uzbek Primary School Corpus”	Khabibulla Madatov	91 - 94

		Shukurla Bekchanov Surayyo Khajibaeva	
1908	Özbek Dili Metinlerindeki Eşgönderge Çözümlemesi Algoritması Algorithm of Coreference Resolution in Uzbek Texts	Elov Botir Boltayevich Abdisalomova Şahlo Abdimurod kızı Karimova Xurriyat Şaripovna	95 - 100
1909	The Process of Lemmatization and Stemming in the Automatic Morphological Analysis of Uzbek Texts	Manzura Abjalova Eşref Adalı Munojot Adilova	101 - 106
1913	Turkish Question - Answer Dataset Evaluated with Deep Learning	Kadir Tutar Olçay Taner Yıldız	107 - 111
1914	Exploring the Semantic Complexity of Adjective-noun Collocations Between Uzbek and English for Improved Machine Translation	Nilufar Abdurakhmonova Nargiza Shamieva Eşref Adalı	112 - 115
1915	Linguistic Support and Algorithm of Phrase Alignment in Aligner Tool	Xoshimovna Shohida Shahabitdinova Matyakubova Noila Shakirjanovna	116 - 120
1946	Evaluating the Performance of PEFT and LoRA Approaches in Transformer-based Architecture for Handwritten Character Recognition	Pinar Savci Bihter Das Resul Das	121 - 126
1947	Evaluating the Performance of Turkish Automatic Speech Recognition Using the Generative AI-based Whisper Model	Tunahan Gokcimen Bihter Das Resul Das	127 - 131
1961	Özbek Dilinde Kelimelerin Kök ve Köklerini Ayırma Algoritmaları Algorithms for Parsing Roots and Stems of Words in Uzbek Language	Elov Botir Boltayevich Xusainova Zilola Yuldashevna Umirova Svetlana Mamurjonovna Normamatov Sultonbek A.Şahlo Abdimurod kızı Mahmadiev Shavkatjon	132 - 136
1970	Özbekçe Deyimlerin Dilsel Modellemesinin DDİ Açısından Önemi The Importance of Linguistic Modeling of Uzbek Idioms in NLP	Gülyamova Shahnoza Kahramonovna Nurboyeva Maftuna Vahobcon kızı	137 - 140
1981	IT Service Desk Ticket Classification via Large Language Models	Ezgi Paket Göksu Şenerkek Fatma Betül Akyol Furkan Salman	141 - 146
1998	Enhancing the Quality of Clustering Job Skills Through the Fine-tuning of the Sentence Transformer Model	V. S. Ramazanova A. S. Yerimbetova M. A. Sambetbaeva Zh. B. Sadirmekova S. K. Serikbayeva	147 - 152
2004	Named Entity Recognition from Kazakh Speech	Bauyrzhan Kairatuly Madina Mansurova	153 -156
2005	Sentiment Analysis of Reviews for E-Commerce Applications	Elif Hanife Aydoğan Feyza Yıldırım Okay	157 - 162
2011	LLM and RAG-Based Question Answering Assistant for Enterprise Knowledge Management	Gürkan Şahin Karya Varol Burcu Kuleli Pak	163 - 168
2018	Named Entities Recognition in Kazakh Text by SpaCy NER Models	Nurzhan Mukazhanov Aigerim Yerimbetova Mussa Turdalyuly Bakzhan Sakenov	169 - 174
2022	Efficient Unsupervised Domain Adaptation with	M. Rawhani D. Karaboğa U. Nalbantoğlu A. Baştürk B. Akay	175 - 180
2032	Aligning Sentences for Kazakh-Turkish Parallel Corpora	Diana Rakhimova Eşref Adalı Aidana Karibayeva	181 - 186
2048	Automated-Computer Based Assessment of Free-Text Exam Answers By Transformers	Burak Keskin Melih Günay	187 - 192

	2073	Knowledge-Augmented Large Language Model	Başak Buluz Kömeçoğlu Burcu Yılmaz	193 - 198
	2092	LLMs for Document-level Text Simplification in Turkish Foreign Language Learning	Fatih Bektaş Kutay Arda Dinç Gülşen Eryiğit	199 - 203
	2105	Mining Software Requirements From Turkish Texts:Techniques and Challenges	Serhat Uzunbayır Senem Kumova Metin	204 - 209
	2108	Using LLMs for Annotation and ML Methods for Comparative Analysis of Large-Scale Turkish Sentiment Datasets	Ercan Ezin Rukiye Savran Kiziltepe Murat Karakus	210 - 215
	2122	Sentiment Analysis for Hotel Reviews in Turkish by using LLMs	Ata Onur Özdemir Efe Batur Giritli Yekta Said Can	216 - 220
	2123	Semantic Knowledge Base for the Emotional Coloring Analysis of Kazakh Texts	Banu Yergesh Dinara Kabdylova Tilekbergen Mukhamet	221 - 224
	2140	Adaptive Batch Budget for LLM Inference	Cağrı Yeşil Berhan Turku Ay Funda Ay Ak Öykü Berfin Mercan Oğuzhan Nefesoğlu	225 - 229
	2154	Open Data Processing for Analysis of Behavior of Users in Digital Environment	Nikolai Prokopyev Pavel Ustin Fail Gafarov	230 - 232
	2164	Evaluating Turkish BERT-based Language Models for Effective Customer Feedback Interpretation in CRM	Can İşcan Muhammet Furkan Özara Ahmet Erkan Çelik Akhan Akbulut	233 - 238
	2175	Does Prompt Engineering Help Turkish Named Entity Recognition?	Arda Serdar Pektezol Ahmet Batuhan Ulugergerli Volkan Öztoklu Seniz Demir	239 - 243
	2188	Araneum Uzbekicum: A Gigaword Web-Crawled Uzbek Corpus	Vladimi-r Benko Radovan Garabik Shahlo Khamroyeva	244 - 2149
	2189	Automatic Topic Detection in Large Text Data in Uzbek using Clustering Methods	Umirova Svetlana Kholmuhamedov Bakhtiyor Karimov Suyun Narzieva Mamura	250 - 255
	2193	Machine Learning-Based Synonymous Word	Diana Rakhimova Madina Mansurova Nurakhmet Matanov Amira Yerik	256 - 260
	2194	Enhancing Text-to-SQL Conversion in Turkish: An Analysis of LLMs with Schema Context	Ferhat D emirkıran Ali Kemal Coşkun Yavuz Kömeçoğlu Başak Buluz Kömeçoğlu Ramazan Güven	261 - 265
	2203	Linguistic Knowledge Graph "Turklang" as Universal Model for Linguistic Resources and Tools in Turkic Languages	Ayrat Gatiatullin Nikolai Prokopyev Lenara Kubedinova Nilufar Abdurakhmonova Rustam Burnashev	266 - 270
	2216	Requirements for the Development of a Website Builder with Adaptive Design	Gulmira Bekmanova Banu Yergesh Assel Omarbekova Laura Orynbay Aiganym Bessembayeva Dinara Kabdylova Altanbek Zulkhazhav Beibarys Sultan	271 - 276
AI	1870	A Modified Marine Predators Algorithm for Numerical Function Optimization	Resul Özdemir Murat Taşyürek	277 - 281

		Veysel Aslantaş	
1884	Document Classification and Key Information Extraction Using Multimodal Transformers	Mehmet Selman Baysan	282 - 287
		Furkan Kızılay Ayşe İrem Özmen Gökhan İnce	
1926	Employee Turnover Prediction on Synthetic and Real Datasets	Azhar Murzaeva Volkan İllik Kadir Yunus Koc	288 - 292
1937	Data Augmentation in Remote Sensing Image Change Captioning	Orkhan Karimli İlyas Mustafazade Ali Can Karaca Fatih Amasyalı	293 - 298
1968	Comparative Evaluation of Word2Vec and Node2Vec for Frequently Bought Together Recommendations in E-commerce	Mustafa Keskin Enis Teper Alya Kurt	299 - 303
1971	Disrupting the Retrieval Phase in RAG-Based LLM Chatbots through Input Manipulation	Felix Pitterling Georgii Koshelev Oliver Haddad Rodrigo Teran	304 - 308
1976	Mutation Testing Reinvented: How Artificial Intelligence Complements Classic Methods	Serhat Uzunbayir Kaan Kurtel	309 - 314
1997	Prompting Large Language Models for Aerial Navigation	Emirhan Balcı Mehmet Sarıgül Barış Ata	315 - 320
2006	Açıklanabilir Yapay Zeka Üzerine Bir İnceleme A Review of Explainable Artificial Intelligence	Samed Al Şeref Sağıroğlu	321 - 326
2020	Investigating the Effects of Pre-trained Deep Learning Models and Fusion Techniques on Fruit Segmentation Performance	Esmâ İbiş Aybars Uğur	327 - 332
2034	Beyin Tümörü Ameliyatlarında Beyin Kayması Probleminin Telifisi: Bilgisayar Destekli Yöntemler Üzerine Bir Derleme Compensating for the Brain Shift Problem in Brain Tumor Surgeries: A Review of Computer-Assisted Methods	Ayşe Gül Eker Meltem Kurt Pehlivanoğlu Nur Banu Albayrak Nevcihan Duru Tolga Turan Dünder	333 - 338
2039	AI-Enhanced Endometrial Cancer Diagnosis System	Raziye Aslıhan Kürkcü Işın Yeşim Yeşilkaya Baylan Erkut Attar Burcu Selçuk Tacha Serif	339 - 343
2040	El Titreme Sinyalinden Transformer Temelli Parkinson Hastalığı Sınıflandırması Transformer Based Parkinson's Disease Classification from Hand Tremor Signal	Murat Atçeken Lütfü Hanoğlu Mehmet Ersin Bitirgen Bahadır K. Güntürk İbrahim Özşeker	344 - 347
2042	Üretken Yapay Zeka Fırsatlar ve Tehditler: Bibliyometrik Analiz Generative Artificial Intelligence Opportunities and Threats: Bibliometric Analysis	Betül Ersöz Şeref Sağıroğlu Halil İbrahim Bülbül	348 - 353
2053	Users Acceptance of Generative Artificial Intelligence Based on ChatBot Use	Ayşe Yeşim Mutlu	354 - 359
2054	MMBAtn: Max-Mean and Bit-wise Attention for CTR Prediction	Hasan Saribas Cagri Yesil Serdarcan Dilbaz Halit Orenbas	360 - 365
2065	Türkçe Otomatik Konuşma Tanıma Sistemleri İçin Sentetik Veri Üretme Yöntemi Synthetic Data Generation Method for Turkish Automatic Speech Recognition Systems	Hilal Tekgöz Harun Uz Muhammed Murat Özbek Tolga Büyüktanır	366 - 370
2066	A Federated Learning Framework for Classifying the Images in Ultrasonic Nondestructive Testing	Abdulkadir Gulsen Hilal Hacilar	371 - 375

		Burak Kolukisa Burcu Bakir-Gungor	
2071	AI Assisted Customer Review Sentiment Analysis and Department Classification Tool	Burcu Seçuk Berat Ilgaz Dursun Tacha Serif	376 - 381
2075	Histopatolojik Görüntülerde Doğru Mitoz Tespiti için Geliştirilmiş Renk Normalleştirme Yöntemi Enhanced Stain Normalization Method for Accurate Mitosis Detection in Histopathological Images	Refik Samet Nooshin Nemati Emrah Hançer Serpil Sak Bilge Ayça Kırmızı	382 - 387
2110	ATGRUVAE: Reducing Noise and Improving Forecasting Performance in Stock Data	Hüseyin Akkas Burak Kolukisa Burcu Bakir-Gungor	388 - 392
2113	Evaluating the Impact of Sentiment Analysis on Deep Reinforcement Learning-Based Trading Strategies	Mustafa Etcil Burak Kolukisa Burcu Bakir-Gungor	393 - 398
2118	PsychSynth: Advancing Mental Health AI Through Synthetic Data Generation and Curriculum Training	Vedanta S P Madhav Rao	399 - 404
2121	Covid-19 Bilgisayarlı Tomografi Görüntülerinin Segmentasyonu için Topluluk Öğrenmesiyle Birleştirilmiş U-Net Modeli U-Net Model Combined with Ensemble Learning for Segmentation of Covid-19 Computed Tomography Images	Furkan Atlan İhsan Pençe	405 - 410
2129	Turkish Sign Language Video Generation from Pose Sequences Using a Conditional GAN Model	Feyza Özkan Hatice Kübra Tekin Hacer Yalim Keles	411 - 416
2131	Statistical Attention Layer for Neural Network Training	Ergun Biçici Hasan Saribaş	417 - 420
2134	Dudak Şekillerinden Kişilik Analizi: Marifetname ve Açıklanabilir Yapay Zeka Yaklaşımı Personality analysis from lip shapes: Marifetnâme and explainable artificial intelligence approach	Semra Çelebi İbrahim Türkoğlu	421 - 426
2148	Breast Cancer Detection Using Deep Learning Models	Erkan Akkaş Mustafa Yapar Şeref Sağıroğlu Ali Öter Betül Ersöz	427 - 430
2191	CGAN Mimarisi ile Kısıtlı Veri Ortamında Kestirimci Bakım Performansının İyileştirilmesi Improving Predictive Maintenance Performance in Limited Data Environment with CGAN	Mehmet Musa Özcan Uğurhan Kutbay	433 - 438
2207	Integrating the Focusing Neuron Model with N-BEATS and N-HITS	Şuayb Talha Özçelik Faik Boray Tek	439 - 442
2217	Advanced Convolutional Neural Networks for Plastic Classification in Recycling Systems	Nazym Alimbekova Ainur Zhumadillayeva Sunggat Aiymbay	443 - 447
ML	1846 Semi-Supervised Learning for Sensor-Based Flash Point Prediction in Oil Industry	Mert Sülük Şule Gündüz Öğüdücü	448 - 452
	1847 Robust Stacked Ensemble Model for Lung Cancer Diagnosis	Ayhan Akbas Gonca Buyrukoglu Selim Buyrukoglu	453 - 457
	1856 Kod Kusurlarının Tespitinde Makine Öğrenmesi Tekniklerinin Başarım Analizi The Performance Analysis of Machine Learning Techniques in Code Smell Detection	Arman Yavuz Oya Kalipsiz	458 - 463
	1860 Enhancing Financial Time-Series Analysis with TimeGAN: A Novel Approach	Cemal Öztürk	464 - 467
	1869 From Traditional to Deep: Evaluating Sentiment Analysis Models on a Large-Scale Tweet Dataset	Alisahib Mammadov Gokhan Bakal	468 - 473

1871	Comparison of Feature Selection Methods for Mechanical Properties of Cold Rolled Products in Flat Steel Manufacturing	Didem Bakiler Ilme Merve Öper E. Fatih Yetkin	474 - 478
1883	A Machine Learning Approach to Steel Sheet Production Surface Quality	Asena Öztürk Mehmet Nafiz Aydın	479 - 484
1938	NLP Tabanlı Özellik Çıkarımına Dayalı Makine Öğrenme Algoritmaları Kullanılarak SQL Enjeksiyonu Ataklarının Tespit Edilmesi Detection of SQL Injection Attacks Using Machine Learning Algorithms Based on NLP-Based Feature Extraction	Hakan Can Altunay	485 - 489
1940	Enhancing Driver Injury Severity Prediction Using Optimized Oversampling and Feature Selection Techniques	Vahide Nida Kılıç Gizen Mutlu Esra Saraç Çiğdem Acı Murat Özen	490 - 495
1972	Predicting Electric Vehicle Adoption in the EU: Analyzing Classification Performance and Influencing Attributes Across Countries, Gender, and Education Level	Mert Kumbasar Gül Tokdemir Thouraya Gherissi Labben Gurdal Ertek	496- 499
1977	Arabam.com Kurumsal Üyelerinin Satın Alma Davranışlarının Analizi	Dilan Oragaz İsmail Duru	500 - 505
1983	Makine Öğrenmesi ve Derin Öğrenme Modelleri Kullanılarak MR Görüntülerinden Demans Sınıflandırma	Meltem Kurt Pehlivanoğlu Onur Varol Osman Aldemir Ata Emir Uncu Nevcihan Duru	506 - 511
1995	Indoor Localization with GravNetConv and Dynamic Graphs	Mert Bayraktar Alper Ozcan Umit Deniz Ulusar	512 - 516
2023	GPS Spoofing Detection on Autonomous Vehicles with XGBoost	Emre İşleyen Şerif Bahtiyar	517 - 522
2025	SmartSniffer: Predicting Food Spoilage Time with an Electronic Nose-based Gas Monitoring Apparatus Utilizing a Two-Stage Pipeline Model	Pandey Shourya Prasad Sreyas Janamanchi Barath S Narayan Madhav Rao	523 - 528
2064	GNF: Generative Natural Faces Dataset for Face Expression Recognition Task	Cansu Özer Volkan Dağlı Mustafa Kağan Gürkan	529 - 533
2081	Incorporating Knowledge Graph Embeddings into Graph Neural Networks for Sequential Recommender Systems	Kazım Emre Yüksel Susan Üsküdarlı	534 - 539
2095	The Classification of Human and Cognitive Architecture Time Estimation Using Machine Learning Methods	Behiye Şahin Sonay Duman	540 - 543
2101	Optimization of Roadside Assistance Tow Truck Services Provided to Insured Individuals with Traffic and Comprehensive Insurance Policies	Serkan Kırca Fatma Yağmur Erbaş Sedat Cebesoy Buse Özkanat Aysun Yıldırım Ceren Yıldırım	544 - 549
2104	Optimizing Recommendation Systems By Fusion of KNN, Singular Value Decomposition, and XGBoost for Enhanced Performance	Mohammed Basim Mohammed Mohammed Erkut Arıcan	550 - 555
2119	Spoken Accent Detection in English using Audio-Based Transformer Models	Oguzhan Ozturk Hasan Kilimci H.Hakan Kilinc Zeynep Hilal Kilimci	556 - 561
2136	Analysis of Adversarial Training for Resilient Image Recognition Against Different Attacks	Emir Can Karakuş Fırat Kutluhan İslim Ece Gelal Soyak	562 - 567

	2150	Business Process Management Anomaly Detection through Semantic Embedding-Integrated Graph Neural Networks	Teoman Berkay Ayaz Ege Gülce Stanley Hsu Alper Özcan Akhan Akbulut	568 - 573
	2155	Multi-Aspect Anomaly Detection with Graph Neural Networks and Kolmogorov-Arnold Networks in Business Process Management	Teoman Berkay Ayaz Ege Gülce Stanley Hsu Alper Özcan Akhan Akbulut	574 - 579
	2156	Comparison of a Deep Learning and a Hybrid Model for Classification of an Unbalanced Urgent Cases Dataset for Human Faces	Faruk Özgür Neslihan Arıkan Özge Öztimur Karadağ	580 - 585
	2171	Estimating the Manufacturing Cost of a Metal Part from Textual and Geometric Features	Talha Rehman Abid Mert Daloğlu Cem Yıldız Ali Erman Erten Kamer Kaya	586 - 591
	2176	Object Detection in Hyperspectral Images with Unsupervised Domain Adaptation	Sinem Aybüke Şakacı Alp Ertürk Erchan Aptoula	592 - 596
	2182	Machine Learning Approaches to Predict Thyroid Cancer Recurrence: A Comparative Study	Candide Ozturk Ozgur Sagir Ulas Vural	597 - 602
	2197	Yangın Söndürme Süreçlerinde Su Tüketiminin Makine Öğrenmesi ile Tahmini Prediction of Water Consumption in Fire Extinguishing Processes Using Machine Learning Approaches	Emin Ölmez Ahsen Usta Orhan Akbulut	603 - 606
	2202	Synthetic Vibration Data Generation and Fault Classification in CNC Machines Using Transformer GANs and ConvLSTM Networks	Özlem Erbay Batrav Erbay	607 - 612
IoT	1862	Design and Implementation of a Management System for Wireless Electronic Combination Locks	Batuhan Kol Metin Bilgin	613 - 618
	1887	Solar IoT: Monitoring the Orientation and Electrical Parameters of the Solar Panel	Hakan Dalkılıç Oğuz Gora	619 - 624
	1973	Multifunctional Smart Spoon for Parkinson's: Stability Enhancement and Diagnostic Tools	Divyansh Singhal Sasi Snigdha Yadavalli Nupur Patil Siddharth Chauhan Madhav Rao	625 - 629
	1988	An AI-Assisted Autonomous IoT Agent for Smart Spaces	Yakup Kayataş Sanem Kabadayı	630 - 635
	1991	Advanced Detection and Prevention of Sinkhole Attacks in 6TiSCH Networks	Burak Aydın Hakan Aydın Sedat Görmüş	636 - 641
	2000	Comparison of Different Weather Data Acquisition Methods	Emre Evcin Yusuf Murat Erten	642 - 647
	2057	Employing Digital Twin to Forest Fire Management Systems	Bugra Aydın Sema Fatma Oktug	648 - 653
	2085	Distributed Key Value Store for IoT Edge Devices	Burak Aslantaş Elif Nurdan Pektaş Şebnem Baydere	654 - 659
EMB	2096	Soft Error Reliability Assessment of TinyML Algorithms on STM32 Microcontroller	Ahmet Selim Karakuş Osman Buğra Gökteş Sadık Akgedik Sanem Arslan	660 - 664
HUM	1999	Development of An Algorithm for Converting Json Formats to Xml by Forming its File Data Structure	Aigul Mukhitova Aigerim Yerimbetova Vladimir Barakhnin	665 - 670

			Elmira Daiyrbayeva	
	2130	Facial Emotion Recognition for Imitation in Human-Robot Interaction	İbrahim Yanç Aykan İpek Selma Yılmazyıldız Kayaarma	671 - 676
HPER	2030	A Glimpse to Scalable LLM Architectures: Building Real-Time Sentiment Analyser with Kafka and FastAPI	Uzay Çetin Yunus Emre Gündoğmuş	677 - 681
	2165	Parallelization of BitColor Algorithm via Multithreading and GPU for Graph Coloring	Burak Kocausta Gizem Sungu Terci Alp Arslan Bayrakci	682 - 687
BCHN	1922	Security Dynamics of Blockchain-Enabled SDN Systems: A Taxonomic Approach	Deniz Dudukcu Murat Karakus	688 - 693
	2056	Comparative Performance Analysis of Ethereum and Optimism Smart Contracts in Health Insurance	Beyhan Adanur Dedetürk Bilge Kagan Dedetürk	694 - 699
BIO	1907	Disease Prediction From Human Microbiome by Utilizing Machine Learning	Bahadır Emin Temel Bora Kocapinar Zerrin Isik	700 - 705
	1923	Medikal Görüntülerin Etkili Bir Şekilde Sınıflandırılması için Dikkat Tabanlı Temsil Öğrenimi	Selim Arslan	706 - 711
		Attention Based Representation Learning for Effective Classification of Medical Images	Kemal Polat	
	1933	Utilizing Tree-Based Algorithms for Genetic Variant Interpretation	Rumeysa Aslıhan Ertürk Mehmet Baysan	712 - 717
	1978	A Test Bench for Replicating Human Breathing: Evaluating Thermal Effects of N95 Filtering Facepiece Respirator Leaks ' Preliminary Findings	Geoffrey Marchais Barthelemy Topilko Mohamed Arbane Jean Brousseau Clothilde Brochot Yacine Yaddaden Ali Bahloul Xavier Maldague	718 - 723
	1979	Development of Multiphysics Models for the Study of Airflow and Thermal Effects During the Use of Filtering Facepiece Respirators	Barthelemy Topilko Geoffrey Marchais Mohamed Arbane Jean Brousseau Ali Bahloul Xavier Maldague Clothilde Brochot Yacine Yaddaden	724 - 730
	1994	Awareness of Gamification as a Non-Pharmacological Intervention in Sleep Health Research: A Systematic Literature Review	Arife Gülah Erol Murat Yılmaz Paul M. Clarke	731 - 736
	2003	Optik Çukur Bölütlemesi Segmentation of the Optic Cup	Saadet Aytaç Arpacı Songül Varlı	737 - 741
	2090	Clinically Significant Prostate Cancer Detection and Diagnosis in Bi-parametric MRI with Deep Learning Models	Clinton Binda Asoh-Itambi Mecit Yüzkat Songül Varlı	742 - 744
	1993	Malaria Incidence Prediction Using Climate Factors with Machine Learning Models	Ayoade Adeyemi Cyrille Mesue Njume Goodnews Akindele Nana Aisha Umar Ayodele James Oyejide	748 - 753
CDSec	1857	Dual-layered Approach for Malicious Domain Detection	Nadide Bilge Doğan Alp Barış, Beydemir Şerif Bahtiyar Umutcan Doğan	754 - 759
	1953	Analysis of the Zero-Day Detection of Metamorphic Malware	Sibel Gulmez Arzu Gorgulu Kakisim İbrahim Sogukpınar	760 - 765
	1960	White-Box Style Intrusion Detection System Integration into Industrial Metaverses	Yasir Kılıç	766 - 771

			Vahide Nida Kılıç Ali İnan	
2037	A Practical Investigation of Spear Phishing Spam Emails: Comparative Analysis and Evaluation	Kendrick Kurt Günter Bollens	772 - 777	
2091	Multipurpose Malware Detection System	Mert Gursimsir Cem Ayar Ibrahim Sogukpinar	778 - 782	
CRYP	1867	A New Method to Detect Malicious DNS over HTTPS via Feature Reduction	Ali K. Bozkurt Halil E. Aköz Ataberk Taşpınar Şerif Bahtiyar	783 - 788
	1939	Generative Adversarial Networks for Synthetic Jamming Attacks on UAVs	Burcu Sönmez Sarıkaya Şerif Bahtiyar	789 - 794
	1958	Detecting Corruptive Noise Rounds for Statistical Disclosure Attacks	Alperen Aksoy Doğan Kesdoğan	795 - 800
	1996	Future Directions of Cybersecurity in Industrial Internet of Things Through Edge Computing	Tamara Zhukabayeva Lazzat Zholshiyeva Nurdaulet Karabayev	801 - 806
	2061	Resource-Efficient Ensemble Learning for Edge IIoT Network Security against OSINT-based Attacks	Mert İlhan Ecevit Zakire Çukur Muhammed Ali İzgün Noor Ul Ain Hasan Dağ	807 - 812
	2074	Transfer Learning for Phishing Detection: Screenshot-Based Website Classification	Furkan Çolhak Mert İlhan Ecevit Hasan Dağ	813 - 818
2145	Blok Şifrelerin Karıştırma ve Yayılım Tabakaları için Yeni Bir Analiz Aracı A New Analysis Tool for Confusion and Diffusion Layers of Block Ciphers	Mehmet Ali Demir Meltem Kurt Pehlivanoğlu Pınar Savaştürk Emir Öztürk Muharrem Tolga Sakallı Sedat Akleylek	819 - 824	
CVIS	1851	Impact of Image Augmentation on Deep Learning-Based Classification of Granite Tiles	Gaye Ediboglu Bartos Sibel Ünaldı Nesibe Yalçın	825 - 828
	1877	Advanced Facial Expression Classification with CNN-Transformer Integration for Human-Computer Interaction	Ali Azmoudeh Cigdem Altin Gumussoy Hazım Kemal Ekenel	829 - 834
	1936	Word Image Representation at Local and Global Levels Based on Vision Transformers	Baha Edine Harrath Mohamed Mhiri Mohamed Cheriet	835 - 840
	1949	Detecting Duplicate Products in E-Commerce Images Using Siamese Networks	Enis Teper Furkan Eseoğlu Mustafa Keskin	841- 846
	1963	Comparative Analysis of Visual Attribute Tagging Models for Upper-Body Clothing Products	Engin Kaya Mert Yanık	846 - 850
	1986	Development of A Model of Kazakh Sign Language Recognition Based on Deep Learning Method	Aigerim Yerimbetova Bakzhan Sakenov Ulmeken Berzhanova Nurzhan Mukazhanov Elmira Daiyrbayeva Mohamed Othman	851 - 856
	1989	Recognising Kazakh Sign Language with Mediapipe	Aigerim Yerimbetova Diana Kaidina Bakzhan Sakenov Elmira Daiyrbayeva Mussa Turdalyuly Ulmeken Berzhanova	857 - 862
	1990	Ultrason Görüntülerinden Meme Kanseri Teşhisi için Lezyon Tespitli Hibrit Derin Öğrenme Modelleri	Osman Doğuş Gülgün	863 - 868

		Hybrid Deep Learning Models with Lesion Detection for Breast Cancer Diagnosis from Ultrasound Images	Hamza Erol	869 - 874
1992		Olumsuz Hava Koşullarında Gemi Tespiti ve Sınıflandırılması Ship Detection and Classification in Adverse Weather Conditions	Yahya İzala Yaşar Becerikli	
2010		Removing Background from Noisy Handwritten Signatures on Banking Documents using GANs	Ege Dinçer Sacide Kalaycı Emre Yurdakul Bilge Köroğlu	875 - 879
2013		A Faster R-CNN Model for Multi-class Classification and Detection of Land, Air, and Sea Vehicles	Enes Güvelioğlu Çiğdem İnan Acı	880 - 885
2098		Çelik Hurdasının Sınıflandırılmasında ResNet ve Görüntü Dönüştürücü Tabanlı Modellerin Başarımı	Sefa Temur Levent Karacan	886 - 891
2117		Advanced Computer Vision Techniques for Reliable Gender Determination in Budgerigars (Melopsittacus undulatus)	Atalay Denknalbant Efe İlhan Cemalcılar Majid Ahangari Abdussamat Saidburkhan Alireza Zirak Ghazani Erkut Arıcan	892 - 897
2151		Teslimat Süresi Tahminlerinde Makine Öğrenmesi Modellerinin Yorumlanabilirliği Interpretability of Machine Learning Models in Delivery Time Predictions	Serhat Agit Satıcı Habil Kalkan	898 - 903
2169		Deep Learning based Order Form Recognition	Enes Alperen Buğaz Orhan Akbulut Aysun Taşyapı Çelebi Uğur Yıldız	904 - 908
2170		Learning Based Photo Management on Smartphones	Beyza Nur Şenay Orhan Akbulut Aysun Taşyapı Çelebi Uğur Yıldız	909 - 912
2177		Retinal Disease Classification Using Optical Coherence Tomography Angiography Images	Omer Faruk Aydın Muhammet Serdar Nazlı F. Boray Tek Yasemin Turkan	913 - 918
2178		Segmentation Based Classification of Retinal Diseases in OCT Images	Öykü Eren F. Boray Tek Yasemin Turkan	919 - 924
2183		Unsupervised Translation from Shortwave Infrared Images to RGB Images: A Comparative Evaluation	Duygu Tasbas Hacer Yalim Keles	925 - 930
DSCI	1882	Automatic Segmentation of Time Series Data with PELT Algorithm for Predictive Maintenance in the Flat Steel Industry	Saygın Kaçar Tuğçe Ballı E. Fatih Yetkin	931 - 936
	1921	Otomobil Kredilerinde Temerrüt Tahmini ve Araç Geri Kazanım Olasılığı Analizi - Bir Segmentasyon Çalışması Default Prediction and Vehicle Recovery Probability Analysis in Auto Loans - A Segmentation Study	Sahin Nicat Anıl Ferdi Kaya	937 - 942
	1955	E-Ticaret Sadakat Programı Müşteri Eğilim Tahmini Customer Propensity Prediction in E-Commerce Loyalty Program	Yunus Emre Gündoğmuş Sinan Keçeci Ege Erdem Emre Rençberoğlu	943 - 946
	1967	Yapay Zeka ve Makroekonomik Göstergeler ile Tüzel Kredilerin Değerlendirilmesi Evaluation of Corporate Loans with Artificial Intelligence and Macroeconomic Indicators	Burak Yüksel Hakkı Berkay Çiçek	947 - 951
	1974	Enhanced Bot Detection on TwiBot-20 Dataset	Mehmet Ali Osman Atik Şevket Umur Çakır Alper Özcan	952 - 956
	1975	Drug-Drug and Drug-Protein Link Prediction on DTINet dataset	Mehmet Ali Osman Atik Yusuf Çelik Alper Özcan	957 - 960
	2029	Perakende Verilerinde Anomali Tespiti ve Döviz Kuru İlişkisi Üzerine ChatGPT Destekli Yorumlama	Şadi Evren Şeker	961 - 966

		ChatGPT Supported Interpretation on Anomaly Detection in Retail Data and Exchange Rate Relationship	Hatice Nizam-Özoğur	
	2033	Text to SQL Transformation Using LLM: a Comparative Research of T5, Seq2Seq, and SQLNet Models	Zhazira Shaikhiyeva Madina Mansurova Gulshat Amirkhanova	967 - 972
	2076	Sağlık Sigortası Sahiplerinin Davranışsal Analizi ve Kümelenmesi Clustering and Behavioral Analysis of Health Insurance Owners	Omer Sezer Koyuncu Seçil Arslan	973 - 978
	2087	On symbolic Prediction of Time Series for Predictive Maintenance Based on SAX-LSTM	Aykut Güler Tuğçe Ballı E. Fatih Yetkin	979 - 983
	2135	Profiling Driver Behaviors Using AI-Based Methods and Deep Learning Techniques for Improving Road Safety: A Comparative Study of Algorithms	Volkan Oban Mustafa Kaya Güzide Safi İrem Nur Çimen Tubanur Çatak Bulut Karadağ Gökhan Gümüş Aslıhan Çandır Fatih Alagöz	984 - 989
IR	1896	ReRag: A New Architecture for Reducing the Hallucination by Retrieval-Augmented Generation	Robin Koç Mustafa Kağan Gürkan Fatoş T. Yarman Vural	990 - 994
	1941	Enhancing Object Detection in Aerial Images Using Transformer-Based Super-Resolution	Aslan Ahmet Haykır İlkay Öksüz	995 - 1000
NET	1985	Proof of Concept Implementation for RSVP TSN Control Plane	Necip Gozuacik	1001 - 1004
	2100	Integrating Blockchain and SDN for Centrality-Aware Virtual Multicast Tree Embedding	Furkan Ayaz Evrin Guler Murat Karakus Davut Hanbay	1005 - 1010
	1969	QoS Aware Routing Approaches in Software Defined Smart Grids	Sedef Demirci	1011 - 1016
	2008	Deep Reinforcement Learning Routing in Mobile Networks	Arif Burak Dikmen Hasari Çelebi	1017 - 1022
RBOT	1942	Endüstriyel Robotik Sistemlerin Güvenlik Doğrulaması Safety Verification of Industrial Robotic Systems	Fatih Furkan Arslan Metin Özkan	1023 - 1028
	2077	EKF Based Localization: Integrating IMU and LiDAR Data in the Hilti SLAM Challenge	Behice Bakır Havvanur Bozömeroğlu Ebu Yusuf Güven	1029 - 1034
SING	1965	Communication (Educational) Kit (HaKi)	Murat Sever Utku Bilgin	1035 - 1038
	2089	Manyetik Parçacık Görüntülemesinde Sistem Matrisi için Farklı Dalgacık Dönüşümlerinin Seyreklik Seviyesi Karşılaştırması Sparsity Level Comparison of Different Wavelet Transforms for the System Matrix in Magnetic Particle Imaging	Vildan Atalay Aydın	1039 - 1043
	2097	Sparse Channel Estimation For M-QAM-Based Underwater Acoustic Communication Systems	Mhd Tahssin Altabbaa Berkay Tekat Emin Tarik Iseri	1044 - 1048
OTH	1858	The 80/20 Principle in Morphemics-Morphology in the Educational Corpus of the Uzbek Language	Shahlo Khamroeva Bakhtiyor Mengliyev Muyassar Kholova	1049 - 1052
	1904	Gamification as a Tool for Personalized Learning in Inclusive Education	Dilaram Baumuratova Tamara Zhukabayeva Mira Rakhimzhanova	1053 - 1058
	1918	A Metaheuristic Algorithm for the Fixed Charge Transportation Problem	Nermin Kartli	1059 - 1062
	2027	Eğitimde Sürükleyici Teknolojilerin Kullanılması Fırsatlar ve Beklentiler	Atamuratov Rasuljon Kadirjanovich Majidova Gulhayo Abdirazzoq qızı Bayjonov Furqat Baxramovich Ongarov Mansurbek Bayrambekovich	1063 - 1068

		Saydullayev Zafar Erkinovich	
	2103	Bilgisayar Mühendisliği Öğrencilerinin Perspektifinden Bilişim Hukukunun Güncel Sorunları ve Çözüm Önerileri Current Challenges and Solution Proposals in IT Law from the Perspective of Computer Engineering Students	Sevda Bora Çınar 1069 - 1075
	2200	A Comparison of shcU-Net Based GAN and U-net Based GAN in Adult Dental Segmentation	Gürdal Altundağ Hakan Öcal 1075 - 1080
	1932	Leveraging Quantum Computing and Optimization to Estimate Financial Crashes in Small and Medium-Sized Enterprises	Ege Dincer Berkay Coskuner Ege Bilaloglu Bilge Koroglu 1081 - 1086
SW	1859	Investigating The Adoption of International Software Quality Standards in Turkey: A Comprehensive Analysis	Sevgi Koyuncu Tunç 1087 - 1093
	1886	Development of the Functional Structure of the Science and Education Information System	Dauletov Adilbek Yusupbayevich Matyakubova Noila Shakirjanovna 1094 - 1098
	1892	React ve Preact Javascript Çerçevesinde Karşılaştırmalı Analiz Comparative Analysis on React and Preact Javascript Frameworks	Muhammed Furkan Uygur Nesibe Yalçın 1099 - 1104
	1917	CAGE: A Tool for Code Assessment and Grading	Ümit Kanoğlu Oğuz Kerem Yıldız Hasan Sözer Olca Taner Yıldız 1115 - 1110
	1957	Extracting Driving Styles from Automotive Sensor Data to Develop Personas	M. Cagri Kaya Tayssir Bouraffa Krzysztof Wnuk 1111 - 1114
	1962	Lojistik Sipariş Dağıtım Entegrasyonu Sürecinde Sipariş Geri Çağırma Süreci Tasarımı ve Yazılım Geliştirme Design and Software Development of The Order Recall Progress in The Logistics Order Distribution Integration Process	İklim Barman Ersin Şengül 1115 - 1120
	2009	The Dimension of Green Coding in Software Quality Control Processes	Volkan Abur 1121 - 1126
	2055	Are We Asking the Right Questions to ChatGPT for Learning Software Design Patterns?	Çağdaş Evren Gerede 1127 - 1132
	2060	Optimizing LLVM IR: Transforming Multiplication to Addition for Enhanced Execution Efficiency	Huseyin Karacalı Efecan Cebel Nevzat Donum 1133 - 1138
	2080	Estimation of Software Integration Test Duration via UML Statecharts	Fehim Göler Tolga Ovatman 1139 - 1144
	2093	DIA4M: A Tool to Streamline DevOps Processes of Distributed Cloud-Native Systems	Eren Tarak H. Hakan Kilinc 1145 - 1150
	2111	Software Industry Perception of Academic Collaboration	Deniz Akdur 1151 - 1156
	2139	Görüntü İşlemeyle Doğrulamalı Robotik Test Otomasyon Kullanımı: POS Cihazları Üzerine Uygulama	Miraç Emektar Harun Kadioğlu Ahmet Efendioğlu Fatih Mehmet Harmancı 1157 - 1161
	2141	VoIP Sistemlerinde Zihin Haritası Tabanlı Test Stratejiler : SIP Pbx Ürünü Üzerine Bir İnceleme Mind Map-Based Testing Strategies in VoIP Systems: A Case Study on SIP Pbx Products	Miraç Emektar Furkan Günaydın Fatih Mehmet Harmancı 1162 - 1167
	2173	A Robust Microservices Framework for Indoor Tracking System Development	Gafur Hayyrbayev Kerem Küçük Mahmut Çavur 1168 - 1172
DM	1927	Unsupervised Pattern Extraction of Time Series Data for Energy Disaggregation	Şirin Azazi Deveci Melih Günay 1173 - 1178
	1944	Topic Modeling Enhanced Tripartite Graph for Recommendation using Metapaths	Yaren Yılmaz Irem İşlek Şule Gündüz Öğüdücü 1179 - 1184
	1948	Community Detection on Software Library Dependency Graphs using Graph Neural Networks	Şevket Umut Çakır 1185 - 1190

		Mehmet Ali Osman Atik Ümit Deniz Uluşar	
2190	Enhancing Mesh and Point Cloud Similarity Detection through Geometric Features and ICP	Talha Rehman Abid Mehtap Öklü Cem Yıldız Ali Erman Erten Kamer Kaya	1191 - 1196
2214	Comparative Analysis and Practical Implementation of Machine Learning Algorithms for Phishing Website Detection	Samad Najjar-Ghabel Shamim Yousefi Payam Habibi	1197 - 1202
2215	A Technical Analysis and Practical Implementation of Machine Learning Algorithms for Predicting Survival in Breast Cancer Patients	Shamim Yousefi Samad Najjar-Ghabel Hamidreza Shafaei	1203 - 1208
BIG 1881	Comparison Between Time Series and Relational Databases	Alpar Türkoğlu Onurcan Ersen İbrahim Onuralp Yiğit Dincer Unal Hatice Golcuk	1209 - 1212
1930	A Performance Evaluation Study on a Data Analytics Platform for Emergency Calls	Engin Yakar H. Hakan Kilinc	1213 - 1218
2079	Adaptive Composite Market Volatility Index (CMVI) for Enhanced Stock Price Forecasting	Rabia Çevik Uğur Barış Özyürek Ali Kanal Vael Kokach Büşra Kocaçınar Oznur Şengel Fatma Patlar Akbulut	1219 - 1223
2142	Hybrid Deep Learning Framework for Stock Price Prediction Incorporating Technical and Macroeconomic Indicators	Ali Can Turan Vael Kokach Büşra Kocaçınar Oznur Şengel Fatma Patlar Akbulut	1224 - 1228
2125	Emotion-Aware Multimodal Biometric Identification by using Biosignals	Yekta Said Can Beyzanur Bektan Fatih Alagöz	1229 - 1235
1854	Özbekçe-Türkçe Otomatik Çeviri Yazılımı için Deyimlerin Veritabanını Teşkil Etmede Karşılaşılan Güçlükler Automatic Translation Software Difficulties in Organizing the Database of Idioms for Uzbek and Turkish	Manzura Abjalova Umida Raşidova Eşref Adalı	1236 - 1240
2028	Reversible Steganographic System for the Transmission of Personal Medical Data	Elmira Daiyrbayeva Ekaterina Merzlyakova Aigerim Yerimbetova Aigul Mukhitova	1241 - 1246

Algorithm for Aligning Paragraphs and Sentences in Aligner Tool

Elov Botir Boltayevich

Tashkent State University of Uzbek Language and Literature named Alisher Navoi,

Tashkent, Uzbekistan
elov@navoiy-uni.uz

Dauletov Adilbek Yusupbayevich

Alfraganus University,

Tashkent, Uzbekistan
davletov--odilbek@mail.ru

Khamroeva Shahlo Mirdjonovna

Tashkent State University of Uzbek Language and Literature named Alisher Navoi,

Tashkent, Uzbekistan
shaxlo.xamrayeva@navoiy-uni.uz

Matyakubova Noila Shakirjanovna

Tashkent State Univ. of Uzbek Language and Literature named after Alisher Navoi,

Tashkent, Uzbekistan
matyakubovanoila@navoiy-uni.uz

Abstract— A parallel corpus is one of the main resources for training and evaluating machine translation systems. By adapting parallel texts, it is possible to improve the translation quality of machine translators, which allow people to use different languages freely. In addition, parallel corpora play an important role in the efficiency of natural language processing tasks such as searching engines, sentiment analysis, and object recognition. There are several stages in the formation of such corpora, one of them is the alignment process. Once the parallel texts are collected, they need to be aligned at the paragraph, sentence, word or phrase level in order to determine the correspondence between segments in different languages. Today, several Aligner tools are available for these tasks, automating this process by aligning and identifying translation equivalents based on neural or statistical models. But not all available tools are equally effective in different languages. This article provides information about the linguistic and software support of the Uzbek-English "Aligner" system, which aligns parallel texts in Uzbek and English, and the stages of its creation.

Keywords— *Parallel corpus, parallel texts, alignment, Aligner, segmentation, source language, target language.*

I. INTRODUCTION

Parallel corpora are a rich source of linguistic information that has far-reaching implications for research, education, and technology development. Sound methodologies are essential for creation and widespread use of parallel corpora because with such corpora we can deepen our understanding of language diversity, advance intercultural communication, and open up new possibilities in the fields ranging from computational linguistics to cross-cultural studies. One of the main steps to ensure that corpora are properly formed is the alignment process, and Aligners are the only tools that correctly distribute texts in parallel. Today there are several aligners and the most commonly used ones are sentence aligner, word aligner and phrase aligner.

II. CONDUCTED SCIENTIFIC RESEARCH

Like many natural language processing tools, alignment tools have gone through several stages of development. There are single-function aligners and hybrid aligners available today, each with its own advantages and disadvantages. **GIZA++**, developed at the University of Aachen in 1999, is a statistical machine translation toolkit that includes word matching tools between parallel corpora[1]. It is widely used in machine translation and natural language processing. **HunAlign**, created at the Budapest University of Technology

and Economics in 2002, is a sentence-level aligner that uses heuristics and statistical methods to align parallel corpora. It has been widely used in various machine translation and corpus linguistics projects[2].

Berkeley Aligner [3] was developed by The Berkeley NLP group in 2010 and it is a word aligner based on IBM Model 1 and IBM Model 2. It provides aligning models for different language pairs and mainly used in machine translation research. UCambridge Aligner, developed in 2011, is a tool used to align parallel corpora at the sentence level[4]. It uses a Bayesian model to estimate matching probabilities and is used in research on machine translation and language modeling.

In addition, MGIZA++ (Multidisciplinary GIZA++), which was developed in 2012, provides versatile capabilities to speed up the alignment process, which allows it to be used for large-scale parallel corpora[5]. Fast_align, created in 2014, is an open source word alignment tool[6]. It stands out due to its speed and accuracy in matching large-scale bilingual corpora. In 2017, researchers at Facebook AI Research created MUSE, a toolkit designed for multilingual unsupervised and supervised word alignment[7]. Although it is not a dedicated alignment tool, it includes features for aligning words across languages, which in turn allows for cross-language analysis and transfer studies.

III. UZBEK-ENGLISH "ALIGNER" SYSTEM

Although aligners are used in various areas of NLP, their main task is to match text segments given in the source language(SL) to text segments in the target language(TL) [8]. Aligners are selected depending on what the matching object is. The aligner that we have created is mainly designed for aligning Uzbek-English parallel corpus, and allows to align the parallel corpus in the following stages:

- Paragraph alignment
- Sentence alignment

IV. ALIGNER FORMATION STAGE.

As an input a lexical dataset of Uzbek and English words, parallel texts in Uzbek and English languages which is formed as a parallel corpus are used. Creation of the corpus is carried out in several stages:

- First of all, the original sources of Uzbek and English texts are compiled. The Resource guide "Preparing World Heritage Nominations" (Second edition, 2011)Published in November 2011 by the United

Identify applicable funding agency here. If none, delete this text box.

Nations Educational, Scientific and Cultural Organization is taken as a research object [9]. Texts were extracted from it and a corpus of more than 1 million sentences was formed.

- Collected texts for the corpus are also processed in several stages:
 - the text is cleaned of excess noise;
 - if there are abbreviations in the English text, they are identified and rewritten in their full form. For example, the word “it’s” in the given English text is a contraction of the pronoun it and the verb “is” or “has”. If we leave them unseparated, these two separate meaning words will be treated as a single token during the tokenization process, leading to large errors in the matching process.
- Processed texts are divided into small segments (sentence form).
- The number of allocated segments is calculated. This process is necessary to know the exact size of aligned texts in the SL and TL, in order to determine whether the number of aligning sentences is the same or how much they differ. Because our main goal is to determine whether there is an identical translation of the SL text in the TL and to highlight the appropriate translation.

If the difference in segments is not so big, we can use as input the texts that have passed all the steps given in the Fig.1.

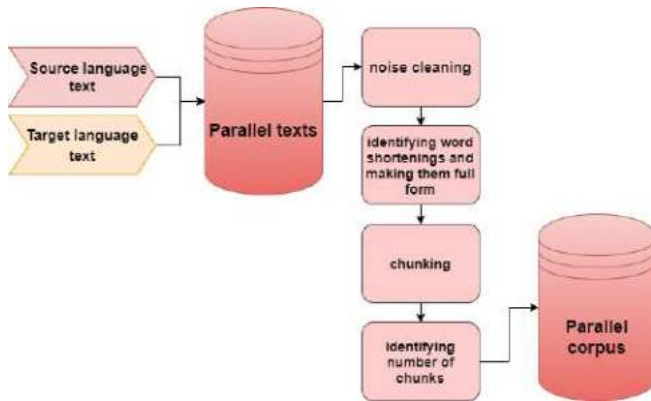


Fig. 1. Stage of preparation of input materials.

B. Paragraph Alignment.

In order to align paragraphs the following steps should be carried:

- Segmentation process: At this stage of alignment, the paragraphs which is going to be aligned are first separated into sentence.
- Counting number of sentence: The number of sentences in both languages is determined. If the number of sentences is the same, the alignment process goes on much easier, but not always the translated sentences of the SL are not at the same number and same form in TL. There are several reasons for this, sometimes it can be technical error while translating the sentence or other cases grammatical structure of the TL sentence. To solve grammatical problems, we first studied the sentence

structures of the source and target languages (see Table I).

For example, When I got on the coach, the driver had not taken his seat, and I saw him talking to the police. // Avtobusga chiqqanimda haydovchining hali o'z joyini egallamaganini, politsiya xodimi bilan gaplashayotganini ko'rdim. The example given in English is the compound-complex sentence, which consists of a compound sentence with an adverbial clause and two main clauses. When translated into Uzbek, it becomes a simple extended sentence.

TABLE I. TYPES OF SENTENCES IN UZBEK AND ENGLISH.

Types of the sentence	The simple sentence	The compound sentence	The complex sentence	The compound-complex sentence	The composite sentence
English language	+	+	+	+	+
Uzbek language	+	+	+	-	-

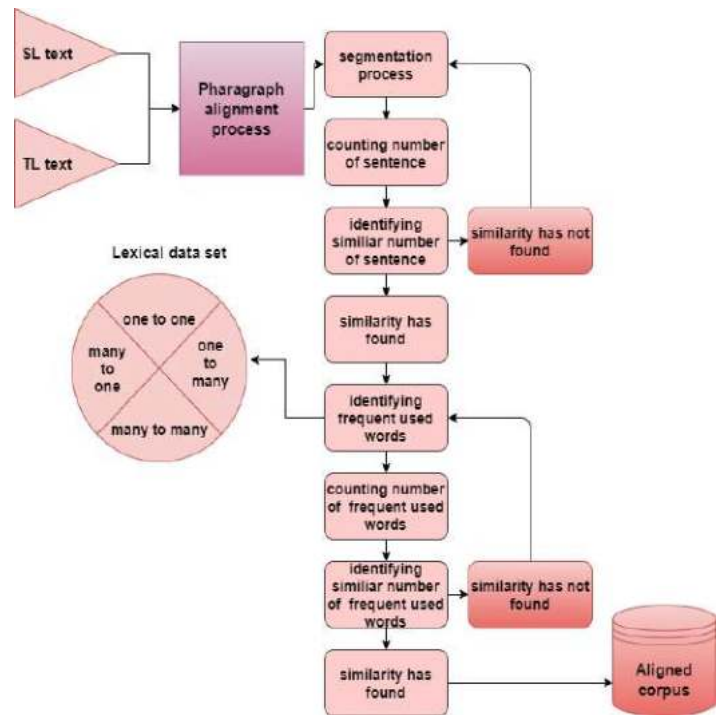


Fig. 2. Paragraph alignment algorithm

Or, One night, as soon as I finished my work at home, I went to get some vegetables from the market. // Bir kuni tunda uydagi ishlarimni yakunlashim bilanoq bozordan biroz sabzavotlashga olishga ketdim. In this example, a compound sentence with a time clause in English is translated into Uzbek as a simple expanded sentence.

During the research, several grammatical forms that do not exist in the Uzbek language or causes of the structural changes of the sentences during the translation process were identified, their linguistic base was formed. We will cover the complete information about this linguistic base in our further articles.

- Identifying frequent used words: At this stage the most frequently used word in SL and their translation in TL are determined. If the number of detected words

and their translations are the same or if the degree of proximity is high, it is taken as an aligned pair. (see: Fig 2.).

In the example given in Table II (see: Table II), the word **nomination** is used six times in the SL and the same in the TL. In addition, the number of matching paragraphs is seven in both languages. Therefore, it is possible to accept a paragraph in the TL as corresponding to a paragraph given in the SL.

TABLE II. STEP TO IDENTIFY THE MOST FREQUENTLY USED WORDS.

Source language	Target language
Lack of preparation time is the biggest enemy of successful nominations . Far too many are prepared against unrealistically short timeframes. It can take at least a year to set up appropriate support mechanisms and gather material, and a further year to write the nomination text and consult stakeholders. When research is needed, protection has to be achieved, and new management systems put in place and documented, so the process might take much longer. If the aim is a successful nomination that leads to inscription on the World Heritage List and long-term conservation and presentation of the property, a realistic timeframe should be allowed. Too often, lack of adequate preparation time leads to deferred or referred nominations , which is frustrating for States Parties, the World Heritage Committee and the Advisory Bodies. Sometimes political commitments are made which set an unrealistic timeframe for preparing a nomination , resulting in a nomination dossier which is inadequate and not ready for evaluation.	Tayyorgarlik ko'rish uchun vaqtning yetishmasligi, nomzod larni muvaffaqiyatli taqdim etishning eng katta dushmani hisoblanadi. Juda qisqa muddat ichida haddan ziyod ko'p nomzodlar tayyorlanadi. Tegishli qo'llab-quvvatlash mexanizmlarini o'rnatish va ma'lumot to'plash uchun kamida bir yil, nomzodlik matnini yozish va manfaatdor tomonlar bilan maslahatlashish uchun yana bir yil kerak bo'ladi. Tadqiqot o'tkazish kerak bo'lganda, yangi boshqaruv tizimlarini himoya qilish, joriy etish, ularni hujjatlashtirish lozim, shuning uchun bu jarayonga ancha uzoq vaqt ketishi mumkin. Agarda maqsad – obyektni Butunjahon merosi ro'yxatiga kiritilishi va obyektni uzoq muddatli muhofaza qilish va uning taqdimotiga olib keluvchi muvaffaqiyatli nomzodlik bo'lsa, buning uchun real muddatlar belgilanishi kerak. Aksariyat hollarda yetarli tayyorgarlik uchun vaqtning yo'qligi nomzodlikka qo'yishning kechiktirilishi yoki qayta ko'rib chiqishga berilishiga olib keladi, bu esa Ishtirokchi-davlatlar, Butunjahon merosi qo'mitasi va Maslahat organlarining umidlarini puchga chiqaradi. Ba'zan nomzodni tayyorlash uchun noreal muddatlarni belgilaydigan siyosiy majburiyatlar olinadi, bu natijada nomuvofiq va baholashga tayyor bo'lmagan nomzodliklarning paydo bo'lishiga olib keladi.

C. Sentence alignment

The sentence alignment process is somewhat more complex than the paragraph alignment process and involves several analytical processes. We will consider them in the order given in the Fig.3 (see Fig.3).

- At the first step the given text is divided into segments in the form of sentences.
- At the next step the length of sentences is determined by counting the number of words in a given sentence. For example: (See Table III.).

- After the length of the sentences of both languages have been identified, the sentence with similar or very close length are determined.

TABLE III. THE STAGE OF DETERMINING THE LENGTH OF SENTENCES.

Sometimes political commitments are made which set an unrealistic timeframe for preparing a nomination, resulting in a nomination dossier which is inadequate and not ready for evaluation	Ba'zan nomzodni tayyorlash uchun noreal muddatlarni belgilaydigan siyosiy majburiyatlar olinadi, bu natijada nomuvofiq va baholashga tayyor bo'lmagan nomzodliklarning paydo bo'lishiga olib keladi.
--	--

There are 27 words in the sentence given in English, and 22 words in the sentence with the same translation in Uzbek. The main reason for the difference between the tokens is the non-use of grammatical forms such as articles, prepositions and auxiliary verbs, which do not exist in Uzbek. During the

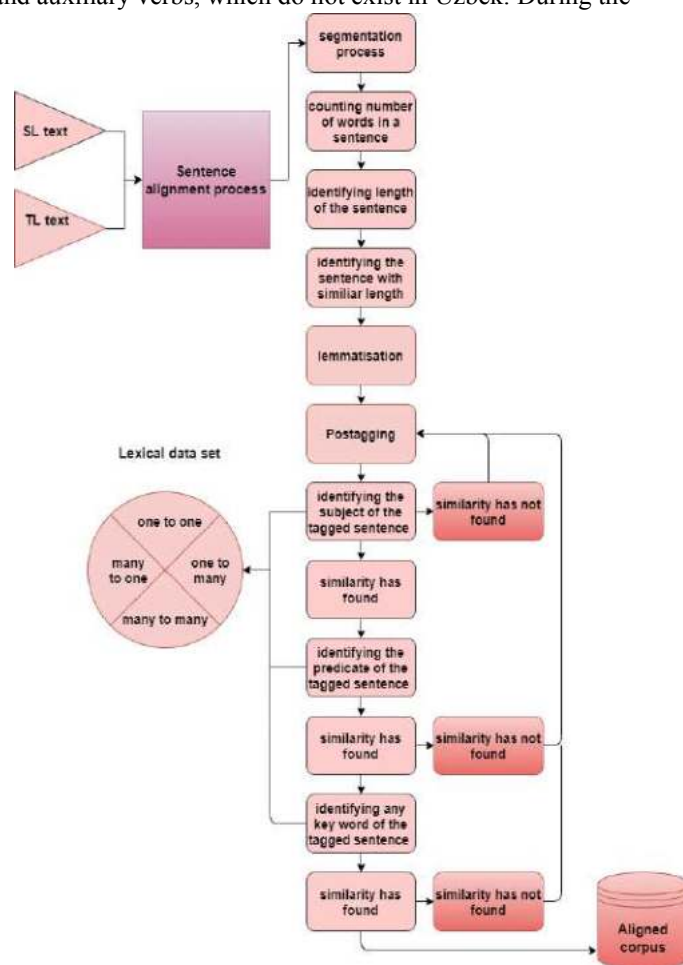


Fig. 3. Sentence alignment algorithm

research process special database of such grammatical forms was formed for Aligner program.

- Before POS Tagging process lemmas of the tokens should be identified[10]. It is very important process in alignment especially when the source language is Uzbek. If the lemma of the word is not identified, some complications arise in the process of POS tagging.

As in the Uzbek language there are sentences with a hidden subject, in which the subject of the sentence can be understood from the suffixes added to the predicate, but in

English, the subject of the sentence must be given separately in order to form the sentence correctly. For example, in the sentence “**Topshiriqlarni tugatdim**”, “topshiriqlarni” is the object, “tugatdim”- is the predicate, subject of the sentence is not given separately on the sentence. But the suffix - **im** added to the verb “tugatmoq (to finish)” indicates that the action was performed by the pronoun “I”, first person singular. So, in English, this sentence is translated as “**I have finished tasks**” using the pronoun “I”, where “I” is the subject of the sentence, “have finished” is the predicate of the sentence, and “tasks” is the object. A database of possessive clauses in such sentences was collected and trained to the Aligner program.

- The subject of the given sentences is determined and matched, if the match is correct, the match is accepted. If the match is not found lemmatization process repeats one more time.
- The predicate of the sentence is identified and compared. At this stage the selected lemma is searched in the dictionary and a match is determined.

This is also a somewhat complicated process, because the structure of the predicate is completely different in both languages. There are simple and complex predicates in the Uzbek language, as well as in English, but their structure is different. Auxiliary verbs that form the complex predicate in English do not exist in Uzbek. Therefore, when translated into Uzbek, the verbs formed in the complex predicate in English becomes the simple one in Uzbek, or complex predicate formed by two independent verbs in Uzbek language is usually translated with a single verb and become simple predicate in English. For example, the complex predicate in English “have been working” turns into the simple predicate in Uzbek “ishlayotgandi”, or vice versa, the complex predicate “qaytib keldi” in Uzbek becomes the simple predicate in English, “visited”. In order to solve such problems, a lexical data set in the form of “one to one”, “one to many”, “many to one”, “many to one” was formed.

- In the next step, any active word in the tagged sentence in SL is determined and compared with the sentences in the TL. If a match is found, the step ends, if not, the word is compared with another synonym form in the lexical database. If a synonym form is found, the stage ends, if not, another active word is selected and the stage is repeated (see: Fig 3).

IV. CONCLUSION

Aligner tools are the most useful and effective tools for working on parallel texts and determining whether the translation of the source language text into the target language with one-to-one correspondence. Although there are many effective aligners available today, they do not perform equally

well and accurately in all languages. This situation is especially common when adapting the translation of texts from languages belonging to different families. When determining their compatibility, if the specially designed aligners for those languages are used, the efficiency indicator will be significantly higher.

A database of many grammatical and lexical rules, which can be separated only by human intervention during the translation process, in Uzbek and English languages has been collected for creation of the Uzbek-English Aligner software and trained. In the future, this software is expected to be used not only for aligning texts in parallel corpus, but also for searching engines and translation tools, and will show effective results.

REFERENCES

- [1] F. Och, H. Ney. "Improved Statistical Alignment Models". Proc. of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 440-447, Hongkong, China, October 2000.
- [2] F. Och, H. Ney. "A Systematic Comparison of Various Statistical Alignment Models". Association for Computational Linguistics, 2023.
- [3] A. Pauls, D. Klein, D. Chiang, K. Knight "Unsupervised Syntactic Alignment with Inversion Transduction Grammars" Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA.
- [4] J. Sánchez "A comparative study of Neural Machine Translation frameworks for the automatic translation of open data resources", Escola Tècnica Superior d'Enginyeria Informàtica Universitat Politècnica de València, 2018.
- [5] M. Junczys-Dowmunt, A. Szal, "SymGiza++: Symmetrized Word Alignment Models for Statistical Machine Translation", International Joint Conference, SIIS 2011, Warsaw, Poland, June 13-14, 2011
- [6] Ch. Dyer, V. Chahuneau, N. Smith, "A Simple, Fast, and Effective Reparameterization of IBM Model 2", North American Chapter of the Association for Computational Linguistics, 1 June 2013
- [7] N. Robinson, N. Carlson, D. Mortensen, E. Vargas, Th. Fackrell, N. Fulda, "Task-dependent Optimal Weight Combinations for Static Embeddings", Northern European Journal of Language Technology, November 2022
- [8] N. Matyakubova, A. Dauletov, Sh. Khamroyeva, B. Mengliyev, E. Adali, "Algorithm of Creating The "Uzbek-English Aligner" Program", 2023 8th International Conference on Computer Science and Engineering UBMK 2023, Mehmet Akif Ersoy University, Burdur – Turkey.
- [9] Preparing World Heritage Nominations. Published in November 2011 by the United Nations Educational, Scientific and Cultural Organization. Second edition, 2011.
- [10] Sh. Sirojiddinov, B. Elov, Sh. Khamroeva, E. Adali, Z. Xusainova. "Pos Tagging of Uzbek Text Using Hidden Markov Mode" 8 th International Conference on Computer Science and Engineering UBMK 2023, Mehmet Akif Ersoy University, Burdur – Turkey.