

INNOVATION SCIENCE AND TECHNOLOGY

ISSUE 10, PART I / OCTOBER 2026

 **Acceptance of papers**



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences



EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhamatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu Simplicie
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfiyeva
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)



Xoliyorova Shoxista Qahramon qizi
PhD in Economics, Assistant Lecturer at the Department TSUE

Nasiba G'anjon qizi Sattorova
PhD, Senior Lecturer at the Department of Banking

CONTENTS

IMPLEMENTATION OF THE IMPROVED METHODOLOGY IN PRACTICE AND ASSESSMENT OF ITS ECONOMIC EFFICIENCY	8
Eshniyazova Yulduz Yuldashbayevna	
EXPERIMENTAL OPTIMIZATION OF DESIGN AND ENERGY PARAMETERS OF A GARLAND-TURBINE MOBILE MICRO HYDROPOWER PLANT FOR FREE-FLOWING WATER	13
Sultonov Sanjarbe, Zokirov Islomjon, Nurillaev Ahadullo, Abdullaev Shavkat	
METHODOLOGY FOR INCREASING THE EFFICIENCY OF CREDIT OPERATIONS IN COMMERCIAL BANKS	19
Djumaniyazov Ikram Karimbaevich	
COMPARATIVE ANALYSIS OF A COMPACT RESIDUAL CONVOLUTIONAL MODEL AND LINEAR REGRESSION FOR IMAGE SUPER-RESOLUTION	27
Husan Arziqulov Normurod o'g'li	
COMPARATIVE EVALUATION OF MACHINE LEARNING ALGORITHMS AND CLASS IMBALANCE HANDLING STRATEGIES FOR CREDIT CARD FRAUD DETECTION	33
Mahmudova Dilnoza Avaz qizi	
DIGITAL TRANSFORMATION OF FINANCIAL RESOURCE MANAGEMENT IN PUBLIC SOCIAL SECTOR INSTITUTIONS	39
Khasanov Tulkin Axmatovich	
WAYS TO ENSURE THE STABILITY OF THE CAPITAL BASE OF COMMERCIAL BANKS	45
Isakov Isobek Janabay's son	
CROSS-DAY EXPERIMENTAL EVIDENCE OF THERMAL-LOAD SHIFTING IN A SOLAR-DRIVEN HYDRONIC RADIANT FLOOR WITH MACROENCAPSULATED RT52 PHASE CHANGE MATERIAL	50
Jurabek Kamolov, Sokhib Temirov, Marjona Makhmudova	
STUDY OF THE TRACTION DYNAMICS OF AN AGGREGATE CONSISTING OF A FRUIT-HARVESTING MACHINE AND ITS TWO-AXLE CONTAINER-CARRYING TRAILER PLATFORM	59
Norchayev Davron, Sardor Nigmatjonov, Norchayev Jaloliddin, Norchayev Rustam	
PRIORITY DIRECTIONS FOR STRENGTHENING THE FINANCIAL STABILITY OF AGRICULTURAL ENTERPRISES	67
Kulmatov Zafar Jumanazarovich	
METHODOLOGICAL APPROACHES TO DEVELOPING UNIVERSITY STUDENTS' DIGITAL COMMUNICATIVE COMPETENCE BASED ON EMOTIONAL INTELLIGENCE	73
Malikakhon Kamiljanova	
SIMULTANEOUS REMOVAL CHARACTERISTICS OF OIL AND HEAVY METAL IONS BY BIOFLOCCULANTS BASED ON THE CELLULOSE-DMAEMA/PEPA COMPOSITE	80
Xayitboeva Dilshoda, Ibragimov Nasrilla, Jumaev Anvarbek	
ECONOMIC ANALYSIS AND DEVELOPMENT STRATEGY OF THE WOOD-POLYMER COMPOSITES MARKET	84
Abdumalikova Zarnigor, Ayubova Indira, Jumaev Anvarbek	
ENERGY UTILIZATION ANALYSIS OF AGRICULTURAL DRYING BASED ON INTEGRATED SOLAR THERMAL AND PHOTOVOLTAIC ENERGY SUPPLY	89
Ibragimov Umidjon, Mirzayarova Sevara, Khujayorova Aziza	
RECENT ADVANCES IN THE SYNTHESIS OF NANOSTRUCTURED MATERIALS FOR ZINC-ION BATTERIES	96
Diyorbek Absattorov, Dilnoza Erimbatova, Murodjon Samadiy	
ENTHALPY-MODEL SELECTION AND PRACTICAL IDENTIFIABILITY OF A MACROENCAPSULATED RT52 RADIANT FLOOR UNDER UNCERTAIN BRANCH FLOW	101
Jurabek Kamolov, Sokhib Temirov, Marjona Makhmudova	
ENTHALPY-MODEL SELECTION AND PRACTICAL IDENTIFIABILITY OF A MACROENCAPSULATED RT52 RADIANT FLOOR UNDER UNCERTAIN BRANCH FLOW	101
Jurabek Kamolov, Sokhib Temirov, Marjona Makhmudova	

THE IMPORTANCE OF THERMAL ENERGY STORAGE SYSTEMS IN ENHANCING THE EFFICIENCY OF SOLAR DRYERS	110
B. A. Hikmatov, M. S. Mirzaev,	
METHODOLOGY AND PRINCIPLES FOR THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN ADAPTIVE ECONOMIC MODELS OF HIGHER EDUCATION.....	117
O'lmasxo'jayev Zoirxo'ja Abbos o'g'li	
FORECASTING ECONOMIC EFFECTS OF GREEN INNOVATION IN MINING THROUGH COST REDUCTION, ENERGY SAVINGS AND LONG-TERM COMPETITIVENESS.....	122
Kurbanova Mehriniso Nematjanovna	
AN INTELLECTUAL HYBRID MODEL AND FEATURE SPACE FOR NAMED ENTITY RECOGNITION IN KARAKALPAK TEXTS	127
Kalmuratov Bekbosin Kushkinbaevich, Uzakbaev Jangeldi Sabit-uli, Rzambetova Suliwxan Baxtiyarovna	

AN INTELLECTUAL HYBRID MODEL AND FEATURE SPACE FOR NAMED ENTITY RECOGNITION IN KARAKALPAK TEXTS

Kalmuratov Bekbosin Kushkinbaevich

Senior Lecturer, Department of Mathematical Modeling, Nukus State Technical University.
ORCID: 0009-0003-7544-6621

Uzakbaev Jangeldi Sabit-uli

Assistant, Department of Mathematical Modeling, Nukus State Technical University.
ORCID: 0009-0002-5375-5805
E-mail: zhanuzakbaev@gmail.com

Rzambetova Suliwxan Baxtiyarovna

Trainee Assistant, Department of Mathematical Modeling, Nukus State Technical University.
ORCID: 0009-0009-8635-075X
E-mail: suliwxanrzambetova@gmail.com

Abstract. An intelligent hybrid model and a multi-source feature space are presented for Named Entity Recognition (NER) in Karakalpak electronic texts. The agglutinative morphological structure of the Karakalpak language, multilayer affixation, inconsistencies between Latin and Cyrillic writing systems, and limited availability of electronic language resources may lead to entity boundary detection errors during subword tokenization in multilingual Transformer models (mBERT and XLM-RoBERTa). To address these challenges, an architecture is proposed that integrates contextual Transformer representations, morphological affix features, and domain-specific gazetteers through a Confidence-Guided Gated Fusion mechanism. A Conditional Random Fields (CRF) layer is employed to constrain invalid transitions between sequence labels. The mathematical formulations of the model, the multidimensional feature space, and the software implementation process in the PyTorch environment are described in detail.

Keywords: Karakalpak language, Named Entity Recognition, intelligent hybrid model, gated fusion mechanism, morphological features, Transformer, gazetteers, Conditional Random Fields.

Annotatsiya. Qoraqalpoq tilidagi elektron matnlarda nomlangan obyektlarni aniqlash (Named Entity Recognition – NER) uchun intellektual gibrid model va ko'p manbali belgilar fazosi taklif etilgan. Qoraqalpoq tilining agglutinativ morfologik tuzilishi, ko'p qatlamli affiksatsiyasi, lotin va kirill yozuvlaridagi nomuvofiqliklar hamda elektron til resurslarining cheklanganligi ko'p tilli Transformer (mBERT, XLM-RoBERTa) modellarida subword tokenizatsiyasi jarayonida obyekt chegaralarini aniqlash bilan bog'liq xatoliklarni yuzaga keltirishi mumkin. Mazkur muammolarni bartaraf etish maqsadida kontekstual Transformer tasvirlari, morfologik affiks belgilari va sohaga oid gazetteer lug'atlarini ishonchlilik darajasiga asoslangan eshikli birlashtirish mexanizmi (Confidence-Guided Gated Fusion) orqali integratsiyalovchi arxitektura ishlab chiqilgan. Ketma-ketlikdagi ruxsat etilmagan teglar o'rtasidagi o'tishlarni cheklash uchun shartli tasodifiy maydonlar (Conditional Random Fields – CRF) qatlami qo'llangan. Modelning matematik ifodalari, ko'p o'lchamli belgilar fazosi hamda PyTorch muhitida dasturiy amalga oshirish jarayoni batafsil yoritilgan.

Kalit so'zlar: qoraqalpoq tili, nomlangan obyektlarni aniqlash, intellektual gibrid model, eshikli birlashtirish mexanizmi, morfologik belgilar, Transformer, gazetteer lug'atlari, shartli tasodifiy maydonlar.

Аннотация. Представлены интеллектуальная гибридная модель и многокомпонентное пространство признаков для распознавания именованных сущностей (Named Entity Recognition – NER) в электронных текстах на каракалпакском языке. Агглютинативная морфологическая структура каракалпакского языка, многоуровневая аффиксация, несогласованность использования латинской и кириллической графики, а также ограниченность электронных языковых ресурсов могут приводить к ошибкам определения границ именованных сущностей при субсловной токенизации в многоязычных моделях Transformer (mBERT, XLM-RoBERTa). Для решения указанных проблем предложена архитектура, интегрирующая контекстные представления Transformer, морфологические аффиксальные признаки и предметно-ориентированные словари именованных сущностей (gazetteers) посредством механизма управляемого объединения на основе уровня достоверности (Confidence-Guided Gated Fusion). Для ограничения недопустимых переходов между метками в последовательности применяется слой условных случайных полей (Conditional Random Fields – CRF). Подробно рассмотрены математические выражения

модели, многомерное пространство признаков и процесс программной реализации в среде PyTorch.

Ключевые слова: каракалпакский язык, распознавание именованных сущностей, интеллектуальная гибридная модель, механизм управляемого объединения, морфологические признаки, Transformer, словари именованных сущностей, условные случайные поля.

INTRODUCTION

In the modern digital information society, Natural Language Processing (NLP) is one of the rapidly developing fields of science and technology. The Strategy “Digital Uzbekistan – 2030” approved by the Decree of the President of the Republic of Uzbekistan No. UP-6079¹ dated October 5, 2020, and the Resolution No. PP-357² dated October 14, 2024, “On the Development of Artificial Intelligence Technologies,” established the creation of linguistic technologies for related languages as a priority direction [1]. Artificial intelligence capabilities were also utilized during the editing process of this text.

One of the initial stages of extracting semantic information from texts is Named Entity Recognition (NER), which automatically extracts and classifies person names (PER), location names (LOC), organization names (ORG), and temporal expressions (DATE) [2]. Belonging to the Kipchak group of the Turkic language family, the Karakalpak language possesses an agglutinative structure where word formation occurs through affixation chains (for example, the toponym “Nókis” appears in the forms “Nókiste”, “Nókistegi”, and “Nókisten”) [3]. Furthermore, expanding electronic resources, ensuring the balanced application of Latin and Cyrillic scripts, and standardizing spelling consistency require continuous refinement of the tokenization process [4].

In multilingual Transformer models (mBERT, XLM-RoBERTa), subword tokenization (BPE, WordPiece) provides opportunities to enhance the processing of Turkic affixes and improve entity boundary detection accuracy [5, 6]. Alongside developing rule-based systems [7], creating a hybrid intelligent model that integrates contextual semantics with morphological rules and domain-specific gazetteers based on confidence levels represents an important and relevant task.

LITERATURE REVIEW

Named Entity Recognition (NER) is one of the key tasks in natural language processing. Modern research extensively utilizes Transformer models such as BERT and XLM-RoBERTa, as well as approaches based on CRF and morphological features [4–8, 11]. In low-resource languages, combining contextual features with morphological and lexical knowledge proves to be an effective strategy [9]. The agglutinative structure and specific characteristics of linguistic resources in the Karakalpak language necessitate incorporating additional morphological and lexical features into NER systems [12]. Artificial intelligence capabilities were also utilized during the editing process of this text. Based on these approaches, this study proposes a hybrid model integrating XLM-RoBERTa, morphological features, Gazetteer dictionaries, and CRF components.

RESEARCH METHODOLOGY

The research employed a hybrid intelligent model to identify named entities in Karakalpak electronic texts. The methodology comprises stages of text tokenization, contextual feature extraction using XLM-RoBERTa, formation of morphological and orthographic characteristics, and matching with Gazetteer dictionaries. The extracted features are fused into a unified vector via a Confidence-Guided Gated Fusion mechanism, and the final label sequence is determined using CRF. The mathematical formulations and software implementation of the model were developed in the PyTorch environment.

ANALYSIS AND RESULTS

Formal Mathematical Model of Named Entities

A text sequence in the Karakalpak language consists of discrete tokens and is represented as follows:

$$X = (x_1, x_2, \dots, x_n), \quad x_i \in V, \quad i = 1, 2, \dots, N \quad (1)$$

Where X is the input text sequence, x_i is the i -th token, N is the number of tokens, and V is the vocabulary.

The BIOES annotation scheme is utilized for sequence labeling:

$$Y = (y_1, y_2, \dots, y_n), \quad y_i \in \mathcal{Y} \quad (2)$$

In the BIOES system: B denotes the beginning of an entity, I — inside an entity, E — end of an entity, S — single-

¹ Decree of the President of the Republic of Uzbekistan, of 05.10.2020 r. № DP-6079 <https://lex.uz/ru/docs/7008256>

² Resolution of the President of the Republic of Uzbekistan, of 14.10.2024 r. № RP-357 <https://lex.uz/ru/docs/7157438>

token entity, and 0 — outside any entity. Artificial intelligence capabilities were also utilized during the editing process of this text (Table 1).

Table 1
Main Classes of Named Entities³

Class	Meaning	Example
PER	Person	Ajiniyaz, Berdaq
LOC	Location	Nókis, Ámúdaryo
ORG	Organization	Nukus STU, Murojaat.uz
DATE	Date/Time	2024-jil, October 14
Class	Meaning	Example

Set of Entity Classes:

$$C = \{PER, LOC, ORG, DATE, MISC\} \quad (3)$$

A named entity is defined by a formal tuple:

$$E_k = (s_k, e_k, c_k, d_k), \quad 1 \leq s_k \leq e_k \leq N \quad (4)$$

Where s_k and e_k are the start and end positions of the k -th entity, c_k is its class, and d_k is the domain (Media, Official, Education, Medicine). Hierarchical relationships are taken into account for nested entities [8]. Artificial intelligence capabilities were also utilized during the editing process of this text.

Multi-Source Feature Space

The model constructs feature vectors for each token across three independent linguistic channels. The primary channels are summarized in Table 2.

Table 2
Primary Channels of the Multi-Source Feature Space⁴

Channel	Feature	Core Information	Output Dimension
Contextual	h_i^{ctx}	Context and semantics via XLM-RoBERTa	$d_{ctx} = 768$
Morphological	h_i^{morph}	POS, affix, case, and orthographic features	d_{morph}
Gazetteer	h_i^{gaz}	Matching with domain-specific dictionaries	d_{gaz}

Contextual Representation Vector

The contextual representation vector h_i^{ctx} is extracted from the multilingual XLM-RoBERTa Transformer layer. When a word is split into multiple subword tokens, their average vector is used:

$$h_i^{ctx} = (1 / M_i) \sum_{m=1}^{M_i} u_{i,m} \quad (5)$$

$$h_i^{ctx} \in \mathbb{R}^{d_{ctx}}, \quad d_{ctx} = 768 \quad (6)$$

Morphological and Orthographic Features

Due to the agglutinative nature of the Karakalpak language, morphological structure is modeled as an independent feature channel:

$$h_i^{morph} = [e_i^{POS} \parallel e_i^{Aff} \parallel e_i^{cAsE} \parallel e_i^{ORTH}] \in \mathbb{R}^{d_{morph}} \quad (7)$$

Where e_i^{POS} denotes part-of-speech, e_i^{Aff} represents affixes, e_i^{cAsE} indicates grammatical cases, and e_i^{ORTH} stands for orthographic feature vectors; $\parallel$ denotes concatenation.

Lexical Knowledge (Gazetteer) Vector

The third channel incorporates additional domain knowledge from gazetteer dictionaries. Token matches with vocabulary items are identified using the Aho–Corasick algorithm:

$$h_i^{gaz} \in \mathbb{R}^{d_{gaz}} \quad (8)$$

The gazetteer vector provides supplemental signals regarding match confidence and entity categories. This plays a crucial role for low-resource languages, serving as a foundational feature layer. Artificial intelligence capabilities were also utilized during the editing process of this text.

Confidence-Guided Gated Fusion Mechanism

³ Compiled by the authors based on sources [1, 2, 7, 8, 11].

⁴ Note: Compiled by the authors based on the developed model.

Morphological and gazetteer vectors are first projected into a shared contextual space:

$$m_i = \tanh(W_m h_i^{\text{morph}} + b_m) \quad (9)$$

$$g_i^g = \tanh(W_g h_i^{\text{gaz}} + b_g) \quad (10)$$

Where W_m and W_g are projection matrices, and b_m and b_g are bias vectors. Subsequently, auxiliary features are merged into a unified representation:

$$h_i^{\text{ling}} = \tanh(W_l [m_i \parallel g_i^g] + b_l) \quad (11)$$

Where h_i^{ling} represents the unified morphological and lexical representation. The gate vector is then computed as follows:

$$\alpha_i = \sigma(W^{\text{gate}} [h_i^{\text{ctxt}} \parallel h_i^{\text{ling}}] + b^{\text{gate}}) \quad (12)$$

Due to the sigmoid activation function, α_i represents the confidence weight assigned to the Transformer representation, while $1 - \alpha_i$ corresponds to the weight given to the linguistic representation. The final token representation is defined as:

$$z_i = \alpha_i \odot h_i^{\text{ctxt}} + (1 - \alpha_i) \odot h_i^{\text{ling}} \quad (13)$$

Where $\odot$ represents element-wise multiplication. The resulting z_i vector is forwarded to the CRF layer. The primary advantage of this mechanism is that higher weight is assigned to the Transformer output when contextual signal is strong, whereas linguistic features take precedence when morphological or gazetteer signals dominate.

Conditional Random Fields (CRF) and Loss Function

The final token representations are converted into globally consistent label sequences via the CRF layer. Unlike standard token-wise classification, the CRF model evaluates transition probabilities between adjacent labels. The overall score of sequence Y is defined as:

$$S(X, Y) = \sum_{i=1}^N P_{i, y_i} + \sum_{i=0}^N A_{y_i, y_{i+1}} \quad (14)$$

Where P_{i, y_i} is the local emission score for label y_i at token i , and $A_{y_i, y_{i+1}}$ is the transition score between consecutive labels. The conditional probability of sequence Y given X is expressed as:

$$P(Y|X) = \exp(S(X, Y)) / \sum_{\tilde{Y} \in \mathcal{Y}^N} \exp(S(X, \tilde{Y})) \quad (15)$$

The objective is to maximize the log-likelihood of the gold-standard sequence Y^* . The loss function is formulated as follows:

$$\mathcal{L}(\theta) = -\log P(Y^*|X; \theta) + \lambda \|\theta\|_2^2 \quad (16)$$

$$\mathcal{L}(\theta) = -S(X, Y^*) + \log \sum_{\tilde{Y} \in \mathcal{Y}^N} \exp(S(X, \tilde{Y})) + \lambda \|\theta\|_2^2 \quad (17)$$

Where θ represents trainable parameters, Y^* is the true label sequence, and λ is the regularization coefficient. The regularization term prevents model overfitting.

Software Module and Overall Execution Flow

The proposed confidence-guided gated fusion module is implemented in PyTorch as follows:

```
import torch
import torch.nn as nn
class ConfidenceGatedFusion(nn.Module):
    def __init__(self, d_ctx=768, d_morph=128, d_gaz=32):
        super().__init__()
        self.W_m = nn.Linear(d_morph, d_ctx)
        self.W_g = nn.Linear(d_gaz, d_ctx)
        self.W_ling = nn.Linear(d_ctx * 2, d_ctx)
        self.W_gate = nn.Linear(d_ctx * 2, d_ctx)
        self.sigmoid = nn.Sigmoid()
        self.tanh = nn.Tanh()
    def forward(self, h_ctx, h_morph, h_gaz):
        h_m = self.tanh(self.W_m(h_morph))
        h_g = self.tanh(self.W_g(h_gaz))
        h_ling = self.tanh(self.W_ling(torch.cat([h_m, h_g], dim=-1)))
        gate = self.sigmoid(self.W_gate(torch.cat([h_ctx, h_ling], dim=-1)))
        z = gate * h_ctx + (1.0 - gate) * h_ling
        return z
```

In this module, h_i^{morph} and h_i^{gaz} are first projected into a d_{ctx} -dimensional space, integrated with h_i^{ctxt} via the gating mechanism, and the resulting z_i vector is passed to the CRF layer (Table 3).

Table 3

Algorithmic Execution Pipeline of the Model⁵

Stage	Primary Operation	Output
1	Text tokenization	$X=(x_1, \dots, x_N)$
2	Contextual representation extraction via XLM-RoBERTa	h_i^{ctx}
3	Morphological and orthographic analysis	h_i^{morph}
4	Lexical matching via Gazetteer	h_i^{gaz}
5	Feature projection into a unified space	m_i, g_i^g
6	Fusion of linguistic features	h_i^{ling}
7	Computation of Confidence-Guided Gate	α_i
8	Extraction of final token representation	z_i
9	Sequence label scoring via CRF	Y^*
10	Optimization via loss function	θ^*

In this pipeline, the Transformer provides global context, the morphological channel supplies grammatical features, and the Gazetteer contributes lexical knowledge. The gating mechanism adaptively determines their relative confidence, while the CRF layer ensures the global logical consistency of the final label sequence. Artificial intelligence capabilities were also utilized during the editing process of this text.

CONCLUSION AND RECOMMENDATIONS

The mathematical and software architecture of an intelligent hybrid model for Named Entity Recognition (NER) in Karakalpak language texts has been developed. The core concept of the model lies in integrating three distinct information sources — contextual representations based on XLM-RoBERTa, morphological and orthographic features, and domain-specific Gazetteer knowledge — within a unified feature space. Artificial intelligence capabilities were also utilized during the editing process of this text.

The Confidence-Guided Gated Fusion mechanism adaptively determines the confidence ratio between contextual and linguistic features for each token. This capability demonstrates high efficacy for morphologically rich languages with distinct resource characteristics, such as Karakalpak. Furthermore, the application of a CRF layer ensures the logical sequence consistency of BIOES labels. The implementation of the model in the PyTorch environment establishes a robust computational foundation for subsequent experimental testing and integration into practical information systems.

Based on the research findings, the following recommendations are proposed:

Construct an annotated Karakalpak text corpus and conduct experimental evaluations using Precision, Recall, and F1-score metrics to assess model performance.

Continuously expand the morphological feature database and domain-specific Gazetteer dictionaries to enhance named entity recognition accuracy.

Integrate the developed model into automated text processing, information retrieval, and electronic document management systems in the Karakalpak language to advance its practical applicability.

REFERENCES

1. Jurafsky, D. and Martin, J. H. (2020) Speech and Language Processing. Pearson, pp. 9–10. Available at: <https://web.stanford.edu/~jurafsky/slp3/>
2. Bird, S., Klein, E. and Loper, E. (2009) Natural Language Processing with Python. Sebastopol: O'Reilly Media, pp. 179–202. Available at: <https://www.nltk.org/book/>
3. Goldberg, Y. (2017) Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers, pp. 65–67. Available at: <https://doi.org/10.2200/S00762ED1V01Y201703HLT037>
4. Conneau, A. et al. (2020) 'Unsupervised Cross-lingual Representation Learning at Scale', in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 8440–8451. Available at: <https://aclanthology.org/2020.acl-main.747/>
5. Devlin, J. et al. (2019) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', in Proceedings of NAACL-HLT 2019, pp. 4171–4186. Available at: <https://aclanthology.org/N19-1423/>
6. Liu, Y. et al. (2019) 'RoBERTa: A Robustly Optimized BERT Pretraining Approach', arXiv preprint arXiv:1907.11692. Available at: <https://arxiv.org/abs/1907.11692>
7. Küçük, D. and Eryiğit, G. (2014) 'Named Entity Recognition in Turkish', Turkish Journal of Electrical Engineering and

⁵ Note: Compiled by the authors based on the developed algorithmic model.

- Computer Sciences, 22(5), pp. 1251–1267.
8. Şeker, E. A. and Eryiğit, G. (2023) 'Extending a CRF-based NER System for Turkish', *Journal of Information Science*, 49(3), pp. 621–636.
 9. Dai, X. et al. (2021) 'Gated Multimodal Fusion for Low-Resource NER', in *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 3218–3228.
 10. Lafferty, J., McCallum, A. and Pereira, F. C. (2001) 'Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data', in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, pp. 282–289. Available at: <https://repository.upenn.edu/handle/20.500.14332/1578>
 11. Lample, G. et al. (2016) 'Neural Architectures for Named Entity Recognition', in *Proceedings of NAACL-HLT 2016*, pp. 260–270. Available at: <https://aclanthology.org/N16-1030/>
 12. Mengliev, D., Ayimbetov, M. and Kojabayev, A. (2023) 'Automated Morphological Analysis for Karakalpak', *Science and Innovative Development*, (3), pp. 78–86.
 13. NLTK Project (2023) Natural Language Toolkit. Available at: <https://www.nltk.org/>

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 10

© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles

Published every
monthly



Directions

Social, economic, political,
technological, scientific



Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:



Contact us
+998 50 737 87 88



Telegram channel
t.me/scopus_IST2100



Journal official website
<https://ist-journal.uz/index.php/IST>