IEEE
TÜRKİYE SECTION

UBMK'25

Bildiriler Kitabı
Proceedings

Editör Eşref ADALI

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

10th International Conference on Computer Science and Engineering

17-18-19 Eylül (September) 2025 İstanbul - Türkiye

IEEE TÜRKİYE SECTION

UBMK'25

Bildiriler Kitabı
Proceedings

Editor Eşref ADALI

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

10th International Conference on Computer Science and Engineering

17-18-19 Eylül (September) 2025 İstanbul - Türkiye

Media type	Part Number	ISBN	Online ISSN
XPLORE COMPLIANT	CFP25L97-ART	979-8-3315-9975-1	2521-1641
CD-ROM	CFP25L97-CDR	979-8-3315-9974-4	

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (UBMK'2025)

10th International Conference on Computer Science and Engineering

17-18-19 Eylül 2025 -İstanbul-Türkiye
17-18-19 September 2025 - İstanbul-Türkiye

IEEE Catalog Number : CFP25L97-ART

ISBN : 978-8-3315-9975-1

Additional copies mey 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

UBMK’2025’ye Hoşgeldiniz
Welcome to UBMK’2025

Sevgili Katılımcılar:

UBMK uluslararası nitelikli konferans serisi, 1990 yılından beri düzenli olarak yapılmakta olan Bilgisayar Mühendisliği Bölüm Başkanları toplantılarında alınan bir kararla on yıl önce başlamıştır. Konferansın 10.su IEEE-UBMK-2025 bu yıl 17-18-19 Eylül, 2025 günlerinde İstanbul Teknik Üniversitesinin ev sahipliğinde düzenlemiştir.

IEEE-UBMK-2025 konferansına bu yıl Almanya, Amerika Birleşik Devletleri, Azerbaycan, Fransa, Irak, İngiltere, İsveç, İtalya, Kanada, Kazakistan, Kırım, Kırgızistan, Rusya, Özbekistan, Tataristan, Taylant, Ürdün ve Türkiye’den 610 dolayında bildiri gönderilmiş ve bu bildiriler Türk ve yabancı 250 hakem tarafından değerlendirilmiştir.

Her bildiri en az iki hakem tarafından incelenmiş ve uzlaşma olmadığı durumlarda üçüncü bir hakemin değerlendirmesine başvurulmuştur. Bildiri başına düşen ortalama hakemlik 2,3 olmuştur. Bu değerlendirmelerin sonunda 327 bildirinin sözlü olarak sunulması uygun bulunmuştur. Kabul edilen ve sunulan bildiriler içerik ve kalite ölçünlerini sağlaması durumunda IEEE Xplore’da yayımlanacaktır.

Konferans çalışmalarında, Bilgisayar Mühendisliği Bölüm Başkanları Danışma Kurulu olarak görev almışlardır. Bildirilerin değerlendirilmesi Bilim Kurulu üyeleri tarafından yapılmıştır. Konferansın düzenlenmesi ise Yürütme Kurulunun önerileri doğrultusunda, Düzenleme Kurulu tarafından yapılmıştır.

Son olarak, konferansın başarılı bir şekilde yürütülmesi için tüm olanaklarını sunan İstanbul Teknik Üniversitesi Rektörü Sayın Prof. Dr. Hasan Mandal’a teşekkür ediyoruz. Ayrıca Düzenleme Kuruluna, bildirileri titizlikle değerlendiren Bilim Kurulu Üyelerine ve değerli araştırmalarının sonuçlarını bilişim camiası ile paylaşan bildiri sahiplerine teşekkürlerimizi iletiriz.

Prof. Dr. Eşref ADALI
UBMK-2025 Konferans Başkanı ve Bildiri Kitabı Editörü

Dear Participants:

The UBMK international conference series started nine years ago with a decision taken at the Computer Engineering Department Heads (BMBB) meetings, which have been held regularly since 1990. The 10th edition of the conference, UBMK'25, was held this year on October 17-18-19, 2025, hosted by İstanbul Technichal University.

This year, approximately 610 papers were submitted to the IEEE-UBMK-2025 conference from Germany, the United States, Azerbaijan, France, Iraq, the United Kingdom, Sweden, Italy, Canada, Kazakhstan, Crimea, Kyrgyzstan, Russia, Uzbekistan, Tatarstan, Thailand, Jordan, and Turkey, and these papers were evaluated by 250 Turkish and foreign referees.

Each paper was evaluated at least by two referees, and in cases where there was no consensus, a third referee was consulted. At the end of these evaluations, 327 papers were accepted for oral presentation. Accepted and presented papers will be submitted for inclusion into IEEE Xplore subject to meeting IEEE Xplore's scope and quality requirements.

During the conference, Heads of Information Engineering Departments took part in the Advisory Board. The evaluation of the papers was made by the members of the Scientific Committee. The conference was organized by the Organizing Committee in line with the recommendations of the Executive Committee.

Finally, we would like to thank İstanbul Technical University Rector Prof. Dr. Hasan Mandal for his continued support for the success of the conference. In addition, we would like to thank the Organizing Committee, the Scientific Committee Members who carefully evaluated the papers, and the owners of the papers who shared the results of their valuable research with the informatics community.

Prof. Dr. Esref ADALI
UBMK'25 Conference Chair and Proceedings Editor

Düzenleyenler / Organizer

Destekleyenler / Sponsors

Models for Automatic Recognition of Onomastic Units (Proper Nouns) in the Uzbek Language

Elov Botir Boltayevich
Dept. of Computational Linguistics and Digital Technologies
Tashkent State University of Uzbek Language and Literature named after Alisher
Tashkent, Uzbekistan
elov@navoiy-uni.uz

Samatboyeva Madina To'lqinjon qizi *Department of Computational Linguistics and Digital Technologies,*
Tashkent State University of Uzbek Language and Literature named after Alisher Navoi
Tashkent, Uzbekistan
samatboyevamadina@navoiy-uni.uz

Norova Mavluda Fayzulloyevna
Department of Uzbek Language and Literature, Russian and English Languages,
Bukhara State Medical Institute named after Abu Ali ibn Sino
Bukhara, Uzbekistan.
norova.mavluda@bsmi.uz

Samatboyeva Madina To'lqinjon qizi
Department of Computational Linguistics and Digital Technologies,
Tashkent State University of Uzbek Language and Literature named after Alisher Navoi
Tashkent, Uzbekistan
samatboyevamadina@navoiy-uni.uz

***Abstract*— This article explores the types of onomastic units, their linguistic analysis, their role within various language units, and their macro and micro scales. The classification of onomastic units, their functional characteristics, and their examination based on modern linguistics—particularly computational linguistics—are discussed. Additionally, the concept of onomastic units and Named Entity Recognition (NER) is studied, compared, and analyzed. The methods for identifying three major types of NER objects in texts are examined. The lexical-semantic features of onomastic units in the Uzbek language corpus are analyzed, and the effectiveness of NER approaches in their automatic extraction is evaluated. The main goal of the article is to study naming units in Uzbek at the intersection of corpus linguistics and computational linguistics and to highlight the possibilities of applying modern technological methods for their identification.**

Keywords—Onomastics, language units, onomastic units, anthroponym, toponym, organization names, abbreviation, indicator.)

I. INTRODUCTION

According to linguists, "language is the cultural code of a nation" [1]. Modern linguistic research has significantly increased the importance of large language corpora in the field of Natural Language Processing (NLP). Particularly for languages like Uzbek, which are rich in morphology and possess unique structural and lexical systems, the in-depth study of language units through corpora has become a pressing issue. One such direction is the identification and automatic processing of onomastic units (such as toponyms, anthroponyms, ethnonyms, and names of organizations and institutions) in language corpora.

In this process, NER (Named Entity Recognition) technology plays a crucial role. It enables the automatic extraction of personal names, place names, organization names, and other specialized units from texts. NER technology, on one hand, helps determine the typology of onomastic units in linguistics and, on the other hand, facilitates the effective use of the Uzbek language in artificial intelligence systems.

This article analyzes the lexical-semantic features of onomastic units in the Uzbek language corpus and evaluates the effectiveness of NER approaches in their automatic extraction. The main goal of the article is to study naming units in Uzbek at the intersection of corpus linguistics and computational linguistics and to highlight the possibilities of applying modern technological methods for their identification.

II. MAIN SECTION

Onomastics is a branch of linguistics that studies proper names (such as personal names, place names, ethnonyms, hydronyms, etc.) and holds an important place in linguistics. This field not only analyzes the system of language units but also reflects the history, cultural heritage, geographical location, and social changes of a people.

The branch of linguistics known as onomastics investigates eponymous nouns, their origins, motivational bases, semantics, and linguistic structure. Onomastics is derived from the Greek word meaning "the art of naming" [2].

Onomastics examines any eponymous noun constituting onomastic language units from linguistic and sociolinguistic perspectives. An onomasticon is a dictionary that represents the general collection of eponymous nouns in a language and reflects the repertoire of eponymous nouns belonging to the language of a particular ethnic group at a specific time [3]. Onomastics is also referred to as the science of names. Under this term, there exists a branch of science studied within the scope of anthropotoponymy, toponymy, ethnonymy, and other fields.

The onomastic system of any language, including Uzbek, is extremely rich in quantity and vast in scope, differing based on the type of named objects while manifesting as a real linguistic phenomenon with certain interconnections and relationships. At the micro-level within a language—for example, in oikonymy, agronomy, and other small fields—specific areas emerge.

Just as the lexical richness of a language consists of specific lexical units (lexemes), the onomastic fund of a language is formed by eponymous nouns included in the scope of onomastic vocabulary. These specific names are likened to terms such as "language units," "lexical units," and "speech units" in onomastics and are referred to as onomastic units. An onomastic unit refers to a specific eponymous noun that is individually named [4].

While language units (phonemes, morphemes, lexemes) are the general structural units of a language, onomastic units are specialized units related only to eponymous nouns. Onomastic units are divided into several main groups. The most commonly used and largest types of onomastic units form a macro-scale. See Table .

TABLE I. TYPE OF ONOMASTİC UNİT

№	Type of Onomastic Unit	Definition	Example
1	Anthroponym	Personal names, surnames, nicknames	Bahodir, Munisa
2	Toponym	Names of geographical objects	Tashkent, Kazakhstan
3	Hydronym	Names of water bodies	Amudarya, Zarafshon
4	Oronym	Names of mountains and flatlands	Hisor, Chatkal
5	Ethnonym	Names of peoples and ethnic groups	Qo'shtamg'ali, Qang'li
6	Mythonym	Names of individually named plants and trees (e.g., Kala, Cactus) or mythical and legendary names found in fairy tales, epics, folk tales, and legends (e.g., Go'ro'g'li, Alpomish)	Kala, Cactus, Go'ro'g'li, Alpomish
7	Cosmonym	Names of celestial objects in space	Saturn, Neptune
8	Demononym	Special names given to demons and jinn	Qora dev (Black Demon), Sariq dev (Yellow Demon)
9	Anemonim	Names of natural phenomena (storms, winds, floods, etc.)	Sarsar shamoli (Sarsar Wind)
10	Politonym	Names of national symbols (flags, emblems, etc.)	"Oytamg'a" (ancient Turkic tribes), "Jilontamg'a" (ancient Turkmen tribes)
11	Ideonym	Names of historical documents or works significant to a nation's history	"Declaration of Independence"
12	Ergonim	Names of organizations, institutions, and enterprises formed by groups of people engaged in activities	"Uzbekistan Red Crescent Society"
13	Faleronim	Names of orders, medals, badges, and crowns	"Friendship" Order
14	Dignitonim	Names of honorary titles and positions	"Hero of Uzbekistan," "Distinguished Scholar of Uzbekistan"
15	Stratonim	Names of geological formations (sedimentary and volcanic) and their parts	Stromboli Volcano, Kilimanjaro Volcano
16	Poreyonim	Names given to individual transport vehicles	Spark, Malibu

Onomastic units serve as markers reflecting the language, culture, and history of a people.

Onomastic units perform three main functions:

1) **Identification** – Identifying individuals, places, or other objects.

2) **Information Carrier** – Preserving historical, cultural, religious, or geographical information.

3) **Emotional-Aesthetic** – Serving as stylistic tools in art and literature.

III. LITERATURE REVIEW

The three largest types of NER objects are: Person names (PER), Location names (LOC), and Organization/Institution names (ORG). All these NER objects are studied in Uzbek linguistics under the topic of eponymous nouns. These units constitute the largest category within onomastics.

Many leading scholars in Uzbek linguistics have contributed to the field of onomastics (the science of names). For instance, Ernest Begmatov, who played a significant role in the formation of Uzbek anthroponymy, deeply studied onomastics, analyzing the formation of anthroponyms, the classification of names by macro and micro scales, and the semantics and social functions of personal names. The scholar began his research in onomastics and defended his doctoral dissertation in 1965 titled "Uzbek Language Anthroponymy" [6] under the supervision of Fahri Kamol. This work was the first to analyze the linguistic and extralinguistic features of eponymous nouns. E. Begmatov studied the lexical-semantic features of Uzbek names, issues of onomastic formations, the ethnography of names, the history of Uzbek surnames and patronymics, and the spelling of Uzbek anthroponyms in Uzbek and Russian [7]. Additionally, the scholar authored several research works on this topic, such as Names and People (1966) [8], The Meaning of Your Name (co-authored, 1968), The Spelling of Personal Names (1970), The Spelling of Uzbek Names (1972) [9], Uzbek Names (1991), The Beauty of Names (1994) [10], and Uzbek Names: Explanations of 14,600 Names (1998) [11]. Ernest Begmatov's scholarly legacy is vast, with nearly 300 published works to date, including 8 monographs and 13 pamphlets.

Uzbek linguists and geographers have also conducted significant research on place names (toponyms), which form a large part of Uzbek onomastics. The data and extensive lexical collections they compiled serve as a database for future NLP applications.

To'ra Nafasov is a linguist and researcher of the toponymy of the Kashkadarya region. In his research, he analyzed the structural and linguistic aspects of place names in the oasis. AT. Nafasov created two explanatory toponymic dictionaries. The first, Explanatory Dictionary of Uzbek Toponyms [12], covers the toponymy of Kashkadarya and Surkhandarya regions at both micro and macro scales, essentially creating a regional dictionary. The second, Educational Explanatory Dictionary of Uzbek Toponyms [13], provides explanations for toponyms (oikonyms, hydronyms, oronyms) from all regions of the republic. Additionally, the scholar has produced significant works such as Uzbek Nomnoma [14] and Uzbek Village Names [15]. T. Nafasov also developed a curriculum for a specialized course titled Toponymy [16] [17].

Among scholars who have researched organization and institution names, M. Saparniyazova is a linguist specializing

in ergonymy. M. Saparniyazova has published over 170 scholarly articles in national and international academic journals and is the author of 4 monographs and more than 20 educational and methodological guides. Saparniyazova Muyassar defended her doctoral dissertation (DSc) on The Structural-Semantic, Linguocultural, and Pragmatic Features of Ergonims in the Uzbek Language [18].

M. Saparniyazova studied ergonyms (names of organizations, enterprises, and institutions) as an onomastic microsystem, analyzing their typology and linguistic features.

IV. Onomastic Units And Ner

In Uzbek linguistics, the concept of onomastic units refers to eponymous nouns. Unlike common nouns, eponymous nouns rename an object, specifying it. This type of object is distinguished from its category and assigned a new name and meaning. In the *Explanatory Dictionary of the Uzbek Language*, a common noun is defined as a noun representing the name of a category of objects [19], while an eponymous noun is defined as "a specific name for a person or object" [20]. Scholars also highlight several other distinctions between them [21]. See Table II.

TABLE II. THE DIFFERENCE OF COMMON NOUN AND EPONYMOUS NOUN

№	Common Noun	Eponymous Noun
1	Can serve as the basis for forming eponymous nouns.	Although derived from common nouns, they lose associative connections and serve to distinguish objects and phenomena from one another.
2	Common nouns are constantly present in consciousness.	Eponymous nouns are memorized mechanically and quickly forgotten.
3	Grammatically used in singular and plural forms.	Eponymous nouns lack this feature.
4	Represent generality.	Represent specificity.

NER – NE (Named Entity) refers to a "named object" within the category of nouns, while R (Recognition) refers to the process of automatic identification. NER is an NLP method used to identify all named entities in a text. If the concept of NER refers to a named entity, then eponymous nouns in Uzbek can be applied under NER objects in NLP. While research worldwide, particularly in English, includes not only nouns but also numbers (cardinal, ordinal) and related combinations (quantity, time, money, date, etc.) as NER objects, this list is based on the structure, grammar, and rules of that language. For Uzbek, it is appropriate to use all onomastic units in texts as NER objects in automatic analysis. Here, onomastic units are identified in texts as NER objects in the following way (the sentence for analysis is taken from the Kun.uz news site [22]):

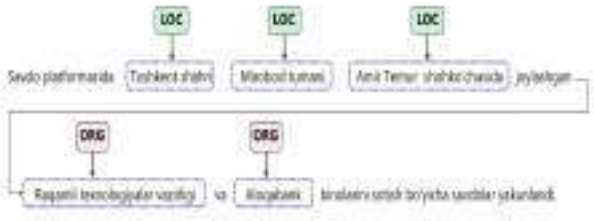

Fig. 1. Identification of NER Objects in Text

NER systems generally operate in two stages:

1) **Identification** – Locating named entities (NER) in the text.

2) **Classification** – Categorizing the identified entities (NER) correctly, such as person, location, organization, etc. [23].

In the example above (Figure 1), nouns, particularly eponymous nouns, are extracted from the sentence. Based on certain models, such as analysis of capitalization rules, eponymous nouns are identified (identification) and categorized into relevant groups (classification).

In Uzbek, the largest categories of eponymous nouns are person names (PER) and place names (LOC), but organization and institution names (ORG) are also frequently encountered. Organization names differ from person and place names in their form and linguistic diversity. The distinctive features of organization names include:

3) **1. Existence in Multiple Languages:** The influx of new words from foreign languages and their assimilation into public consciousness means that the names of organizations, enterprises, and institutions may exist in various languages (English, Russian, Arabic, Turkish, etc.), depending on the nature of their activities or products. For example: See Table III

TABLE III. ORGANIZATION NAMES AND TYPES

№	Organization Name	Type of Organization
1	"Gold step"	Construction company
2	"Artel"	Household appliances and electronics manufacturer
3	"Nestle"	Food products company
4	"MyTaxi"	Transportation service company
5	"Click"	Financial services company

4) **Abbreviations**: Often, organization names appear in abbreviated form (from Italian abbreviature – abbreviation; Latin abbrevio – to shorten). For example: See Table IV

TABLE IV. ABBREVIATION

№	Abbreviated Form	Full Form
1	BMT	United Nations (Birlashgan Millatlar Tashkiloti)
2	O'zMU	National University of Uzbekistan (O'zbekiston Milliy Universiteti)
3	DBQ	State Customs Committee (Davlat bojxona qo'mitasi)
4	MKB	Microcredit Bank (Mikrokreditbank)
5	HET	Regional Electric Networks (Hududiy elektr tarmoqlari)
6	O'TY	Uzbekistan Railways (O'zbekiston temir yo'llari)
7	OS	Supreme Court (Oliy sud)

5) **Word Combinations**: Organization names are not always direct abbreviations but can also be formed from word combinations. For example:

- **Agrobank** – Agro + bank
- **UzAuto** – Uzbekistan + Auto(mobile)

6) **Use of Indicators**: Organization names are often accompanied by indicators. For example: See Table V

TABLE V. USE OF ORGANIZATION NAMES WITH INDICATORS

№	Organization Name with Indicator
1	Ministry of Internal Affairs (Ichki ishlar vazirligi)
2	State Tax Committee (Davlat soliq qo‘mitasi)
3	Intellectual Property Agency (Intelektual mulk agentligi)
4	State Testing Center (Davlat test markazi)
5	Tashkent City Health Department (Toshkent shahar sog’ligni saqlash boshqarmasi)
6	State Technical Control Inspectorate (Davlat texnik nazorat inspeksiyasi)
7	UzAuto Motors Joint-Stock Company (UzAuto Motors aksiyadorlik jamiyati)
8	Uzbekistan Airways Company (Uzbekistan Airways kompaniyasi)

Additionally, indicator words specifying the type of organization include: vazirlik (ministry), qo‘mita (committee), agentlik (agency), markaz (center), boshqarma (department), idora (office), universitet (university), bank (bank), kompaniya (company), jamiyat (society), korxona (enterprise), firma (firm), inspeksiya (inspectorate), fermer-xo’jaligi (farm), AJ (joint-stock company), MCHJ (limited liability company), NTT (non-governmental non-profit organization), YATT (sole proprietorship), and others.

Below is another diagram illustrating the identification of organization names (ORG) in text (the sentence for analysis is taken from the Kun.uz news site [25]):

Fig. 2. Identification of Organization Names (ORG) in Text

V. AUTOMATIC IDENTIFICATION OF ORGANIZATIONAL AND INSTITUTIONAL NAMES AS ONOMASTIC UNITS IN UZBEK USING TWO NLP METHODS: DICTIONARY-BASED AND RULE-BASED APPROACHES

In this study, we explore the automatic identification of organizational and institutional names—classified as onomastic units in the Uzbek language—within texts using two natural language processing (NLP) methods: a dictionary-based approach and a rule-based approach.

Initially, a list of frequently occurring, publicly known, and commonly used organizational and institutional names in speech and written texts was compiled. This list serves as a foundational database and was subjected to a normalization phase. Currently, our lexical database contains nearly 1,000 names of organizations and institutions.

However, we believe that this method does not provide a fully comprehensive solution, as manually collecting such a database is time-consuming and labor-intensive.

The second method involves analyzing two features of organization names: word formation (derivational characteristics) and orthographic conventions.

In terms of word formation, special attention is given to the abbreviation-based structure of organizational names. See Table VI

TABLE VI. FORMATION TYPES OF ABBREVIATIONS IN UZBEK LANGUAGE

Method	Description	Example
Based on initials	Uses the initial letters of each word	BMT — Birlashgan Millatlar Tashkiloti (United Nations)
Based on syllables	Uses the first syllables of words	UzMU — O‘zbekiston Milliy Universiteti (National University of Uzbekistan)
Mixed or partial	Some names are fully or partially abbreviated	IIV — Ichki Ishlar Vazirligi (Ministry of Internal Affairs)

The spelling conventions for organizational abbreviations are as follows:

1. Abbreviations are generally written in uppercase letters.

2. Dots are not used between the capital letters in abbreviations (e.g., BMT, UNESCO).

3. According to Uzbek orthographic rules, when a full form is provided, it is given in parentheses after the first use of the abbreviation: YXHT (Yevropada Xavfsizlik va Hamkorlik Tashkiloti).

4. Suffixes are directly attached to abbreviations: SamDU — SamDUning.

5. Adjacent abbreviations are written separately: O‘zXDP MK (O‘zbekiston Xalq Demokratik Partiyasi Markaziy Kengashi).

6. In the names of state or international organizations, each word starts with a capital letter: Jahon Tinchlik Kengashi (World Peace Council).

7. Abbreviations formed from compound names also begin with uppercase letters: AQSH (Amerika Qo‘shma Shtatlari).

8. Some non-proper compound nouns are also abbreviated using uppercase letters: AES (atomic power station).

9. When part of the abbreviation corresponds to a syllable, only the first letter is capitalized: ToshDTU (Toshkent Davlat Texnika Universiteti).

Based on these linguistic rules, it is possible to construct models for the automatic identification of organization and institution names. For example, regular expressions can be used to detect abbreviations that follow the “capital letters rule”:

- `\b[A-Z]{2,}\b` — identifies abbreviations consisting solely of uppercase letters.
- `\b[A-Z][a-z]*[A-Z][a-z]*\b` — identifies mixed-case abbreviations.

Data Resources and Corpus

A large-scale annotated dataset has been created for this task:

- NER objects: 76,850 entities (collected from mass media texts)
- Human names: 26,788 (from dictionaries and online resources)
- Geographical names: 7,127 (from lexical sources)

This dataset is publicly available in the Uzbek Language Corpus (https://uznatcorpara.uz/), which currently contains 5 billion sentences. See Table VII.

TABLE VII. AUTOMATIC NER ANALYSIS OF 77,821 SENTENCES

Label	Precision	Recall	F1-Score	Support
N	0.894	0.945	0.919	49,368
VB	0.923	0.928	0.925	45,627
JJ	0.940	0.891	0.915	12,907
NUM	0.864	0.873	0.869	2,967
RR	0.940	0.921	0.931	10,283
P	0.976	0.971	0.974	12,844
II	0.969	0.990	0.974	6,206
C	0.976	0.998	0.987	4,771
Prt	0.990	0.998	0.994	2,581
MD	0.963	0.967	0.965	2,539
IM	0.952	0.364	0.526	55
UH	0.947	0.838	0.889	494
NER	0.907	0.872	0.887	113,866
IB	0.687	0.599	0.640	6,106
PUNCT	1.000	1.000	1.000	37,725

OVERALL NER PERFORMANCE

Metric	Value
Precision	0.907 (90.7%)
Recall	0.872 (87.2%)
F1-Score	0.887 (88.7%)
Total NER objects	113,866

The NER system achieved a precision of 90.7%, indicating a high level of accuracy. In earlier evaluations with smaller datasets, the precision scores were lower (e.g., 87.3% for 17,038 sentences and 89.5% for 56,616 sentences). A recall of 87.2% indicates that about 12.8% of NER entities were not successfully detected. The F1-score of 88.7% demonstrates a balanced and effective overall performance. A total of 113,866 named entities were identified across the corpus.

VI. CONCLUSION

This article thoroughly examines the types of onomastic units, their linguistic features, their role in the language system, and their significance in modern linguistics, particularly computational linguistics. The concepts of onomastic units and Named Entity Recognition (NER), a crucial direction in natural language processing, are compared, and their differences and commonalities are highlighted. The methods for identifying the three main types of naming units—person names, place names, and organization names—in NER systems are discussed. Research shows that the accurate identification of onomastic units is important not only for linguistic analysis but also for enhancing the efficiency of artificial intelligence and computational linguistics systems. In this regard, the relationship between onomastic units and NER technologies serves as a key factor in shaping new, integrative directions in linguistics. In the article, two methods for the automatic identification of organization names were analyzed: the dictionary-based approach and the rule-based approach. Existing linguistic rules of the Uzbek language were presented. Real results obtained through automatic analysis based on corpus texts were provided, along with the dataset compiled specifically for the Uzbek language.

REFERENCES

[1] V. A. Maslova. Linguocultural Studies. Moscow: Academia, 2001, 208 p.

[2] N. V. Podolskaya. Dictionary of Russian Onomastic Terminology. Moscow: Nauka, 1988, 96 p.

[3] N. V. Podolskaya. Cited Dictionary, Moscow: Nauka, 1988, 123 p.

[4] E. Begmatov. Uzbek Language Anthroponymy. Tashkent: Fan, 2013, 36 p.

[5] A. V. Superanskaya, V. E. Staltmane, N. V. Podolskaya, A. Kh. Sultanov. Theory and Methodology of Onomastic Research. Edited by A. P. Nepokupny. Moscow: Nauka, 1986, 206 p

[6] Begmatov, E. Uzbek Language Anthroponymy. Ph.D. dissertation, Philological Sciences, Tashkent, 1965, 480 p.

[7] E. Begmatov. On the Meanings of Personal Names, Uzbek Language and Literature, no. 2, 1992, p. 5.

[8] E. Begmatov. Names and People. Tashkent: Fan, 1966, 65 p.

[9] E. Begmatov. The Spelling of Uzbek Names. Tashkent: Fan, 1972, 308 p.

[10] E. Begmatov. The Beauty of Names. Tashkent: Fan, 1994.

[11] E. Begmatov. Uzbek Names. Tashkent: National Encyclopedia of Uzbekistan, 1998.

[12] T. Nafasov. Explanatory Dictionary of Uzbek Toponyms. Tashkent: O‘qituvchi, 1988.

[13] T. Nafasov, V. Nafasova. Educational Explanatory Dictionary of Uzbek Toponyms. Tashkent: Yangi Asr Avlodi, 2018.

[14] T. Nafasov. Uzbek Nomnoma. Karshi: Nasaf, 1993.

[15] T. Nafasov. Uzbek Village Names. Karshi: Nasaf, 1994.

[16] T. Nafasov. Toponymy of Uzbekistan. Tashkent, 1988, p. 22.

[17] T. Sodigova. Toponyms of the Syrdarya Region. Tashkent, 2018, p. 11.

[18] Nafasov T. – Biography, ScienceWeb. Online. Available: https://scienceweb.uz/researcher/6840/biography

[19] Explanatory Dictionary of the Uzbek Language, Vol. II. Moscow: Russian Language, 1981, p. 225.

[20] Explanatory Dictionary of the Uzbek Language, vol. I. Moscow: Russian Language, 1981, p. 62.

[21] B. Qilichev. Onomastics: Textbook. Bukhara, 2023, p. 115.

[22] AKT vazirligi va Aloqabank binosi Aloqabankdan kredit olgan kompaniyaga sotildi, Kun.uz, Apr. 7, 2023. Online. Available: https://kun.uz/news/2023/04/07/akt-vazirligi-va-alogabank-binosi-alogabankdan-kredit-olgan-kompaniyaga-sotildi

[23] B. Elov, M. Samatboyeva. NER: Tagging Issues, in Proceedings of Modern Technologies of Computational Linguistics (CTCL 2025), Tashkent, 2025, p. 248.

[24] Abbreviatura, National Encyclopedia of Uzbekistan (Letter A). Tashkent: State Scientific Publishing House of the National Encyclopedia of Uzbekistan, 2000–2006. Online. Available: https://uz.wikipedia.org/wiki/Abbreviatura