

IDENTIFYING NER (NAMED ENTITY RECOGNITION) OBJECTS IN UZBEK LANGUAGE TEXTS

¹Elov Botir Boltayevich, ²Samatboyeva Madina To‘lqinjon qizi

¹Philosophy Doctor (PhD) on technical sciences, associate professor

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

²Graduate student of Computational linguistics faculty

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

<https://doi.org/10.5281/zenodo.7834009>

Abstract. *This article discusses NER (Named Entity Recognition) features and detection methods important to the field of NLP. The article describes the concepts of quick word recognition, categorization, understanding of the content of the text, noun, and named object. Methods for automatic identification of named objects from Uzbek texts were presented. NER objects, types and examples corresponding to these types are given in detail on the basis of the table. The opinions of scientists who presented models for identifying NER objects from text abroad, engaged in this scientific work and carried out analysis processes through language corpora were studied. Examples of NER types and classifications were presented. The systems of identifying NER objects in Uzbek texts were analyzed on the basis of dictionaries and grammatical rules, and the ideas were proved by examples. “IOB” and “BILUO” schemes were studied for determining NER objects and their boundaries, and examples were presented. Models for automatically processing the text and identifying NER objects from its content have been presented. At the end of the article, an approximate interface of the program for identifying NER objects in Uzbek language texts (“Uzbek NER analyzer”) was presented. The principle of operation of this program, its content is explained in the interface-based case.*

Keywords: *object, noun, name, named object identification, object categorization, NER, NLP, POS tagging, token, tokenization, part of speech, grammar rules, dictionary.*

INTRODUCTION

Since ancient times, people have learned to name everything in existence. They have called “names” to natural creatures, events, inanimate objects, destinations, and even themselves. Through this, they tried to distinguish them from each other, categorize them and keep them in mind quickly.

The “name” (proper noun, named object) in the text serves as one of the elements that determine the main content of the text. This article examines the phenomenon of named entity recognition (NER), which is one of the important issues of NLP. Also, the methods of determining NER and its practical importance are discussed.

REVIEW OF LITERATURE

Many scientists have been involved in the process of identifying NER objects. In particular, *Rowen Brown* mentioned in his articles about the methods of determining NER and anthology (8).

In his scientific works, *Yujian Tang* has given recommendations on the detection of NER objects through the Python programming language and the use of the SpaCy library (32).

Several Indian scientists have expressed their opinion on the identification of NER objects. In particular, *Mehal Gupta* commented on the process of identifying NERs and the proper classification of objects (26). *Dipanjari Sarkar* provides valuable information on NER types,

detection models and Stanford NER Tagger and NLTK. In addition, this scientist lists NER objects in the SpaCy library and provides examples of them (27).

The identification of objects named in the Tatar language has been thoroughly studied. In particular, *Olga Nevzorova*, *Damir Mukhamedshin* and *Alfiya Galieva*, members of the Academy of Sciences of Tatarstan, use cases of electronic language corpora to recognize named objects in their work, algorithms aimed at identifying NER objects, direct and recognition of NER objects based on search queries using inverse search and cited these studies as the example of the national corpus of the Tatar language “Tugan Tel” (25).

Oguzhan Ozcelik and *Cagri Toraman* achieved effective results in identifying NER objects in the Turkish language. In particular, they identified Turkish language NERs based on 20 models using the Transformer model. The program can not only identify NERs, but also work on errors. The program shows the highest F1 score of up to 96.1% (23).

Brahim Ait Benalia, *Soukaina Mihia*, *Nabil Laachfoubia*, *Addi Ait Mloukb* tried to identify NER objects in the Arabic language in their scientific works. They studied six BERT-based models (Bidirectional Encoder Representations from Transformers) and used a bi-LSTM-CRF architecture to identify NER objects in dialectal Arabic. As examples, they used information from media and mass media (22).

METHODS

The process of identifying NER objects is carried out in the following steps:

1. Extraction of information – the first step in determining NER is to extract the objects indicated in the sentence, paragraph, text. At this stage, the whole text is marked and the text border is defined. Here, the text is divided into sentences according to the capital letter system.

2. Tokenization process – chunked sentences are now tokenized within themselves.

3. Determining the limit of tokens according to the “IOB” or “BILUO” scheme (token tagging format) and “assembling” them again – at this point, tokens of several content NERs are “merged” based on the model.

4. Searching for objects – the next process in NER is searching for NER objects in tokens.

5. Assign the correct category to identified NER objects.

Explaining the above process differently, the identified nouns are analyzed not only morphologically, but also semantically. In the process of analysis, attention is also paid to the form of nouns. The nouns separated from the text are defined according to their “proper noun” characteristics. Named nouns in a sentence are separated. Proper nouns are analyzed according to “NER features” (in the text, capitalization is a lexicon outside the dictionary, suffixes are added after it, the object is called by “name” again...). The dictionary also helps to identify NER objects. But even this base cannot be a perfectly effective solution.

In addition, it is not possible to fully identify NER objects in texts by creating a huge list of all common nouns that exist in our language. Not all NER objects in the text can be identified even using grammatical rules. Identifying NER objects in texts generated by speech recognition or software applications (not normalized - text with various spelling errors, redundant characters and words) creates several difficulties. In order to effectively solve this problem, it is advisable to use not only *grammatical rules*, *a dictionary base*, but also *machine learning* and *deep learning* (Machine learning - ML; Deep learning - DL) tools.

THE MAIN PART

NER is a named object

Whenever we hear a word or read a text, we naturally have a habit of identifying and categorizing the word according to people, place, location, numbers, etc. Through this, it is possible to quickly recognize the word, memorize it, categorize it, and understand the content of the text (13). For example, when you hear the word “Samarkand”, we immediately think of three or four attributes associated with it, and it comes to mind that its main category is “place name”. This method is extracting named objects from text content. Named object recognition is one of the main methods of identifying objects in NLP. A named object is a word or phrase that clearly identifies one thing from many things (28).

NER is an NLP technique that can extract key features from text and classify them into predefined categories. The process of identifying personal names, location names, company names, and similar named objects from text that do not exist in the dictionary is an important step in solving many NLP tasks. Named object recognition in NLP is also commonly referred to as object identification, object extraction, or object segmentation. NER object detection algorithms are the following models (18):

- *rule-based analysis*;
- *dictionary based*;
- *POS tagging (Part of Speech – morphological tagging)*;
- *Parsing (syntactic tagging)*.

The purpose of NER is to identify named objects in the text and assign corresponding categories to them. Three main approaches are important for NER: *lexical-based*, *rule-based*, and *machine learning-based*. However, the NER system can combine several of these categories (20).

To understand the process of identifying NER objects from text content, consider the following sentence:



(Tashkent is the capital of Uzbekistan and the largest city in Central Asia by population)

Figure 1. Identification of NER objects (example) (33)

Here are words from the vocabulary of blue-colored people. Some of these nouns represent real objects that exist in the world. For example, from the above, the following nouns represent existing places on the map: “**Toshkent**” (“Tashkent”); “**O‘zbekiston**” (“Uzbekistan”); “**Markaziy Osiyo**” (“Central Asia”).

If we can find nouns in a text, specifically named nouns, with such accuracy, we can use this information to automatically identify a list of named objects in the text in NLP. So the goal of NER is to identify and label these nouns with relevant real-world concepts. (For example: UNICEF is an organization (ORG); Alisher is a person's name (PER)).

NER systems do more than just simple dictionary lookup. Perhaps they use a statistical model to determine how a word appears in a sentence in the text and what kind of noun that word represents (18).

Uzbek texts also contain NER objects. In particular, NER objects represent proper nouns - named nouns. However, in programs aimed at identifying named objects within the English language, only proper nouns are not selected as NER objects. They also contain number units (day, date, year, percentage, amount...). NER objects help to add more meaning to the content of the

sentence. The nouns in the text complement each other in terms of meaning and content and are connected to each other. NER objects help to further enrich the content of the sentence and clarify the main meaning.

NER and its properties

If there are numerical NER objects in the texts, it is not difficult to automatically identify them. But when nouns are involved, it is difficult to distinguish them from other non-NER units. In this case, it is necessary to pay attention to the features of NER objects:

1. Usually always written with a capital letter (Capitalization).

E.g.: Shahar sifatida Toshkent haqidagi birinchi ma'lumotlar eramizdan avvalgi II asrdagi qadimgi-sharqiy manbalarda uchraydi (33).

(The first information about Tashkent as a city can be found in ancient Eastern sources of the 2nd century BC.)

2. Out of Vocabulary (OOV).

E.g.: Birlashgan Millatlar Tashkiloti (BMT) — dunyoda tinchlikni mustahkamlash va xavfsizlikni ta'minlash, davlatlarning o'zaro hamkorligini rivojlantirish maqsadida tashkil etilgan xalqaro tashkilot (3).

(The United Nations (UN) is an international organization established to strengthen peace and security in the world, and to develop mutual cooperation between states.)

3. Morphological analysis almost always has a proper noun (POS Tag);

E.g.: Navoiy yoshligidan Xurosonning (Transoxaniya) bo'lajak hukmdori Husayn Boyqaro bilan do'st bo'lgan (1).

(Navoi was a friend of the future ruler of Khurasan (Transoxania) Husayn Boyqaro from his youth.)

4. Addition of suffixes.

E.g.: Navoiyning "Hamsa"si O'rta Osiyoda yuqori o'ringa ega bo'lgan (1).

(Navoi's "Hamsa" was ranked high in Central Asia.)

5. Clarification of the meaning of the words connected to him in the sentence.

E.g.: Rim tarixchisi Kvint Kursiy Rufning (miloddan avval I asr oxiri — milodiy I asr) yozishicha, Samarqand qal'asi devorining aylanasi taxminan 10,5 km bo'lgan (29).

(According to the Roman historian Quintus Curcius Rufus (late 1st century BC - 1st century AD), the circumference of the Samarkand fortress wall was approximately 10.5 km.)

6. Freedom of placement in the sentence.

E.g.: Xorazm viloyati cho'l zonasida, Xorazm vohasining g'arbiy qismida, o'rtacha 100 m balandlikda joylashgan (37).

Cho'l zonasida, Xorazm vohasining g'arbiy qismida, o'rtacha 100 m balandlikda Xorazm viloyati joylashgan.

(Located in the desert zone of the Khorezm region, in the western part of the Khorezm oasis, at an average height of 100 m.

Khorezm region is located in the desert zone, in the western part of the Khorezm oasis, at an average height of 100 m.)

NER objects

To know what an object is, the NER model must be able to identify the word or string of words that make up the object and determine which category it belongs to. An object is a part of a

sentence that can be identified and separated. Hence, the basis of any NER model is a two-step process:

- *identifying the named object;*
- *categorize the object;*

First of all, it is necessary to create object categories such as location, event, organization, etc., and provide relevant information to the NER model. Then, by tagging words and phrases with their corresponding objects, the NER model identifies the objects and categorizes them.

Table 1.

NER objects

No	English	Uzbek	Composition	Example
1	Person	Inson ismi	Name Surname Father's name Nickname Nickname	Maftuna Alimova (Ahroriy) Alisher qizi (Alisherovna/ binni Alisher/ ibn Alisher/ Shayxzoda) “qora soch” Navoiy
2	Location		State name Province name City name Avenue name Neighborhood name Village name Name of village Street name valley, field, hill, land plot name grave, cemetery, holy place name name of the island the name of the glacier field, desert, meadow, grove, ravine, roads name field, registan, avenue, park, amusement park, the name of the bridges ancient castle, fortress, fortification, wall, the name of caravanserais	Uzbekistan Navoi region Sangal village Neighborhood of peace Baharistan street Nurafshon National garden Kyzylkum desert Sangardak spring Ahmad father's cemetery Islam Karimov named airport Father of Islam Mosque Borijar mountain “Fedchenko” glacier Aviazozol station Alisher Navoi university “Efendi” restaurant “Life” hotel “Abdullah Qadiri” creative school Mount Everest “Kokan railway station”

			residence, camp, market, settlements name, the name of mosques and madrasahs railway name name of intersections train station, bus station, name of airports name of stations mountain and mountain the name of the ranges university, school name cafe, restaurant, bar names hotel, hostel, motel names hospital, salon names	“Dostlik” bridge “Abu Sakhi” market
3	Geo-political entity	Geo-siyosiy hudud		Eurasia Great Britain
4	Organization	Tashkilot		UNICEF
5	Time	Vaqt	Calendar day/month name/hour/minute/second	March 25 April 06:00 5 minutes 1 second
6	Quantity	Miqdor	Percent Age Length Weight Volume Liter Temperature	50% (50 percent) 5 years old 50 meters 100 kg (kilogram, tons, quintals...) 300m ² (square meter) 400 l (liter) 25C (25 degrees, +25C, +25 degrees, -25C, -25 degrees)
7	Money	Pul birligi	Monetary(currency) Count	So‘m – monetary 1000 so‘m - count
8	Hydronymy	Suv ishootlari nomi	The name of the ocean The name of the sea The name of the river The name of the lakes The name of the waterfalls	The Pacific Ocean The Black sea The Nile River Lake Balkhash Niagara Falls

			The name of the springs Name of wells	
9	Work of art	San'at asari	Movie name Title of the artwork Name of artwork Name of sculptural works The name of weaving works...	Titanic "Khamsa" "Mona Lisa" photo "Statue of Liberty" "The king speaks"
10	Live things	jonli mavjudotlar	Animals – hayvon nomlari Plants – o'simlik nomlari	Reks – name of dogs Aloe – name of flowers
11	Language	Tillar		English German
12	Disease	Kasallik		Anemia Dysplasia...
13	Artifact	Yodgorlik		Registon Minorai Kalon
14	Event	Voqea	A historical event A modern event	"Crusades" is a historical event "Independence Day" is a modern event

Object categories can also be provided in abbreviated or non-abbreviated form within the program (19).

Table 2.

NER objects in the SpaCY library provided in abbreviated form

№	NER objects	Abbreviation	Types
1	Person	PER	Name, surname, nick
2	Location	LOC	mountain ranges, bodies of water, etc
3	Nationalities, religious and political groups	NORP	Nationalities, religious and political groups
4	Organization	ORG	organizations
5	Geo-political Entity	GPE	countries, cities, etc
6	Facility	FAC	buildings, airports, etc.

The process of determining NER objects from the sentence structure based on the table is as follows:

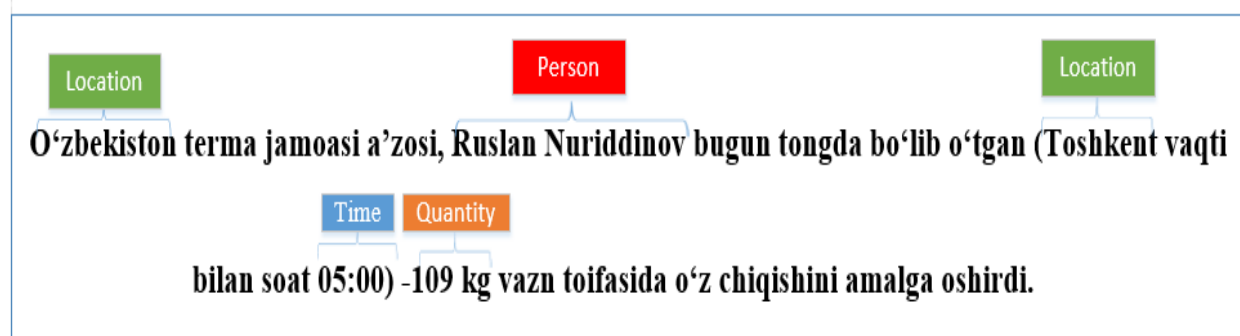


Figure 2. Example of NER objects

The NER objects in this sentence are:

Ruslan Nuriddinov	Odamning ismi (Person)
O‘zbekiston, Toshkent	Joylashuv (Location)
05:00	Vaqt (Time)
109 kg	Og‘irlik - miqdor (Quantity).

NER objects are essential parts of a particular sentence, including a noun or verb phrase (or both). In the above sentence, NER objects and their semantic groups were distinguished.

NER detection process

After reading a certain text, naturally named objects such as people, places, etc. are identified. This process is the process of identifying NER objects. For example, consider the following sentence:

Namanganda quvvati 150 MVt bo‘lgan quyosh fotoelektr stansiyasini qurish bo‘yicha Xitoyning GD Power – Powerchina kompaniyasi g‘olib deb topilgandi (24).

(The Chinese company GD Power - Powerchina was declared the winner for the construction of a 150 MW solar photovoltaic power plant in Namangan.)

From the above sentence, we can identify three types of (named) objects:

№	NER turi	Misol
1	Joylashuv (LOC)	Namangan, Xitoy (Namangan, The Chinese)
2	Miqdor (QUANTITY)	150 MVt (150 MW)
3	Tashkilot (ORG)	GD Power – Powerchina (GD Power – Powerchina)

But in order to automatically perform the same work through computers, it is necessary to create models that help to recognize objects so that the computer can classify them. Machine learning and natural language processing (NLP) are used for this.

NLP: It studies the structure and rules of language, forming intelligent systems capable of extracting meaning from text and speech.

Machine learning: It helps to train the machine based on the given data and improve it.

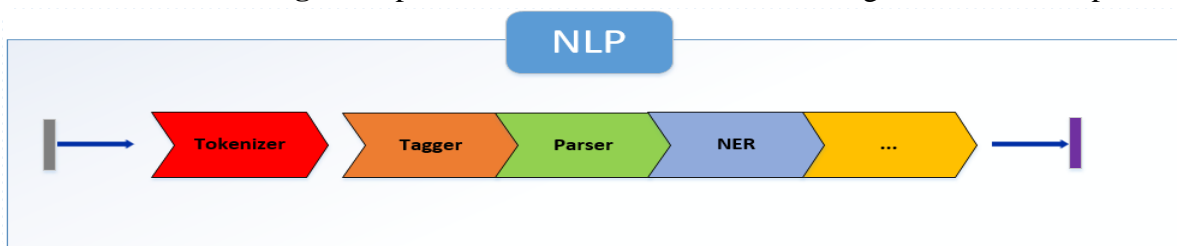


Figure 3. NER detection process(14).

1. **Tokenizer** - speech is divided into tokens.
2. **Tagger** - determined based on the model according to the specified tags.
3. **Parser** - syntactic analysis is performed and the appropriate category is determined.
4. **Result** – NER object is defined.

The process of identifying NER objects is also widely used in morphological text analysis systems. As a result of the morphological analysis of the text, it is necessary to perform the process of PosTagging (categorization) on the words that are not in the dictionary. This means that words of this type are most likely NER objects. So, when we identify each token in the sentence using the NER tagging model, our sentence looks like this:

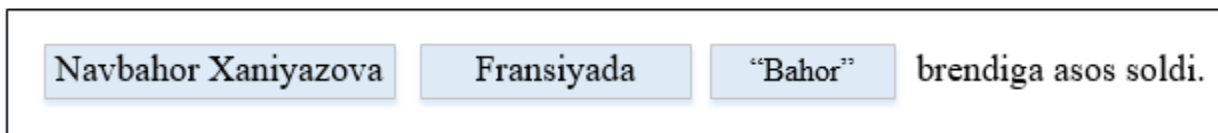


Figure 4. Identification of NER objects by morphological analysis (example) (7).
 (Navbahor Khaniyazova founded the “Bahor” brand in France.)

NERs in this sentence:

Navbahor Xaniyazova	PERSON (odam ismi)
Fransiyada	LOCATION (joylashuv)
Bahor	ORGANIZATION (tashkilot)

The tokenization process is important in identifying NER objects. Tokenization is one of the most common ways to work with text data. But what does the term “tokenization” actually mean? Tokenization is essentially dividing a phrase, sentence, paragraph, or entire text document into smaller units, such as individual words or terms. Each of these small units are called tokens (16). Some words may be split into two or more words during tokenization. This is standard practice because some words can be tokenized in a way that doesn't depend on their meaning. In such cases, byte-pair encoding (BPE) tokenization according to **BERT** models can be implemented (11). In addition, the NER objects limit is broken if the text or sentence content is divided into tokens. If there is an NER object in the sentence, the content of which consists of several tokens, how is the analysis process carried out in this case? It uses “**IOB**” or “**BILUO**” (format used for tagging tokens) (10) schemes (12). These schemes are effective methods for identifying single-component and multi-component NER objects. When text content is divided into sentences, each sentence is divided into tokens within itself.

We analyze the following sentence using the “IOB” scheme (**Figures 5-6**):

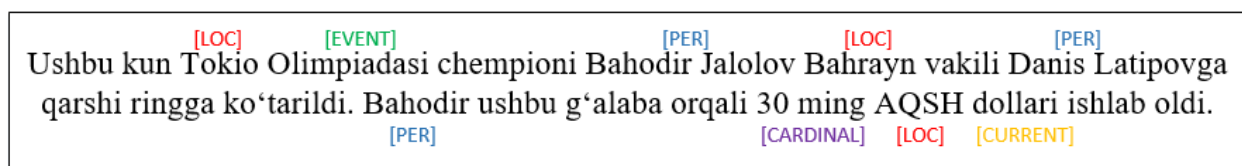


Figure 5. Analysis using the “IOB” scheme (2).

Ushbu	O
kun	O
Tokio	B-LOC
Olimpiadasi	I-EVENT
chempioni	O
Bahodir	B-PER
Jalolov	I-PER
Bahrayn	B-LOC
vakili	O
Danis	B-PER
Latipovga	I-PER
qarshi	O
ringga	O

ko'tarildi	O
.	O
Bahodir	B-PER
ushbu	O
g'alaba	O
orqali	O
30	B-CARDINAL
ming	I-CARDINAL
AQSH	I-LOC
dollari	I-CURRENT
ishlab	O
oldi	O
.	O

“IOB” means:

- “**B**” – “beginning”,

- **“I”** – “inside”,
- **“O”** – “outside”

When we divide this sentence into tokens, NER objects are also divided into tokens. But in the case that the composition of NER objects consists of several words, the objects can be reassembled in this way using the “IOB” or “BILUO” scheme.

[EVENT] Ushbu kun Tokio Olimpiadasi chempioni Bahodir Jalolov Bahrayn vakili Danis Latipovga qarshi ringga ko'tarildi. Bahodir ushbu g'alaba orqali 30 ming AQSH dollari ishlab oldi.
 [PER]
[LOC]
[PER]

[PER]
[CARDINAL]
[MONEY]

Figure 6. The result of the analysis using the “IOB” scheme

(In this day, the champion of the Tokyo Olympics, Bahadir Jalolov, entered the ring against the representative of Bahrain, Danis Latipov. Bahadir earned 30,000 US dollars through this victory.)

Here, for example, the unit “Tokio Olimpiadasi” (“Tokyo Olympics”) is two tokens, but one NER object is an event. When we divide this NER content into a token, “Tokio” (“Tokyo”) is the name of **the place** (LOC) and we designate it as **“B” - begin**; “Olimpiadasi” (“Olympiad”) is an **event** and we define it as **“I” - inside**. (In the “BILUO” scheme, it is denoted by **“L” - last**). It is through this scheme that the boundaries of all NER objects are determined.

“BILUO” scheme is one of the most effective schemes for NER object detection. We pay attention to the following sentence:

28-dekabr
2022-yil
“Finanse TSI” MChJ
 birja savdolarida “Kapitalbank” ATB aksiyalarining 50 foizini sotib oldi.

Figure 7. Identification of NER objects using the “BILUO” scheme (11).

(December 28, 2022 “Finanse TSI” MCJ bought 50% of the shares of “Kapitalbank” ATB in the stock market.)

28	B
-	I
dekabr	L
2022	B
-	I
yil	L
Finanse	B
TSI	I
MChJ	L
birja	O

savdolarida	O
“Kapitalbank”	B
ATB	L
aksiyalarining	O
50	B
foizini	L
sotib	O
oldi	O
.	O

“BILUO”:

- **“B”** – “beginning”,
- **“I”** – “inside”,
- **“L”** – “last”,
- **“U”** – “unit” (Unit - for NER objects with one content)
- **“O”** – “outside” (Outside - to represent non-NER objects).

As a result of our observations, we have researched that the BILUO scheme is a more effective method for more precisely demarcating the boundaries of NER objects.

Approaches to the detection of NER

The process of identifying NER objects in texts is carried out in four ways:

1. Dictionary-based approach – to identify NER objects in the Uzbek language, the following dictionary database is needed:

- “Uzbek names” (6)
- “Explanation of Uzbek names” (5)
- “The meaning of 505 names” (20)
- “Introduction to toponymy” (35)
- “Short Toponymic Dictionary” (34) (“Краткий топонимический словарь” – Nikonov explained the origin of the names of about 4,000 large geographical objects in these dictionaries.)
- “Brief explanatory dictionary of place names” (38)
- “Learning-explanatory dictionary of toponyms of the Uzbek language” (31)
- “Phytonyms in the Uzbek language” (9).

2. Rule-based approach - philological rules are based on *grammar (lexicology, morphology, syntax)* and *models* are formed based on these rules. The words in the sentence are most likely to be NER objects in the following cases:

– “**who?**”, “**what?**”, “**where?**” in the sentence adverbial nouns that answer questions, “how many?” numbers and “when?” if there are words denoting the time when the question is answered;

– when proper nouns come in the form of abbreviations - **abbreviations**, in most cases, if several capital letters are included in their structure;

– if the composition includes units that form a person's surname and first name: “**ibn**” (son - Ahmad ibn Muhammad), “**bint (binni)**” (daughter) (Zuhra bin Abdullah); “**son (o’g’li)**”, “**daughter (qizi)**” (son of Ahmad Fazil, daughter of Hakima Fazil); “**zoda**” (Hamza Hakimzada, Turgun Sharifzada), “**iy**”, “**viy**”, “**iya**”, “**via**” (Abdulla Qadiri, Abdulla Alavi, Mirzakalon Ismaili, Muzayyana Alaviya); **-ov, (-ova, -yev (-yeva)** (Alisherova, Aliyev); **-ovna,(-yevna), -ovich(-yevich)** (Mustafoyevna, Erkinovich); sometimes there are cases of **0** forms, in which two words start with a capital letter (Parda Tursun, Sultan Jora, Ilyas Muslim).

– “Name + words denoting blood kinship” are present in this form; (Aunt Nigora, Aunt Salima.

3. Machine learning is an approach based on machine learning.

4. Deep learning is an approach based on deep learning.

One way to identify names and personal information is to use the Pingar online application and the Google Maps API. These applications identify NER objects such as people, organizations, addresses, emails, ages, phone numbers, URLs, dates, times, money, and amounts (4). In addition, several electronic libraries have also focused on NER detection (36)

NER object detection algorithms can be classified as follows(12):

1. Based on traditional ML:

- *Conditional Random Fields (CRF)*
- *Maximum-entropy Markov model..*

2. Based on neural networks:

- *LSTM*

- CNN
- Transformers

Table 3.

NER Detection Libraries(17)

№	The name of libraries	What programming language is used
1	SpaCY	Python
2	GATE	Java
3	OpenNLP	Java
4	CoreNLP	Java
5	NLTK	Python
6	CogcompNLP	Java

A program that identifies NER objects in Uzbek language texts:

It is very important to identify NER objects in Uzbek language texts. This program is necessary not only for the morphoanalyzer, but also for tokenization, lemmatization, semantic analysis corpus, and homonym detection programs in the future. It is also a valuable resource for fast search engine-based applications (mainly searching libraries for books by author or title), editing applications, address input applications, electronic maps, and web applications based on them. Below we present the approximate interface of the NER object detection program:



Figure 8. Approximate interface of a program that identifies NER objects in Uzbek language texts

In this program, text is entered by the user in the text input part (left side of the interface) and the search function (magnifying glass icon) is started. Below the text input area will appear the “tag(s)” we entered in the “model” section. On the right side of the interface, the names of various NER objects are listed (the number of these objects may increase), and the user specifies the names of the objects he needs (want to be identified). At the bottom of the interface, the sentence entered by the user is displayed again in the “result” state. The identified NER objects

appear proportionally (the same) as the color of the objects identified above and also show the name (category) of the object.

Summary

NER is the process of automatically classifying words in a sentence into noun categories. When identifying NERs, the desired result cannot be achieved using a dictionary search or grammar rules. But based on these approaches, initial indicators can be achieved. Therefore, various traditional ML and deep learning-based algorithms are used to solve this problem. Identifying named objects from the text helps to easily understand the content of the text and to advance the main idea in the content. The object named in the article, the NER object, was analyzed and its properties were studied. NER objects and their types are presented on the basis of the table. Ideas are proved with examples. NER object detection approaches (rule-based, dictionary-based, machine-based, and deep learning-based) were analyzed one by one. Conclusions were made on the identification of NER objects in Uzbek texts. Identification of NER objects in Uzbek language texts is one of the automatic analysis processes performed on the text. The NER object detection program not only automatically recognizes the text, names it, searches for non-dictionary lexicons, assigns them a category, but also serves as a valuable resource for the Uzbek language morphoanalyzer and semantic corpora.

REFERENCES

1. Alisher Navoiy - Vikipediya (wikipedia.org)
2. Bahodir Jalolov 30 ming AQSh dollari ishlab oldi (xabar.uz)
3. Birlashgan Millatlar Tashkiloti - Vikipediya (wikipedia.org)
4. David Nettleton. Commercial Data Mining. Processing, Analysis and Modeling for Predictive Analytics Projects. Book. 2014. P-172.
5. E. Begmatov. Explanation of Uzbek names. Publishing House of the National Encyclopedia of Uzbekistan. 2016
6. E. Begmatov. Uzbek names. Encyclopedia publishing house. 2007
7. Fransuzlarda ayolning jamiyatdagi o'rni, kelin va qaynona obrazi: Fransiyaga kelin bo'lgan xorazmlik tikuvchi qiz hikoyasi (kun.uz)
8. How Does Named Entity Recognition Work: NER Methods? | by Roger Brown | Cogito Tech LLC | Medium
9. <https://arxiv.uz/uz/documents/diplom-ishlar/tilshunoslik/o-zbek-tilida-fitonimlar>
10. [https://en.wikipedia.org/wiki/Inside%20%93outside%20%93beginning_\(tagging\)](https://en.wikipedia.org/wiki/Inside%20%93outside%20%93beginning_(tagging))
11. <https://kun.uz/uz/news/2023/01/07/kapitalbank-atb-uzum-ekotizimining-bir-qismi-boladi>
12. <https://medium.com/swlh/a-beginners-introduction-to-named-entity-recognition-ner-2002b1a010c1>
13. <https://ru.shaip.com/blog/named-entity-recognition-and-its-types/>
14. <https://spacy.io/usage/processing-pipelines>
15. <https://towardsdatascience.com/named-entity-recognition-with-deep-learning-bert-the-essential-guide-274c6965e2d>
16. <https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/>
17. <https://www.analyticsvidhya.com/blog/2021/06/nlp-application-named-entity-recognition-ner-in-python-with-spacy/>

18. <https://www.analyticsvidhya.com/blog/2021/06/part-10-step-by-step-guide-to-master-nlp-named-entity-recognition/>
19. <https://www.geeksforgeeks.org/python-named-entity-recognition-ner-using-spacy/>
20. https://www.researchgate.net/profile/Azamat_Primov/publication/341649652_O'zbek_tili_teonimlarini_o'rganish_masalalariga_doir_Ob_issledovanii_teonimov_v_uzbekskom_azyke_About_learning_Uzbek_language/links/5ecd104c45851529451051ca/Ozbek-tili-teonimlarini-organish-masalalariga-doir-Ob-issledovanii-teonimov-v-uzbekskom-azyke-About-learning-Uzbek-language.pdf
21. https://www.researchgate.net/publication/261760150_A_Hybrid_Model_for_Named_Entity_Recognition_Using_Unstructured_Medical_Text
22. <https://www.sciencedirect.com/science/article/pii/S1877050922007141/pdf?md5=0093d48942e307bf95a86b73c3cc0b88&pid=1-s2.0-S1877050922007141-main.pdf>
23. Information Processing & Management | Vol 59, Issue 6, November 2022 | ScienceDirect.com by Elsevier.
24. Namanganda FESqurishbo'yicha Xitoykompaniyasitomonidan berilgantaklifqaytako'ribchiq iladi (kun.uz)
25. Named Entity Recognition in Tatar: Corpus-Based Algorithm (ceur-ws.org)
26. Named Entity Recognition(NER) using Conditional Random Fields (CRFs)in NLP | by Mehul Gupta | Data Science in your pocket | Medium ;
27. Named Entity Recognition: A Practitioner's Guide to NLP - KDnuggets
28. Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty. Named Entity Recognition in English Text// Journal of Physics: Conference Series.2020. 15-p
29. Samarqand - Vikipediya (wikipedia.org)
30. Shubham Singh. How to Get with NLP – 6 Unique Methods to Perform Tokenization. 2019. 2022
31. T. Nafasov, V. Nafasova. An explanatory dictionary of toponyms of the Uzbek language. New Century Generation Publishing House.
32. The Best Way to do Named Entity Recognition (NER) | by Yujian Tang | Dev Genius (medium.com)
33. Toshkent - Vikipediya (wikipedia.org)
34. V. A. Nikonov. Qisqacha toponimik lug'at ("Краткийтопонимическийсловарь. 1966
35. V. A. Nikonov. Toponimikaga kirish ("Введениевтопонимики"). 1965
36. Venkat N. Gudivada, C.R. Rao. Handbook of statistics. Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications. 2018. P-32
37. Xorazm Region - Wikipedia
38. Z. Dosimov, H. Egamov. Brief explanatory dictionary of place names. Teacher's publishing house. 1977