

NER (NAMED ENTITY RECOGNITION – NOMLANGAN OBYEKTNI TANIB OLISH) OBYEKT LARI TARKIBIDA TOPONIMLARNI AVTOMATIK ANIQLASH

Samatboyeva Madina To'liqinjon qizi

ToshDO'TAU,

Kompyuter lingvistikasi mutaxassisligi magistranti

msamatboyeva@gmail.com

Annotatsiya: Ushbu maqolada NER (Named Entity Recognition) obyektlari tarkibiga kiruvchi toponim (joy nomi)lar haqida fikr yuritildi va ularning turlari tadqiq qilindi. Maqolada atoqli va turdosh ot tushunchasiga ham ta'rif berildi. Toponim turlari jadvallar asosida misollar bilan birgalikda keltirib o'tildi.

Kalit so'zlar: *NER, obyekt, nom, nomlangan obyekt, turdosh ot, atqli ot, joy nomi, toponim, leksika.*

Abstract: In this article, toponyms included in NER (Named Entity Recognition) objects were considered and their types were studied. The article also defined the concept of common nouns and common nouns. Types of toponyms are given together with examples on the basis of tables.

Key words: *NER, object, name, named object, cognate noun, proper noun, place name, toponym, lexicon.*

Аннотация: В этой статье рассматриваются топонимы (географические названия), которые являются частью объектов NER (Named Entity Recognition), и исследуются их типы. В статье также дается определение понятия имя собственное и имя существительное нарицательное. Типы топонимов приведены вместе с примерами на основе таблиц.

Ключевые слова: *NER, объект, имя, названный объект, имя существительное нарицательное, имя собственное, топоним, лексика.*

Bizni o'rab turgan har qanday geografik obyektning nomi bor va bu “nom” obyektlarni bir-biridan farq qilish uchun xizmat qiladi. Shu sababdan, tildagi ko'pgina so'zlar nomlarga (atoqli otga) aylangan [1]. Kundalik hayotimizdagi “kutubxona”, “maktab”, “bog'cha” so'zlari ham obyektни bildiruvchi so'zlar, ammo ular turdosh ot (appellyativ – lotincha “atoqli ot aksi”)lardir. Turdosh va atoqli ot tushunchalari o'zi anglatib kelgan obyektning “nomlanish” xususiyatiga ko'ra bir-biridan farq qiladi. Ya'ni, nomlangan hamda nomlanmagan obyekt mazmunning kengayishi va qayta nomlanish sifati bilan ajralib turadi.

NE (Named Entity) – ot so'z turkumi doirasida nomlangan obyekt – “atoqli ot”ni ifodalaydi [2]. Nomlangan obyektlarni avtomatik aniqlovchi jarayon – **NER (Named Entity Recognition)** – nomlangan obyektни tanib olish bo'lib, matndagi

barcha nomli obyektlar aniqlanadi. NER **tabiiy tilni qayta ishlash (NLP – Natural Language Processing)** ning ko‘plab sohalarida qo‘llaniladi [3]. NER obyektlaridan biri – *joy nomlari (toponim)*dir.

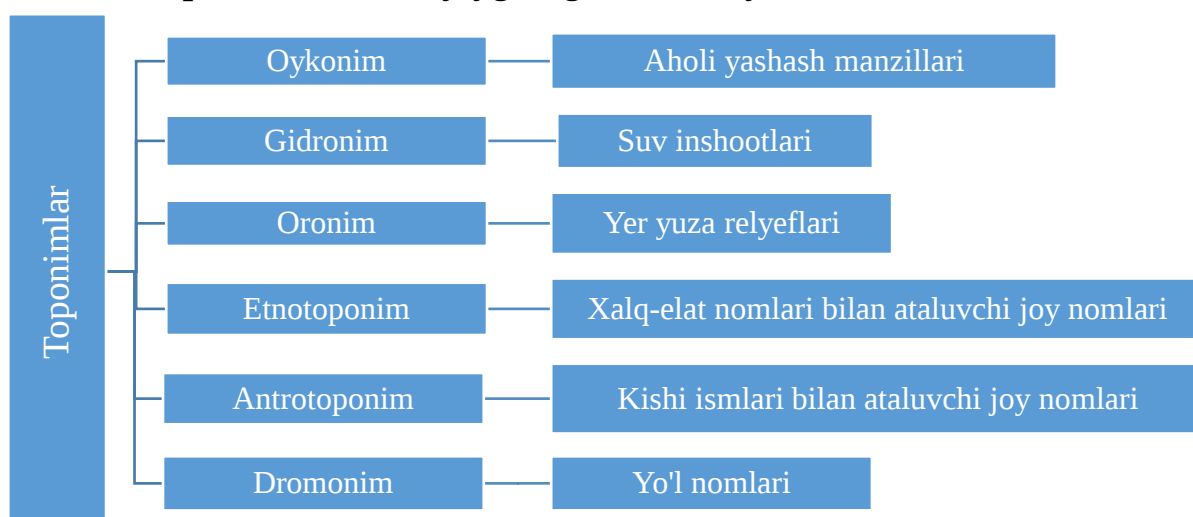
Toponimlar

Yuqoridagi turdosh ot sifatida keltirilgan misollar (“kutubxona”, “maktab”, “bog‘cha”) “joy”ni bildiruvchi obyektlarni ifodalaydi. Ushbu so‘zlarni eshitganimizda, ko‘z oldimizda umumiy “bino” ko‘rinishiga ega obyektlar tasvirlanadi. Agar ushbu joyni bildiruvchi obyektlar qayta nom bilan nomlansa (masalan, “Alisher Navoiy kutubxonasi”, “Avloniy maktabi”, “Lolazor bog‘chasi”), ular atoqli otga aylanadi va “toponimlar” deb yuritiladi. Joy nomlari, **geografik nomlar** yoki **toponimlar** deb ataladi. Toponimlarni **toponimika** fani o‘rganadi. **Toponimika** yunoncha **topos** - joy va **onoma (yoki onima)** - nom so‘zlaridan tarkib topgan.

Toponimlar til lug‘at tarkibining bir qismi bo‘lib, til qonuniyatlariga bo‘ysunadi [4]. Toponimlar tilshunoslik (lingvistika)ning onomastika (yunoncha “onomastike” - nomlash, nom qo‘yish san’ati [5] bo‘limida o‘rganiladi.

Toponimlar joy xususiyatini, uning tarixini, ushbu obyekt bilan bog‘liq voqea-hodisani anglatgan oddiy so‘z va nomning birikuvidan hosil bo‘lgan atoqli otdir. Toponimlar doirasida uchta tushunchani farqlab olish lozim: **“Toponimika”**, **“Toponimiya”**, **“Toponim”**.

- **Toponimika** – joy nomlari, ularning tuzilishi, geografik nomlarning hosil bo‘lishi va rivojlanishi bilan shug‘ullanuvchi *fan*;
- **Toponimiya** – *joy nomlari yig‘indisi* (masalan, Buxoro toponimiyasi – Buxoro joy nomlari yig‘indisi);
- **Toponim** – bu bitta joyga tegishli *xususiy nom*.



1-rasm. Toponim turlari

Toponimlar tilshunoslik nuqtayi nazari bilan o‘rganilganda nomning paydo bo‘lishi, uning tarixi, yaratilishi, leksik-semantik tasnifi, lingvistik strukturasi, so‘z

turkumi, uning atoqli otligi, qaysi tilga mansub leksika ekanligi, toponim bilan bog'liq til qonuniyatlari o'rganiladi. Ammo, rivojlanib borayotgan asrimizda, lingvistik qarashlar kompyuter tiliga moslashtirilib, zamonaviy elektron dasturlar yaratilayotgan bir davrda, qo'lda joy nomlarini aniqlash va ular bilan bog'liq statistik ilmiy tadqiqot ishlarini olib borish mushkul.

Ma'lum bir hududdagi geografik nomlarning yig'indisi shu hududda yashovchi xalqlarning asrlar davomida, nomlar yaratishdagi ijod mahsuli hisoblanadi. Dunyoda qancha geografik nom borligini hech kim aniq bilmaydi. Taxminiy hisoblarga ko'ra, butun yer sharida yarim milliarddan ortiq geografik nom bor. Holbuki, ular qatoriga soy, jilg'a, buloq, quduq, jar, qir, mahalla, guzar, ko'cha kabi mikrotoponimlar kirmaydi. Agar, yer yuzidagi barcha katta-kichik joy nomlarini hisobga olish imkoni bo'lganda, son-sanoqsiz raqam hosil bo'lar edi [1]. Ushbu zaruratlar matnni qayta ishlash jarayonida joy nomlarini avtomatik aniqlash ishini amaliyotga joriy etish kerakligini taqozo etmoqda.

1. **Oykonim** – aholi yashash manzillari bo'lib, “polinimiya”, “urbanomiya” atamallari bilan ham yuritiladi. Aholi yashash manzillari tuzilish tarkibi turli leksikalardan iborat bo'ladi.

Toponimlarning *oykonim (aholi yashash manzillari)*lar turi hosil bo'lishida ishtirok etadigan leksikalar:

- Suvga oid leksikalar;
- Yer yuza relyeflariga oid leksikalar;
- O'simlik, daraxtlarga oid leksikalar;
- Hayvonlarga oid leksikalar;
- Kasb-hunarga oid leksikalar.

1-jadval. Oykonimlar hosil bo'lishida ishtirok etadigan leksikalarga misollar

№	Leksika turi	Oykonimni hosil qiluvchi leksika	Hosil bo'lgan toponim(lar)
1	Suvga oid leksikalar	Daryo	Daryobo'yi, Daryoliq, Sirdaryo, Qoradaryo
		O'zak – “suvi ba'zan qurib qoladigan kichik ariq”	Ko'ko'zak, Injenero'zak, Sario'zak
		Ro'd, ro'dbor – “daryo”, “kanal”	Ro'dak, Labiro't, Ro'dasoy
		Kanal	Kanalboyi, Kanalobod, Kanalpasti
2		Adir	Bo'lakadir, Adirmoq

	Yer yuza relyeflariga oid leksikalar	Bel	Beltoʻv, Beltepa
		Dara	Omondara, Palandara
		Doʻng	Doʻng, Doʻngariq
3	Oʻsimlik, daraxtlarga oid leksikalar	Paxta	Paxta, Paxtakor, Paxtakon, Paxtakesh
		Ajriq	Ajriqti
		Arpa	Arpa, Arpakoʻl, Arpalikoʻl, Arpapoya
		Archa	Archazor, Yakkaarcha
4	Hayvonlarga oid leksikalar	Aygʻir	Aygʻir, Aygʻirtepa
		Baliq	Baliqli, Baliqlikoʻl
		Burga	Burgali, Burgalik
		Boʻri	Boʻrisoy, Boʻritepa
5	Kasb-hunarga oid leksikalar	Aravachi	Aravakashlar, Aravachilar
		Baxshi	Baxshilar, Baxshiyon, Baxshitepa
		Baliqchi	Baliqchi, Baliqchilar
		Boʻzchi	Boʻzachi, Boʻzchi

2. **Gidronimlar** – (yunoncha “hidro” — suv, “nom” — nom) suv inshootlari nomi boʻlib, okean, dengiz, koʻl, soy, jilgʻa, quduqning har biri suv obyektidir [4]. Hidronimlarning ham tuzilish tarkibi suv inshootiga tegishli boʻlgan leksikalardan iborat boʻladi.

2-jadval. Hidronimlar hosil boʻlishida ishtirok etadigan leksikalarga misollar

№	Gidronimlarga oid leksikalar	Hosil boʻlgan toponim(lar)
1	Dengiz	Dengizkoʻl
2	Daryo	Sirdaryo, Qashqadaryo, Qoradaryo
3	Ariq	Yangiariq, Toshariq, Moyariq, Ariqboʻyi, Beshariq
4	Koʻl	Koʻliyon, Koʻlob, Labikoʻl, Luqlikoʻl, Oqkoʻl
5	Rud	Shahrud (shahr i rud), Bolorud, Labirud (lab i rud)

6	Nahr (nar, nor)	Nahr ash-Shosh, Ulug'nor, Norin, Minor,
7	Quduq	Sho'rquduq, Uchquduq, Jalolquduq, Oqquduq

Bundan tashqari yana bir qancha topotermin – gidronimlarni bildiruvchi (yoki ularga ishora qiluvchi) leksikalar mavjud. Masalan, *seldara*, *akvatoriya*, *ob*, *qayirma*, *obshor*, *chashma*, *havza*, *quvur*, *hovuz*...

3. **Oronimlar** – tog'-u toshlarga emas, balki re'lyefning salbiy shakllari - vodiylar, daralar, jarliklar, soyliklar, o'yiq-o'ypotlar, shuningdek, tekislik, past tekisliklar va qumliklar nomlaridir. [4]

5-jadval. Oronimlar hosil bo'lishida ishtirok etadigan leksikalarga misollar

№	Oronimlarga oid leksikalar	Hosil bo'lgan toponim(lar)
1	Ranglar (Oq, qora, sari(q), qo'ng'ir...)	Qo'ng'iro't, Qo'ng'irobod, Qoraqo'rg'on, Saribozor
2	Bel – Baxmal, Zomin tumanlarida “dovon”	O'rtabel, Ayribel, Beltov, Oqbel, Qumbel
3	Dovon – “tog'dan oshaverishga yaqin dam olinadigan tekisroq maydoncha”	Kandirdovon
4	Egar (Navoiyda va qurama shevasida – “iyar”)	Egarchi, Iyartosh, Egarbozor
5	Zov (zovlin)	Zovurak, Avezov, Surxzov
6	Qir (qirang, qirpang)	Qaqir, Qirguli, Qoraqir, Uzunqir

4. **Etnotoponim** – xalq-elat nomlari bilan ataluvchi toponim (joy nomlari)dir. Toponim – joy nomiga nisbat berilganda, nomning kelib chiqishi ko'p hollarda tarixiy aspektlarga asoslanadi. Buni O'zbekiston toponimiyasi jihatidan tadqiq etsak, mamlakatimiz xududida qadimdan turli qabila-urug'lar birgalikda yashab kelishgan. Tarixiy manbalarda ham O'zbek xalqining o'zi 99 urug'dan iborat bo'lganligi haqida ma'lumotlar keltiriladi. Qaysi bir urug' yangi bir joyda o'troq o'rinishsa, shu joy nomini o'z urug'i nomi bilan atashgan. Hozirgi kunda joy nomlari etnotoponim bo'lsa, ushbu hududda qachonlardir ushbu urug'ga mansub xalq yashagan deb taxmin qilinadi.

6-jadval. Etnotoponimlar hosil bo'lishida ishtirok etadigan

leksikalarga misollar

№	Etnotoponim oid leksikalar	Hosil bo'lgan toponim(lar)
1	Qozoq	Qozoq, Qozoqon
2	O'zbek	O'zbekon, Yuqori O'zbekon
3	Dovon – “tog‘dan oshaverishga yaqin dam olinadigan tekisroq maydoncha”	Kandirdovon
4	Egar (Navoiyda va qurama shevasida – “iyar”)	Egarchi, Iyartosh, Egarbozor
5	Zov (zovlin)	Zovurak, Avezov, Surxzov
6	Qir (qirang, qirpang)	Qaqir, Qirguli, Qoraqir

6-jadval. Dramonim (yo‘l, ko‘cha nomlari)larga misollar

№	Etnotoponim oid leksikalar	Hosil bo'lgan toponim(lar)
1	“Yo‘l” so‘zi qo‘shib yoziladigan dramonimlar	Yangiyo‘l, Toshyo‘l, Eskiyo‘l
2	“Yo‘l” so‘zi ajratib yoziladigan dramonimlar	Samarqand xalqa yo‘li, Buyuk ipak yo‘li, Yangi yo‘l
3	“Yo‘l” so‘zi chiziqcha(defis) bilan yoziladigan dramonimlar	Toshkent-Jizzax yo‘li, Toshkent-Guliston yo‘li

Xulosa qilib shuni aytish mumkinki, toponimlarni o‘rganish va ularni avtomatik aniqlovchi dastur modellarini yaratish matnlar ustidagi kompyuter yordamida bajariladigan ishlarni osonlashtiradi. Matnlar qayta ishlanadi va qidiruvga asoslangan dasturlarda ham ushbu modellar keng qo‘llaniladi. Toponimlar aniqlanganda matn mazmuniga asoslangan asosiy fikrni avtomatik olishga ham yordam beradi.

Foydalanilgan adabiyotlar:

1. Q. Hakimov. Toponimika. “Mumtoz so‘z” nashriyoti. Toshkent. 2016. 5 b, 47 b.
2. Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty. Named Entity Recognition in English Text. Journal of Physics: Conference Series. 2020
3. Susan Li. Named Entity Recognition with NLTK and SpaCy. 2018.1p
<https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>



4. S.Qorayev. Toponimika. O'zbekiston faylasuflari milliy jamiyati nashriyoti Toshkent. 2006. 5-6 b, 47 b, 66 b
5. <https://qomus.info/oz/encyclopedia/o/onomastika>