

O'ZBEK TILIDAGI MATNLARDA ATOQLI OTLARNI AVTOMATIK ANIQLASH NE VA NER TUSHUNCHASI

Abdullayev Abdulla Quranbayevich
“Urganch innovatsion universiteti” NTM
Oliy ta’lim muassasasi Ta’limni kredit
tizimini boshqarish bo‘lim boshlig‘i.

Elov Botir Boltayevich
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti.
Texnika fanlari falsafa doktori(PhD), dotsent.

Samatboyeva Madina To‘lqinjon qizi
Kompyuter lingvistikasi mutaxassisligi magistranti
Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili
va adabiyoti universiteti.

Annotatsiya: Ushbu maqolada o‘zbek tilidagi matnlarda atoqli otlar va ularni avtomatik aniqlash metodlari to‘g‘risida fikr yuritilgan. Xususan, atoqli otlarning NLPda ifodalanishi va ularning lingvistik xususiyatlari o‘rganilgan. NE va NER tushunchalariga atroflicha ta’rif berilib, ularning xususiyatlari, aniqlash usullari va matnlar tarkibida kelish holatlari tadqiq qilingan.

Kalit so‘zlar: Ot, turdosh ot, atoqli ot, nomlash, nomlangan obyekt, NE, NER

Abstract: This article discusses proper nouns in Uzbek texts and methods for their determination. It studies the expression of common nouns in NLP and their linguistic features. Detailed definitions of the concepts of NE and NER, their properties, methods of definition and their use in texts are given.

Keywords: Noun, common nouns, proper nouns, naming, named object, NE, NER.

Аннотация: В данной статье рассматриваются собственные имена существительные в узбекских текстах и методы их определения. В нем изучается выражение нарицательных имен существительных в НЛП и их лингвистические особенности. Даны подробные определения понятия НЭ и НЭР, их свойства, способы определения и употребление их в текстах.

Ключевые слова: Существительное, нарицательные существительные, собственные имена существительные, именование, названный предмет, НЭ, НЭР.

Narsa-hodisalarga nom qo‘yish va uni nomlashning asosiy tamoyillari, atalish motivlari masalasi, olimlarni, shu jumladan, tilshunoslarni ham juda qadimdayoq qiziqtirgan. Geraklit, Demokrit, Platon va Aristotellarning shu to‘g‘ridagi qarashlari buning isbotidir. Narsa va hodisalarga nom qo‘yish masalasi umumiy va nazariy tilshunoslikning yirik vakillari bo‘lmish A.A.Potebnya, I.A.Boduen de Kurtene, F.de Sossyur kabi olimlarni ham qiziqtirgan. Ular

o'zlarining nazariy qarashlari orqali nomlanish muommosini hal etishga katta hissa qo'shganlar. [Суперанская, 1973:45-90]

Qadimdan insonlar borliqdagi barcha narsani nomlab o'rganib kelishgan. Tabiatdagi jonzoatlarni, voqealarni, hodisalarni, jonli va jonsiz narsalarni, manzillarni, hattoki o'zlarini ham "nom"lar bilan atashgan. Bu orqali ularni bir-biridan farqlash, toifaga ajratish va yodda tez saqlab qolishga intilishgan.

Atrofimizdagi barcha obyektlar, hattoki subyektlar ham nomlanadi. Jumladan, inson ismlari, joy nomlari, suv ishootlari, hayvonlar, o'simliklar nomlari bunga misoldir. Bu nomlar juda ko'pki, ularning barchasini bir qarashda anglab olish qiyin. To'g'ri, tilshunoslik nuqtayi nazar bilan qaraganda nomlangan obyektlar matn tarkibida bosh harfda yoziladi, ammo barcha obyektlar ham emas. Bunda matnlarda nomlangan obyektlarni mashinaviy o'qitish yordamida aniqlash ishimizni osonlashtiradi.

"O'zbek tilining izohli lugati"da turdosh otni bir turdagi predmetlarning nomini anglatadigan otlar [Ўзбек тилининг изоҳли луғати. II том, 1981:62], "Atoqli ot bir shaxs yoki predmetga xos nom" [Ўзбек тилининг изоҳли луғати. II том, 1981:225], deb izohlangan. Ushbu tushunchalar onomastika nuqtayi nazaridan ham o'ziga xos izohlariga ega. Masalan, N.V.Podolskaya "Собственное имя" (atoqli ot)ni "*onim*" terminiga havola qiladi va "*onim*" termini shunday talqin qilingan: "Onim (onoma - atoqli ot) – o'zi nomlayotgan obyekttni boshqa obyektlardan farqlashga, individuallashtirish va identifikatsiyalashga xizmat qiluvchi so'z, so'z birikmasi yoki gapdir"[Подольская, 1988:91]. Ushbu muallif turdosh otni apellyativ terminiga havola qilgan holda quyidagicha tushuntiradi: "Turdosh ot atoqli otning teskarisidir". Ushbu izohda turdosh ot tushunchasining mohiyati ochilgan emas. Keltirilgan izoh: "turdosh ot – apellyativdir, apellyativ – turdosh otdir" deyilgan aksiomaga aylanib qolgan. Buning ustiga apellyativ tushunchasini faqat turdosh ot, turdosh otlardan iborat, turdosh ot tushunchasiga teng deb bilish ham noto'g'ridir. Atoqli otlarning lug'aviy asoslari faqatgina turdosh otlardan iborat bo'lmaydi. Chunki atoqli otlarga boshqa so'z turkumlari ham lug'aviy asos bo'lishi mumkin.

Atoqli (antroponimik) va turdosh otlar orasidagi ma'noviy munosabatlar. O'zbek tilshunosligida atoqli otlarning ma'nosiga oid muammolar, ayniqsa, atoqli ot bilan unga asos bo'lgan turdosh so'z (apellyativ)ning ma'nosini farqlash, chegaralash masalasi yetarli tadqiq qilingan emas. Shunga qaramay, turdosh va atoqli ot ma'nolari haqida mulohazalar aytish mumkin.[Берматов, 2013:90]

1. Atoqli ot turdosh otlardan hosil bo'ladi, ammo ular ma'nosi bilan assosiativ aloqani yo'qotib, narsa va hodisalarni bir-biridan farqlash uchun xizmat qiladi.

2. Turdosh ot bir tildan ikkinchi tilga tarjima qilinadi, onimlar tarjima qilinmaydi.

3. Turdosh otlar ongda doimiy turadi, onimlar mexanik ravishda esda saqlanadi, tezda esdan chiqadi.

4. Turdosh ot grammatik jihatdan birlik va ko'plikda qo'llanadi. Onimlarda bu xususiyat yo'q.

5. Turdosh otlar umumiylikni anglatadi, atoqli otlar xususiy bo'ladi: o'quvchi – sinfdagi har bir yosh; Tolibjon shulardan biri. [Qilichev, 2023:145]

Atoqli otlar – bu ot so'z turkumining funksional-semantik sinfi bo'lib, ularning asosiy vazifasi obyektlarni tanlash va aniqlash, ushbu mavzu bo'yicha yagona va to'liq tushunchalarni ifodalashdir.[Фонякова, 1990:3]

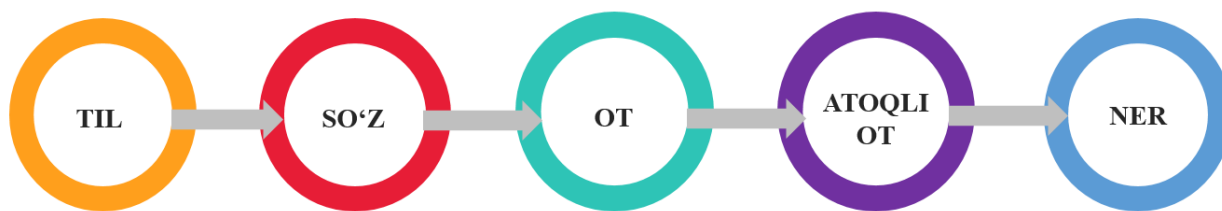
Atoqli otlar bu nomga atab qayta “yangi nom” qo’yishdir. Atoqli otlarda *aniqlik, xoslik, yagonalik, mazmunlilik* xususiyatlari mavjud bo’ladi.

Matnlar tarkibidagi “nom” (atoqli ot, nomlangan obyekt)lar esa matnning asosiy mazmunini belgilovchi unsurlardan biridir. Shuningdek, NLP (Natural Language Processing – Tabiiy tilni qayta ishlash)ning muhim masalaridan biri bo’lgan *nomlangan obyekt*ni tanib olish (NER – *Named Entity Recognition*) hodisasidir. Matnlarni avtomatik qayta ishlab, matn tarkibidagi nomlangan obyektlarni aniqlash jarayoni – NER obyektlarini aniqlash deb nomlanadi. NER obyektlari matndagi lug’atdan tashqari leksika, ya’ni atoqli otlarni ifodalaydi. NER obyektlari, kishi ismlari, joy nomlari, tashkilot nomlari, yo’nalishlar...kabi bir qancha atoqli otlarni ifodalaydi, nomni aniqlashtiradi, aniqlangan nomlarni to’g’ri toifalarga ajratadi. Ilk bora NER va uning vazifalariga doir ilmiy ishlar 1996-yil Message Understanding (MUC-6)” konfrensiyasida e’lon qilingan. Shundan so’ng jahondagi ko’pgina kompyuter lingistikasidagi ilmiy ishlar matnlardagi nomlangan obyektlarni aniqlash(NER)ga qaratildi.

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.[Rachna Jain..., 2020:5] 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.

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.[Susan Li, 2018:4]

NER obyektlari – atoqli otlarni ifodalaydi. Atoqli otlar o’z nomi bilan nomlangan obyektlarga ishora qiladi. Matnlar tarkibidagi nomlangan obyektlar modellar yordamida avtomatik aniqlanganda keng doirada matn tarkibidagi barcha ot so’z turkumidagi so’zlar ko’zdan kechirilib, so’ngra tor doirada faqat atoqli otlar aniqlanadi. Ot so’z turkumining eng birinchi til xususiyati bu – uning uch obyekt(kishi (kim?), narsa (nima?), joy (qayer?))ni ifodalashi. Bu xususiyat otlarning boshqa so’z turkumlari(sifat, fe’l, olmosh...)dan farqlanishini ko’rsatadi. NER obyektlari esa atoqli otlarga xos eng birinchi xususiyat “bosh harflilik”ga ko’ra aniqlanadi. NER atamasi ingliz tilidan abbreviatura (qisqartma) shaklida hosil bo’lgan termin hisoblanadi. Ingliz tilida amalga oshirilgan ilmiy ishlarni kuzatganimizda, NER obyektlari ro’yxati ko’p ekanligini kuzatamiz. Ammo biz o’zbek tili doirasida ushbu ro’yxatni qo’llay olmaymiz. Chunki, ingliz tilida NER obyektlari tarkibiga otlardan tashqari, sonlar, sifatlar, miqdor o’lchovlari, vaqt o’lchovlari...kabi birliklar ham kiritilgan. Bu esa o’zbek tili grammatikasi uchun xos emas. Bunga sabab, ikki til (o’zbek va ingliz) ikki xil til oilasiga mansubligidadir. Xususan, til oilalari turlicha bo’lgach, ularning grammatikasi (morfologiyasi, leksikologiyasi) tubdan farq qiladi.



1-rasm. Tilning grammatik xususiyatiga ko'ra NER obyektlarini aniqlash

NER obyektlari matn tarkibiga ko'proq ma'no yuklashga yordam beradi. Bu orqali esa matndan ma'lumot olish ishi amalga oshiriladi va u bog'lanib kelayotgan so'z NER orqali yanada aniqlashtiriladi. NER obyektlarini aniqlash, asosan, izohsiz matn bloklarini tahlil qilishdan boshlanadi. Ya'ni, bu yerda tayyor matn olinadi va uning tarkibidagi atoqli otlar(proper noun)lar ajratiladi.

Lekin atoqli ot tushunchasini faqat o'zbek tili matnlarida NER obyektlarini topish uchun qo'llash mumkin. Chet elda mavjud bo'lgan NER obyektlarini aniqlashga yo'naltirilgan dasturlar va ushbu yo'nalishda qilingan ilmiy ishlar o'rganilganda, mashinaviy o'qitish tizimida faqat atoqli otlar NER obyekti sifatida olinmagan. Bu ro'yxatda atoqli otlardan tashqari yillar, miqdorlar, foizlar, pul qiymatlari, fasllar, ranglar, qismlar, dizaynlar, voqealar, tibbiy kodlar ham kiritilgan. Bundan ko'rinadiki, o'zbek tilidagi matnlar uchun atoqli otlar – NER obyekti, lekin NLP uchun NER obyekti bu faqat atoqli ot emas. Atoqli otlar – NER, NER bu – atoqli otgina emas! NER tushunchasi qamrovi kengdir. Unda atoqli otlardan tashqari yana ko'pgina obyektlarni qamrab oladi. Ammo biz til oilalari farqliligi, ularning grammatikasi, leksikasi, morfologiyasi turlichaligini hisobga olib, o'zbek tili uchun faqat atoqli otlarni NER obyekti sifatida olish maqbul yo'l ekanligiga amin bo'ldik.

NER obyektlari gap mazmuniga ko'proq ma'no yuklashga yordam beradi. Matn tarkibidagi otlar ma'no va mazmun jihatdan bir-birini to'ldiradi va bir-biriga bog'lanib keladi. NER obyektlari esa gap mazmunini yanada boyitishga va asosiy ma'noni aniqlashtirishga yordam beradi.

Foydalanilgan adabiyotlar:

1. Суперанская А.В. общая теория имени собственного. М., “Наука”,1973, с.45-90; Жаббаров Э. Ўзбек халқ ўйинлари лексикаси – Филол. фанлари номзоди, дисс. автореф., Тошкент, 1998, 8-бет
2. Ўзбек тилининг изоҳли луғати. II том. М.: Рус тили, 1981.225-бет. Уэбек тилининг изоҳли луғати. 1 том. М.: Рус тили, 1981.62-бет.
3. Уэбек тилининг изоҳли луғати. 1 том. М.: Рус тили, 1981.62-бет.
4. Подольская Н. В. Словарь русской ономастической терминологии, М.: Наука, 1988.
5. Э.А.Бегматов. Ўзбек тили антропонимикаси. Ўзбекистон Республикаси Фанлар академияси “Фан” нашриёти Тошкент. 2013. 90б
6. Вағрамали Қиличев. Ономастика. О'қув қo'лланма. Вухоро. 2023. 115
7. Фoнякова О.И. Имя собственное в художественном тексте: Учеб. пособие / О.И. Фoнякова. Л.: ЛГУ, 1990. – С. 3.
8. Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty. Named Entity Recognition in English Text. Journal of Physics: Conference Series. 2020
9. Susan Li. Named Entity Recognition with NLTK and SpaCy. 2018.1p