

O‘ZBEK TILIDAGI TERMINLARNI SUN’IY INTELLEKT ASOSIDA AVTOMATIK ANIQLASH VA TASNIFLASH

Shahzoda Turahodjaeva Ibrohimadhamxo‘djayevna

ORCID: 0009-0002-9170-3918

Shahzoda1790@mail.ru

Annotatsiya: Mazkur maqolada o‘zbek tilidagi terminlarni sun’iy intellekt (SI) texnologiyalari asosida avtomatik aniqlash va tasniflashning lingvistik hamda texnologik imkoniyatlari tahlil qilinadi. Tadqiqotda tabiiy tilni qayta ishlash (NLP) texnologiyalari, xususan, zamonaviy til modellari va mashinaviy o‘qitish usullaridan foydalangan holda terminologik birliklarni avtomatik identifikatsiya qilish, ularni semantik va tematik guruhlarga ajratish masalalariga e’tibor qaratiladi. O‘zbek tilining agglutinativ tabiati, morfologik qurilishi, so‘z yasaliş xususiyatlari hamda terminlarning ko‘p komponentli strukturasi avtomatik terminologik tahlil jarayonida muhim lingvistik omil sifatida baholanadi. Shuningdek, sun’iy intellekt asosida terminlarni aniqlashda korpus ma’lumotlarining hajmi va sifati, terminologik resurslarning yetarli darajada shakllanmaganligi, ko‘pma’nolilik, kontekstual farqlanish hamda morfologik variantlarning mavjudligi bilan bog‘liq muammolar yoritiladi. Tadqiqot natijalari o‘zbek tilining raqamli resurslarini takomillashtirish, terminologik bazalarni shakllantirish, avtomatik matn tahlilini rivojlantirish hamda o‘zbek tilining sun’iy intellekt tizimlaridagi lingvistik reprezentatsiyasini kengaytirishda muhim ilmiy-amaliy ahamiyat kasb etishi asoslanadi.

Kalit so‘zlar: sun’iy intellekt, o‘zbek tili, termin, terminologiya, avtomatik termin aniqlash, terminlarni tasniflash, tabiiy tilni qayta ishlash, mashinaviy o‘qitish, lingvistik model, terminologik korpus, semantik tahlil, raqamli lingvistika

AUTOMATIC DETECTION AND CLASSIFICATION OF UZBEK TERMS BASED ON ARTIFICIAL INTELLIGENCE

Shahzoda Turahodjaeva Ibrohimadhamxo‘djayevna

ORCID: 0009-0002-9170-3918

Shahzoda1790@mail.ru

Abstract: This article analyzes the linguistic and technological potential of automatic identification and classification of terms in the Uzbek language based on artificial intelligence (AI) technologies. The study focuses on the automatic identification of terminological units and their classification into semantic and thematic groups using natural language processing (NLP) technologies, particularly modern

language models and machine learning methods. The agglutinative nature of the Uzbek language, its morphological structure, word-formation features, and the multi-component structure of terms are considered important linguistic factors in the process of automatic terminological analysis. The article also examines the challenges associated with the use of artificial intelligence for terminology identification, including the volume and quality of corpus data, the insufficient development of terminological resources, polysemy, contextual variation, and the existence of morphological variants. The findings substantiate the scientific and practical significance of the research for improving digital resources of the Uzbek language, developing terminological databases, advancing automatic text analysis, and expanding the linguistic representation of the Uzbek language in artificial intelligence systems.

Keywords: artificial intelligence, Uzbek language, term, terminology, automatic term identification, term classification, natural language processing, machine learning, linguistic model, terminological corpus, semantic analysis, digital linguistics

KIRISH. So‘nggi yillarda sun‘iy intellekt (SI) texnologiyalarining jadal rivojlanishi ta‘lim, iqtisodiyot, axborot texnologiyalari bilan bir qatorda tilshunoslik sohasida ham yangi tadqiqot yo‘nalishlarining shakllanishiga olib kelmoqda. Xususan, tabiiy tilni qayta ishlash (NLP) texnologiyalari asosida matnlarni avtomatik tahlil qilish, lingvistik birliklarni aniqlash, semantik tasniflash va terminologik ma‘lumotlarni qayta ishlash imkoniyatlari kengayib bormoqda. Bu jarayon o‘zbek tilining raqamli resurslarini rivojlantirish, terminologik birliklarni avtomatik qayta ishlash hamda milliy tilning sun‘iy intellekt tizimlaridagi lingvistik reprezentatsiyasini takomillashtirish masalalarini dolzarb ilmiy muammo sifatida kun tartibiga olib chiqmoqda.

O‘zbek tilining agglutinativ tabiati, morfologik shakllarining rang-barangligi, so‘z yasashining mahsuldorligi hamda terminlarning bir va ko‘p komponentli tuzilishga ega ekanligi ularni avtomatik aniqlash va tasniflash jarayonida o‘ziga xos lingvistik yondashuvni talab etadi. Ayniqsa, muayyan terminning umumiste‘mol leksikasidan farqlanishi, uning kontekstual ma‘nosi, sohaviy xususiyati va semantik munosabatlarini avtomatik aniqlash sun‘iy intellekt asosidagi terminologik tahlilning muhim masalalaridan biridir.

Hozirgi kunda sun‘iy intellekt asosidagi til modellari va mashinaviy o‘qitish algoritmlaridan foydalanish orqali terminologik birliklarni matn tarkibidan avtomatik ajratib olish, ularni semantik, mavzuviy va funksional jihatdan tasniflash imkoniyati mavjud. Bunday yondashuv terminologik korpuslarni shakllantirish, sohaviy lug‘atlarni takomillashtirish, avtomatik matn tahlili tizimlarini rivojlantirish hamda

o'zbek tilidagi ilmiy-texnik axborotni raqamlashtirish uchun muhim nazariy va amaliy asos yaratadi.[1:14]

Shu nuqtayi nazardan, o'zbek tilidagi terminlarni sun'iy intellekt asosida avtomatik aniqlash va tasniflash masalasini tadqiq etish zamonaviy raqamli lingvistikaning dolzarb yo'nalishlaridan biri hisoblanadi. Mazkur maqolada o'zbek tilidagi terminologik birliklarni avtomatik aniqlashning lingvistik mezonlari, sun'iy intellekt va tabiiy tilni qayta ishlash texnologiyalarining ushbu jarayondagi imkoniyatlari, terminlarni semantik va tematik guruhlariga ajratish usullari hamda avtomatik tasniflash jarayonida yuzaga keladigan lingvistik va texnologik muammolar tahlil qilinadi.

Mavzuga doir adabiyotlar tahlili. O'zbek tilida tabiiy tilni qayta ishlash (NLP) yo'nalishida bir qator ilmiy tadqiqotlar va amaliy loyihalar amalga oshirilmoqda. Xususan, o'zbek tilidagi matnlarda bazis so'zlarni aniqlash, so'z turkumlarini avtomatik teglash, morfologik tahlil va lemmatizatsiya kabi masalalarga bag'ishlangan tadqiqotlar tilning raqamli qayta ishlanishi uchun muhim lingvistik va texnologik asos yaratmoqda [2; 3; 5]. Mazkur yo'nalishdagi tadqiqotlar o'zbek tilining morfologik murakkabligi, agglutinativ xususiyatlari va so'z shakllarining variantlilikligi kabi omillarni hisobga olgan holda avtomatik lingvistik tahlil modellarini takomillashtirishga qaratilgan.

Sun'iy intellektga asoslangan zamonaviy til modellari inson tilini avtomatik anglash, qayta ishlash va generatsiya qilish imkoniyatlarini kengaytirib, NLPning muhim tarkibiy yo'nalishlaridan biriga aylanmoqda [4]. Ushbu texnologiyalardan terminologik birliklarni avtomatik aniqlash va tasniflashda foydalanish o'zbek tilining sohaviy leksikasini tizimli ravishda o'rganish imkonini beradi. Ayniqsa, terminlarning morfologik shakllari, so'z yasaliş modellari, kontekstual qo'llanishi va semantik xususiyatlarini avtomatik aniqlash terminologik ma'lumotlarni qayta ishlash samaradorligini oshirishi mumkin.

O'zbek tilining morfologik jihatdan murakkabligi sharoitida sun'iy intellekt asosida ishlab chiqilgan lemmatizatsiya va morfologik tahlil vositalarining takomillashuvi terminlarni avtomatik aniqlash uchun ham muhim ahamiyatga ega [5]. Chunki terminlar turli grammatik shakllarda qo'llanishi, derivatsion affiksalar yordamida yangi terminologik birliklar hosil qilishi yoki ko'p komponentli birikmalar tarkibida namoyon bo'lishi mumkin. Shu sababli terminlarni avtomatik aniqlash jarayonida leksik, morfologik, sintaktik va semantik belgilarni kompleks hisobga olish zarur.

Shuningdek, mavjud NLP tadqiqotlarida shakllantirilgan morfologik tahlil, lemmatizatsiya va so'z turkumlarini avtomatik aniqlash kabi texnologik yechimlar terminologik birliklarni avtomatik identifikatsiya qilishning lingvistik asosini tashkil etadi. Bunda sun'iy intellekt modellaridan foydalanish orqali terminlarning matndagi

kontekstual xususiyatlarini aniqlash, ularni umumiste'mol leksikasidan farqlash hamda muayyan fan va soha bo'yicha semantik-tematik guruhlarga ajratish imkoniyatlari kengayadi. Natijada o'zbek tilidagi terminologik resurslarni avtomatlashtirish va raqamli terminologik bazalarni shakllantirish uchun yangi imkoniyatlar yuzaga keladi.

Tadqiqot metodologiyasi. Bundan tashqari, mahalliy va xalqaro hamkorlik doirasida o'zbek tilining raqamli resurslarini, jumladan, matn korpuslari, lingvistik ma'lumotlar bazalari, avtomatik tarjima tizimlari hamda tabiiy tilni qayta ishlashga yo'naltirilgan modellarni rivojlantirish bo'yicha ilmiy-amaliy loyihalar amalga oshirilmoqda. Mazkur jarayonlar o'zbek tilidagi terminologik birliklarni avtomatik aniqlash, ularning qo'llanish xususiyatlarini belgilash va muayyan sohalar kesimida tasniflash uchun zarur bo'lgan lingvistik va texnologik infratuzilmani shakllantirishga xizmat qiladi.

Sun'iy intellekt texnologiyalarining o'zbek tiliga tatbiqi terminologik birliklarni avtomatik qayta ishlash uchun ham keng imkoniyatlar yaratmoqda. Xususan, mashinaviy o'qitish, chuqur o'rganish va tabiiy tilni qayta ishlash texnologiyalari yordamida matn tarkibidagi terminlarni avtomatik aniqlash, ularning morfologik va sintaktik xususiyatlarini belgilash, kontekstual ma'nosini aniqlash hamda semantik-tematik guruhlarga ajratish mumkin. Bunda terminlarni aniqlash jarayonida ularning shakliy tuzilishi bilan bir qatorda, muayyan sohada qo'llanish chastotasi, kontekstual barqarorligi va boshqa leksik birliklar bilan semantik munosabatlari ham hisobga olinadi.

O'zbek tiliga moslashtirilgan zamonaviy til modellaridan, xususan, UzRoBERTa kabi modellar asosida terminologik matnlarni qayta ishlashda foydalanish mumkin. Mazkur modellar kontekstual axborotni hisobga olish imkoniyati sababli bir xil shaklga ega bo'lgan, biroq turli kontekstlarda har xil ma'noni ifodalovchi birliklarni farqlashda muhim ahamiyat kasb etadi. Terminlarni avtomatik aniqlash jarayonida model matndagi so'z va so'z birikmalarining kontekstual xususiyatlarini tahlil qilish orqali ularning terminologik birlik sifatidagi ehtimolini belgilashi mumkin. Keyingi bosqichda aniqlangan terminlarni tibbiyot, farmatsevtika, huquq, iqtisodiyot, texnika, axborot texnologiyalari kabi sohaviy guruhlarga ajratish imkoniyati yuzaga keladi.

BERTbek modeli ham o'zbek tilining grammatik va morfologik xususiyatlarini hisobga oluvchi til modeli sifatida terminologik tahlil uchun muhim texnologik asoslardan biri bo'lishi mumkin. Xususan, mazkur modelning kontekstual tahlil imkoniyatlaridan foydalanish orqali terminlarning matndagi qo'llanish holatlarini aniqlash, ko'p ma'noli birliklarning terminologik ma'nosini farqlash va terminologik birikmalarni avtomatik identifikatsiya qilish mumkin. Bunda modelning tabiiy tilni tushunish imkoniyatlari terminlarning shakliy, semantik va kontekstual belgilarini kompleks tahlil qilishga yo'naltiriladi.

Terminlarni avtomatik aniqlash jarayonida Named Entity Recognition (NER), matn tasnifi, semantik tahlil va kontekstual modellashtirish kabi NLP vazifalarining imkoniyatlaridan ham foydalanish mumkin. Biroq terminologik birliklar shaxs, tashkilot yoki geografik nomlardan farqli ravishda muayyan fan yoki kasbiy faoliyat sohasiga xos tushunchalarni ifodalashi sababli ularni aniqlashda maxsus terminologik korpus va belgilangan ma'lumotlar to'plamining mavjudligi muhim hisoblanadi.

Shu jihatdan, sun'iy intellekt asosida o'zbek tilidagi terminlarni avtomatik aniqlash va tasniflash tizimini yaratish bir necha bosqichni qamrab oladi: terminologik korpusni shakllantirish, matnlarni belgilash (annotatsiyalash), terminlarni avtomatik ajratib olish, ularning morfologik va semantik xususiyatlarini aniqlash hamda sohaviy guruhlariga tasniflash. Mazkur bosqichlarning izchil amalga oshirilishi o'zbek tilining raqamli terminologik resurslarini boyitish, sohaviy lug'atlarni avtomatlashtirish va zamonaviy NLP tizimlarining o'zbek tilidagi imkoniyatlarini kengaytirishga xizmat qiladi.

Olingan natijalar va ularning tahlili. Sun'iy intellekt asosidagi til modellari o'zbek tilining morfologik va sintaktik xususiyatlarini hisobga olgan holda terminologik birliklarni avtomatik aniqlash va tasniflash imkoniyatlarini kengaytirmoqda. Xususan, NLP texnologiyalari yordamida matn tarkibidagi terminlarni ajratib olish, ularning morfologik shakllarini aniqlash, kontekstual ma'nosini belgilash hamda muayyan soha bo'yicha guruhlash mumkin. Bunda terminlarning shakliy tuzilishi bilan bir qatorda, ularning qo'llanish chastotasi, kontekstual barqarorligi va semantik xususiyatlarini ham hisobga olish muhimdir.

O'zbek tilining agglutinatib tabiati terminlarni avtomatik aniqlash jarayonini murakkablashtiruvchi asosiy omillardan biridir. So'zlarning turli affikslar yordamida shakllanishi terminlarning ko'plab morfologik variantlarda namoyon bo'lishiga sabab bo'ladi. Shu bois avtomatik terminologik tahlilda lemmatizatsiya, morfologik tahlil va kontekstual modellashtirish vositalaridan kompleks foydalanish talab etiladi. Sun'iy intellekt modellari ushbu jarayonlarda terminlarni umumiste'mol leksikadan farqlash, ko'p komponentli terminologik birikmalarni aniqlash va ularni semantik-tematik guruhlariga ajratishda qo'llanishi mumkin.

Terminlarni avtomatik aniqlash va tasniflashning samaradorligi, avvalo, sifatli terminologik korpus va annotatsiyalangan ma'lumotlar bazasining mavjudligiga bog'liq. O'zbek tilida bunday raqamli resurslarning yetarli darajada shakllanmaganligi, terminlarning ko'pma'noliligi, morfologik variantliligi hamda sohaviy terminologiyaning muntazam yangilanib borishi mazkur yo'nalishdagi asosiy muammolardan hisoblanadi. Shuningdek, sun'iy intellekt modellarini o'qitish uchun katta hajmdagi ishonchli va muvozanatli ma'lumotlarning zarurligi ham muhim omildir.

Shu nuqtayi nazardan, sun'iy intellekt asosida o'zbek tilidagi terminlarni avtomatik aniqlash va tasniflash tizimlarini takomillashtirish terminologik korpuslarni yaratish, sohaviy lug'atlarni raqamlashtirish, ilmiy-texnik matnlarni avtomatik qayta ishlash hamda o'zbek tilining raqamli lingvistik resurslarini boyitishga xizmat qiladi.

Muhokama. O'zbek tilidagi terminlarni sun'iy intellekt asosida avtomatik aniqlash va tasniflash masalasi zamonaviy raqamli lingvistikaning istiqbolli yo'nalishlaridan biridir. Tadqiqot natijalari shuni ko'rsatadiki, tabiiy tilni qayta ishlash texnologiyalaridan foydalanish terminologik birliklarni katta hajmdagi matnlar tarkibidan tezkor aniqlash va ularni muayyan mezonlar asosida guruhlash imkonini beradi. Ayniqsa, mashinaviy o'qitish va chuqur o'rganish modellarining kontekstual tahlil imkoniyatlari terminlarning umumiste'mol birliklaridan farqlanishida muhim ahamiyat kasb etadi.

O'zbek tilining agglutinatив xususiyati terminlarni avtomatik aniqlashda alohida lingvistik yondashuvni talab qiladi. Bir o'zakka bir necha grammatik va derivatsion affikslarning qo'shilishi terminlarning turli morfologik shakllarda namoyon bo'lishiga sabab bo'ladi. Shu bois terminlarni avtomatik aniqlashda faqat yuzaki shaklga asoslanish yetarli emas. Morfologik tahlil, lemmatizatsiya, so'z yasalishi va kontekstual ma'lumotlarni birgalikda hisobga olish aniqlik darajasini oshirishi mumkin.

Terminlarni tasniflash jarayonida ularning semantik va tematik xususiyatlari ham muhim hisoblanadi. Masalan, tibbiyot, farmatsevtika, huquq, iqtisodiyot, texnika va axborot texnologiyalari kabi sohalarga oid terminlarni avtomatik guruhlash orqali sohaviy terminologik bazalarni shakllantirish imkoniyati yuzaga keladi. Bunda terminning faqat shakliy belgisi emas, balki uning matndagi konteksti, boshqa birliklar bilan semantik aloqasi va muayyan sohada qo'llanish xususiyati ham hisobga olinishi zarur.

Tadqiqotning muhim jihatlaridan biri shundaki, sun'iy intellekt modelining samaradorligi bevosita o'qitish ma'lumotlarining sifati va hajmiga bog'liq. O'zbek tilida terminologik birliklar bilan maxsus belgilangan yirik korpuslarning yetarli emasligi avtomatik aniqlash jarayonining aniqligiga ta'sir qilishi mumkin. Bundan tashqari, terminlarning ko'pma'noliligi, sinonimik variantlari, qisqartmalar va ko'p komponentli terminologik birikmalarni aniqlash ham muayyan murakkabliklarni yuzaga keltiradi.

Shu sababli o'zbek tilidagi terminlarni sun'iy intellekt asosida avtomatik aniqlash va tasniflash uchun lingvistik bilimlar va texnologik yondashuvlarning integratsiyasi zarur. Terminologik korpuslarni yaratish, ularni morfologik va semantik jihatdan annotatsiyalash, sohaviy belgilar asosida tasniflash hamda zamonaviy til modellaridan foydalanish mazkur yo'nalishning ilmiy va amaliy samaradorligini oshiradi. Bunday

tizimlar kelgusida elektron lugʻatlar, avtomatik tarjima, ilmiy matnlarni qayta ishlash va raqamli terminologik bazalarni shakllantirishda qoʻllanishi mumkin.

XULOSA. Oʻzbek tilidagi terminlarni sunʼiy intellekt asosida avtomatik aniqlash va tasniflash zamonaviy raqamli lingvistikaning muhim yoʻnalishlaridan biridir. Tadqiqot natijalari sunʼiy intellekt va tabiiy tilni qayta ishlash texnologiyalari yordamida terminologik birliklarni matn tarkibidan avtomatik ajratish, ularning morfologik va semantik xususiyatlarini aniqlash hamda sohaviy guruhlariga tasniflash imkonini berishini koʻrsatdi.

Oʻzbek tilining agglutinatib tabiati, morfologik variantlilik, terminlarning koʻpmaʼnoliligi va koʻp komponentli shakllari avtomatik tahlil jarayonida muayyan murakkabliklarni yuzaga keltiradi. Shu sababli terminlarni aniqlashda morfologik, leksik-semantik va kontekstual belgilarni kompleks hisobga olish muhimdir.

Kelgusida sifatli terminologik korpuslar va annotatsiyalangan maʼlumotlar bazasini shakllantirish, sunʼiy intellekt modellarini oʻzbek tilining lingvistik xususiyatlariga moslashtirish terminlarni avtomatik aniqlash va tasniflash samaradorligini oshiradi. Mazkur yondashuv oʻzbek tilining raqamli resurslarini boyitish, elektron lugʻatlarni takomillashtirish, sohaviy terminologik bazalarni yaratish va ilmiy-texnik matnlarni avtomatik qayta ishlash uchun muhim ilmiy-amaliy imkoniyatlar yaratadi.

Foydalanilgan adabiyotlar

1. Abdurahmonova N.Z. Kompyuter lingvistikasi: darslik. - Toshkent: Nodirabegim, 2021. - 394 b.
2. Rahimov A. Kompyuter lingvistikasi asoslari: oʻquv qoʻllanma. - Toshkent: Akademnashr, 2011. - 160 b.
3. Xolmanova Z. Kompyuter lingvistikasi: oʻquv qoʻllanma. - Toshkent: Shafolat Nur Fayz, 2021. - 160 b.
4. Elov B.B., Abdullayeva O., Primova M. Oʻzbek tilida tabiiy tilni qayta ishlashdagi morfologik, sintaktik va semantik tahlil metodlari // Uzbekistan: Language and Culture. - 2025. - №1. - DOI: 10.52773/tsuull.uzlc.2025.1/WRHQ5488.
5. Fayziyeva A.S. Oʻzbek tilida sunʼiy intellekt terminlarining shakllanish manbalari va usullari // Qoʻqon DPI. Ilmiy xabarlar. - 2026. - T. 8, № 06. - B. 84-90. - DOI: 10.70728/b.series.fil.v08.i06.010.
6. Quldashev N.A., Vahobova O.A. Sunʼiy intellekt terminologiyasining oʻzbek tiliga oʻzlashishi // World of Philology. - 2026. - T. 5, № 3. - B. 232-236. - DOI: 10.70728/phil.v05.i03.036.