

O'ZBEK TILIDA TABIIY TILNI QAYTA ISHLASH (NLP)
TEXNOLOGIYALARI

Turabova Ibodat Ikromjon qizi

SamDCHTI Sharq tillari fakulteti

ikromjonovna08@gmail.com

Ilmiy rahbari: Rustamov R.T

Samarqand davlat chet tillar institute

ИНФОРМАЦИЯ О
СТАТЬЕ

АННОТАЦИЯ:

ИСТОРИЯ СТАТЬИ:

Received: 08.10.2026

Revised: 09.10.2026

Accepted: 10.10.2026

**КЛЮЧЕВЫЕ
СЛОВА:**

*kompyuter lingvistikasi,
tabiiy tilni qayta ishlash,
NLP, o'zbek tili, sun'iy
intellekt, morfologik tahlil,
sintaktik tahlil, semantik
tahlil, BERT, lingvistik
korpus.*

Mazkur maqolada tabiiy tilni qayta ishlash Natural Language Processing - NLP texnologiyalarining o'zbek tiliga tatbiq etilishi, uning asosiy yo'nalishlari va mavjud imkoniyatlari tahlil qilinadi. O'zbek tilining agglutinativ tuzilishi, morfologik jihatdan murakkabligi, so'z shakllarining ko'pligi, nisbatan erkin so'z tartibi hamda lingvistik resurslarning cheklanganligi NLP tizimlarini yaratishda o'ziga xos muammolarni yuzaga keltiradi. Maqolada tokenizatsiya, morfologik tahlil, sintaktik va semantik tahlil hamda transformer modellarining o'zbek tilidagi imkoniyatlari yoritiladi. Shuningdek, o'zbek tili uchun yaratilayotgan datasetlar, morfologik analizatorlar va BERT asosidagi modellar tahlil qilinadi. Tadqiqot natijasida sifatli lingvistik korpuslar va tilga moslashtirilgan neyron modellar o'zbek tilida NLP texnologiyalarini rivojlantirishning muhim omillari ekani asoslanadi.

KIRISH

Bugungi kunda axborot texnologiyalarining jadal rivojlanishi inson va kompyuter o'rtasidagi muloqotni tabiiy til asosida tashkil etish masalasini yanada dolzarb qilib qo'ymoqda. Insonlar kundalik hayotida axborotni asosan tabiiy tillar orqali qabul qiladi, yaratadi va uzatadi. Kompyuter esa ushbu ma'lumotlarni qayta ishlashi uchun tabiiy tildagi matn va nutqni raqamli shaklda tahlil qilishi zarur.

Tabiiy tilni qayta ishlash - Natural Language Processing yoki qisqacha NLP - kompyuter lingvistikasi va sun'iy intellektning muhim yo'nalishi bo'lib, inson tilidagi matn va nutqni

kompyuter yordamida qayta ishlashga qaratilgan. NLP texnologiyalari avtomatik tarjima, ovoqli yordamchilar, qidiruv tizimlari, matnni avtomatik tasniflash, imlo tekshirish, chatbotlar, sentiment tahlili va savol-javob tizimlarining asosini tashkil etadi.

O'zbek tili ham raqamli texnologiyalar rivojlanishi jarayonida NLP tadqiqotlarining muhim obyektiga aylanmoqda. Biroq o'zbek tili ko'plab xalqaro NLP tadqiqotlarida resurslari nisbatan cheklangan tillardan biri sifatida qaraladi. O'zbek tili uchun katta hajmdagi sifatli belgilangan korpuslar, morfologik va sintaktik resurslar hamda yuqori sifatli til modellari sonini yanada oshirish zarur.

Mazkur maqolaning maqsadi o'zbek tilida tabiiy tilni qayta ishlash texnologiyalarining asosiy yo'nalishlarini ko'rib chiqish, mavjud ilmiy ishlanmalarni tahlil qilish va ushbu yo'nalishning istiqbollarini ko'rsatishdan iborat.

1. TABIIY TILNI QAYTA ISHLASH TUSHUNCHASI

Tabiiy tilni qayta ishlash inson tomonidan yaratilgan tabiiy tildagi ma'lumotlarni kompyuter yordamida qayta ishlash, tahlil qilish va ulardan ma'no chiqarishga qaratilgan texnologiyalar majmuasidir. NLP tilshunoslik hamda kompyuter texnologiyalari va sun'iy intellektni birlashtiradi.

NLP tizimining umumiy ishlash jarayonini quyidagicha tasavvur qilish mumkin: Matn, tokenizatsiya, morfologik tahlil, sintaktik tahlil, semantik tahlil, natija. Masalan, "Talabalar yangi kitoblarni o'qidilar" gapini kompyuter qayta ishlaganda dastlab alohida tokenlar aniqlanadi. Keyin har bir so'zning grammatik xususiyatlari, so'ngra so'zlar o'rtasidagi sintaktik munosabatlar va umumiy mazmun tahlil qilinadi.

2. O'ZBEK TILINING NLP UCHUN O'ZIGA XOS XUSUSIYATLARI

O'zbek tili turkiy tillar oilasiga mansub bo'lib, agglutinativ tillar qatoriga kiradi. Agglutinativ tillarda grammatik ma'nolar ko'pincha o'zakka turli qo'shimchalarning ketma-ket qo'shilishi orqali ifodalanadi. Masalan: kitob, kitoblar, kitoblarimiz, kitoblarimizdan.

Bu holat NLP tizimlari uchun ham imkoniyat, ham murakkablik tug'diradi. Bir tomondan, qo'shimchalarning ma'lum tartibda kelishi algoritmik tahlil uchun qoidalar yaratish imkonini beradi. Ikkinchi tomondan, bitta o'zakdan ko'plab so'z shakllari hosil bo'lishi mumkin.

O'zbek tilida lotin va kirill yozuvlarining amaliyotda uchrashi ham raqamli qayta ishlashda hisobga olinishi kerak. Bundan tashqari, so'z ma'nosining kontekstga bog'liqligi faqat lug'at asosidagi tizimlarning imkoniyatini cheklaydi.

3. O'ZBEK TILIDA MORFOLOGIK TAHLIL

Morfologik tahlil NLPning asosiy bosqichlaridan biri bo'lib, so'zning grammatik tuzilishini aniqlashga xizmat qiladi. Unda o'zak, qo'shimchalar, so'z turkumi, son, kelishik, zamon, shaxs-son va egalik kabi xususiyatlar aniqlanishi mumkin. Masalan, "kitoblarimizdan" so'zini kitob + lar + imiz + dan shaklida tahlil qilish mumkin. Bu yerda kitob - o'zak, -lar - ko'plik, -imiz - egalik, -dan - chiqish kelishigi.

Stemming soʻzning qoʻshimchalarini olib tashlab, asosiy stemini aniqlash jarayonidir. Lemmatizatsiya esa soʻzning lugʻaviy, normal shaklini aniqlashga qaratiladi. Oʻzbek tili uchun bu ikki jarayon boʻyicha alohida tadqiqotlar olib borilmoqda.

3. OʻZBEK TILIDA SINTAKTIK TAHLIL

Sintaktik tahlil gapdagi soʻzlarning oʻzaro grammatik munosabatlarini aniqlash jarayonidir. Masalan, “Talaba kitobni kutubxonada oʻqidi” gapida talaba harakat bajaruvchisi, kitobni obyekt, kutubxonada joy holi, oʻqidi esa kesim vazifasida keladi. Sintaktik tahlilda constituency parsing va dependency parsing kabi yondashuvlardan foydalaniladi. Dependency parsing soʻzlar oʻrtasidagi bogʻlanishlarni daraxt shaklida ifodalashga imkon beradi.

Oʻzbek tilida sintaktik tahlil soʻz tartibining nisbatan erkinligi va morfologik shakllarning murakkabligi sababli oʻziga xos yondashuvni talab qiladi. Universal Dependencies standartidagi belgilangan korpuslar bunday tizimlarni oʻqitish va baholash uchun muhim resurs hisoblanadi.

4. SEMANTIK TAHLIL VA SENTIMENT ANALYSIS

Semantik tahlil matn yoki gapning maʼnosini avtomatik aniqlashga qaratiladi. Masalan, “Bu kitob menga juda yoqdi” gapida ijobiy munosabat ifodalangan boʻlsa, “Bu xizmatdan koʻnglim qolgan” gapida salbiy munosabat mavjud. Sentiment analysis ijtimoiy tarmoqlar monitoringi, mijozlar fikrlarini tahlil qilish, mahsulot va xizmatlar haqidagi sharhlarni oʻrganish hamda elektron savdo kabi sohalarda qoʻllanishi mumkin.

Oʻzbek tilida sentiment analysis uchun sifatli va katta hajmdagi belgilangan maʼlumotlar zarur. Shu sababli oʻzbekcha sharhlar va ijtimoiy tarmoq matnlari asosida maxsus datasetlar yaratish muhim.

5. BERT VA ZAMONAVIY NEYRON MODELLAR

Soʻnggi yillarda NLP texnologiyalarida transformer arxitekturasiga asoslangan modellar katta ahamiyat kasb etmoqda. BERT — Bidirectional Encoder Representations from Transformers — matndagi soʻzlarni ularning konteksti bilan birgalikda ifodalash imkonini beradi. Oʻzbek tili uchun BERTbek kabi maxsus modellar ishlab chiqilgan. Bunday modellar sentiment analysis, mavzularni tasniflash va named entity recognition kabi vazifalarda foydalanilishi mumkin.

Oʻzbek tiliga mos modellarni yaratishda katta hajmdagi va turli janrdagi matnlardan iborat sifatli korpus zarur. Bu esa oʻzbek tilining raqamli resurslarini rivojlantirish masalasini yanada muhim qiladi.

6. OʻZBEK TILI UCHUN LINGVISTIK DATASETLARNING AHAMIYATI

NLP modelini yaratish uchun algoritmnining oʻzi yetarli emas. Modelni oʻqitish, sinash va baholash uchun katta va sifatli maʼlumotlar toʻplami zarur.

Datasetlar morfologik tahlil, POS-tagging, sintaktik parsing, sentiment analysis va boshqa vazifalarda ishlatiladi. O'zbek tili uchun so'nggi yillarda morfologik, leksik va sintaktik jihatdan belgilangan datasetlar yaratilmoqda.

Bunday resurslarning ko'payishi o'zbek tilida NLP tadqiqotlarini rivojlantirish, turli modellarni taqqoslash va amaliy dasturiy vositalar yaratish imkoniyatlarini kengaytiradi.

7. O'ZBEK TILIDA NLP TEXNOLOGIYALARINING AMALIY QO'LLANILISHI

NLP texnologiyalarini o'zbek tilida avtomatik tarjima, imlo tekshirish, ovozli yordamchilar, chatbotlar, ta'lim texnologiyalari va axborot qidiruv tizimlarida qo'llash mumkin. Avtomatik tarjimada morfologik, sintaktik va semantik tahlil muhim. Imlo tekshiruvchilar esa so'z shakllarini tanish va xatolarni aniqlash uchun morfologik resurslardan foydalanadi. Nutqni matnga va matnни nutqqa aylantirish texnologiyalari o'zbekcha ovozli yordamchilar yaratish uchun asos bo'la oladi.

Ta'limda NLP o'quvchilarning yozma ishlarini tekshirish, grammatik xatolarni aniqlash, matn mazmunini tahlil qilish va individual topshiriqlar yaratishda yordam berishi mumkin.

8. O'ZBEK TILIDA NLP TEXNOLOGIYALARINING MUAMMOLARI

O'zbek tilida NLPni rivojlantirishda bir qator muammolar mavjud. Birinchi muammo - lingvistik resurslarning yetarli emasligi. Ikkinchi muammo - o'zbek tilining agglutinatив va morfologik jihatdan murakkab tuzilishi. Uchinchi muammo - lotin va kirill yozuvlari bilan bog'liq normalizatsiya masalalari. Shuningdek, qo'lda belgilangan katta hajmdagi datasetlarning yetishmasligi, internet tilining tez o'zgarishi, sheva va uslubiy xilma-xillik ham modellarni o'qitishda qiyinchilik tug'diradi.

Mazkur muammolarni hal qilish uchun tilshunolar, dasturchilar va sun'iy intellekt mutaxassislarining hamkorligi, ochiq lingvistik resurslarni ko'paytirish va yagona standartlardan foydalanish muhimdir.

9. O'ZBEK TILIDA NLP TEXNOLOGIYALARINING ISTIQBOLLARI

O'zbek tili uchun NLP texnologiyalarining istiqbollari keng. Avvalo, katta hajmdagi milliy lingvistik korpus yaratish zarur. Bunday korpusga badiiy, ilmiy, rasmiy, yangilik va internet matnlarini kiritish mumkin.

Ikkinchi yo'nalish - o'zbek tiliga maxsus moslashtirilgan katta til modellarini yaratish. Uchinchi yo'nalish - morfologik va sintaktik analizatorlarni takomillashtirish. To'rtinchi yo'nalish - o'zbekcha chatbot, avtomatik tarjima, sentiment analysis va ovozli texnologiyalarni rivojlantirish. O'zbek tilining raqamli resurslari qanchalik ko'p va sifatli bo'lsa, NLP modellarining ishlash samaradorligi ham shunchalik yuqori bo'ladi.

XULOSA

Tabiiy tilni qayta ishlash texnologiyalari bugungi kunda sun'iy intellekt va kompyuter lingvistikasining muhim yo'nalishlaridan biriga aylanmoqda. Ushbu texnologiyalar inson tilidagi katta hajmdagi ma'lumotlarni avtomatik qayta ishlash, tahlil qilish va ulardan zarur axborotni ajratib olish imkonini beradi.

O‘zbek tilida NLP texnologiyalarini rivojlantirish alohida lingvistik yondashuvni talab qiladi. O‘zbek tilining agglutinatib tabiati, morfologik shakllarning ko‘pligi, so‘z tartibining nisbatan erkinligi, turli yozuv shakllari va lingvistik resurslarning cheklanganligi mavjud NLP tizimlari uchun muayyan qiyinchiliklarni yuzaga keltiradi.

Tadqiqot davomida morfologik tahlil, stemming, lemmatizatsiya, POS-tagging, sintaktik va semantik tahlil o‘zbek tili NLP tizimlarining asosiy tarkibiy qismlari ekanligi ko‘rsatildi. Shuningdek, BERT kabi transformer modellarining o‘zbek tiliga moslashtirilishi yangi imkoniyatlarni yuzaga keltirmoqda.

Kelgusida o‘zbek tilining katta hajmdagi va ko‘p janrli lingvistik korpuslarini yaratish, morfologik va sintaktik resurslarni kengaytirish, o‘zbek tiliga mos katta til modellarini ishlab chiqish hamda mavjud datasetlarni yagona standartlar asosida birlashtirish muhim vazifalardan hisoblanadi.

Umuman olganda, o‘zbek tilida NLP texnologiyalarining rivojlanishi ta’lim, avtomatik tarjima, elektron hukumat, axborot qidiruvi, ovozli texnologiyalar, chatbotlar va boshqa raqamli xizmatlarning rivojlanishiga katta hissa qo‘shadi.

FOYDALANILGAN ADABIYOTLAR:

1. Xolmanova Z.T.-”Kompyuter lingvistikasi” Toshkent-2024.-234b.
2. Elov B., Abdullayev O., Primova M.- “O‘zbek tilida tabiiy tilni qayta ishlashdagi morfologik, sintaktik va semantik tahlil metodlari”2025.-1-jild,1-son
3. Rustamov.R .T maqolalari
Sayfiddinov Z.A NLP-X respublika ilmiy-amaliy to‘plami-2025