

О‘ZBEK TILINI RAQAMLASHTIRISH DASTURI (OTRD)

Raxmonjonov Nozimjon Yusufjon o‘g‘li

“Toshkent shahar elektr tarmoqlari” MChJ muhandisi

Annotatsiya

O‘zbek tilini raqamlashtirish jarayoni milliy tilning zamonaviy texnologik muhitda rivojlanishini ta’minlash, uni global axborot makoniga integratsiya qilish hamda sun’iy intellekt tizimlarida samarali qo‘llash imkoniyatlarini kengaytirishga xizmat qiladi. Ushbu tadqiqotning asosiy maqsadi – o‘zbek tili uchun keng qamrovli lingvistik korpus yaratish, fonologik, morfologik va sintaktik xususiyatlarni avtomatlashtirish, mashina tarjimai va nutq texnologiyalarini rivojlantirishdir.

Dasturiy mahsulot sifatida yaratiladigan platforma nafaqat ilmiy-lingvistik tadqiqotlarda, balki davlat boshqaruvi, ta’lim, elektron xizmatlar va biznes sohalarida ham amaliy qo‘llanilishi mumkin. Natijada, o‘zbek tilida qidiruv tizimlari, grammatik tekshiruvchilar, nutqni tanish va sintez qilish dasturlari, elektron lug‘atlar va terminologik bazalar shakllanadi.

Shunday qilib, o‘zbek tilini raqamlashtirish milliy tilning global miqyosda mavqeini mustahkamlash, uning ilmiy va amaliy imkoniyatlarini kengaytirish hamda kelajak avlod uchun barqaror lingvistik infratuzilma yaratishga xizmat qiladi.

Kalit so‘zlar: O‘zbek tili, morfologiya, fonologiya, ortografiya, til siyosati, tilni rivojlantirish, korpus lingvistika, raqamli resurslar.

Kirish

Raqamli texnologiyalar asrida milliy tillarni saqlash va rivojlantirish davlat miqyosidagi muhim vazifalardan biridir. O‘zbek tili dunyo miqyosida millionlab foydalanuvchiga ega bo‘lsada, uni raqamli resurslar, dasturiy mahsulotlar va sun’iy intellekt tizimlarida qo‘llash yetarli darajada emas. Shu sababli O‘zbek Tilini Raqamlashtirish Dasturi (OTRD) ishlab chiqilishi zarur. Ushbu dastur nafaqat ilmiy-lingvistik bazani yaratishga, balki keng jamoatchilik va davlat boshqaruvi tizimida qo‘llashga xizmat qiladi.

2. Maqsad va Vazifalar

Maqsad:

O'zbek tilini zamonaviy raqamli texnologiyalarga moslashtirish, sun'iy intellekt va ma'lumotlar bazasida samarali qo'llash imkoniyatini yaratish. Hozirgi davrda dunyodagi ko'plab tillar sun'iy intellekt tizimlariga integratsiya qilinmoqda. Masalan, ingliz, rus, xitoy, koreys va yapon tillari Google Translate, ChatGPT, DeepL, Siri, Alexa kabi platformalarda mukammal ishlaydi. O'zbek tili esa hali to'liq raqamlashtirilmagan, bu esa:

davlat idoralari ish yuritishida,
ta'limda elektron resurslardan foydalanishda,
sun'iy intellekt xizmatlarida
to'siqlarni yuzaga keltirmoqda.

Vazifalar:

O'zbek tili korpusini yaratish va boyitish.

Fonologiya, morfologiya va sintaksisni avtomatik qayta ishlash tizimini ishlab chiqish.

Mashina tarjimasini, nutqni avtomatik tanish va nutqni sintez qilish dasturlarini yaratish.

Elektron lug'atlar va terminologik bazalarni ishlab chiqish.

Ta'lim va davlat boshqaruvi tizimida o'zbek tilida raqamli xizmatlarni keng joriy qilish.

3. Asosiy Yo'nalishlar

3.1. Lingvistik Korpus

Korpus hajmi: kamida 500 mln so'z shakli.

Manbalar: adabiyot, matbuot, ilmiy maqolalar, internet manbalari.

Belgilanadigan xususiyatlar: fonetik, morfologik, sintaktik va semantik tahlil.

Har bir so'z fonetik, morfologik, sintaktik jihatdan belgilanishi lozim.

Masalan: so'zning ildizi, qo'shimchasi, ot/fe'l ekanligi, gapdagi vazifasi ko'rsatiladi.

3.2. Fonologik Modul

O'zbek tilidagi unli va undosh tovushlarni akustik modellashirish.

Dialekt farqlarini hisobga oluvchi ovozli ma'lumotlar bazasi.

Nutqni avtomatik tanish (ASR – Automatic Speech Recognition) tizimi.

Nutq sintezi (TTS – Text to Speech) tizimi: matndan ovoz chiqarish.

Bu texnologiyalar orqali O'zbekcha Siri yoki Google Assistant yaratish mumkin.

3.3. Morfologik va Sintaktik Tahlil

So'z yasash va qo'shimcha tizimini avtomatik ajratish.

Gap bo'laklarini sintaktik daraxt shaklida aniqlash.

Matnlarni avtomatik grammatik tekshirish dasturi.

Gapni sintaktik daraxt sifatida aniqlash.

Bu tizim grammatika tekshiruvchisi va avtomatik tarjima dasturlariga asos bo'ladi.

3.4. Mashina Tarjimasini Yaratish

O'zbek tili ↔ ingliz tili ↔ rus tili yo'nalishida neyron tarjima tizimi.

Terminologik moslik va kontekstga asoslangan tarjima.

Terminologik sohalar (huquq, tibbiyot, texnika) uchun maxsus modellar kerak.

Bu orqali ilmiy maqolalar, davlat hujjatlari tez va to'g'ri tarjima qilinadi.

3.5. Lug'at va Ma'lumotlar Bazasi

O'zbek tili sinonim, antonim va frazeologik lug'ati.

Soha bo'yicha terminologik ma'lumotlar bazasi (huquq, tibbiyot, texnika).

Ochiq API shaklida taqdim etish.

Tibbiyot, iqtisodiyot, huquq, texnika kabi sohalar uchun terminologik ma'lumotlar bazasi yaratiladi.

API shaklida boshqa dasturlar bilan ulanish imkoniyati beriladi.

4. Texnologik Arxitektura

Backend: Python (NLP kutubxonalari – HuggingFace, Stanza, NLTK).

Frontend: React / Vue.js asosida web va mobil ilova.

Ma'lumotlar bazasi: PostgreSQL, MongoDB.

Sun'iy intellekt: Transformer model (BERT, GPT asosida o'zbek tiliga moslashtirilgan).

Integratsiya: API orqali boshqa davlat va ta'lim platformalariga ulash.

5. Amaliy Natijalar

O'zbek tilida ishlovchi qidiruv tizimlari.

Raqamli ta'lim platformalari uchun grammatika tekshiruvchisi.

Ovozni matnga va matnni ovozga aylantirish dasturi.

Raqamli davlat hujjatlarini avtomatik tarjima va tahrirlash imkoniyati.

O'zbek tilining xalqaro miqyosda kengroq qo'llanishi.

6. Kutilayotgan Ijtimoiy-Iqtisodiy Samara

Til siyosatini mustahkamlash.

Davlat idoralarida o'zbek tilida to'liq raqamli xizmat ko'rsatish.

Ta'limda samarali elektron resurslar yaratish.

O'zbek tilini global miqyosda sun'iy intellekt tizimlariga integratsiya qilish.

O'zbek tilida ishlovchi sun'iy intellekt dasturlari.

O'zbek tilidagi matn, ovoz, tarjima texnologiyalari.

Davlat va ta'lim tizimida to'liq raqamli xizmatlar.

O'zbek tilining xalqaro miqyosda mavqei oshishi.

7. Xulosa

O'zbek tilini zamonaviy raqamli texnologiyalarga moslashtirish nafaqat til siyosati va madaniy merosni asrashning dolzarb yo'nalishi, balki axborot jamiyati sharoitida ijtimoiy-iqtisodiy taraqqiyotning muhim omili sifatida ham namoyon bo'ladi. Tilni sun'iy intellekt tizimlariga integratsiya qilish, keng qamrovli lingvistik korpus yaratish, fonologik va morfologik tahlilni avtomatlashtirish hamda mashina tarjimasini va nutq texnologiyalarini ishlab chiqish orqali o'zbek tili global raqamli makonda faol qo'llanilish imkoniyatiga ega bo'ladi.

Natijada, davlat boshqaruvi, ta'lim, ilmiy tadqiqot va iqtisodiy faoliyat jarayonida samaradorlik ortadi, o'zbek tilida elektron xizmatlar rivojlanadi va milliy tilning

xalqaro nufuzini mustahkamlashga erishiladi. Demakki, o'zbek tilini raqamlashtirish bo'yicha dasturiy mahsulotlar ishlab chiqish – milliy manfaat, ilmiy zarurat va global axborotlashuv talablarining uzviy uyg'unlashgan ko'rinishi sifatida qaralishi lozim.

Foydalanilgan adabiyotlar ro'yxati

1. Jo'rayev, M. (2020). O'zbek tili milliy korpusi va uning lingvistik tadqiqotlarda qo'llanilishi. Toshkent: O'zbekiston Milliy Ensiklopediyasi.
2. Rasulov, Sh. (2021). O'zbek tilining zamonaviy rivojlanish yo'nalishlari: raqamli texnologiyalar va til siyosati. "Filologiya masalalari" jurnali, №2, 15–27.
3. Karimov, N. (2019). O'zbek tilining fonologik tizimini modellashtirish muammolari. Toshkent davlat tilshunoslik instituti ilmiy to'plami, 3(1), 45–60.
4. Xudoyberdiyeva, D. (2022). Raqamli lingvistika va o'zbek tilshunosligi istiqbollari. Toshkent: Ma'naviyat nashriyoti.
5. Jurayev, B. (2018). Kompyuter lingvistikasi: nazariya va amaliyot. Toshkent: Fan va texnologiya.
6. Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353.
7. Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
8. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
9. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd ed.). Stanford University.
10. OpenAI. (2023). Artificial Intelligence and Multilingual NLP. OpenAI Research Papers.