



O'ZBEK TILI TARAQQIYOTINING KOMPYUTER LINGVISTIKASI UCHUN HISSASI

Tursunova Madina Mo'minovna.

Navoiy shahar 11-maktab o'qituvchisi

Annotatsiya. Ushbu maqolada o'zbek tilining taraqqiyoti va uning kompyuter lingvistikasi sohasiga qo'shgan hissassi tahlil qilinadi. Tadqiqot davomida o'zbek tilining morfologik xususiyatlari, til korpuslarining yaratilishi, mashina tarjimasi tizimlarining rivojlanishi va tabiiy til ishlov berish texnologiyalarining joriy etilishi kabi masalalar o'rganilgan. Maqola zamonaviy axborot texnologiyalari sharoitida o'zbek tilining raqamli resurslarini boyitish va kompyuter lingvistikasi sohasidagi ilmiy-tadqiqot ishlarini rivojlantirish yo'nalishlarini belgilaydi.

Kalit so'zlar: o'zbek tili, kompyuter lingvistikasi, til korpusi, tabiiy til ishlov berish, mashina tarjimasi, morfologik tahlil, raqamli til resurslari

Kirish. Zamonaviy dunyoda axborot texnologiyalarining jadal rivojlanishi milliy tillarni raqamli muhitga moslashtirishni taqozo etmoqda. Kompyuter lingvistikasi sohasi XX asrning 50-yillaridan boshlab matematik tilshunoslik nazariyasi asosida shakllangan bo'lib, bugungi kunda tabiiy tillarni kompyuter yordamida qayta ishlash bo'yicha keng qamrovli tadqiqotlarni o'z ichiga oladi [1]. O'zbek tili kabi agglutinatativ tillar uchun kompyuter lingvistikasi ayniqsa muhim ahamiyat kasb etadi, chunki bu tillarning boy morfologik tizimi va semantik xususiyatlari maxsus yondashuvlarni talab qiladi. O'zbekiston Respublikasi Prezidentining 2020-yil 20-oktabrdagi PF-6084-son farmoniga muvofiq, o'zbek tilini rivojlantirish va til siyosatini takomillashtirish bo'yicha kompleks chora-tadbirlar amalga oshirilmoqda, jumladan, oliy ta'lim muassasalarida "Kompyuter lingvistikasi" yo'nalishi ochilmoqda [2]. Ushbu tizimli yondashuv o'zbek tilining kompyuter lingvistikasi sohasidagi ilmiy-tadqiqot ishlarini yangi bosqichga ko'tarishga xizmat qilmoqda.

Asosiy qism. O'zbek tilining kompyuter lingvistikasi sohasidagi taraqqiyoti bir necha asosiy yo'nalishda amalga oshirilmoqda. O'zbek tili uchun zamonaviy til korpuslarining yaratilishi bu sohaning asosiy ustuvorliklaridan biridir. Til korpusi - bu muayyan til yoki uning variantlariga oid matnlarning elektron shaklda to'plangan, tartibga solingan va lingvistik jihatdan izohlab berilgan yig'indisi bo'lib, kompyuter lingvistikasining turli vazifalarini hal qilishda asosiy manba bo'lib xizmat qiladi [3]. Sharipov va boshqa tadqiqotchilar tomonidan amalga oshirilgan ishlar natijasida o'zbek tili uchun morfologik va sintaktik belgilangan korpus yaratilgan, bu esa til bilan bog'liq turli muammolarni yechishda muhim vosita bo'lmoqda [4]. Korpus yaratish jarayonida Python dasturlash tilidan foydalangan holda maxsus dasturiy ta'minot ishlab chiqilgan va turli janrlardagi matnlar to'plangan. Bunday korpuslar nafaqat fundamental lingvistik tadqiqotlar uchun, balki amaliy



dasturlar - mashina tarjima tizimlari, matn tahlili dasturlari va til o'rganish platformalari yaratish uchun zarurdir.

O'zbek tili uchun tabiiy til islov berish (NLP) texnologiyalarining rivojlanishi katta ahamiyat kasb etmoqda. Chunki o'zbek tili resurs jihatidan cheklangan tillar qatoriga kiradi va murakkab lingvistik tuzilishga ega, NLP sohasida noyob qiyinchiliklarni keltirib chiqaradi [5]. Ushbu muammolarni hal qilish uchun tadqiqotchilar lemmalashtirish, stemming, tovush tanish va semantik tahlil kabi sohalarga e'tibor qaratmoqdalar. Sattarova va boshqa mualliflar tomonidan o'tkazilgan tadqiqotlar o'zbek tilida NLP texnologiyalarini rivojlantirish, interdistsiplinar hamkorlik, tilshunoslik tajribasi va ilg'or hisoblash texnologiyalaridan foydalanish orqali bu muammolarni yengish strategiyalarini taklif qiladi. Bunday yondashuvlar natijasida o'zbek tilga oid matnlarni vektorlash uchun TF-IDF metodi, kosinus o'xshashlik algoritmi va Word2Vec kabi zamonaviy usullar muvaffaqiyatli qo'llanilmoqda.

O'zbek tili uchun nutq tanish tizimlari va til modellarining yaratilishi kompyuter lingvistikasining istiqbolli yo'nalishlaridan biri hisoblanadi. 2023-yilda nashr etilgan tadqiqotda o'zbek tilida uzluksiz nutqni tanish uchun statistik va neyron tarmoqlarga asoslangan til modellari yaratilgan [6]. Ushbu tadqiqot doirasida 80 million o'zbek so'zidan iborat matn korpusi yordamida UzLM nomli o'zbek til modeli ishlab chiqilgan. Yaratilgan til modellarida perplexity ko'rsatkichi 3-gramma til modelida 7.2 ni, LSTM-LM til modelida esa 2.5 ni tashkil etdi. Til modellarini avtomatik nutq tanish tizimiga integratsiya qilish natijasida tanish aniqligi sezilarli darajada oshgani isbotlangan. Bunday til modellari nutq tanish tizimlaridan tashqari, kompyuter lingvistikasining boshqa muammolarini yechishga mo'ljallangan ilovalar - mashina tarjima, matn generatsiyasi va sentiment tahlil dasturlarida ham qo'llanilishi mumkin.

O'zbek tilining morfologik xususiyatlarini hisobga olgan stemming algoritmlari va lemmatizatsiya usullarining ishlab chiqilishi muhim yutuq hisoblanadi. Agglutinativ til bo'lgan o'zbek tilida ko'plab so'zlar qo'shimcha va oksomalash yo'li bilan hosil bo'ladi, bu esa an'anaviy matn qayta ishlash algoritmlarini qo'llashda qiyinchiliklar tug'diradi [7]. UzbekStemmer nomli qoidaga asoslangan stemming algoritmi yaratilgan bo'lib, bu algoritm o'zbek tilining grammatik va morfologik qoidalarini hisobga olgan holda chekli holatlar mashinasi (finite state machine) texnologiyasi asosida ishlab chiqilgan. Bunday algoritmlar matn qazib olish va axborot qidiruv tizimlarida so'zlarning turli shakllarini ularning asosiy (ildiz) shakllariga keltirish uchun zarur bo'lib, bu esa qidiruv samaradorligini sezilarli darajada oshiradi.

O'zbek tili uchun matn tasnifi va sentiment tahlil sohasidagi tadqiqotlar ham kompyuter lingvistikasi rivojiga katta hissa qo'shmoqda. Kuriyozov va hamkasblarining tadqiqotida o'zbek tilidagi matnlarni tasnif qilish uchun maxsus ma'lumotlar to'plami yaratilgan bo'lib, u 10 xil yangiliklar va matbuot veb-saytlaridan to'plangan va 15 toifali yangiliklar, matbuot va huquqiy matnlarni qamrab oladi [8]. Tadqiqot natijasida Recurrent Neural Network (RNN) va Convolutional Neural Network (CNN) ga asoslangan modellar qoidaga





asoslangan modellardan yaxshiroq natija berishi aniqlangan. Eng yaxshi natija o'zbek korpusida o'qitilgan transformer asosidagi BERT modeli bo'lgan BERTbek modeli tomonidan erishildi. Bu natijalar o'zbek tilida matn tasnifi bo'yicha keyingi tadqiqotlar uchun yaxshi asos yaratmoqda va til texnologiyalarining rivojlanishiga muhim hissa qo'shmoqda.

Xulosa. O'zbek tilining kompyuter lingvistikasi sohasidagi taraqqiyoti ko'p qirrali va sistemli xarakter kasb etmoqda. Til korpuslarining yaratilishidan tortib to nutq tanish tizimlari va mashina tarjimai dasturlarigacha bo'lgan keng spektrdagi ishlar o'zbek tilining raqamli muhitdagi mavqeini mustahkamlashga xizmat qilmoqda. Davlat tomonidan qabul qilingan strategik qarorlar va ilmiy tadqiqot muassasalarining faol faoliyati natijasida bu sohada erishilgan yutuqlar nafaqat nazariy, balki amaliy ahamiyatga ham ega. O'zbek tilining agglutinativ xususiyatlari, boy morfologik tizimi va semantik murakkabligi kompyuter lingvistikasi uchun qiyinchiliklar tug'dirsa-da, ayni paytda bu xususiyatlar tilning raqamli texnologiyalar bilan integratsiyasi uchun noyob imkoniyatlar yaratadi. Kelajakda o'zbek tili uchun yanada takomillashgan AI-asoslangan til modellari, ko'p tilli mashina tarjimai tizimlari va intellektual matn tahlil dasturlari yaratilishi, shuningdek, xalqaro hamkorlik asosida o'zbek tilining jahon axborot tarmog'idagi o'rnini mustahkamlash istiqbolda turmoqda.

Adabiyotlar ro'yxati

1. Kompyuter lingvistikasi: muammolar, yechimlar, istiqbollar. 1-jild, 1-son. Navoiy nomidagi O'zbekiston Davlat Universiteti, 2021.
2. Jo'rayev, J. (2023). Haydar Xorazmiy "Gulshan ul-asror" asari qo 'lyozmalariga ishlangan miniatyuralar xususida. Uzbekistan: Language and Culture, 1(1).
3. Jo'raev, A. (2025). Improving the methodology for improving students' physical abilities using wrestling elements. Academic Journal of Science, Technology and Education, 1(3), 16-19.
4. Жураев, А. Б. (2024). Формирования развития спортсменов дзюдоистов к подготовке профессиональной педагогической деятельности. Science and Education, 5(6), 254-258.
5. Juraev, A. B. (2024). MORPHOFUNCTIONAL STATE OF JUDOKA ATHLETES FOR THE PREPARATION OF PROFESSIONAL TEACHING ACTIVITIES. Multidisciplinary Journal of Science and Technology, 4(5), 579-583.
6. Жураев, А. Б. (2024). Обучения к нагрузкам физической подготовки как деятельность подготовке школьников. Science and Education, 5(2), 574-579.
7. Sennett, R. (2008). The Craftsman. Yale University Press. [Yale University PressGoogle Books](#)



8. Tulanova, M. (2025). METHODS OF PAIN MANAGEMENT IN PEDIATRIC DENTISTRY: MODERN APPROACHES AND CLINICAL RECOMMENDATIONS. International Journal of Artificial Intelligence, 1(4), 686-691.

9. Tulanova, M. (2024). Dental Implantation Planning: Instrumentation and Medication Support. Spanish Journal of Innovation and Integrity, 37, 88-90.

10. Tulanova, M. (2024). TYPES OF IMPLANTS: CLASSIFICATION, INSTRUMENTATION, AND MEDICATION SUPPORT. Modern World Education: New Age Problems–New solutions, 1(3), 67-69.

11. Файзиев, О. Р. (2017). Потребность населения в растительном масле в условиях Узбекистана и методические аспекты её определения. Экономика и предпринимательство, (12-2), 305-309.

12. Davletov, I., Fayziyev, O., & Умарова, З. (2023). Moliyaviy hisobotning xalqaro standartlariga muvofiq moliyaviy hisobotlarni konsolidatsiyalash tamoyillari. Центральноазиатский журнал образования и инноваций, 2(5), 148-151.

13. Fayziyev, O., Saidov, M. A., & Abduraxmonova, N. (2022). Agrar sektorni rivojlantirishda resurslar salohiyatidan foydalanishning hozirgi holati tahlili. Евразийский журнал права, финансов и прикладных наук, 2(13), 275-279.

14. Файзиев, О. Р. (2015). Социально-экономическое значение возделывания масличных культур в Узбекистане. World science, 3(3 (3)), 10-11.

15. Caiger-Smith, A. (1985). Lustre Pottery: Technique, Tradition and Innovation in Islam and the Western World. Faber & Faber.

16. Sherzodovna, T. S., & Mamatkulovna, D. N. FEATURES OF CURRICULUM IN SECONDARY EDUCATION.

17. Mamatkulovna, D. N., Sherzodovna, T. S., & Raxmonovna, B. M. (2023). IMPORTANCE OF A COMMUNICATIVE METHOD FOR TEACHING FOREIGN LANGUAGES. INTERNATIONAL JOURNAL OF SOCIAL SCIENCE & INTERDISCIPLINARY RESEARCH ISSN: 2277-3630 Impact factor: 8.036, 12(06), 200-203.

18. Тагаева, Т. (2022). Индивидуальные особенности художественного стиля УС Моэма. Общество и инновации, 3(11/S), 132-137.

19. Тагаева, Т. Б. (2018). ЛИНГВОКУЛЬТУРОЛОГИЧЕСКАЯ ИНФОРМАЦИЯ В СЕМАНТИКЕ АНГЛИЙСКИХ ФРАЗЕОЛОГИЗМОВ. In Культурные инициативы (pp. 205-207).

20. ULUGBEKOVNA, R. S. (2020). The First Period of Washington Irving's.

21. Rasulova, S. U. (2021). INTERPRETATION OF THE EASTERN SUBJECT IN THE WORK OF WASHINGTON IRVING. ББК 74.48 K90, 282.

22. O'zbekiston Respublikasi Prezidentining 2020-yil 20-oktabrdagi PF-6084-son "O'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida"gi farmoni.

