

O‘ZBEK TILI DIALEKTLARI UCHUN SUN’IY INTELLEKT ASOSIDAGI IZOHLI LUG‘ATNI YARATISH

Abidova Shaxnoza Baxodirovna

TATU v.b dotsenti

Samadov Dilshodbek Usmon o‘g‘li

TATU magistranti

Raximova Madina Dilshod qizi

Tatu magistranti

Annotatsiya: Mazkur maqolada o‘zbek tilining Surxondaryo va Qashqadaryo hududlariga xos dialektlarini raqamlashtirish, tizimlashtirish va izohlash muammolari hamda ularni zamonaviy axborot texnologiyalari asosida yechish yo‘llari yoritilgan. Tadqiqot doirasida sun’iy intellekt (NLP) asosida ishlovchi dialekt izohli lug‘at tizimi ishlab chiqilib, unda dialekt birliklarni yig‘ish, normalizatsiya qilish, semantik tahlil qilish va adabiy til bilan moslashtirish jarayonlari amalga oshiriladi.

Tizim foydalanuvchilarga dialekt so‘zlarni qidirish, tahrirlash, kontekst asosida izohlash hamda ovoz orqali izlash imkoniyatlarini taqdim etadi. Shuningdek, strukturali ma’lumotlar bazasi va CRUD amallariga asoslangan arxitektura orqali ma’lumotlarni samarali boshqarish va kengaytirish imkoniyati yaratilgan.

Natijada, yaratilgan tizim o‘zbek tilining dialekt boyligini saqlash, ilmiy tadqiqotlar olib borish va ta’lim jarayonini zamonaviy texnologiyalar asosida rivojlantirishda muhim amaliy ahamiyat kasb etadi.

Kalit so‘zlar: dialekt, o‘zbek tili, sun’iy intellekt, NLP (tabiiy tilni qayta ishlash), dialekt lug‘ati, semantik tahlil, ovozli qidiruv, ma’lumotlar bazasi, raqamlashtirish, CRUD amallari

Аннотация: В данной статье рассматриваются вопросы цифровизации, систематизации и интерпретации диалектов узбекского языка, в частности диалектов Сурхандарьинского и Кашкадарьинского регионов, а также пути их решения с использованием современных информационных технологий. В

рамках исследования разработана система толкового диалектного словаря на основе искусственного интеллекта (NLP), обеспечивающая сбор, нормализацию, семантический анализ и сопоставление диалектных единиц с литературным языком.

Система предоставляет пользователям возможности поиска, редактирования, контекстного анализа, а также голосового ввода запросов. Кроме того, использование структурированной базы данных и архитектуры с поддержкой CRUD-операций обеспечивает эффективное управление и расширение данных.

В результате разработанная система имеет важное практическое значение для сохранения диалектного богатства узбекского языка, проведения научных исследований и развития образовательных процессов на основе современных технологий.

Ключевые слова: диалект, узбекский язык, искусственный интеллект, NLP (обработка естественного языка), диалектный словарь, семантический анализ, голосовой поиск, база данных, цифровизация, CRUD операции

Abstract: This paper addresses the issues of digitization, systematization, and interpretation of Uzbek language dialects, particularly those of the Surkhandarya and Kashkadarya regions, and proposes solutions based on modern information technologies. Within the framework of the study, an AI-based (NLP) explanatory dialect dictionary system is developed, enabling the collection, normalization, semantic analysis, and mapping of dialectal units to the standard literary language. The system provides users with functionalities such as search, editing, contextual interpretation, and voice-based query input. In addition, the use of a structured database and a CRUD-based architecture ensures efficient data management and scalability.

As a result, the proposed system has significant practical importance for preserving the dialectal richness of the Uzbek language, supporting linguistic research, and enhancing educational processes through modern technological approaches.

Keywords: dialect, Uzbek language, artificial intelligence, NLP (natural language processing), dialect dictionary, semantic analysis, voice search, database, digitization, CRUD operations

Kirish

O‘zbek tili turkiy tillar oilasiga mansub bo‘lib, o‘zining boy tarixiy taraqqiyoti va hududiy rang-barangligi bilan ajralib turadi. Tilning dialektal qatlamlari esa uning ichki xilma-xilligini namoyon etuvchi muhim lingvistik hodisa hisoblanadi. Xususan, Surxondaryo va Qashqadaryo viloyatlariga xos shevalar o‘zining fonetik, leksik va semantik xususiyatlari bilan o‘zbek adabiy tilidan ma’lum darajada farqlanadi hamda alohida ilmiy tadqiqot obyekti tashkil etadi. Ushbu hududiy shevalarda uchraydigan ko‘plab so‘z va iboralar xalqning tarixiy turmush tarzi, urf-odatlarini va madaniy qadriyatlarini aks ettiradi [1].

Dialektologiya sohasida olib borilgan an’anaviy tadqiqotlar asosan dala materiallari yig‘ish, yozma lug‘atlar tuzish va tavsifiy tahlil qilish bilan cheklanib kelgan. Biroq bunday yondashuvlar ma’lumotlarning tarqoqligi, yangilanmasligi va keng foydalanuvchilar uchun yetarli darajada qulay emasligi bilan xarakterlanadi. Shu bois zamonaviy axborot texnologiyalaridan foydalanish, xususan sun’iy intellekt va tabiiy tilni qayta ishlash (NLP) usullarini joriy etish orqali dialektlarni raqamlashtirish va tizimlashtirish dolzarb masalaga aylanmoqda.

Mazkur tadqiqotda Surxondaryo va Qashqadaryo shevalariga oid shakllantirilgan ma’lumotlar bazasiga tayangan holda, sun’iy intellekt asosida o‘zbek tili dialekt izohli lug‘at tizimini ishlab chiqish masalasi ko‘rib chiqiladi. Taklif etilayotgan tizim dialekt birliklarini yig‘ish, saqlash, qayta ishlash va izohlashni avtomatlashtirishga qaratilgan bo‘lib, unda semantik tahlil, normalizatsiya va lug‘at birliklarini guruhlash kabi jarayonlar amalga oshiriladi [2].

Tizim foydalanuvchilarga dialekt soʻzlarni tezkor qidirish, ularning adabiy ekvivalentlarini aniqlash, kontekstual izohlarini koʻrish hamda hududiy xususiyatlarini tahlil qilish imkonini beradi. Bu esa nafaqat lingvistik tadqiqotlar samaradorligini oshiradi, balki taʼlim jarayonida ham amaliy ahamiyat kasb etadi.

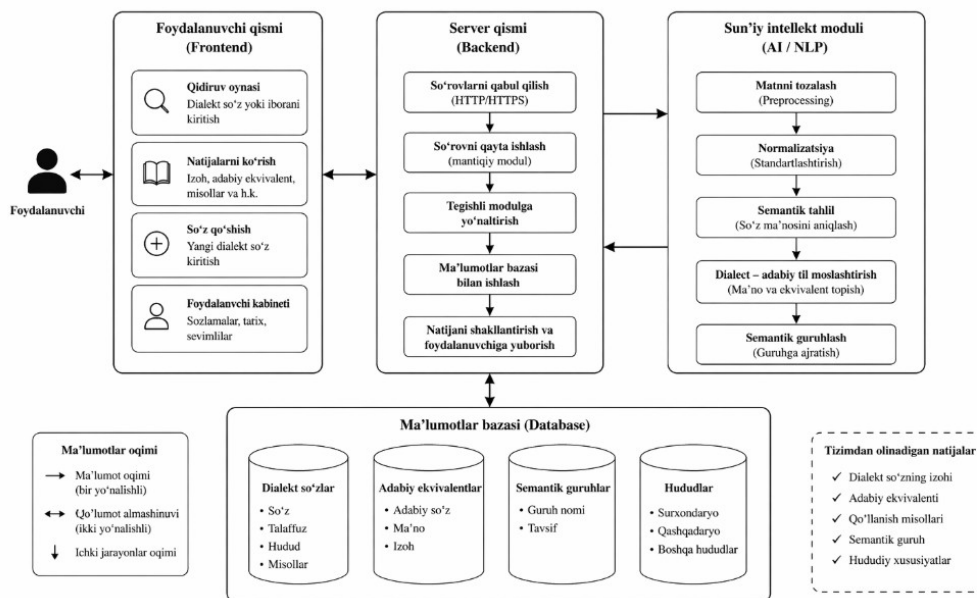
Tadqiqotning dolzarbligi oʻzbek tilining dialekt boyligini saqlash, uni raqamli muhitga olib chiqish va zamonaviy texnologiyalar bilan integratsiya qilish zarurati bilan belgilanadi. Mazkur yondashuv jahon miqyosida olib borilayotgan raqamli lingvistika va NLP asosidagi tadqiqotlar bilan uygʻun boʻlib, oʻzbek tilshunosligida yangi bosqichni boshlab berishi mumkin.

Oʻzbek tili dialekt izohli lugʻat tizimning ishlash prinsipi va umumiy tuzilmasi

Zamonaviy axborot texnologiyalarining jadal rivojlanishi natijasida lingvistik tizimlarni raqamlashtirish va avtomatlashtirish muhim ahamiyat kasb etmoqda. Oʻzbek tilining hududiy dialektlarini tizimli ravishda saqlash, qayta ishlash va foydalanuvchilarga qulay shaklda taqdim etish uchun samarali dasturiy yechimlar zarur hisoblanadi.

Mazkur tadqiqot doirasida ishlab chiqilgan dialekt izohli lugʻat tizimi koʻp bosqichli arxitekturaga asoslanib, foydalanuvchi soʻrovlarini qabul qilish, ularni sunʼiy intellekt yordamida tahlil qilish hamda maʼlumotlar bazasi orqali aniq natijalarni shakllantirish jarayonlarini oʻz ichiga oladi. Tizimning umumiy tuzilmasi uning funksional imkoniyatlari va ishlash samaradorligini taʼminlovchi asosiy komponentlar oʻrtasidagi oʻzaro bogʻliqlikni aks ettiradi.

Quyida ushbu tizimning umumiy arxitekturasi va ishlash prinsipi diagramma koʻrinishida keltirilgan.



1-rasm. O‘zbek tili dialekt izohli lug‘at tizimining umumiy arxitekturasini va asosiy komponentlari o‘rtasidagi o‘zaro bog‘liqlik tasvirlangan.

Rasmda o‘zbek tili dialekt izohli lug‘at tizimining umumiy arxitekturasini aks ettirilgan bo‘lib, unda asosiy komponentlar — foydalanuvchi interfeysi, server qismi, sun‘iy intellekt moduli va ma’lumotlar bazasining o‘zaro bog‘liq holda ishlashi ko‘rsatilgan. Tizim foydalanuvchi tomonidan kiritilgan dialekt so‘zni qabul qiladi va uni server orqali qayta ishlash bosqichiga uzatadi.

Keyingi bosqichda sun‘iy intellekt moduli yordamida matn ustida semantik tahlil amalga oshirilib, so‘zning ma’nosi va adabiy ekvivalenti aniqlanadi. Shundan so‘ng tizim ma’lumotlar bazasiga murojaat qilib, tegishli natijani topadi va uni foydalanuvchiga tushunarli shaklda taqdim etadi. Ushbu arxitektura tizimning tezkor, aniq va samarali ishlashini ta’minlaydi.

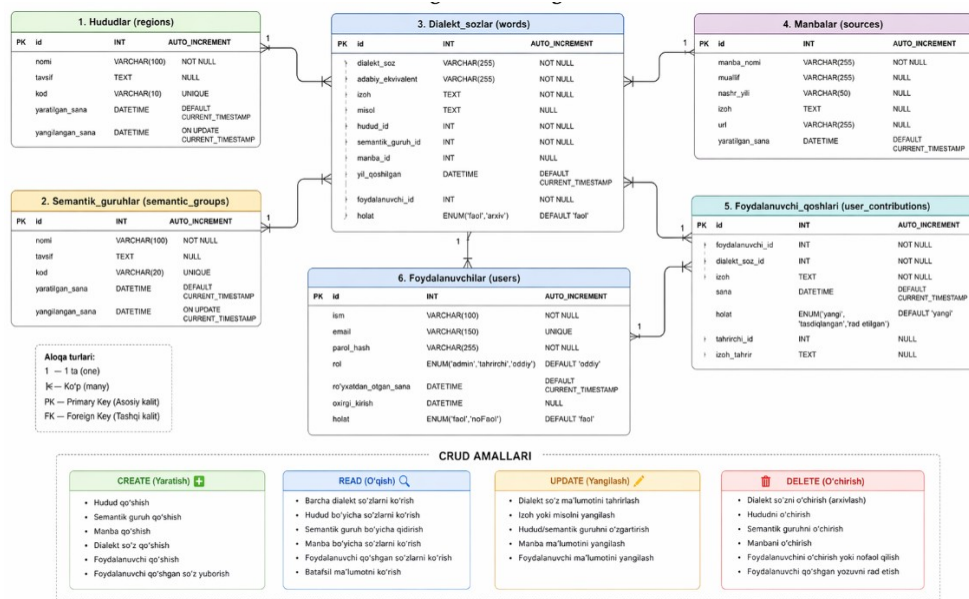
Shu bilan birga, taklif etilayotgan ma’lumotlar bazasi modeli kengaytirilgan holda loyihalashtirilgan bo‘lib, unda qo‘shimcha jadvallar orqali tizimning funksional imkoniyatlari yanada oshirilgan. Xususan, manbalar (sources) jadvali

dialekt soʻzlarning kelib chiqishiga oid maʼlumotlarni saqlashga xizmat qiladi, foydalanuvchilar (users) jadvali esa tizimdan foydalanish va boshqaruv jarayonlarini tashkil etadi.

Foydalanuvchi tomonidan kiritilgan maʼlumotlarni alohida nazorat qilish maqsadida foydalanuvchi qoʻshgan yozuvlar jadvali kiritilgan boʻlib, bu tizimda verifikatsiya va tahrirlash jarayonlarini amalga oshirish imkonini beradi. Shu orqali maʼlumotlarning ishonchliligi va sifatini nazorat qilish mexanizmi shakllantiriladi.

Mazkur maʼlumotlar bazasi tuzilmasi barcha javdallar oʻrtasida oʻzaro bogʻlanishlarni taʼminlab, CRUD (Create, Read, Update, Delete) amallarini toʻliq bajarish imkonini yaratadi. Natijada tizimda maʼlumotlarni kiritish, oʻqish, yangilash va oʻchirish jarayonlari samarali va izchil amalga oshiriladi.

Quyida ushbu tizimning maʼlumotlar bazasi tuzilmasi diagramma koʻrinishida keltirilgan.

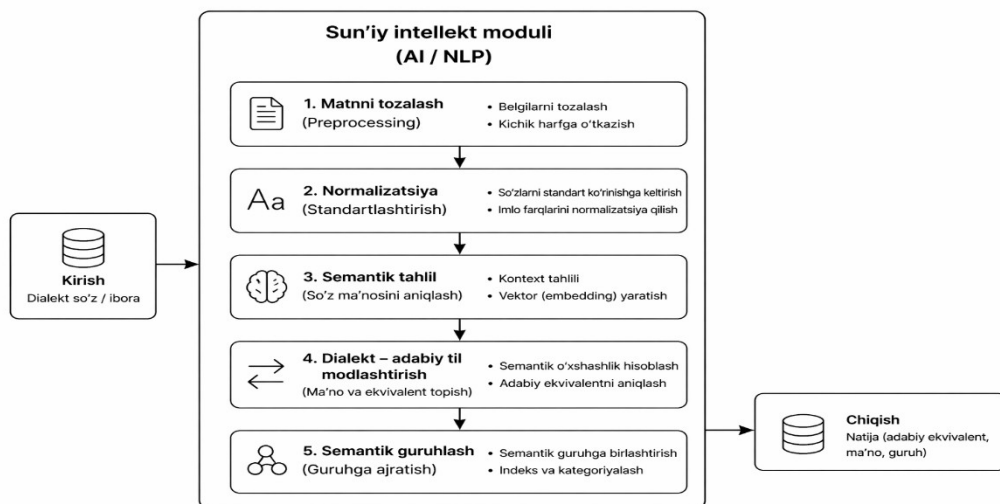


2-rasm. Dialekt izohli lugʻat tizimining kengaytirilgan maʼlumotlar bazasi tuzilmasi va CRUD amallari aks ettirilgan.

Ushbu 2-rasmda keltirilgan ma'lumotlar bazasi tuzilmasi dialekt izohli lug'at tizimini dasturiy jihatdan samarali tashkil etishga yo'naltirilgan bo'lib, unda barcha asosiy obyektlar alohida jadvallar ko'rinishida modellashtirilgan. Markaziy jadval sifatida dialekt_so'zlar jadvali ajratilgan bo'lib, u boshqa jadvallar bilan tashqi kalitlar (foreign key) orqali bog'langan. Bu yondashuv ma'lumotlarning takrorlanishini kamaytirish va normalizatsiyani ta'minlashga xizmat qiladi.

Tizimda hududlar va semantik_guruhlar jadvallari orqali dialekt birliklarining geografik va mazmuniy tasnifi amalga oshiriladi. Manbalar jadvali esa har bir so'zning kelib chiqishi va ilmiy asosini ko'rsatishga xizmat qiladi. Foydalanuvchi bilan bog'liq jarayonlar esa foydalanuvchilar va foydalanuvchi_qo'shlari jadvallari orqali boshqariladi, bu esa tizimda rollar, verifikatsiya va tahrirlash imkoniyatlarini yaratadi.

Mazkur tuzilma relyatsion ma'lumotlar bazasi tamoyillariga asoslangan bo'lib, jadvallar o'rtasidagi bog'lanishlar orqali ma'lumotlar yaxlitligi (data integrity) ta'minlanadi. Tizimda CRUD amallarini to'liq bajarish imkoniyati mavjud bo'lib, foydalanuvchi tomonidan ma'lumotlarni yaratish, o'qish, yangilash va o'chirish jarayonlari samarali amalga oshiriladi. Bu esa dasturiy tizimning barqaror ishlashi va kengaytirish imkoniyatlarini ta'minlaydi. Mazkur imkoniyatlar bilan bir qatorda tizimda sun'iy intellekt moduli ham muhim o'rin egallaydi. Ushbu modul kiritilgan dialekt so'zlarni dastlab tozalash va normalizatsiya qilish orqali standart ko'rinishga keltiradi. Keyingi bosqichda semantik tahlil yordamida so'zning mazmuni aniqlanib, uning adabiy til bilan mosligi baholanadi. Shuningdek, tizim semantik o'xshashlik asosida eng yaqin ekvivalentni tanlab, dialekt birliklarni guruhlariga ajratadi. Natijada foydalanuvchiga aniq, tizimli va mazmunli natija taqdim etiladi.



3-rasm. Sun'iy intellekt moduli orqali dialekt so'zlarni qayta ishlash va semantik tahlil qilish jarayoni tasvirlangan.

Ushbu 3-rasmda sun'iy intellekt moduli (AI/NLP) ning ishlash jarayoni bosqichma-bosqich ko'rsatilgan bo'lib, unda dialekt so'zlar ustida amalga oshiriladigan asosiy lingvistik va hisoblash jarayonlari aks ettirilgan. Dastlab kiritilgan so'z matnni tozalash va normalizatsiya bosqichlaridan o'tkazilib, standart ko'rinishga keltiriladi. Keyingi bosqichda semantik tahlil orqali so'zning ma'nosi aniqlanadi hamda vektor ko'rinishida ifodalanadi.

Shundan so'ng tizim semantik o'xshashlik asosida dialekt so'zga eng yaqin adabiy ekvivalentni aniqlaydi va uni tegishli guruhga ajratadi. Yakuniy bosqichda esa qayta ishlangan ma'lumotlar asosida foydalanuvchiga izohlangan va tizimlashtirilgan natija taqdim etiladi.

Mazkur tadqiqotning ilmiy yangiligi shundan iboratki, unda o'zbek tilining Surxondaryo va Qashqadaryo shevalariga oid dialekt birliklarni qayta ishlash uchun sun'iy intellekt asosidagi integratsiyalashgan yondashuv taklif etilgan. Tizimda dialekt so'zlarni oddiy kalit so'z orqali emas, balki semantik tahlil va

vektor o'xshashlik asosida aniqlash mexanizmi joriy etilgan bo'lib, bu qidiruv aniqligini sezilarli darajada oshiradi.

Bundan tashqari, tizimda matn bilan bir qatorda ovoz orqali qidiruv imkoniyatining mavjudligi uni an'anaviy dialekt lug'atlardan tubdan farqlaydi. Ushbu yondashuv foydalanuvchiga yanada qulay interfeys yaratib, til birliklarini tezkor va intuitiv tarzda izlash imkonini beradi.

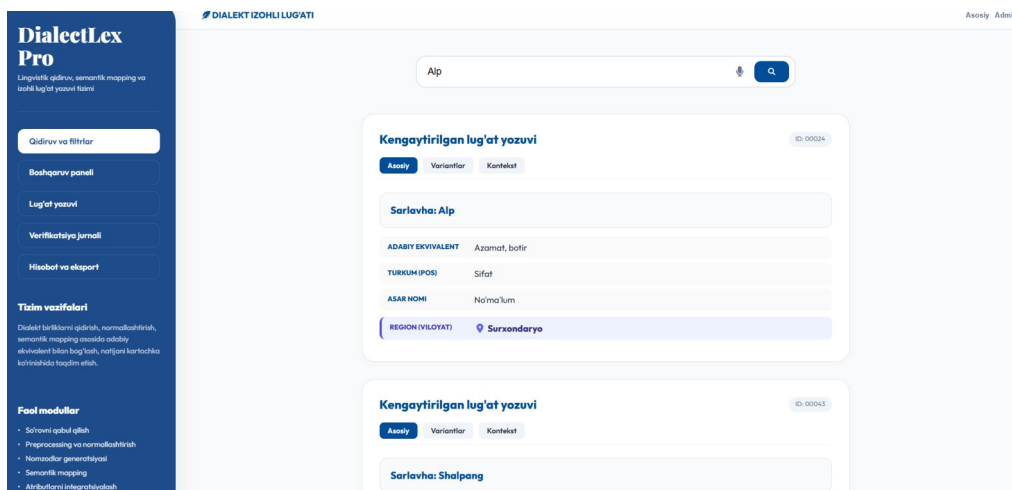
Yana bir muhim jihat sifatida dialekt birliklarni avtomatik normalizatsiya qilish, ularni adabiy til bilan moslashtirish va semantik guruhlariga ajratish funksiyalari ishlab chiqilgan. Bu esa lingvistik ma'lumotlarni chuqur tahlil qilish va tizimli boshqarish imkonini yaratadi.

Shuningdek, tadqiqot doirasida ishlab chiqilgan ma'lumotlar bazasi modeli va CRUD amallarini to'liq qo'llab-quvvatlovchi arxitektura tizimning kengaytiriluvchanligi va barqaror ishlashini ta'minlaydi. Natijada ushbu ish o'zbek tilshunosligida dialektlarni sun'iy intellekt asosida o'rganish va raqamlashtirish bo'yicha yangi ilmiy-amaliy yondashuvni taklif etadi.

Taklif etilayotgan dialekt izohli lug'at tizimining amaliy interfeysi foydalanuvchi uchun sodda va qulay tarzda ishlab chiqilgan. Tizim orqali foydalanuvchi dialekt so'zlarni qidirish, ularning kengaytirilgan izohlarini ko'rish hamda semantik jihatdan boy ma'lumotlarga ega bo'lish imkoniyatiga ega.

Interfeysda har bir dialekt so'z uchun alohida kartochka shaklida ma'lumotlar taqdim etilib, unda so'zning adabiy ekvivalenti, turkumi, manbasi va hududiy tegishliligi aks ettiriladi. Bu esa foydalanuvchiga ma'lumotlarni tizimli va tushunarli shaklda qabul qilish imkonini beradi.

Shuningdek, tizimda qidiruv va filtrlash funksiyalari mavjud bo'lib, ular foydalanuvchiga kerakli ma'lumotlarni tezkor topish imkonini yaratadi. Ushbu yondashuv tizimning amaliy samaradorligini oshiradi va foydalanuvchi tajribasini yaxshilaydi.



4-rasm. Dialekt izohli lug‘at tizimining foydalanuvchi interfeysi va qidiruv natijalari ko‘rsatilgan.

4-rasmda DialectLex Pro tizimining asosiy interfeysi tasvirlangan bo‘lib, unda foydalanuvchi tomonidan kiritilgan so‘z asosida qidiruv natijalari kengaytirilgan lug‘at yozuvi shaklida aks ettirilgan. Har bir yozuvda dialekt so‘zning adabiy ekvivalenti, turkumi va hududiy mansubligi ko‘rsatilgan bo‘lib, bu ma’lumotlar semantik mapping asosida shakllantirilgan. Shuningdek, tizim menyusi orqali foydalanuvchi boshqa funksiyalarga ham oson o‘ta oladi.

Tizimda foydalanuvchilar tomonidan yangi dialekt birliklarni kiritish imkoniyati ham yaratilgan bo‘lib, bu lug‘at bazasini doimiy ravishda boyitib borishga xizmat qiladi. Maxsus forma orqali foydalanuvchi dialekt so‘z, uning normalizatsiyalangan shakli, adabiy ma’nosi, semantik guruhi va hududiy tegishliligi kabi muhim atributlarni kiritadi.

Mazkur yondashuv ma’lumotlarni strukturali tarzda yig‘ish va keyinchalik ularni sun‘iy intellekt yordamida qayta ishlash uchun qulay sharoit yaratadi. Shuningdek, foydalanuvchi tomonidan kiritilgan ma’lumotlar tizimda saqlanib, keyinchalik verifikatsiya va tahlil jarayonlarida foydalaniladi. Bu esa tizimning dinamik rivojlanishini va ma’lumotlar bazasining kengayib borishini ta’minlaydi.

The image shows a mobile application interface for 'DialectLex Pro'. A central white form titled 'Yangi soʻz qoʻshish' (Add new word) is overlaid on a blurred background of the app's main menu. The form contains several input fields and a dropdown menu, each with a label and a placeholder example:

- SHEVA SOʻZI**: Masalan: boʻzola
- NORMALLASHTIRILGAN SHAKL**: Masalan: boʻzola
- ASAR NOMI (MANBA)**: Masalan: Surxon surati
- TURKUM (POS)**: Ot (dropdown menu)
- SEMANTIK GURUH**: Masalan: Shaxs bildiruvchi belgilar
- REGION (VILOYATI)**: Masalan: Surxondaryo, Gashqadaryo
- ADABIY MANOSI**: Adabiy dagi xarakteristika (dropdown menu)
- HISOL (KONTEKST)**: Gap ichida ishlatilishi (dropdown menu)

At the bottom of the form is a blue button labeled 'SAQLASH' (Save) and a smaller link 'Bekor qilish' (Cancel).

5-rasm. Dialekt soʻzlarni tizimga kiritish uchun moʻljallangan foydalanuvchi interfeysi tasvirlangan.

Rasmda yangi dialekt soʻz qoʻshish oynasi koʻrsatilgan boʻlib, unda foydalanuvchi tomonidan kiritiladigan barcha asosiy lingvistik atributlar mavjud. Ushbu forma orqali kiritilgan maʼlumotlar toʻgʻridan-toʻgʻri maʼlumotlar bazasiga uzatiladi va tizimning umumiy lugʻat bazasini kengaytirishga xizmat qiladi. Kelgusida ushbu modulni yanada takomillashtirish orqali kiritilgan maʼlumotlarni avtomatik tekshirish, semantik jihatdan boyitish va sunʼiy intellekt yordamida tavsiyalar berish imkoniyatlarini joriy etish rejalashtirilmoqda. Xususan, tizim foydalanuvchi tomonidan kiritilgan soʻzlarni mavjud baza bilan solishtirib, takroriy yoki oʻxshash birliklarni aniqlay oladi hamda optimal variantlarni taklif qiladi.

Avtomatik tarjima va koʻp hududli dialektlarni qoʻllab-quvvatlash imkoniyatlarini kengaytirish orqali tizimning funksional imkoniyatlari yanada oshiriladi. Bu esa nafaqat lugʻat bazasini boyitish, balki oʻzbek tilining dialekt boyligini keng qamrovda oʻrganish va saqlashga xizmat qiladi.

Xulosa

Mazkur maqolada o‘zbek tilining Surxondaryo va Qashqadaryo hududlariga xos dialektlarini raqamlashtirish va tizimli boshqarishga qaratilgan dasturiy yechim ishlab chiqildi. Tadqiqot jarayonida sun’iy intellekt (NLP) asosida ishlovchi dialekt izohli lug‘at tizimi yaratilib, unda dialekt so‘zlarni qidirish, semantik tahlil qilish va adabiy til bilan moslashtirish mexanizmlari amaliy jihatdan joriy etildi.

Amaliy natija sifatida ishlab chiqilgan DialectLex Pro tizimi orqali foydalanuvchi interfeysi, qidiruv moduli, ma’lumot kiritish (forma), boshqaruv paneli hamda verifikatsiya jarayonlari to‘liq integratsiyalashgan holda amalga oshirildi. Xususan, foydalanuvchi tomonidan dialekt so‘zlarni qo‘shish, tahrirlash va qidirish imkoniyatlari yaratilgan bo‘lib, bu tizimda CRUD amallarining to‘liq ishlashini ta’minlaydi.

Tizimda semantik qidiruv va ovoz orqali izlash funksiyalarining mavjudligi uni an’anaviy lug‘atlardan farqlab, qidiruv aniqligi va foydalanuvchi qulayligini sezilarli darajada oshiradi. Ma’lumotlar bazasining strukturali tashkil etilishi esa dialekt birliklarni tizimli saqlash, tezkor izlash va keyinchalik tahlil qilish imkonini beradi.

Taklif etilgan tizim o‘zbek tilining hududiy boyligini saqlash, ilmiy tadqiqotlar olib borish hamda ta’lim jarayonini zamonaviy texnologiyalar asosida rivojlantirish uchun samarali vosita sifatida xizmat qiladi. Ushbu yechim kelgusida boshqa hududiy dialektlarni ham qamrab olish orqali kengaytirilishi mumkin bo‘lgan innovatsion platforma hisoblanadi. Hamda dasturda keying bosqichlarda shevalarni bazasini kengaytirib borishimiz mumkim.

Foydalanilgan adabiyotlar

1. Karimov B., Tursunov S. *O'zbek tilida tabiiy tilni qayta ishlash masalalari*. – Toshkent: Innovatsion rivojlanish nashriyoti, 2022.
2. Xolmatov T., Raxmatullayev Sh. *O'zbek tilshunosligida zamonaviy tadqiqotlar va dialektologiya*. – Toshkent: Fan, 2023.
3. Jurafsky D., Martin J.H. *Speech and Language Processing* (3rd ed., draft). – Stanford University, 2023.
4. Eisenstein J. *Introduction to Natural Language Processing*. – MIT Press, 2023.
5. Lewis M., Liu Y., Goyal N. et al. *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation*. – arXiv, 2022.
6. OpenAI. *GPT-4 Technical Report*. – OpenAI, 2023.
7. Devlin J., Chang M.W., Lee K., Toutanova K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. – Updated applications and studies, 2022.
8. Brown T.B. et al. *Language Models are Few-Shot Learners*. – Updated research usage and implementations, 2022.