Materiallar to'plami · Conference Proceedings
"ZAMONAVIY TADQIQOTLAR VA BARQAROR RIVOJLANISH" respublika ilmiy-amaliy konferensiyasi
04 August 2026
UDK 811.512.133
FILOLOGIYA

O'ZBEK TILI UCHUN SAVOL-JAVOB TIZIMLARINI YARATISHDAGI MUAMMOLAR VA ULARNI HAL QILISH YO'LLARI

Gulxayyo G'ofurjon qizi Po'latova*
PhD, o'qituvchi
✉ gpulatova2409@gmail.com

ANNOTATSIYA

Ushbu maqolada o'zbek tili uchun savol-javob (QA) tizimlarini yaratishdagi asosiy muammolar – ma'lumotlar bazasining yetishmasligi, tilning agglyutinativ morfologiyasi va ko'p tilli modellarning o'zbek tilidagi cheklangan samaradorligi tahlil qilingan. Shuningdek, RAG (Retrieval-Augmented Generation) arxitekturasi, T5/Flan-T5/mT5/Llama modellarini moslashtirish va gibrid til modellari asosidagi yechim yo'llari ko'rib chiqilgan.

Kalit so'zlar

savol-javob tizimlari, kam resursli tillar, RAG, til modeli, morfologik murakkablik.

DOI: 10.5281/zenodo.21743658

KIRISH

So'nggi yillarda sun'iy intellekt asosidagi savol-javob (QA) tizimlari ta'lim, davlat xizmatlari va biznes sohalarida keng qo'llanilmoqda. Biroq bunday tizimlar, asosan, ingliz va boshqa yuqori resursli tillar uchun rivojlangan. O'zbek tili uchun QA tizimlarini yaratish murakkab vazifa hisoblanadi, chunki bu sohada tayyor ma'lumotlar bazasi va modellar hali yetarli darajada shakllanmagan. Bu esa bunday tizimlarni o'zbek tili uchun ishlab chiqishda yuzaga keladigan qiyinchiliklar va ularni bartaraf etish yo'llarini ishlab chiqish bugunning dolzarb vazifalaridan biri ekanligini ko'rsatadi.

So'nggi yillarda sun'iy intellekt texnologiyalarining jadal rivojlanishi inson faoliyatining deyarli barcha sohalariga sezilarli ta'sir ko'rsatmoqda. Hisoblash quvvatining ortishi, katta hajmdagi ma'lumotlarni qayta ishlash imkoniyatlarining kengayishi hamda mashinaviy o'qitish usullarining takomillashuvi natijasida ilgari ko'p vaqt va mehnat talab qilgan jarayonlar qisqa muddatda, yuqori aniqlik va samaradorlik bilan amalga oshirilmoqda. Bugungi kunda sun'iy intellekt dasturlash, ta'lim, sog'liqni saqlash, davlat boshqaruvi, soliq tizimi, moliya va boshqa ko'plab yo'nalishlarda qaror qabul qilishni qo'llab-quvvatlash, axborotni qayta ishlash hamda xizmatlar sifatini oshirishning muhim vositasiga aylandi.

Tabiiy tilni qayta ishlash (Natural Language Processing, NLP) sun'iy intellektning eng istiqbolli yo'nalishlaridan biri bo'lib, u inson tilini avtomatik tahlil qilish, tushunish va qayta ishlashga xizmat qiladi. Ushbu yo'nalish doirasida yaratilayotgan savol-javob tizimlari foydalanuvchilarga tabiiy tilda berilgan savollarga mazmunan to'g'ri va aniq javoblarni taqdim etish orqali axborot izlash jarayonini sezilarli darajada soddalashtiradi. Bunday tizimlar elektron ta'lim, raqamli kutubxonalar, davlat xizmatlari, tibbiyot, huquq va boshqa ko'plab sohalarda keng qo'llanilib, foydalanuvchilar uchun qulay va samarali axborot almashinuvini ta'minlamoqda.

Sun'iy intellekt va tabiiy tilni qayta ishlash (Natural Language Processing – NLP) texnologiyalarining rivojlanishi natijasida savol-javob (Question Answering – QA) tizimlari ilmiy tadqiqotlarning muhim

yo‘nalishlaridan biriga aylandi. So‘nggi yillarda transformer arxitekturasiga asoslangan yirik til modellari ushbu sohada yuqori natijalarni namoyon etmoqda. Ayniqsa, kam resursli tillar uchun savol-javob tizimlarini yaratish masalasi dolzarb ilmiy muammolardan biri hisoblanadi. Tukeyev va hammualliflar qozoq tili misolida morfologik jihatdan boy va kam resursli tillar uchun savol-javob tizimlarini ishlab chiqish masalasini tadqiq qilganlar [1]. Tadqiqotda FEMSeg_kaz morfologik segmentatsiyasi, MAPP normallashtirish usuli va domen bo‘yicha fine-tuning birgalikda qo‘llanilib, 50386 ta dasturlashga oid savol-javob juftligidan iborat korpus yaratilgan. Natijalar morfologik oldindan qayta ishlash (morphology-aware preprocessing) javoblarni tanlash aniqligini sezilarli darajada yaxshilashini ko‘rsatgan. Mualliflar agglutinativ tillarda affikslarning ko‘pligi va morfologik o‘zgarishlar semantik o‘xshashlikni aniqlashni murakkablashtirishini ta’kidlaydilar. Mazkur yondashuv o‘zbek tili uchun ham muhim metodologik asos bo‘lib xizmat qilishi mumkin, chunki o‘zbek va qozoq tillari tipologik jihatdan o‘xshash turkiy tillar hisoblanadi.

Bi va hammualliflar an’anaviy xitoy tabobati (Traditional Chinese Medicine – TCM) ta’limi uchun TCM-VQA IoTNet nomli intellektual vizual savol-javob tizimini taklif qilganlar. Mualliflar tomonidan o‘tkazilgan ablyatsiya (ablation) tajribalari VisualBERT, GRU va Attention mexanizmlarining tizim samaradorligiga sezilarli ta’sir ko‘rsatishini isbotlagan. Ayniqsa, VisualBERT moduli olib tashlanganda model aniqligi eng katta darajada pasayganligi multimodal semantik tasvirlashning savol-javob tizimlaridagi muhim o‘rnini ko‘rsatadi. Shuningdek, GRU moduli foydalanuvchi holatini hisobga olgan holda shaxsiylashtirilgan javoblarni shakllantirish imkonini bergan.

Mazkur tadqiqotning muhim jihati shundaki, unda savol-javob tizimi faqat matn asosida emas, balki tasvir, matn va qo‘shimcha kontekst ma'lumotlarini birlashtiruvchi multimodal yondashuv asosida ishlab chiqilgan. Bu yondashuv bilimga boy (knowledge-intensive) sohalarda savol-javob tizimlarining aniqligini oshirishini ko‘rsatadi. Biroq ushbu tadqiqot asosan an'anaviy xitoy tibbiyoti sohasiga yo‘naltirilgan bo‘lib, o‘zbek tili kabi kam resursli tillar uchun lingvistik muammolar, morfologik tahlil, korpus yaratish yoki Retrieval-Augmented Generation (RAG) texnologiyalarini qo‘llash masalalari ko‘rib chiqilmagan. Shu sababli mazkur maqola o‘zbek tili uchun savol-javob tizimlarini yaratishda to‘g'ridan-to‘g'ri yechim bermasa-da, multimodal arxitektura, VisualBERT modeli va attention mexanizmini qo‘llash bo‘yicha muhim metodologik asos sifatida xizmat qiladi.

Ushbu yo‘nalish bo‘yicha O‘zbekistonda ham tadqiqot ishlari boshlangan bo‘lib, Muhamadiyev va hammualliflar o‘z tadqiqotlarida o‘zbek tili uchun UzLM til modelini ishlab chiqib, avtomatik nutqni tanish tizimlarida qo‘llash imkoniyatlarini o‘rganganlar [3]. Mualliflar 80 milliondan ortiq so‘z va 15 million gapdan iborat korpusni shakllantirib, statistik hamda neyron til modellarini taqqoslaganlar. Tadqiqot natijalariga ko‘ra, LSTM asosidagi til modeli an'anaviy n-gram modellarga nisbatan yuqori samaradorlikni ko‘rsatgan. Shuningdek, mualliflar o‘zbek tilining agglutinativ tuzilishi, erkinroq so‘z tartibi va resurslarning kamligi til modellari yaratishda asosiy muammolardan biri ekanligini ta'kidlaydilar. Ushbu tadqiqot savol-javob tizimlarini yaratishga bevosita bag'ishlanmagan bo‘lsa-da, sifatli til modeli va katta korpusning mavjudligi QA tizimlarining asosiy komponentlaridan biri ekanligini ko‘rsatadi.

Xo‘jayarov va hammualliflar esa bevosita o‘zbek tili uchun savol-javob tizimlarini yaratishga qaratilgan tadqiqotni amalga oshirgan [4]. Mualliflar T5, Flan-T5, mT5 va Llama modellarini Retrieval-Augmented Generation (RAG) texnologiyasi bilan birgalikda moslashtirib, 10 mingdan ortiq kontekst-savol-javob juftliklaridan tashkil topgan ma’lumotlar bazasida fine-tuning jarayonini o‘tkazgan. Modellar BLEU, ROUGE va METEOR ko‘rsatkichlari orqali baholangan. Tadqiqotda RAG texnologiyasining tashqi bilim bazalaridan kerakli ma’lumotlarni izlab topish va generativ model bilan birlashtirish orqali javob sifatini sezilarli oshirishi ko‘rsatib berilgan. Shu bilan birga, mualliflar o‘zbek tilida annotatsiyalangan QA korpuslarining yetarli emasligi, yopiq soha (closed-domain) savol-javob tizimlarini yaratishdagi qiyinchiliklar hamda mos til modellarining cheklanganligini asosiy muammolar sifatida qayd etgan.

Yuqorida tahlil qilingan ilmiy tadqiqotlar shuni ko‘rsatadiki, o‘zbek tili uchun savol-javob tizimlarini yaratish jarayoni qator lingvistik va texnologik muammolar bilan tavsiflanadi. Birinchidan, yuqori sifatli va keng qamrovli annotatsiyalangan savol-javob korpuslari hamda bilim bazalarining yetarli emasligi zamonaviy sun’iy intellekt modellarini samarali o‘qitish imkoniyatini cheklaydi. Ikkinchidan, o‘zbek tilining agglutinativ

tuzilishi, murakkab morfologik xususiyatlari va so‘z yasash imkoniyatlarining kengligi tabiiy tilni qayta ishlash hamda semantik tahlil jarayonlarini murakkablashtiradi. Uchinchidan, mavjud ko‘p tilli (multilingual) yirik til modellarining o‘zbek tilidagi ishlash samaradorligi resurslarga boy tillarga nisbatan past bo‘lib, bu savol-javob tizimlarining aniqligi va ishonchliligiga salbiy ta’sir ko‘rsatadi. Shu bois o‘zbek tili uchun maxsus til modellari, sifatli korpuslar va lingvistik resurslarni yaratish, shuningdek, tilning morfologik va semantik xususiyatlarini hisobga oluvchi sun’iy intellekt algoritmlarini ishlab chiqish mazkur yo‘nalishdagi ustuvor ilmiy vazifalardan biri hisoblanadi.

XULOSA

O‘zbek tili uchun samarali savol-javob tizimlarini yaratish, avvalo, yuqori sifatli annotatsiyalangan savol-javob korpuslari va bilim bazalarini shakllantirishni talab etadi. Shuningdek, tilning agglutinativ xususiyatlarini hisobga oluvchi morfologik va semantik tahlil usullarini takomillashtirish hamda o‘zbek tiliga mos yirik til modellarini ishlab chiqish muhim ahamiyat kasb etadi. Mazkur yo‘nalishdagi ilmiy va amaliy ishlarning rivojlantirilishi o‘zbek tilida yuqori aniqlikdagi savol-javob tizimlarini yaratish, ularning ta’lim, davlat xizmatlari va boshqa sohalarda samarali qo‘llanilishiga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR:

1. Tukeyev U., Rysbek B., Hnatkowska B. Question–Answer Models for Teaching Programming in Kazakh: A Morphology-Aware Controlled Study. Applied Sciences. 2026, 16, 6156.
2. Bi W., Xiong Q., Chen X., Du Q., Wu J., Zhuang Z. TCM-VQA IoTNet: An intelligent visual question answering system for traditional Chinese medicine education based on multimodal learning and IoT // *Alexandria Engineering Journal*. – Egypt, 2025. – Vol. 118. – P. 325-336.
3. Mukhamadiyev A., Mukhiddinov M., Khujayarov I., Ochilov M., Cho J. Development of Language Models for Continuous Uzbek Speech Recognition System. Sensors. 2023, 23, 1145.
4. Khujayarov I., Ochilov M., Abdullayeva M., Kholmatov O., Nasimov R. Development of Uzbek Question-Answering Models through Fine-Tuning Approaches. ICFNDS'25. Dubai, 2025.