

HUQUQIY DISKURS ASOSIDA O‘ZBEK TILIDA SAVOL-JAVOB TIZIMLARI UCHUN LINGVISTIK MODEL YARATISH

Hamroyeva Shahlo Mirdjonovna

Filologiya fanlari doktori (DSc), prof.v.b.,

shahlo.xamrayeva@navoiy-uni.uz

ToshDO‘TAU

Umarova Nozimaxon Qobil qizi

Kompyuter lingvistikasi mutaxassisligi

1-bosqich magistranti

nozimaumarova333@gmail.com

ToshDO‘TAU

Annotatsiya. Mazkur tezisda o‘zbek tilidagi huquqiy diskursning lingvistik xususiyatlari tahlil qilinib, shu asosda savol-javob (Question Answering – QA) tizimlari uchun lingvistik model taklif etiladi. Huquqiy matnlarning normativligi, terminologik aniqligi va sintaktik murakkabligi ularni avtomatik qayta ishlashda muhim omil hisoblanadi. O‘zbek tilining agglyutinativ tabiati esa savol va javob o‘rtasidagi moslikni aniqlash jarayonini yanada murakkablashtiradi. Tadqiqot natijasida huquqiy savol-javob tizimlari uchun model ishlab chiqildi. Ushbu model o‘zbek tilida huquqiy QA tizimlarini yaratishda nazariy asos sifatida xizmat qiladi.

Kalit so‘zlar: *huquqiy diskurs, savol-javob tizimi, lingvistik model, agglyutinativ til, semantik moslik, NLP.*

Abstract. This thesis analyzes the linguistic features of legal discourse in the Uzbek language and, on this basis, proposes a linguistic model for question-answering (QA) systems. The normativity, terminological precision, and syntactic complexity of legal texts are important factors in their automatic processing. The agglutinative nature of the Uzbek language further complicates the process of identifying the correspondence between questions and answers. As a result of the study, the issue of developing a model for legal question-answering systems is discussed in detail. This model serves as a theoretical basis for creating legal QA systems in the Uzbek language.

Keywords: *legal discourse, question-answering system, linguistic model, agglutinative language, semantic similarity, NLP.*

Kirish. So‘nggi yillarda tabiiy tilni qayta ishlash (Natural Language Processing, NLP) sohasida inson tilini kompyuter orqali tushunish, qayta ishlash va unga mos javob qaytarish masalalari dolzarb yo‘nalishlardan biriga aylandi. Xususan, savol-javob tizimlari foydalanuvchi tomonidan tabiiy tilda berilgan savollarni tahlil qilish, kerakli axborotni aniqlash va tegishli manbalar asosida aniq javob shakllantirish imkonini beradi. Bunday tizimlar ta‘lim, tibbiyot, davlat boshqaruvi, moliya va huquq kabi sohalarda amaliy ahamiyat kasb etmoqda. Ayniqsa, huquqiy sohada avtomatlashtirilgan savol-javob tizimlariga ehtiyoj tobora ortib bormoqda. Chunki qonun hujjatlari, kodekslar, qarorlar va boshqa normativ-huquqiy matnlar hajman katta, mazmunan murakkab va terminologik jihatdan aniq tuzilmaga ega. Oddiy foydalanuvchi ko‘pincha kerakli huquqiy normani mustaqil topish, uni to‘g‘ri talqin qilish va amaliy vaziyatga moslashtirishda qiyinchilikka duch keladi. Shu nuqtayi nazardan, huquqiy savol-javob tizimlari foydalanuvchining tabiiy tilda berilgan savoliga huquqiy hujjatlar asosida aniq, ishonchli va manbaga tayangan javob qaytarishni maqsad qiladi [6:9-11]. Biroq huquqiy matnlar bilan ishlash oddiy axborot-qidiruv tizimlaridan farq qiladi. Huquqiy diskurs normativlik, rasmiylik, terminologik qat‘iylilik, mantiqiy izchillik va sintaktik murakkablik kabi xususiyatlarga ega. Unda har bir so‘z, qo‘shimcha yoki grammatik birlik ma‘no aniqligiga ta‘sir ko‘rsatishi mumkin. Ayniqsa, o‘zbek tilining agglutinatativ tabiati, ya‘ni grammatik ma‘nolarning asosan qo‘shimchalar orqali ifodalanishi savol va javob o‘rtasidagi lingvistik moslikni aniqlashda qo‘shimcha murakkablik tug‘diradi.

O‘zbek tilida huquqiy savol-javob tizimlarini yaratish masalasi hali yetarli darajada nazariy va amaliy jihatdan tadqiq etilmagan. Mavjud yondashuvlarda ko‘pincha texnologik komponentlarga e‘tibor qaratiladi, biroq savolni lingvistik tahlil qilish, huquqiy terminlarni aniqlash, savol turini belgilash, matndagi semantik

moslikni topish kabi masalalar alohida model sifatida yetarlicha yoritilmaydi. Holbuki, huquqiy QA tizimining samaradorligi bevosita uning lingvistik asoslari qanchalik puxta ishlab chiqilganiga bog'liq.

Shu sababli mazkur ishning maqsadi – huquqiy diskursning o'ziga xos xususiyatlarini tahlil qilish hamda ular asosida o'zbek tilida savol-javob tizimlari uchun lingvistik model ishlab chiqishdan iborat. Huquqiy NLP bo'yicha so'nggi tadqiqotlarda legal question answering, legal text classification, legal summarization va legal information extraction kabi vazifalar alohida yo'nalish sifatida ajratiladi [1:7-8]. Tadqiqotda huquqiy matnlarning terminologik, morfologik, sintaktik va semantik jihatlarini ko'rib chiqilib, savolni tahlil qilishdan javobni shakllantirishgacha bo'lgan jarayonda lingvistik birliklarning o'rni aniqlanadi. Mazkur yondashuv kelgusida o'zbek tilidagi huquqiy axborot tizimlari, elektron maslahat xizmatlari va sun'iy intellektga asoslangan huquqiy yordamchi dasturlarni yaratish uchun nazariy asos bo'lib xizmat qilishi mumkin.

Asosiy qism. Huquqiy diskurs rasmiy uslubning eng murakkab va mas'uliyatli ko'rinishlaridan biri hisoblanadi. Chunki huquqiy matnlar uzun hujjatlar, murakkab sintaksis, qat'iy terminologiya va maxsus diskurs xususiyatlari bilan ajralib turadi [1:9]. Shu sababli huquqiy diskursda har bir so'z, atama va grammatik birlik ma'no aniqligini ta'minlashda muhim o'rin tutadi. Huquqiy til oddiy foydalanuvchi uchun qiyin tushuniladigan maxsus til ko'rinishi bo'lib, undagi terminologiya va normativlik avtomatik qayta ishlashda alohida muammo tug'diradi [8:219-5221]. Avvalo, huquqiy diskursning asosiy belgilaridan biri normativlikdir. Ya'ni huquqiy matnlar jamiyatdagi muayyan munosabatlarni tartibga soluvchi qonun va qoidalarni ifodalaydi. Bunday matnlarda shaxslarning huquqlari, majburiyatlari, javobgarligi hamda muayyan harakatlarni bajarish yoki bajarmaslik shartlari aniq belgilab beriladi. Shuning uchun huquqiy diskursda noaniqlik, ikki xil talqin yoki hissiy ifoda imkon qadar cheklanadi.

Huquqiy diskursning yana bir muhim xususiyati terminologik aniqlik bilan bog'liq. Huquqiy matnlarda qo'llanadigan atamalar odatiy nutqdagi so'zlardan farqli ravishda maxsus ma'noga ega bo'ladi. Masalan, “mehnat shartnomasi”, “ish beruvchi”, “xodim”, “majburiyat”, “javobgarlik”, “modda”, “band” kabi birliklar ma'lum huquqiy tushunchani ifodalaydi. Bunday terminlar matnda izchil va bir xil ma'noda qo'llanishi zarur, chunki terminning noto'g'ri talqin qilinishi huquqiy mazmunning buzilishiga olib kelishi mumkin. Shuningdek, huquqiy diskurs sintaktik jihatdan ham murakkab tuzilishga ega. Qonun hujjatlarida ko'pincha uzun, bir nechta shart, izoh va aniqlovchi qismlardan tashkil topgan gaplar uchraydi. Bunday gaplar huquqiy normani to'liq, cheklovlari va qo'llanish shartlari bilan birga ifodalashga xizmat qiladi. Shu sababli huquqiy matnlarni avtomatik qayta ishlash jarayonida faqat alohida so'zlarni emas, balki gapning umumiy tuzilishi, tarkibiy qismlari va ular o'rtasidagi mantiqiy bog'lanishlarni ham hisobga olish zarur.

Bundan tashqari, huquqiy diskursda standartlashgan ifodalar keng qo'llanadi. Qonun va normativ hujjatlarda “belgilangan tartibda”, “qonunchilikka muvofiq”, “ushbu moddaning birinchi qismida nazarda tutilgan”, “mehnat shartnomasiga asosan” kabi tayyor konstruksiyalar tez-tez uchraydi. Bunday ifodalar huquqiy matnlarning rasmiyligini, izchilligini va aniqligini ta'minlaydi. Shu bilan birga, ular savol-javob tizimlari uchun muhim lingvistik signal vazifasini bajarishi mumkin.

Demak, huquqiy diskurs normativlik, terminologik aniqlik, sintaktik murakkablik va standartlashgan ifodalar bilan ajralib turadi. Mazkur xususiyatlar huquqiy matnlarni oddiy matnlardan farqlaydi va ularni avtomatik tahlil qilishda maxsus lingvistik yondashuv zarurligini ko'rsatadi. Ayniqsa, savol-javob tizimlarida foydalanuvchi savolini huquqiy matndagi tegishli norma bilan bog'lash uchun ushbu lingvistik belgilarni aniqlash, tahlil qilish va modelga kiritish muhim ahamiyat kasb etadi.

O'zbek tili agglyutinatив tillar qatoriga kiradi. Bu til tizimida grammatik ma'nolar, asosan, so'z negiziga turli qo'shimchalar qo'shilishi orqali ifodalanadi. Natijada birgina leksik birlik turli grammatik shakllarda namoyon bo'lishi mumkin. Masalan, ish so'zidan ishchi, ishchilar, ishchilarning, ishchilarga, ishchilardan kabi ko'plab shakllar hosil bo'ladi. Bunday holat tabiiy tilni qayta ishlash jarayonida, ayniqsa, savol-javob tizimlarida muhim lingvistik muammo sifatida namoyon bo'ladi. O'zbek tili NLP nuqtayi nazaridan kam resursli va morfologik jihatdan murakkab tillardan biri hisoblanadi. So'nggi yillarda o'zbek tili uchun BERTbek va UzBERT kabi transformer modellar ishlab chiqilgan bo'lsa-da, lemmatizatsiya, stemming, semantik tahlil va annotatsiyalangan korpuslarning yetishmasligi hanz dolzarb masalalardan biri bo'lib qolmoqda [2:6-8], [4:15-17], [7:12]. Ayniqsa, huquqiy matnlar bilan ishlashda bu muammo yanada murakkablashadi, chunki huquqiy terminlarning qo'shimchalar orqali o'zgarishi savol va matn o'rtasidagi yuzaki moslikni kamaytiradi. Masalan, foydalanuvchi “ish beruvchi qanday majburiyatlarga ega?” degan savolni bersa, huquqiy matnda bu mazmun “ish beruvchining majburiyatlari”, “ish beruvchi zimmasiga yuklatiladi” yoki “ish beruvchi majburdir” kabi shakllarda ifodalanishi mumkin. Bunday holatda tizim faqat aynan bir xil so'z shakliga tayanadigan bo'lsa, to'g'ri javobni topa olmasligi mumkin. Shu sababli o'zbek tilidagi huquqiy savol-javob tizimlarida morfologik normalizatsiya, lemmatizatsiya va semantik moslikni aniqlash bosqichlari alohida ahamiyat kasb etadi. Demak, o'zbek tilining agglyutinatив tabiati savol va huquqiy matn o'rtasidagi bevosita leksik moslikni murakkablashtiradi. Bu esa QA tizimlarida faqat kalit so'zga asoslangan qidiruv yetarli emasligini, balki so'zning asosiy shakli, grammatik vazifasi va semantik bog'liqligini hisobga oluvchi lingvistik model zarurligini ko'rsatadi.

Agglyutinatив tillar uchun POS teglash va stemming masalalari o'zbek tilidagi huquqiy savol-javob tizimi uchun ham bevosita ahamiyatga ega. Tadqiqotlarda

agglyutinativ tillarda soʻz shakllari nazariy jihatdan juda koʻp boʻlishi, bu esa lugʻatdan tashqari soʻzlar va morfologik variantlarni aniqlash muammosini yuzaga keltirishi taʼkidlanadi [2:6-8]. Huquqiy matnlarda bu holat yanada sezilarli koʻrinadi: “shartnoma”, “shartnomaning”, “shartnomaga”, “shartnomadan” kabi shakllar bir xil huquqiy tushunchaga tegishli boʻlsa-da, tizim ularni alohida birlik sifatida qabul qilsa, javobni topish aniqligi pasayadi. Shu sababli lingvistik model tarkibida morfologik normalizatsiya, oʻzakni aniqlash, termin variantlarini yagona kanonik shaklga keltirish va savol bilan matn oʻrtasidagi grammatik moslikni tekshirish alohida bosqich sifatida ajratilishi lozim.

Huquqiy savol-javob tizimlari foydalanuvchi tomonidan tabiiy tilda berilgan savolni tahlil qilish, tegishli huquqiy manbani aniqlash va javobni manbaga tayangan holda shakllantirishga xizmat qiladi. Huquqiy QA tizimlari boʻyicha tadqiqotlarda savolni lingvistik tahlil qilish, mos huquqiy segmentni topish hamda javobni ishonchli manba asosida taqdim etish asosiy bosqichlar sifatida qaraladi [6:23], [1:15]. Ayniqsa, kam resursli va morfologik jihatdan murakkab tillar uchun korpus yaratish, savol-javob juftliklarini shakllantirish va semantik moslikni aniqlash muhim vazifalardan biri hisoblanadi [7:5].

Mazkur tadqiqot doirasida oʻzbek tilidagi huquqiy savol-javob tizimlari uchun quyidagi lingvistik model taklif etiladi:

Savol → lingvistik tahlil → semantik moslik → javob

Ushbu modelning birinchi bosqichi savolni tahlil qilishdan iborat. Bu jarayonda savolning turi aniqlanadi. Masalan, faktual savollar aniq maʼlumotni talab qiladi, normativ savollar muayyan huquqiy qoida yoki majburiyatni aniqlashga qaratiladi, izohli savollar esa huquqiy tushunchaning mazmunini tushuntirishni talab qiladi. Shu bilan birga, savol tarkibidagi asosiy huquqiy tushunchalar, yaʼni entity birliklar ajratib olinadi. Masalan, xodim, ish beruvchi, mehnat shartnomasi, ish

vaqti, ta'til, majburiyat kabi birliklar huquqiy javobni topishda asosiy signal vazifasini bajaradi.

Ikkinchi bosqich lingvistik qayta ishlashni o'z ichiga oladi. Bu bosqichda savol va huquqiy matn tarkibidagi so'zlar lemmatizatsiya qilinadi, morfologik jihatdan tahlil qilinadi hamda gapning sintaktik tuzilishi aniqlanadi. Ayniqsa, o'zbek tilida qo'shimchalar orqali hosil bo'lgan grammatik shakllarni umumiy asosiy shaklga keltirish savol va javob o'rtasidagi moslikni kuchaytiradi.

Lingvistik qayta ishlash bosqichida tokenizatsiya, lemmatizatsiya, POS teglash va sintaktik bog'lanishlarni aniqlash kabi amallar bir-biri bilan bog'liq holda bajariladi. NLP vositalari yordamida matndan token, lemma, POS, dependency va entity kabi atributlarni ajratish mumkinligi ko'rsatilgan [3:41-43]. Huquqiy QA tizimi uchun bu atributlar savolning markaziy terminini, so'ralayotgan huquqiy munosabatni va javob olinishi kerak bo'lgan normativ segmentni belgilashga xizmat qiladi. Masalan, “xodimning yillik ta'til olish huquqi” birikmasida “xodim” subyekt, “ta'til” huquqiy obyekt, “olish huquqi” esa normativ munosabat sifatida ajratiladi.

Uchinchi bosqichda savol va huquqiy matn o'rtasidagi semantik moslik aniqlanadi. Bu jarayonda tizim faqat bir xil so'zlarning uchrashiga emas, balki savol mazmuni bilan huquqiy norma mazmuni o'rtasidagi ma'no yaqinligiga e'tibor qaratadi. Masalan, “xodimning dam olish huquqi” haqidagi savol “ta'til berish tartibi” yoki “ish vaqtidan tashqari dam olish” haqidagi normativ matnlar bilan semantik jihatdan bog'liq bo'lishi mumkin.

To'rtinchi bosqich javobni shakllantirishdan iborat. Bunda eng mos normativ matn segmenti tanlanadi, uning manbasi aniqlanadi va foydalanuvchiga tushunarli shaklda taqdim etiladi. Huquqiy savol-javob tizimida javob faqat umumiy izohdan iborat bo'lmasligi, balki tegishli huquqiy hujjat, modda yoki norma bilan bog'langan bo'lishi lozim.

Shunday qilib, taklif etilayotgan model huquqiy savol-javob tizimlarida lingvistik komponentning markaziy rolini ko'rsatadi. O'zbek tilining agglyutinatив xususiyati, huquqiy diskursning terminologik aniqligi va normativ matnlarning sintaktik murakkabligi bunday tizimlarda maxsus lingvistik yondashuv zarurligini asoslaydi.

Xulosa. Tadqiqot natijalari shuni ko'rsatadiki, huquqiy diskursning o'ziga xos lingvistik xususiyatlari savol-javob tizimlarini yaratishda hal qiluvchi ahamiyatga ega. O'zbek tilining agglyutinatив tabiati esa lingvistik modellashtirishni yanada murakkablashtiradi. Taklif etilgan model huquqiy matnlar bilan ishlovchi QA tizimlari uchun nazariy asos bo'lib xizmat qiladi va kelgusida texnologik yechimlar ishlab chiqishda foydalanilishi mumkin.

Foydalanilgan adabiyotlar ro'yxati

1. Ariai F., Demartini G. Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges // ACM Computing Surveys. – 2025.
2. Elov B.B., Hamroyeva Sh.M., Abdullayeva O.X., Husainova Z.Y., Xudayberganov N.U. Agglyutinatив tillar uchun POS teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida) // O'zbekiston: til va madaniyat. – 2023. – № 2.
3. Elov B.B. Tabiiy tilni qayta ishlash (NLP)da spaCy modulidan foydalanish // Ilm-fan va innovatsion rivojlanish. – 2022. – № 4.
4. Kuriyozov E., Vilares D., Gómez-Rodríguez C. BERTbek: A Pretrained Language Model for Uzbek // Proceedings of SIGUL / LREC-COLING. – 2024.
5. Mansurov B., Mansurov A. UzBERT: Pretraining a BERT Model for Uzbek // arXiv preprint. – 2021. – arXiv:2108.09814.
6. Martinez-Gil J. A survey on legal question-answering systems // Computer Science Review. – 2023. – Vol. 48. – Article 100552.



7. Rakhimova D., Turarbek A., Karyukin V., Sarsenbayeva A., Alieyev R. Legal AI in Low-Resource Languages: Building and Evaluating QA Systems for the Kazakh Legislation // Computers. – 2025. – Vol. 14. – No. 9. – Article 354.
8. Zhong H., Xiao C., Tu C., Zhang T., Liu Z., Sun M. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020.