

O'ZBEK TILIDAGI RASMIY HUJJATLAR ASOSIDA SAVOL-JAVOB TIZIMINI YARATISHDA RAG YONDASHUVINI QO'LLASH VA BAHOLASH METODOLOGIYASI

Muallif: **Qurtov M.X.** | Tashkilot: uzinfocom | E-pochta:

kurtov.mansur@gmail.com

Maqola turi: ilmiy-amaliy konferensiya uchun metodologik maqola

Annotatsiya: *Katta til modellari (LLM) hujjatlar bilan tabiiy tilda muloqot qilish imkonini beradi, biroq ularning ichki bilimlariga tayangan javoblari eskirgan, noaniq yoki manba bilan tasdiqlanmagan bo'lishi mumkin. Retrieval-Augmented Generation (RAG) yondashuvi javob yaratishdan oldin tashqi hujjatlar to'plamidan tegishli parchalarni qidirib, javobni shu dalillarga tayantirishga imkon beradi. Ushbu maqolada o'zbek tilidagi rasmiy hujjatlar asosida savol-javob tizimini yaratish va baholash metodologiyasi taklif etiladi. Yondashuv hujjatlarni tayyorlash, matnni bo'laklash, leksik va semantik qidiruv, manbaga asoslangan javob yaratish hamda natijalarni qidiruv sifati, javob sifati va manbaga sodiqlik bo'yicha alohida baholashni qamrab oladi. Qidiruv uchun Recall@k, MRR va nDCG; javob uchun to'g'rilik va dolzarblik; manbaga tayantirish uchun esa da'volarning keltirilgan parchalar bilan tasdiqlanishi tavsiya etiladi. Maqola takrorlanuvchi tajriba tashkil etishga yordam beradigan metodologik taklifdir; unda bajarilmagan tajribalarning sonli natijalari da'vo qilinmaydi.*

Kalit so'zlar: *sun'iy intellekt, katta til modellari, RAG, o'zbek tili, savol-javob tizimi, hujjat qidiruvi, manbaga sodiqlik.*

1. Kirish

Tashkilotlarda to'planadigan me'yoriy hujjatlar, yo'riqnomalar, texnik tavsiflar va ichki reglamentlardan kerakli ma'lumotni tez topish ko'p vaqt talab qilishi mumkin. Oddiy kalit so'zli qidiruv foydalanuvchi savolni boshqa so'zlar bilan ifodalaganida kerakli parchani topmasligi mumkin. Katta til modellari savolni tabiiy tilda tushunib, javobni ravon shakllantiradi, biroq tashqi manbaga ulanmagan

model mavjud bo'lmagan ma'lumotni ishonchli ohangda bayon qilishi ehtimoli bor.

Lewis va hammualliflar taklif qilgan RAG yondashuvi generativ modelni tashqi, yangilanishi mumkin bo'lgan ma'lumotlar omboridan olingan dalillar bilan birlashtiradi [1]. Tizim avval savolga tegishli matn parchalarini qidiradi, keyin ularni modelga kontekst sifatida uzatadi. Biroq qidiruv noto'g'ri bo'lsa yoki model dalildan chetga chiqsa, RAG tizimi ham xato javob berishi mumkin.

O'zbek tilidagi rasmiy hujjatlarda kirill va lotin yozuvi, imlo variantlari, qisqartmalar, huquqiy atamalar va turli fayl formatlari qidiruv sifatiga ta'sir qilishi mumkin. Shuning uchun tizimni faqat javobning ravonligi bilan baholash yetarli emas: qidiruv, javobning to'g'riligi va manbaga moslik alohida o'lchanishi lozim.

Maqolaning maqsadi — o'zbek tilidagi rasmiy hujjatlar uchun RAG asosidagi savol-javob tizimini loyihalash va uni takrorlanuvchi tarzda baholash metodologiyasini taklif etish. Tadqiqot savollari: (1) leksik va semantik qidiruv tegishli hujjatlarni topishda qanday taqqoslanadi? (2) javob sifati qidiruv parametrlariga qay darajada bog'liq? (3) manba bilan tasdiqlangan javoblarni qanday mezonlar bilan tekshirish mumkin?

2. Mavjud ishlar tahlili

RAG tashqi hujjatlar bazasidan olingan ma'lumotlarni generativ modelga uzatib, bilim talab qiladigan vazifalarda javobni dalillarga tayantirishni ko'zlaydi [1]. Uning ikki asosiy moduli — qidiruv va matn yaratish. Qidiruv tegishli parchalarni topadi; generativ model esa shu kontekst asosida javob beradi.

BEIR benchmarki axborot qidiruv modellarini turli ma'lumotlar to'plami va vazifalarda baholash zarurligini ko'rsatadi [2]. Bu tamoyil o'zbekcha hujjatlar uchun ham muhim: bitta turdagi hujjatda yaxshi ishlagan qidiruv usuli boshqa savol yoki hujjatda shunday natija bermasligi mumkin.

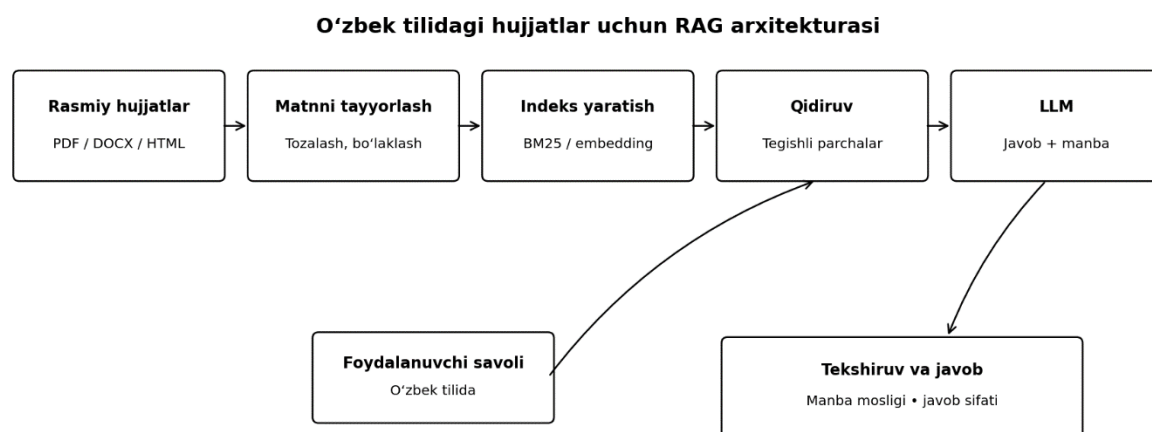
RAGAs tizimi qidirilgan kontekstning dolzarbligi, javobning kontekstga sodiqligi va javobning savolga mosligi kabi o'lchamlarni ajratadi [3]. Bu qidiruv va generatsiya bosqichlaridagi xatolarni alohida ko'rishga yordam beradi. Ushbu maqola mazkur tamoyillarni o'zbek tilidagi rasmiy hujjatlar uchun amaliy baholash

protokoliga moslashtiradi. Yangilik haqida qat'iy xulosa qilishdan oldin o'zbekcha RAG tizimlari bo'yicha qo'shimcha adabiyot qidiruvi zarur.

3. Materiallar va metodlar

3.1. Tizim arxitekturası

Taklif etilayotgan tizim besh bosqichdan iborat: hujjatlarni yig'ish; matnni ajratib olish va tozalash; mazmuniy bo'laklarga ajratish; leksik yoki semantik qidiruv; topilgan parchalar asosida javob yaratish va manbalarni ko'rsatish. Tizim faqat foydalanishga ruxsat etilgan hujjatlar bilan ishlashi, hujjat versiyasi va sanasini saqlashi kerak.



Taklif etilgan sxema; amaliy natija yoki ishga tushirilgan tizim skrinshoti emas.

1-rasm. O'zbek tilidagi rasmiy hujjatlar uchun RAG tizimining taklif etilayotgan arxitekturası.

3.2. Hujjatlarni tayyorlash

Ma'lumotlar to'plamiga ochiq va foydalanish shartlari ruxsat beradigan rasmiy hujjatlar kiritiladi. Har bir hujjat uchun manba URL manzili, sarlavha, sana yoki versiya, hujjat turi va olingan vaqt qayd etiladi. PDF, DOCX va HTML fayllardan matn ajratiladi; sahifa raqamlari, bo'lim nomlari va band raqamlari imkon qadar saqlanadi. Skanerlangan sahifalardan OCR orqali olingan matn alohida tekshiriladi.

Faqat belgilar soniga asoslangan bo'laklash huquqiy band yoki texnik ro'yxatni o'rtasidan uzib qo'yishi mumkin. Shu sababli sarlavha, band, paragraf va

ro'yxat chegaralarini saqlaydigan usulni oddiy token chegarasiga asoslangan usul bilan taqqoslash tavsiya etiladi. Bo'lak uzunligi va qo'shni bo'laklar orasidagi takrorlanish tajriba parametrlari sifatida qayd etiladi.

3.3. Qidiruv usullari

Boshlang'ich taqqoslash uchun BM25 kabi leksik qidiruv va embeddinglarga asoslangan semantik qidiruvdan foydalanish mumkin. Leksik qidiruv aniq atama, raqam yoki hujjat bandi so'ralganda foydali; semantik qidiruv esa savol va hujjatda bir xil so'z ishlatilmaganida yordam berishi mumkin. Ularning qaysi biri tanlangan hujjatlar to'plamida samaraliroq ekanini tajriba ko'rsatishi kerak. Qo'shimcha variant sifatida ikki usul natijalarini birlashtiruvchi gibril qidiruv ham tekshiriladi.

Topilgan har bir parcha uchun hujjat identifikatori, bo'lim yoki sahifa raqami va matn saqlanadi. Tizim manbani ochish imkonini berishi kerak. Agar ishonchli dalil topilmasa, taxminiy javobni fakt sifatida ko'rsatish o'rniga yetarli ma'lumot topilmaganini bildirishi lozim.

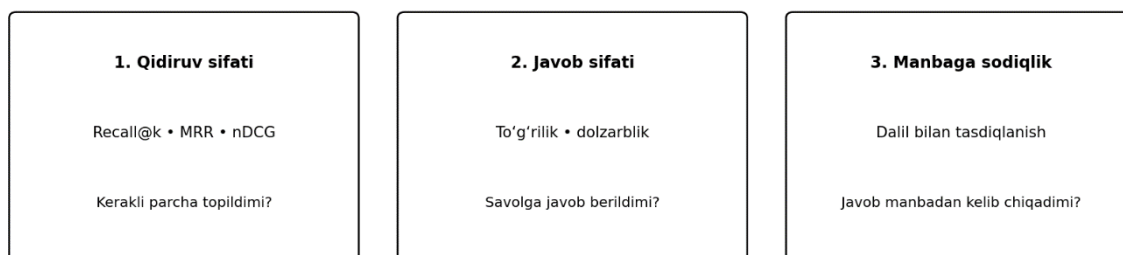
3.4. Javob yaratish va manbaga tayantirish

LLMga beriladigan ko'rsatmada faqat taqdim etilgan parchalar asosida javob berish, muhim da'volarni tegishli manbaga bog'lash va kontekstda javob bo'lmasa buni ochiq aytish talabi bo'ladi. Havolalar haqiqiy hujjat va tegishli bo'limga olib borishi tekshiriladi. Manba havolasi mavjudligi javob to'g'ri degani emas: keltirilgan parcha javobni haqiqatan tasdiqlashi zarur.

3.5. Baholash dizayni

Savol-javoblar to'plami aniq faktni topish, hujjat bandini aniqlash, ikki bandni solishtirish, bir nechta parchadan xulosa chiqarish va hujjatda javobi mavjud bo'lmagan savollarni qamrab oladi. Har bir savol uchun tegishli hujjat yoki parchalar, etalon javob va baholash mezonini ekspert tomonidan tekshiriladi. Yakuniy test savollari tizim sozlamalarini moslashtirish uchun ishlatilmasligi kerak.

RAG tizimini uch qatlamda baholash



Yagona umumiy ball turli xatolarni yashirishi mumkin; qatlamlar alohida hisobot qilinadi.

2-rasm. RAG tizimini qidiruv, javob sifati va manbaga sodiqlik bo'yicha alohida baholash.

3.6. Ko'rsatkichlar

Qidiruv sifati Recall@k, Mean Reciprocal Rank (MRR) va nDCG@k orqali o'lchanishi mumkin. Recall@k kerakli hujjat yoki parcha dastlabki k natija ichida topilgan holatlar ulushini ko'rsatadi. MRR birinchi tegishli natijaning o'rnini hisobga oladi. nDCG natijalarning tartibi va tegishlilik darajasini baholaydi.

Javob sifati etalon javob bilan solishtirish va ekspert bahosi orqali tekshiriladi. Manbaga sodiqlik javobdagi faktik da'volarning keltirilgan parchalar bilan tasdiqlanishi bo'yicha baholanadi. Dolzarblik savolga to'g'ridan-to'g'ri javob berilganini, to'g'rilik esa mazmunning etalon ma'lumotga mosligini o'lchaydi. Bu tushunchalarni bitta ko'rsatkichga birlashtirib yubormaslik muhim.

LLM-as-a-judge usulidan foydalanilsa, baholovchi model, ko'rsatmalar va mezonlar ochiq ko'rsatiladi. Avtomatik baholash natijalarining bir qismini inson ekspertlari tekshirishi lozim, chunki baholovchi model ham xato qilishi mumkin.

3.7. Takrorlanuvchanlik va axborot xavfsizligi

Tajriba natijalarini takrorlash uchun hujjatlar ro'yxati va versiyalari, matnni bo'laklash parametrlari, embedding modeli, qidiruv konfiguratsiyasi, LLM identifikatori, so'rov matni va baholash skriptlari qayd etiladi. Maxfiy yoki shaxsiy ma'lumotlar mavjud bo'lsa, ularni tashqi xizmatga yuborishdan oldin huquqiy va axborot xavfsizligi talablari tekshiriladi. Amaldagi va bekor qilingan hujjat versiyalari aniq farqlanishi kerak.

4. Natijalar va ularni taqdim etish rejasi

Ushbu maqola metodologik ish bo'lib, tizimning amaliy tajribalari bajarilmagan. Shu sababli sonli natijalar yoki bir usulning boshqasidan ustunligi haqida da'vo keltirilmaydi. Keyingi empirik bosqichda kamida uch konfiguratsiya — BM25, semantik qidiruv va gibrid qidiruv — bir xil savollar to'plamida taqqoslanishi, so'ng javob yaratish sifati va manbaga sodiqlik alohida baholanishi tavsiya etiladi.

Konfiguratsiya	Qidiruv ko'rsatkichlari	Javob sifati	Manbaga sodiqlik
BM25 (leksik)	Recall@k, MRR, nDCG	Baholanadi	Baholanadi
Embedding (semantik)	Recall@k, MRR, nDCG	Baholanadi	Baholanadi
Gibrid qidiruv + RAG	Recall@k, MRR, nDCG	Baholanadi	Baholanadi

1-jadval. Tajriba konfiguratsiyalari va baholash rejasi; bu natijalar jadvali emas.

Natijalar bo'limida ma'lumotlar to'plamining hajmi va tarkibi, har bir qidiruv usulining ko'rsatkichlari, savol turlari bo'yicha xatolar va manba bilan tasdiqlanmagan javoblar ulushi berilishi kerak. Xatolar tahlilida muammo hujjat topilmagani, noto'g'ri parcha topilgani, javob manbadan chetga chiqqani yoki hujjat eskirganligi bilan bog'liqligi ajratiladi.

5. Muhokama

Taklif etilgan metodologiya RAG tizimini faqat yakuniy javobning ravonligi bilan baholashdan ko'ra kengroq yondashuvni qo'llaydi. Qidiruv xatosi va javob yaratish xatosini ajratish takomillashtirish yo'nalishini aniqroq belgilaydi. Agar kerakli parcha topilmagan bo'lsa, faqat promptni o'zgartirish muammoni hal qilmasligi mumkin; qidiruv yoki hujjatni tayyorlash jarayonini yaxshilash talab etiladi.

RAG manbaga sodiqlikni avtomatik kafolatlamaydi. Hujjat noto'g'ri yoki eskirgan bo'lsa, tizim shu ma'lumotni manbaga mos ravishda qaytarishi mumkin. Shu sababli hujjat versiyasini boshqarish, rasmiy manbani ko'rsatish, dalil yetarli

bo'lmaganda javob berishdan tiyilish va zarur holatlarda inson nazoratini saqlash muhim.

Cheklovlar qatoriga hujjatlar to'plamining hajmi, kirill-lotin yozuvi farqlari, OCR xatolari, sohasiy terminlar va ekspert bahosining subyektivligi kiradi. Natijalarni boshqa tillar yoki tashkilotlarga umumlashtirishdan oldin qo'shimcha sinovlar o'tkazilishi kerak.

6. Xulosa

Maqolada o'zbek tilidagi rasmiy hujjatlar asosida savol-javob tizimini yaratish uchun RAG yondashuvini qo'llash va baholash metodologiyasi taklif etildi. Tizim hujjatlarni tayyorlash, mazmuniy bo'laklarga ajratish, leksik va semantik qidiruv, manbaga tayangan javob yaratish hamda natijalarni alohida qatlamlarda baholash bosqichlarini qamrab oladi. Qidiruv uchun Recall@k, MRR va nDCG; javob uchun to'g'rilik va dolzarblik; manbaga tayantirish uchun esa da'volarning keltirilgan parchalar bilan tasdiqlanishi tavsiya etildi. Metodologiyaning amaliy samaradorligi hali tajribada tekshirilmagan. Keyingi ish sifatida ochiq rasmiy hujjatlar asosida savol-javoblar to'plamini tuzish, qidiruv konfiguratsiyalarini solishtirish va natijalarni ekspertlar bilan tekshirish rejalashtiriladi.

Foydalanilgan adabiyotlar

- [1] Lewis, P. va boshq. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33. <https://arxiv.org/abs/2005.11401>
- [2] Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. *NeurIPS 2021 Datasets and Benchmarks Track*. <https://openreview.net/forum?id=wCu6T5xFjeJ>
- [3] Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). RAGAs: Automated Evaluation of Retrieval Augmented Generation. *Proceedings of EACL 2024: System Demonstrations*, 150–158. <https://doi.org/10.18653/v1/2024.eacl-demo.16>