ELEKTRON KORPUSDA QIDIRUV TIZIMINING IMKONIYATLARI

Rustamova Zulfiya Xolmirzayevna
Nizomiy nomidagi Ozbekiston milliy pedagogika universiteti
Boshlang‘ich ta’lim nazariyasi va metodikasi kafedrasi dotsenti

Annotasiya: Raqamli texnologiyalarning paydo bo‘lishi bizning ma’lumotni izlash, olish va ulardan foydalanish usulini inqilob qildi. Raqamli ma’lumotlarning ulkan ombori bo‘lgan elektron korpus turli fanlar bo‘yicha olimlar, tadqiqotchilar va mutaxassislar uchun ajralmas vositaga aylandi. Ushbu maqola elektron korpusdagi qidiruv tizimining imkoniyatlarini o‘rganib chiqadi, uning asosiy xususiyatlari, kuchli tomonlari va bilimlarni kashf qilish va ulardan foydalanish uchun ta’sirini ta’kidlaydi.

Kalit so‘zlar: CORP, qidiruv tizimlari, interfeys, ma’lumotlar bazasi, mexanizmlari.

Jarayon CORP formatidagi matnlarni izohli korpusdan ajratib olish va uni qidiruv tizimiga biriktirish mexanizmini ishlab chiqishdan iborat. Ushbu mexanizmni optimallashtirish va uni foydalanuvchi interfeysi bilan integratsiya qilish jarayonida Intellect lingvokognitiv-semantik platformasining bir qismi sifatida izohli korpuslarni qurishning yangi yo‘nalishi paydo bo‘ldi.

Elektron annotatsiya korpusini yaratish doimo duch kelgan asosiy muammolardan biri foydalanuvchilarning ularga samarali kirishini ta'minlash vazifasidir. Imkoniyatlarning keng doirasi ko‘zda tutilgan: ma'lum xususiyatlarga ega bo‘lgan butun matnlarga misollar, matnlarning alohida qismlariga misollar, nuqtai nazarlar va ularning kontekstual cheklovlari. Izohlangan korpusni yaratishning dastlabki g‘oyasi nafaqat tegishli nomlar va belgilangan standart iboralar bilan sintaktik tarzda ajratilgan matnlarni o‘z ichiga olgan boshqariladigan ma’lumotlar bazasini olish, balki qidiruv mexanizmini yaratish edi. Kalit so‘zlar bilan aniqlangan matn bo‘laklarini izlash boshqa narsa, so‘zlar va gaplardagi fragmentlar uzunligi mezonlari, so‘roq gaplarning deklarativ jumlalarga belgilangan nisbati kabi muayyan talablarga javob beradigan bo‘laklarni qidirish boshqa narsa hisoblanadi. Oxirgi qidiruv ichki sintaktik va semantik shartlarni va muayyan til ob’ektlari va namunaviy atributlar o‘rtasidagi kontekstdagi iboralar va belgilarning o‘ziga xos formatlari bo‘yicha cheklovlarni o‘z ichiga olishi mumkin.

Elektron korpusga ruxsatsiz kirishni istisno qilish va ma’lumotlar bazasi fayllarini taqdim etilgan grafik va ovozli rasmlar bilan bog‘lash uchun biz Elektron Matnlar Korpusida (SPEC) Qidiruv tizimini ishlab chiqdik. U javon dasturlari tomonidan taqdim etilgan barcha ma’lumotlar uchun kuchli qidiruv buyruqlarini beradi. SPEC matnni matn qismlariga va berilgan matn maydoniga bo‘ladi; SPEC so‘ralgan so‘zni yoki belgilangan so‘zlar ketma-ketligini topadi va ularni ajratib ko‘rsatadi, topilgan ma’lumotlarning avtomatik ravishda yaratilgan teskari havolalar ro‘yxatini tayyorlaydi, bu esa topilgan so‘zning chap va o‘ng tomonidagi belgini asl kontekstda ko‘rish va jadvallarni topish, so‘z shakllari, lemmata va illyustrativ ilovalarga havolalarni ko‘rsatadi.

Elektron korpusda muallif matni yoki lug‘at maqolasi ko‘rgazmali fayl ko‘rinishida chop etilmaydi, balki tizimning ma’lumotlar bazasiga avtomatik ravishda kirish imkoniyati bilan lemma, ibora, so‘z, ma’no, ma’noning miniatyura qismlariga bo‘lingan holda kiritiladi. Havolalarning teskari ro‘yxati va matnning o‘zida bo‘lmasligi mumkin bo‘lgan, ammo dastur ishlashini ta'minlash uchun zarur bo‘lgan qo‘shimcha ma’lumotlar yaratilgan. Elektron korpusning materiali turli tillardagi matnlarni o‘z ichiga olishi mumkin. Lug‘atlar uchun KittyCat veb-sayti ishlab chiqilgan, unda elektron lug‘atlar qulay formatda taqdim etiladi.

Bunday ishlarni bajarishda inson fanlararo guruhi o‘z e'tiborini tegishli faktlarga qaratadi va semantik faktlarni bir vaqtning o‘zida kompyuter uskunalari yordamida ajratib olish mumkin bo‘lgan iqtisodiy vaqtni tejash kuzatiladi.

Taqdim etilgan ishning o‘ziga xosligi shundaki, boshqa muammolarni semantik faktlarni qidirish qobiliyatini amalga oshirmasdan hal qilib bo‘lmaydi. Qidiruv tizimining rivojlangan imkoniyatlari faktik ma’lumotlar yordamida biz ba'zi umume'tirof etilgan umumiy yondashuvlarni tasdiqlashimiz, shuningdek, ulardan ba'zilarini sinovdan o‘tkazishimiz mumkin.

Elektron korpusdagi bugungi qidiruvlar faqat so‘zlar, so‘z birikmalari va jumlalarning turli qismlari bilan cheklangan. Elektron korpusni qidirish tizimini ishlab chiqishda biz nashrlarimizda keltirilgan bir qator o‘ziga xos vazifalarni hal qilishimiz kerak. Elektron korpusdagi qidiruv tizimi ilg‘or algoritmlar va tabiiy tillarni qayta ishlash texnikasi asosida qurilgan. Ushbu yangiliklar tizimga katta hajmdagi ma’lumotlarni tahlil qilish, tegishli kalit so‘zlarni aniqlash va natijalarni ajoyib aniqlik bilan olish imkonini beradi [4]. Qidiruv tizimining eng muhim imkoniyatlaridan biri bu mantiqiy qidiruv operatorlarini qo‘llab-quvvatlash qobiliyatidir, bu foydalanuvchilarga aniq natijalar beradigan murakkab so‘rovlarni tayyorlashga imkon beradi. Kalit so‘zlar, iboralar va mantiqiy operatorlarni birlashtirib, foydalanuvchilar aniq ma’lumotlarni olish uchun qidiruv doirasini toraytirib, kashfiyot jarayonini yanada samarali qilishlari mumkin. [1, 2]

Qidiruv tizimining yana bir muhim qobiliyati-bu ilg‘or filtrlash va tasniflash usullarini qo‘llab-quvvatlash. Ushbu xususiyat foydalanuvchilarga qidiruv natijalarini sana, muallif, nashr turi va dolzarbligi kabi parametrlar asosida takomillashtirishga imkon beradi. Bundan tashqari, qidiruv tizimining qidiruv natijalarini klasterlar va qirralarga ajratish qobiliyati darhol ko‘rinmasligi mumkin bo‘lgan naqsh va munosabatlarni aniqlashni osonlashtiradi.

Qidiruv tizimining imkoniyatlari turli xil ma’lumotlar qazib olish va matnni tahlil qilish vositalari bilan birlashishi bilan yanada kengayadi. Ushbu vositalar foydalanuvchilarga katta ma’lumotlar to‘plamidan tushunchalarni olish, inson kuzatuvidan yashirin bo‘lishi mumkin bo‘lgan tendentsiyalar, korrelyatsiyalar va anomaliyalarni aniqlash imkonini beradi. Muddatli chastota-teskari hujjat chastotasi (TF-IDF) tahlili kabi usullarni qo‘llash orqali foydalanuvchilar o‘zlarining tadqiqot mavzusini chuqurroq tushunib, ma’lumotlar ichidagi yashirin naqshlar va munosabatlarni aniqlashlari mumkin. Korpus tilshunosligi bir nechta asosiy xususiyatlar bilan ajralib turadi, bu uni tilni o‘rganishda qimmatli va aniq yondashuvga aylantiradi. Bularga quyidagilar kiradi: [1, 5, 6]

-Haqiqiy til ma’lumotlaridan foydalanish
-Miqdoriy naqshlarni o‘rganish
-Sifat jihatlarini o‘rganish
-Dalillarga asoslangan tahlil

Elektron korpusdagi qidiruv tizimining eng muhim afzalliklaridan biri bu uning miqyosi va moslashuvchanligi. Tizim katta hajmdagi ma’lumotlarni osonlikcha boshqarishi, so‘rovlarni ajoyib tezlik va samaradorlik bilan qayta ishlashi mumkin. Ushbu o‘lchovlilik zamonaviy tadqiqot muhitida ayniqsa muhimdir, bu yerda ma’lumotlarning katta hajmi eksponent ravishda o‘sib bormoqda. Tarqatilgan hisoblash va bulutga asoslangan infratuzilmaning kuchidan foydalangan holda, qidiruv tizimi murakkab so‘rovlarni samarali qayta ishlay oladi va foydalanuvchilarga kerakli ma’lumotlarga o‘z vaqtida kirishini ta'minlaydi.

Qidiruv tizimining imkoniyatlari juda keng va chuqurdir. Foydalanuvchilarga tegishli ma’lumotlarga tezkor kirishni ta’minlash orqali tizim turli fanlar bo‘yicha bilimlarni kashf etish va ulardan foydalanishni osonlashtiradi. Olimlar va tadqiqotchilar mavjud bilimlardagi bo‘shliqlarni aniqlash, tadqiqot loyihalarini xabardor qilish va farazlarini tasdiqlash uchun qidiruv tizimidan foydalanishlari mumkin. Professionallar va siyosatchilar tizimdan siyosat qarorlarini xabardor qilish, ilg‘or tajribalarni aniqlash va resurslarni taqsimlashni optimallashtirish uchun foydalanishlari mumkin. Qidiruv tizimining imkoniyatlari innovatsiyalarni tezlashtirish, qarorlar qabul qilishni yaxshilash va iqtisodiy o‘sishni ta’minlash imkoniyatiga ega.

Xulosa qilib aytganda, elektron korpusdagi qidiruv tizimi foydalanuvchilarga navigatsiya qilish va katta ma’lumotlar to‘plamlaridan qimmatli tushunchalarni olish imkoniyatini beradigan kuchli vositadir. Ilg‘or algoritmlar, tabiiy tilni qayta ishlash texnikasi va ma’lumotlarni qazib olish vositalaridan foydalangan holda, tizim foydalanuvchilarga qidiruv doirasini yaxshilash, tegishli ma’lumotlarni olish va yashirin naqsh va munosabatlarni ochish imkonini beradi. Raqamli ma’lumotlarning hajmi va murakkabligi o‘sishda davom etar ekan, qidiruv tizimining imkoniyatlari tobora muhim ahamiyat kasb etadi, innovatsiyalarni rag‘batlantiradi, bilimlarni kashf qilishni osonlashtiradi va turli sohalarda qaror qabul qilishni oson qiladi.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI:

1. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
2. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
3. Abdurakhmonova N, Tuliyev U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
4. Q. Po‘latov, Kompyuter tilshunosligi. Toshkent, Akademnashr, 520 (2011) 22. V. V. Rikov, Korpus tilshunosligidan ma'ruzalar kursi (2005) http://rykov-cl.narod.ru/c.html
5. Rustamova, Z. (2022). Linguistic basis of creation of dialectal corpus of uzbek language. *Science and Innovation*, *1*(8), 1255-1258.
6. Rustamova, Z. (2022). O‘zbek tili dialektal korpusini yaratishning lingivistik asoslari. *Science and innovation*, *1*(B8), 1255-1258.