O‘ZBEK TILIDAGI MATNLARNI AVTOMATIK TAHRIRLASHDA KONTEKSTGA ASOSLANGAN XATOLARNI ANIQLASH MASALALARI

Qo‘chqarova Madinabonu Po‘latjon qizi
Andijon viloyati Andijon shahar 14-umumta’lim maktabining
oliy toifali o‘zbek tili fani o‘qituvchisi

Annotatsiya: Ushbu tezisda o‘zbek tilidagi matnlarni avtomatik tahrirlash jarayonida kontekstga asoslangan xatolarni aniqlash masalasi ko‘rib chiqiladi. Lug’atga asoslangan imlo tekshirgichlar, an’anaviy yondashuvlar va kontekstni hisobga oluvchi tabiiy tilni qayta ishlash usullarining imkoniyatlari tahlil qilinadi. O‘zbek tilining agglutinativ xususiyati, qo‘shimchalarning ko‘pligi hamda grammatik va mazmuniy munosabatlar avtomatik tahrirlash tizimlari uchun muhim omillar sifatida yoritiladi. Shuningdek, xatolarni aniqlashda morfologik tahlil, kontekstual til modellari va belgilangan matnlar korpusidan birgalikda foydalanish bo‘yicha amaliy takliflar ilgari suriladi.

Kalit so‘zlar: amaliy kompyuter lingvistikasi, avtomatik tahrirlash, kontekstual tahlil, imlo xatosi, morfologik tahlil, tabiiy tilni qayta ishlash, o‘zbek tili.

Raqamli texnologiyalarning rivojlanishi matnlarni avtomatik qayta ishlash va tahrirlash vositalariga bo‘lgan ehtiyojini oshirmoqda. Matn muharrirlari, electron ta’lim platformalari va turli axborot tizimlarida imlo hamda grammatik xatolarni aniqlash muhim vazifalardan hisoblanadi. Biroq matndagi barcha xatolarni faqat so‘zning lug‘atda mavjud yoki mavjud emasligiga qarab aniqlab bo‘lmaydi. Ayrim hollarda so‘z imlo jihatdan to‘g‘ri bo‘lsa-da, gapning grammatik tuzilishi yoki umumiy mazmuniga mos kelmasligi mumkin.

Shu sababli avtomatik tahrirlashda kontekstni hisobga olish muhim ahamiyat kasb etadi. Kontekst deganda so‘zning atrofidagi birliklar , gapning grammatik tuzilishi va mazmuniy munosabatlari tushuniladi. Kontekstual yondashuv matndagi shubhali birlikni aniqlash bilan birga uning o‘rniga mos kelishi mumkin bo‘lgan variantlarni baholash imkonini beradi.

Kontekstga asoslangan imlo tuzatish masalasi tabiiy tilni qayta ishlash sohasida avvaldan o‘rganib kelinadi. Elmi va Evens(1998) so‘zning to‘g‘ri yozilishini aniqlashda ma’lumotlaridan foydalanish masalasini tadqiq etgan. Zhang va hammualliflar (2020) esa xatoni aniqlash va uni tuzatish bosqichlarini birgalikda hisobga oluvchi Soft-Masked BERT yondashuvini taklif qilgan. Li, LIU va HUANG(2020) tadqiqotida ham xatolarni aniqlash va tuzatishda so‘zning yozilishi shakli bilan bir qatorda umumiy kontekst tasviridan foydalanish masalasi ko‘rib chiqilgan.

Ushbu ishlar kontekstual ma’lumotlarning avtomatik tahrirlashdagi ahamiyatini ko‘rsatadi. Biroq boshqa tillar uchun ishlab chiqilgan usullarni o‘zbek tiliga bevosita ko‘chirishdan oldin tilning grammatik va leksik xususiyatlarini hisobga olish zarur.

Bugungi kunda o‘zbek tilidagi matnlarni avtomatik tahrirlashda uchraydigan muammolarni bir necha turga ajratish mumkin. Birinchi turga lug‘at asosida aniqlanishi mumkin bo‘lgan imlo xatolari kiradi. Ikkinchi tur grammatik shakllarning o‘zaro mos kelmasligi bilan bog‘liq. Uchinchi turda esa so‘zning yozilishi to‘g‘ri bo‘lsa ham, tanlangan birlik gap mazmuniga mos kelmasligi mumkin.

Masalan, “Men bozorga bordim” gapida ega va kesim birinchi shaxs birlik shaklida moslashgan. “Men bozorga bordik” gapida esa ega va kesim o‘rtasida grammatik nomuvofiqlik mavjud. Bunday xatoni aniqlash uchun tizim faqat alohida so‘zlarni lug‘at bilan solishtirmasdan, gapdagi grammatik munosabatlarni ham tahlil qilishi kerak. “Men maktabga bordik” gapida ham ega va kesim o’rtasidagi shaxs-son moslashuvi buzilgan. Gapning egasi birinchi shaxs birlikdagi “men” kishilik olmoshi bilan ifodalangan. Biroq “bordik” kesimi birinchi shaxs ko’plik shaklida qo‘llangan shu sababli ega va kesim

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

grammatik jihatdan moslashmagan. “Ustozlar darsga kiryapti”, adabiy til me’yoridagi odatiy shaklda bu gapni “Ustozlar darsga kiryaptilar” deb ham ifodalash mumkin, lekin kesimning ko‘plikdagi ega bilan moslashuvi o‘zbek tilida har doim ham majburiy emas, shuning uchun bu misolni avtomatik tahrirlashda xato deb belgilash to‘g‘ri bo‘lmaydi.

Kontekstual esa so‘zlarning gapdagi o‘zaro munosabatlarini va mazmuniy muvofiqligini aniqlashga xizmat qiladi. Masalan, “Men yangi ko’ylak sotib oldim va uni tanavvul qildim” gapidagi so‘zlar grammatik jihatdan bog‘langan bo‘lsa-da, odatiy vaziyatda “ko‘ylak” va “tanavvul qilmoq” birliklari o‘rtasida mazmuniy nomuvofiqlik yuzaga keladi. Bunday holatni aniqlashda tizim faqat so‘zlarning imlo shakliga emas, balki ularning konteksdagi ma’nosiga ham e’tibor qaratishi kerak. Shu nuqtayi nazardan, o‘zbek tilidagi matnlarni avtomatik tahrirlashda imlo tekshiruvi, morfologik tahlil, sintaktik munosabatlarni aniqlash va kontekstual ma’no tahlilini birgalikda qo‘llash istiqbolli yondashuv hisoblanadi. Biroq mazkur yondashuvning samaradorligini aniqlash uchun o‘zbek tilidagi haqiqiy matnlar asosida alohida tajribalar o‘tkazish zarur.

O‘zbek tili agglutinativ tillar qatoriga kiradi: so‘zlarga turli grammatik qo‘shimchalar qo‘shilishi orqali yangi shakllar hosil bo‘ladi. Shu sababli avtomatik tahrirlash tizimi so‘zning o‘zak va qo‘shimchalarini tahlil qilish, grammatik shakllarning maqbulligini tekshirish hamda ularning gapdagi vazifasini aniqlash imkoniyatiga ega bo‘lishi lozim.

Kontekstual tahlilning yana bir vazifasi — shaklan to‘g‘ri birliklar orasidan gap mazmuniga mos kelmaydiganini aniqlashdir. Masalan, “U ruchka bilan daftariga yozdi” gapida “ruchka” so‘zi mazmunan mos keladi. “U vilka bilan daftariga yozdi” gapida esa grammatik tuzilish saqlandi lekin odatiy mazmuniy munosabat buzildi.

Bunday holatlarda xatoni aniqlash oddiy imlo tekshiruvidan murakkabroq bo‘lib, semantik va kontekstual ma’lumotlarni talab qiladi. Biroq bunday gaplar har doim xato hisoblanmaydi: noodatiy vaziyat yoki badiiy matn ularni mazmunan asosli qilishi mumkin.

Muammoni hal etish uchun bir nechta usulni birgalikda qo‘llash taklif etiladi. Birinchidan, imlo lug‘ati orqali so‘z shakllarini dastlabki tekshiruvdan o‘tkazish mumkin. Ikkinchidan, morfologik tahlil yordamida o‘zak, qo‘shimcha va grammatik shakllar aniqlanadi. Uchinchidan, kontekstual til modeli so‘zning gapdagi qo‘llanishini baholashga xizmat qiladi. To‘rtinchidan, tizim shubhali birlik uchun bir nechta tuzatish variantini taklif qilib, ularni grammatik va mazmuniy mezonlar asosida saralashi mumkin.

Bunday tizimni ishlab chiqishda o‘zbek tilidagi to‘g‘ri va xatoli gaplardan tashkil topgan belgilangan ma’lumotlar to‘plami muhim o‘rin tutadi. Unda xatolar turi, xato birlik, tavsiya etilgan tuzatish va zarur kontekst qayd etilishi mumkin. Tizim samaradorligini baholashda xatolarni aniqlash aniqligi, tuzatishning to‘g‘riligi va xatosiz matnga asossiz o‘zgartirish kiritish holatlari alohida tekshirilishi lozim.

Xulosa qilib aytganda, o‘zbek tilidagi matnlarni avtomatik tahrirlashda kontekstga asoslangan xatolarni aniqlash amaliy kompyuter lingvistikasining muhim masalalaridan biridir. Lug‘at asosidagi tekshiruv imlo xatolarini aniqlashda foydali bo‘lsa-da, grammatik nomuvofiqlik va mazmuniy xatolarni aniqlash uchun qo‘shimcha tahlil usullari talab etiladi.

Mazkur muammoni hal etish yo‘llaridan biri — imlo tekshiruvi, morfologik tahlil va kontekstual til modellarini birlashtiruvchi yondashuvni ishlab chiqishdir. Kelgusidagi amaliy tadqiqotlarda o‘zbek tiliga xos xatolar korpusini yaratish, tuzatish variantlarini baholash mezonlarini belgilash va tizimni haqiqiy matnlarda sinash maqsadga muvofiq. Taklif etilgan yondashuvning samaradorligi esa alohida tajriba orqali tekshirilishi kerak.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Elmi, M. A., Evens, M. (1998). Spelling Correction using Context. Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics, 360–364. https://doi.org/10.3115/980845.980906
2. Zhang, S., Huang, H., Liu, J., Li, H. (2020). Spelling Error Correction with Soft-Masked BERT. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 882–890. https://doi.org/10.18653/v1/2020.acl-main.82
3. Li, X., Liu, H., Huang, L. (2020). Context-aware Stand-alone Neural Spelling Correction. Findings of the Association for Computational Linguistics: EMNLP 2020, 407–414. https://doi.org/10.18653/v1/2020.findings-emnlp.37

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.