

КОМПЬУТЕР LINGVISTIKASI AVTOMATIK TAHRIRNING NAZARIYASI VA AMALIYOTI

G.M. Ibragimova. f.f.f. (PhD), dotsent

Andijon davlat chet tillari instituti

Mamadaliyeva Guldon Erkinovna

Andijon davlat chet tillari inistituti magistranti

Annotatsiya. Ushbu maqolada kompyuter lingvistikasi sohasida avtomatik tahrir texnologiyalarining nazariy asoslari, amaliy ahamiyati va zamonaviy yondashuvlari tahlil qilinadi. Avtomatik grammatik va uslubiy tahrir qilish tizimlari, ularning tabiiy tilni qayta ishlash (Natural Language Processing – NLP) texnologiyalaridagi o‘rni, shuningdek, sun’iy intellekt yordami bilan matn sifati va aniqligining oshirilish jarayoni ilmiy nuqtai nazardan yoritiladi.

Kalit so‘zlar: kompyuter lingvistikasi, avtomatik tahrir, NLP, sun’iy intellekt, grammatik analiz, korreksiya, til modellari.

Kirish. Kompyuter lingvistikasi – tabiiy tildagi axborotni avtomatik tarzda qayta ishlash, tushunish va generatsiya qilish bilan shug‘ullanuvchi fanlararo soha bo‘lib, u tilshunoslik va informatikaning tutash nuqtasida joylashgan. So‘nggi yillarda sun’iy intellekt va chuqur o‘rganish (deep learning) texnologiyalarining rivojlanishi bu sohada yangi bosqich – avtomatik tahrir jarayonining takomillashtirilishiga olib keldi¹.

Avtomatik tahrir deganda matndagi grammatik, orfografik, uslubiy va mantiqiy xatolarni aniqlash va tuzatish uchun mo‘ljallangan dasturiy vositalar majmuasini tushunamiz. Bugungi kunda Grammarly, Microsoft Editor, Google Docs Suggestion kabi tizimlar yordamida matnni real vaqt rejimida avtomatik tahrirlash mumkin bo‘lib, bu nafaqat yozma muloqotning sifatini oshiradi, balki foydalanuvchi savodxonligini ham rivojlantiradi.

Nazariy jihatdan yoritadigan bo‘lsak, kompyuter lingvistikasida avtomatik tahrir bu kompyuter tizimlari yordamida yozma matnlardagi grammatik, orfografik,

¹ Goldberg Y. *Neural Network Methods for Natural Language Processing*. Morgan & Claypool, 2017.

punktuatsion, uslubiy va ba'zida semantik xatolarni avtomatik aniqlash va tuzatish jarayonidir. Bu jarayon, asosan, *tabiiy tilni qayta ishlash (NLP – Natural Language Processing)*² texnologiyalariga asoslanadi. Avtomatik tahrir matnni tahlil qilish, undagi xatoliklarni aniqlash va ularni tuzatish bo'yicha foydalanuvchiga tavsiyalar berish imkonini yaratadi.

Metodologiya. Kompyuter lingvistikasi doirasida **avtomatik tahrirning nazariyasi va amaliyoti** – bu yozilgan matnlardagi xatoliklarni aniqlash va tuzatish bo'yicha algoritmik yondashuvlarni ishlab chiqish va qo'llash sohasidir.

Bu yo'nalishning **nazariy asoslari** bosqichida avtomatik tahrir jarayonining nazariy poydevori belgilanadi. **Formal tillar va grammatika, xatolik turlari tasnifi, til modellarining roli o'rganiladi.**

Amaliy metodologiya bosqichi avtomatik tahrir tizimini yaratish va test qilish jarayonlarini o'z ichiga olgan bo'lib, unda **ma'lumotlar tayyorlash, xatoni aniqlash (Error Detection), xatoni tuzatish (Error Correction), baholash (Evaluation) jarayonlari mavjud.**

Eksperimental Metodologiya bosqichida esa modellarni turli matn turlarida test qilish (ilmiy, badiiy, ijtimoiy tarmoqlar matnlari) hamda foydalanuvchi sinovlari orqali real dunyo samaradorligini o'lchanadi.

Dasturiy vositalar va platformalar bosqichida quyidagi platformalar va dasturlar bilan ishlash tavsiya qilinadi:

NLTK, spaCy, Stanza – tilni qayta ishlash uchun kutubxonalar

LanguageTool, Grammarly, GECToR – avtomatik tahrir tizimlari

Hugging Face Transformers – neyron modellar bilan ishlash

Muhokama va natijalar. Zamonaviy avtomatik tahrir tizimlari quyidagi bosqichlarda ishlaydi³:

² Jurafsky D. & Martin J. H. *Speech and Language Processing* (3rd ed.). Stanford University, 2021.

³ Abjalova M.A. Matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik bosqichlari/Yosh olim va talabalarning respublika ilmiy-amaliy anjumani materiallari. –T., 2015, 7-bet.

1. Tokenizatsiya – matnni soʻz va gap birliklariga ajratish demakdir. *Tokenizatsiya* — bu tabiiy tilni qayta ishlash (NLP) jarayonining dastlabki bosqichlaridan biri boʻlib, matnni kichik birliklarga — odatda soʻzlar (tokens), tinish belgilar, gaplar yoki iboralar koʻrinishidagi segmentlarga ajratish amaliyotidir. Tokenizatsiya orqali kompyuter dasturlari matnni tahlil qilishga yaroqli shaklga keltiradi. Masalan, “Oʻqituvchi darsga kechikdi.” degan gap tokenlarga quyidagicha ajratiladi: [“Oʻqituvchi”, “darsga”, “kechikdi”]⁴.

2. Morfologik tahlil esa soʻzlarning grammatik shaklini aniqlash, va tabiiy tilni qayta ishlash (NLP) jarayonida har bir soʻzning grammatik shakli, lugʻaviy asos (leksik forma) va unga qoʻshilgan qoʻshimchalar orqali grammatik kategoriyasini aniqlash bosqichidir. Bu tahlil orqali kompyuter dasturi har bir tokenning soʻz turkumi (ot, feʼl, sifat, ravish va h.k.), zamon, kishilik, son, egalik, kelishik kabi grammatik xususiyatlarini aniqlaydi.

3. Sintaktik tahlil – gap tuzilmasini tekshirish va bunda tabiiy tilni qayta ishlashdagi muhim bosqich boʻlib, matndagi gaplarning strukturaviy tuzilmasini aniqlash, yaʼni soʻzlar oʻrtasidagi grammatik va sintaktik munosabatlarni tahlil qilishni anglatadi. Sintaktik tahlil orqali har bir soʻzning gapdagi funksiyasi (egalik, kesimlik, aniqlovchi, toʻldiruvchi, hol) va boshqa soʻzlarga boʻlgan bogʻliqligi aniqlanadi.

Masalan, quyidagi gapni olaylik:
“Talaba kitob oʻqiyapti.”

Sintaktik tahlil natijasida quyidagi tuzilma aniqlandi:

“Talaba” — ega (ot, birlik, kishilik)

“kitob” — toʻldiruvchi (ot, bevosita)

“oʻqiyapti” — kesim (feʼl, hozirgi zamon, uchinchi shaxs)

4. Xatolik aniqlash esa grammatik yoki orfografik nomuvofiqliklarni topishdir. Xatolik aniqlash tabiiy tilni qayta ishlashda (NLP) matnni tahlil qilishning muhim

⁴ Bobojonov S. “Oʻzbek tilida avtomatik matn tahlili va imlo tekshiruv vositalari.” *Filologiya masalalari* №1, 2020. 45–52-b.

bosqichlaridan biri bo'lib, matndagi grammatik, orfografik, punktuatsion va ba'zan stilistik xatolarni aniqlashni anglatadi. Bu jarayon orqali kompyuter tizimlari matnni avtomatik ravishda tekshirib, xatoliklarni topish va tuzatish uchun mos takliflar kiritadi.

Xatolik aniqlash ikki asosiy yo'nalishda amalga oshiriladi. *Birinchisi*, orfografik xatoliklarni aniqlash – bu bo'limda tizim so'zlarning to'g'ri yozilishiga (orfografiyasiga) e'tibor qaratadi. Bunga imlo, tinish belgilari, katta-kichik harflar kabi yozuv qoidalariga rioya etish kiradi. Masalan:

“Yozuvchilar ko'plab kitoblarni yozishdi.” → “Yozuvchilar ko'plab kitoblarni yozdilar.”

“Kechirasiz, sizni chalg'itdim.” → “Kechirasiz, sizni chalg'itdim.”

Ikkinchisi, Grammatik xatoliklarni aniqlash⁵ – bu bosqichda tizim so'zlarning grammatik shaklini tekshiradi, ya'ni so'z turkumlari, zamon, fe'l kelishiklari, ravishlar va boshqa grammatik elementlarning to'g'ri qo'llanilishini aniqlaydi. Misol tariqasida keltirishimiz mumkinki,

“Ular kutubxonaga borganlar.” → “Ular kutubxonaga borgan edi.”

“Men do'stinga xat yozdim, lekin u olmasdi.” → “Men do'stinga xat yozdim, lekin u olmagan.”

Xatolik aniqlash jarayonining asosiy vazifasi — grammatik xatolarni aniqlash va yozuvni aniq, tushunarli va to'g'ri qilish. Bu vazifa ko'pincha nig'orish (rules-based) va statistik (machine learning) yondashuvlaridan foydalangan holda amalga oshiriladi.

Nig'orish yondashuvida xatoliklarni aniqlash uchun oldindan belgilangan grammatik qoidalar va lug'atga asoslanadi.

Statistik yondashuv esa matnni tahlil qilishda kompyuter modelini o'rgatishga asoslanadi, bu esa matnni yuqori darajada tushunishga yordam beradi.

5. Kontekstual tahrir – bu tanlangan variantlar orasida eng mosini taklif qilish. Kontekstual tahrir — bu tabiiy tilni qayta ishlashda matnni tahrirlash jarayonida eng

⁵ Абжалова М.А. Ўзбек тилидаги матнларни автоматик таҳрирлашда услуб масаласи// Таълим тизимидаги ижтимоий-гуманитар фанлар, №2, Т.: 2013, 141-145-бетлар; L. Abduhamidova . Kompyuter lingvistikasining avtomatik tahrir yo'nalishi. BMI.-T.,2016. 33-36

mos va kontekstga mos variantlarni tanlashni anglatadi. Avtomatik tahrir tizimlari matnda aniqlangan xatoliklarni tuzatishda, masalan, grammatika, orfografiya yoki uslubiy xatolarni tuzatishda, kontekstni hisobga olishga katta e'tibor qaratadi. Bu, asosan, matnning o'zgarma tarkibini, uning ma'no va ifoda uslubini saqlagan holda to'g'ri va aniq tahrirni amalga oshirishga imkon beradi.

Kontekstual tahrir tizimlari, ko'pincha, sun'iy intellekt va tabiiy tilni qayta ishlashning chuqur o'rganish usullaridan foydalanadi. Bunday tizimlar bir nechta variantlar orasidan eng to'g'ri va mantiqiy tanlovni qilish uchun matnning ma'nosini va stilistik nuqtai nazarini tushunishadi. Masalan, quyidagi jumlada kontekstual tahrir jarayoni orqali to'g'ri variant tanlanadi:

Noto'g'ri: "U kitobni ko'rdi."

Kontekstual tahrir: "U kitobni o'qidi."

Agar tizim matnning kontekstini aniqlasa va gapni mantiqiy va grammatik jihatdan yaxshiroq tuzishni xohlasa, "ko'rdi" so'zi o'rniga "o'qidi" so'zini tanlashi mumkin. Bunda matnning ma'nosi va konteksti hisobga olinadi, chunki kitobni odatda o'qish mumkin, ko'rish emas.

6. Korreksiya — bu tabiiy tilni qayta ishlash jarayonida matndagi aniqlangan xatoliklarni tuzatish va matnni to'g'rilash bo'yicha avtomatik yoki yarim-avtomatik tizim tomonidan takliflar berish jarayonidir. Korreksiya, asosan, grammatik, orfografik, stilistik yoki semantik xatoliklarni tuzatish uchun amalga oshiriladi. Ushbu jarayon, avtomatik tahrir tizimlari tomonidan ishlatiladigan asosiy qadamlaridan biridir.

Korreksiya jarayoni quyidagi bosqichlarda amalga oshiriladi:

Xatoliklarni aniqlash — bu bosqichda tizim matndagi xatolarni aniqlaydi, masalan, orfografik xatolar, grammatik nomuvofiqliklar, so'z turkumlari orasidagi xatolar, tinish belgilari noto'g'ri qo'llanilishi va hokazo.

Xatoliklarni tahlil qilish — tizim aniqlangan xatoliklarning turini, sababini va ta'sirini tahlil qiladi. Bunda morfologik, sintaktik yoki semantik xatolarni aniqlashda o'rganilgan modellar ishlatiladi.

Korreksiya takliflarini yaratish — tizim xatoliklarni tuzatish uchun bir nechta variantlar yaratadi. Masalan, grammatik xatolarni tuzatish uchun yangi soʻz shakllarini yoki sintaktik oʻzgarishlarni taklif qiladi. Orfografik xatolar uchun toʻgʻri yozilgan soʻzlar keltiriladi.

Eng mos variantni tanlash — tizim yaratilgan takliflar orasidan matnning konteksti va uslubiga mos boʻlgan eng toʻgʻri variantni tanlaydi.

Masalan, quyidagi xatolikni tahlil qilaylik:

“Men yangi kitobni sotib olgandim.” Bu jumlada feʼlning “olgan” shakli notoʻgʻri ishlatilgan, chunki oʻtmish zamon uchun kerakli shakl “olgan edim” boʻlishi kerak. Biroq avtomatik tahrir mutlaqo inson ishtirokini toʻliq almashtira olmaydi. Tilning nozik stilistik, konnotativ va kontekstual jihatlarini hali ham inson tafakkuri aniqroq tushuna oladi.

Xulosa. Kompyuter lingvistikasi doirasida avtomatik tahrir texnologiyalari zamonaviy yozma muloqot madaniyatini shakllantirishda muhim rol oʻynaydi. Tabiiy tilni qayta ishlash, sunʼiy intellekt va chuqur oʻrganish metodlarining uygʻunlashuvi bu tizimlarning yanada takomillashuviga zamin yaratmoqda. Oʻzbek tilidagi avtomatik tahrir vositalarini rivojlantirish tilimizni raqamli muhitda kengroq qoʻllash imkonini beradi.

FOYDALANILGAN ADABIYOTLAR

1. Abjalova M.A. Matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik bosqichlari//Yosh olim va talabalarning respublika ilmiy-amaliy anjumani materiallari. –T., 2015, 7-bet.
2. L. Abduhamidova. Kompyuter lingvistikasining avtomatik tahrir yoʻnalishi. BMI.-T.,2016. 33-36
3. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Stanford University.
4. Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Morgan & Claypool.

5. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
6. Turovskij, M. Y. (2010). *Vvedenie v komp'yuternuyu lingvistiku*. Moskva: URSS.
7. Bobojonov, S. (2022). “O‘zbek tilida avtomatik matn tahlili va imlo tekshiruv vositalari.” *Filologiya masalalari* №1, 45–52
9. Ibragimova, G. Tilshunoslikda paranteza xodisasining tadqiqi xususida. Редакционная коллегия, 260, (2024).
10. Ibragimova, N. Issues of teaching methods of uzbek literature on the example of the life and work of abdulla oripov. *International Journal of Artificial Intelligence*, 1(2), 1824-1828, (2025).
11. Ibragimova, G. M. Study of the Parentheses by World Linguistics. *Excellencia: International Multi-Disciplinary Journal of Education* (2994-9521), 2(3), 5-7. <https://doi.org/10.5281/> (2024).