Materiallar to’plami · Conference Proceedings

TIL, TARJIMA VA TA’LIM: GLOBAL KOMMUNIKATSIYA VA RAQAMLI INNOVATSIYALAR DAVRIDA YANGI YONDASHUVLAR

03 October 2026

UDK 811.512.133

ZAMONAVIY LINGVISTIKA VA GLOBAL KOMMUNIKATSIYA

TABIIY TILLARDAGI GAPLARNI TAHLIL QILISH VA O‘ZBEK TILIDA PEREFRAZLASH MASALASINING RIVOJLANISH TENDENSIYALARI

Khusniya Xusanovna Akhmedova*

✉ khusniya150586@gmail.com

ANNOTATSIYA

Mazkur maqolada tabiiy tillardagi gaplarni kompyuter yordamida avtomatik tahlil qilishning rivojlanish tendensiyalari hamda tabiiy tilga ishlov berish (Natural Language Processing - NLP) doirasida perefrazlash masalasi tahlil qilinadi. Lingvistik analizatorlarning morfologik, sintaktik va semantik bosqichlari, ularning matnni avtomatik qayta ishlashdagi o‘rni yoritiladi. Shuningdek, jahon tajribasida shakllangan perefrazlash tizimlari va yondashuvlari, perefrazni ajratish, aniqlash va generatsiya qilish vazifalari ko‘rib chiqiladi. O‘zbek tilining agglyutinativ xususiyati, leksik sinonimiya va frazeologik birliklarning boyligi perefrazlash jarayoniga ta’sir etuvchi asosiy lingvistik omillar sifatida tavsiflanadi

Kalit so‘zlar

tabiiy tilga ishlov berish, NLP, kompyuter lingvistikasi, lingvistik analizator, morfologik analizator, sintaktik analizator, semantik analizator, perefraz, perefrazni aniqlash, perefrazni generatsiya qilish, sinonimiya, semantik o‘xshashlik, o‘zbek tili.

DOI: 10.5281/zenodo.23277873

Kirish

So‘nggi yillarda axborot texnologiyalari, sun’iy intellekt va tabiiy tilga ishlov berish sohalarining jadal rivojlanishi natijasida inson tilini kompyuter yordamida avtomatik tahlil qilish va qayta ishlashga bo‘lgan ehtiyoj sezilarli darajada oshdi. Tabiiy tilga ishlov berish tizimlari inson nutqini matn yoki audio ko‘rinishida qabul qilish, uni tushunish, talqin qilish va qayta ishlash imkonini beruvchi texnologiyalar majmuasi sifatida qaraladi.

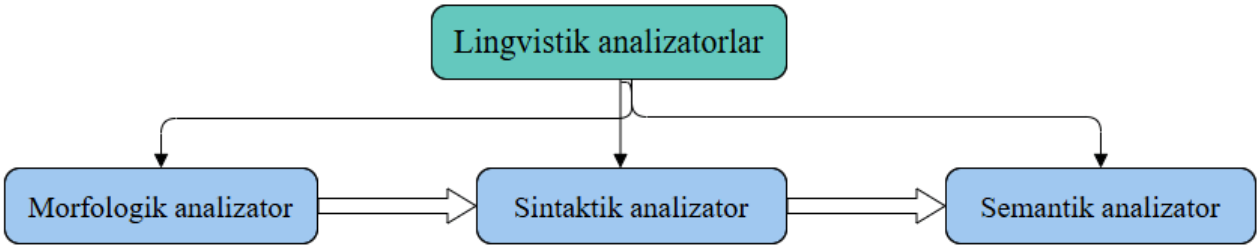
Tabiiy tilni avtomatik qayta ishlashning murakkabligi tilning ko‘p qirraliligi va uni to‘liq tavsiflovchi yagona universal modelning mavjud emasligi bilan bog‘liq. Shu sababli matnni qayta ishlashda lingvistik analizatorlardan foydalanish muhim ahamiyatga ega. Lingvistik analizatorlar tabiiy tilni formallashtirish, bilimlarni ajratib olish, matnning semantik modelini shakllantirish, vektorli reprezentatsiyalarni hosil qilish va keyingi intellektual qayta ishlash bosqichlarini ta’minlashga xizmat qiladi [1].

Mazkur masalalar ichida perefrazlash alohida o‘rin tutadi. Perefrazlashda gapning asosiy mazmuni saqlanadi, biroq uning leksik tarkibi, sintaktik tuzilishi yoki ayrim semantik ifoda vositalari o‘zgaradi. Raqamli axborot hajmining ortishi, takroriy mazmunni aniqlash, plagiatni tekshirish, qidiruv tizimlarini takomillashtirish va matnlarni avtomatik qayta ishlash ehtiyoji perefrazni aniqlash va generatsiya qilish masalalarining dolzarbligini oshirmoqda.

1. Tabiiy tillarda gaplarni tahlil qilish tizimlari

Lingvistik analizatorlar aqlli matn tizimlari arxitekturasining asosiy tarkibiy qismlaridan biri hisoblanadi. Ular funksional vazifalariga ko‘ra morfologik, sintaktik va semantik analizatorlarga ajratiladi.

Mazkur bosqichlar o‘zaro bog‘langan holda ishlaydi va murakkab matn tahlilining ketma-ket jarayonini tashkil etadi.



1-rasm. Tabiiy tillarga ishlov berish jarayoni

Morfologik analizator so‘zlarning grammatik xususiyatlarini aniqlashga xizmat qiladi. Bunda so‘z shakli lug‘atdagi birliklar bilan taqqoslanadi, o‘zak va qo‘shimchalar aniqlanadi hamda so‘zning morfologik belgilari shakllantiriladi. Turkiy tillar uchun morfologik analizatorlarni yaratish masalasi uzoq vaqt davomida tadqiq etib kelinmoqda. O‘zbek tilida ham morfologik analizator va uning lingvistik ta’minotiga bag‘ishlangan tadqiqotlar mavjud.

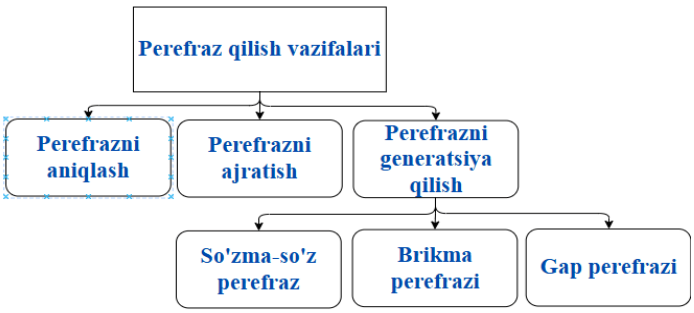
Sintaktik analizator kiruvchi gapning tuzilishini aniqlaydi va undagi sintaktik bog‘lanishlarni modellashtiradi. Sintaktik tahlilda yuqoridan pastga (top-down) hamda pastdan yuqoriga (bottom-up) yondashuvlardan foydalaniladi. O‘zbek tilida sintaktik tahlil grammatik va morfologik ma’lumotlarga tayangan holda amalga oshirilishi mumkin. Bu jarayonda gapdagi so‘zlarning morfologik xususiyatlari keyingi sintaktik munosabatlarni aniqlash uchun muhim manba bo‘lib xizmat qiladi [2].

Semantik analizator esa matn va gaplarning mazmuniy xususiyatlarini aniqlashga yo‘naltiriladi. Semantik tahlilning sifati lingvistik resurslar, xususan, lug‘atlar, tezauruslar, ontologiyalar va sinonimik ma’lumotlarning to‘liqligiga bevosita bog‘liq. Perefrazlash jarayonida aynan semantik tahlil gaplar o‘rtasidagi mazmuniy yaqinlikni aniqlash, sinonim yoki ekvivalent birliklarni tanlash va natijaviy gapning ma’nosini saqlash uchun muhim ahamiyatga ega.

2. Tabiiy tilga ishlov berishda perefrazlash masalasi

Perefraz kompyuter lingvistikasi va tabiiy tilga ishlov berishdagi murakkab masalalardan biridir. Amaliy nuqtai nazardan perefrazni bir xil yoki o‘xshash ma’noni ifodalovchi muqobil lingvistik ifodalar sifatida qarash mumkin. Bunda gaplarning yuzaki leksik tarkibi farq qilishi mumkin, biroq ularning asosiy semantik mazmuni saqlanib qoladi.

Perefrazlashning amaliy ahamiyati bir nechta yo‘nalishlarda namoyon bo‘ladi. U savol-javob tizimlarida turli shakldagi savollarni qayta ishlash, axborotni ajratib olish, mashina tarjimasi, matnni qisqartirish, tabiiy til generatsiyasi, qidiruv tizimlari va plagiatni aniqlash kabi vazifalarda qo‘llanilishi mumkin. Ayniqsa, leksik tarkibi turlicha bo‘lgan, ammo mazmuni bir xil yoki yaqin gaplarni aniqlash oddiy satr yoki so‘z mosligiga asoslangan usullarga nisbatan murakkabroq hisoblanadi.



2-rasm. Gaplarni perefraz qilish vazifalari

Perefrazni aniqlashda asosiy muammolardan biri semantik jihatdan yaqin gaplarda umumiy leksemalarning mavjud bo‘lmasligidir. Aksincha, so‘zlari deyarli bir xil bo‘lgan gaplar turli mazmunni ifodalashi ham mumkin [5]. Shuning uchun faqat yuzaki leksik moslikdan foydalanish yetarli emas. Semantik o‘xshashlikni hisobga oluvchi vektorli va mashinali o‘qitish yondashuvlari ushbu muammoni hal etishda muhim vosita sifatida qaraladi.

Jahon tajribasida perefrazni aniqlash uchun mantiqqa asoslangan, vektorga asoslangan, yuzaki matn o‘xshashligiga asoslangan, sintaktik o‘xshashlikka asoslangan, ramziy ma’no ifodalariga asoslangan, mashinali o‘qitish va dekodlashga asoslangan usullar taklif qilingan.

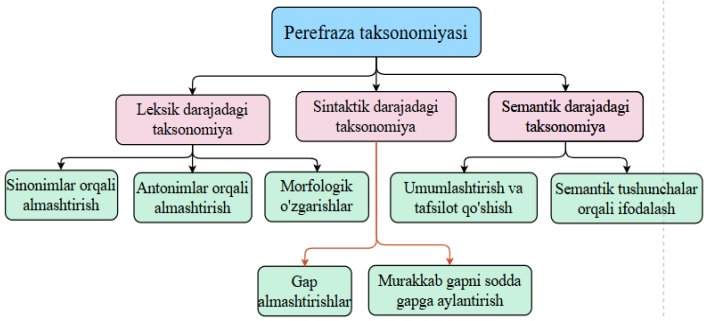
Perefraz bilan bog‘liq vazifalarni shartli ravishda uch guruhga ajratish mumkin: perefrazni ajratish, perefrazni aniqlash va perefrazni generatsiya qilish. Perefrazni ajratish katta korpus tarkibidan o‘zaro perefraz bo‘lishi mumkin bo‘lgan gap yoki iboralarni topishga qaratiladi. Perefrazni aniqlash berilgan gaplar juftligining perefraz ekanligi yoki emasligini belgilaydi. Perefraz generatsiyasi esa berilgan gap asosida yangi, mazmunan ekvivalent gapni hosil qilish jarayonidir.

3. Jahon tajribasida perefrazlash tizimlari

Tabiiy tilga ishlov berish sohasida turli funksional imkoniyatlarga ega bo‘lgan ko‘plab perefrazlash vositalari ishlab chiqilgan. Masalan QuillBot, MyEssaywriter.ai, Ahrefs’ Paraphrasing, WriteHuman, Grammarly, CleverSpinner, Paraphrzr.io, Spin Rewriter va Prepostseo Paraphrasing Tool kabi tizimlarni keltirish mumkin. Ularning ayrimlari bir nechta tillarda ishlaydi, matnni turli darajalarda qayta yozish, sinonimlarni almashtirish va matnning uslubiy ko‘rinishini o‘zgartirish kabi funksiyalarni taqdim etadi. Biroq keltirilgan tahlilga ko‘ra, mazkur tizimlarda o‘zbek tilidagi matnlarni to‘liq va maxsus moslashtirilgan tarzda perefrazlash imkoniyati cheklangan yoki mavjud emas.

5. Perefrazalar taksonomiyasi

Perefrazalarni tizimlashtirishda leksik, sintaktik va semantik darajalar asosida taksonomik yondashuvdan foydalanish maqsadga muvofiq. Leksik perefrazlar so‘z darajasidagi o‘zgarishlarni, sintaktik perefrazlar gap tuzilishidagi o‘zgarishlarni, semantik perefrazlar esa gap mazmuni va kontekstidagi o‘zgarishlarni qamrab oladi.



3-rasm. Perefraz taksonomiyasining iyerarxiyasi

Leksik darajada oddiy sinonimik almashtirish, kontekstga bog‘liq sinonim, metaforik sinonim, antonimik o‘zgarish, so‘z turkumining o‘zgarishi va qisqartmalarni to‘liq shaklga keltirish kabi holatlar kuzatiladi. Sintaktik darajada so‘z tartibini o‘zgartirish, murakkab gapni soddalashtirish yoki gap qismlarining xronologik tartibini qayta tashkil etish mumkin. Semantik darajada esa tafsilot qo‘shish, tafsilotni olib tashlash, mazmuniy umumlashtirish, aniqroq ifodalash va ayrim qarama-qarshi mazmuniy o‘zgarishlar ko‘rib chiqiladi [6].

Chuqur taksonomiyadan foydalanish avtomatik tizim uchun muhim bo‘lib, u perefrazlash jarayonini aniqroq tavsiflash va mashinali o‘qitish algoritmlariga turli o‘zgarishlarni parametr sifatida kiritish imkonini beradi. Bunda o‘zbek tilining lingvistik xususiyatlari bilan bir qatorda, kontekst va semantik ekvivalentlik darajasini ham hisobga olish zarur.

6. Perefrazlash usullarining rivojlanish tendensiyalari

Perefrazlash usullari dastlab lug‘at va sinonimlarni almashtirishga asoslangan yondashuvlardan boshlangan bo‘lsa, keyinchalik statistik va mashinali o‘qitish usullari, undan so‘ng chuqur neyron tarmoqlar va transformer arxitekturalariga asoslangan modellar bilan boyidi. An’anaviy sinonim almashtirish usulida WordNet yoki boshqa tezauruslardan foydalanilib, kontekstga mos sinonimni tanlashga harakat qilinadi. Bannard va Callison-Burch tadqiqotida sinonim ekvivalentlarini tanlash va internet statistikasi asosida mos variantlarni aniqlash yondashuvi ko‘rib chiqilgan [11].

Neyron yondashuvlarda esa Seq2Seq va stacked residual LSTM kabi modellar qo‘llanilgan. Ushbu modellar tabiiy til generatsiyasi uchun samarali bo‘lsa-da, hosil bo‘ladigan perefrazlarning so‘z xilma-xilligi bilan bog‘liq cheklovlar ham qayd etilgan. Keyingi bosqichda kontekstga asoslangan sinonimlarni tanlash usullari takomillashtirilib, tezaurus, tokenizatsiya va POS-teglash kabi jarayonlar bilan birlashtirildi [12].

Muhokama

O‘tkazilgan tahlil shuni ko‘rsatadiki, tabiiy tilga ishlov berish tizimida perefrazlashni samarali tashkil etish uchun faqat bitta tahlil bosqichiga tayanish yetarli emas. Morfologik tahlil so‘z shakllarini aniqlashga, sintaktik tahlil gap tuzilishini ochib berishga, semantik tahlil esa gaplar mazmuniy yaqinligini aniqlashga xizmat qiladi. Ularning ketma-ket va o‘zaro bog‘langan holda ishlashi perefrazlashning lingvistik asosini shakllantiradi.

O‘zbek tili uchun asosiy muammolardan biri resurslarning yetarli darajada shakllanmaganligi bilan bog‘liq. To‘liq qamrovli WordNet yoki keng ko‘lamli perefraz korpusining yetishmasligi sinonimik, frazeologik va semantik munosabatlarni avtomatik aniqlashni murakkablashtiradi. Shu sababli o‘zbek tiliga mos lingvistik ma’lumotlar bazasini shakllantirish, sinonimik va frazeologik birliklarni tizimlashtirish hamda ularni zamonaviy semantik modellar bilan birlashtirish muhim ilmiy vazifa hisoblanadi.

Tahlil natijalariga ko‘ra, o‘zbek tilidagi perefrazlashni uch asosiy funksional yo‘nalishda ko‘rish mumkin: mavjud matndan perefrazlarni ajratish, berilgan gaplar juftligining perefraz ekanligini aniqlash va berilgan gap uchun yangi perefraz variantini generatsiya qilish. Mazkur funksiyalarni bir tizim doirasida birlashtirish o‘zbek tilidagi matnlarni semantik qayta ishlash imkoniyatlarini kengaytiradi.

Xulosa

Tabiiy tillardagi gaplarni avtomatik tahlil qilish tizimlarining rivojlanishi morfologik, sintaktik va semantik analizatorlarning izchil takomillashuvi bilan chambarchas bog‘liq. Tahlil qilingan manbalar asosida lingvistik analizatorlar tabiiy tilni kompyuter yordamida qayta ishlashning muhim vositasi ekanligi aniqlandi. Ayniqsa, perefrazlash jarayonida semantik tahlil, sinonimik munosabatlar va kontekstual moslikni aniqlash alohida ahamiyat kasb etadi.

Perefrazlashning zamonaviy rivojlanish tendensiyasi qoidaga asoslangan sinonim almashtirish usullaridan ma’lumotlarga asoslangan, mashinali o‘qitish va chuqur neyron modellar tomon yo‘nalayotganini ko‘rsatadi. O‘zbek tili uchun esa tilning agglyutinativ xususiyati, sinonimik va frazeologik boyligi hamda maxsus resurslarning cheklanganligi alohida hisobga olinishi zarur.

O‘zbek tilidagi perefrazlash masalasini hal etishda lingvistik resurslarni shakllantirish, leksik, sintaktik va semantik perefrazlarni tasniflash, semantik o‘xshashlikni aniqlash va zamonaviy mashinali o‘qitish modellaridan foydalanishni yagona metodologik yondashuv doirasida birlashtirish maqsadga muvofiq. Mazkur yo‘nalish o‘zbek kompyuter lingvistikasi uchun ilmiy va amaliy jihatdan istiqbolli hisoblanadi.

Foydalanilgan adabiyotlar

[1] Hamroyeva Sh. O‘zbek tili morfologik analizatorining lingvistik ta’minoti. Monografiya. – GlobeEdit, 2020.

[2] Hamroyeva Sh. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filologiya fanlari bo‘yicha falsafa doktori (PhD). – Qarshi, 2018. – 250 b.

[3] Abduraxmonov G., Sulaymonov A., Xoliyorov X., Omonturdiyev J. Hozirgi o‘zbek adabiy tili. Sintaksis. – Toshkent: O‘qituvchi, 1979.

[4] G‘ulomov A., Asqarov M. Hozirgi o‘zbek adabiy tili. Sintaksis. – Toshkent: O‘qituvchi, 1965.

[5] Maxmudov N., Nurmonov A. O‘zbek tilining nazariy grammatikasi. – Toshkent, 1995.

[6] Hamroyeva Sh. O‘zbek tili morfologik analizatorining lingvistik ta’minoti: Filologiya fanlari doktori (DSc) avtoreferati. – Farg‘ona, 2021. – 34 b.

[7] Rahmatullayev Sh., Mahmudov N., Xolmanova Z., O‘razova I., Rixsiyeva K. O‘zbek tili frazeologik lug‘ati. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022. – 636 b.

[8] Androutsopoulos I., Malakasiotis P. Perefrazni aniqlash va tabiiy tilga ishlov berish vazifalariga oid tadqiqotlar.

[9] Bannard C., Callison-Burch C. Paraphrasing with bilingual parallel corpora. In: Proceedings of the 43rd Annual Meeting of the ACL. – 2005. – P. 597–604.

[10] Bolshakov I.A., Gelbukh A. Synonymous paraphrasing using WordNet and internet. In: Lecture Notes in Computer Science. – P. 312–323.

[11] Prakash A., Hasan S.A., Lee K., Datla V.V., Qadir A., Liu J., Farri O. Neural Paraphrase Generation with Stacked Residual LSTM Networks. – 2016. – P. 2923–2934.

[12] Barmawi A.M., Muhammad A. Paraphrasing Method Based on Contextual Synonym Substitution. Journal of ICT Research and Applications. – 2019. – Vol. 13(3). – P. 257–282