

O‘ZBEK TILI MATNLARIDA TINISH BELGILARNI AVTOMATIK QO‘YISH

Matyakubova Noila Shakirjanovna,

Tayanch doktorant

matyakubovanoila@navoiy-uni.uz

ToshDO‘TAU

Avloqulova Iqbol,

Kompyuter lingvistikasi yo‘nalishi 3-bosqich talabasi

iqbolaavloqulova2@gmail.com

ToshDO‘TAU

Annotatsiya. Tinish belgilarini avtomatik tarzda tiklash – matnning mazmunini to‘g‘ri anglash, gaplar chegarasini aniqlash va nutqni o‘qish jarayonini yengillashtirishda muhim ahamiyat kasb etadi. Ayniqsa, nutqni matnga aylantirish, ijtimoiy tarmoqlardagi punktuatsiyasiz yozuvlar va raqamlashtirilayotgan hujjatlar bilan ishlashda tinish belgilarining avtomatik aniqlanishi tabiiy tilni qayta ishlash (NLP) sohasining dolzarb yo‘nalishlaridan biri hisoblanadi. Ushbu maqolada o‘zbek tili matnlarida tinish belgilarini avtomatik tiklash muammosi chuqur yoritilib, mavjud usullar tahlil qilingan. Xususan, transformer arxitekturasiga asoslangan BERT va ELECTRA modellarining samaradorligi, shuningdek, statistik va qoidaviy yondashuvlarning afzalliklari ko‘rib chiqilgan. Tadqiqotda modal so‘zlar va so‘roq yuklamalari ishtirokidagi gaplar asosida Python dasturlash tilida amaliy tajriba o‘tkazilib, NLP jarayonida tilning semantik va sintaktik xususiyatlarini hisobga olgan holda tinish belgilarini aniqlashga qaratilgan fikrlar ilg‘or surilgan. Mazkur yondashuv natijalari esa kelgusidagi tadqiqotlar uchun mustahkam asos bo‘lib xizmat qilishi mumkin.

Аннотация. Автоматическое восстановление знаков препинания имеет важное значение для обеспечения точного понимания содержания текста, определения границ предложений и упрощения процесса чтения. Особенно в случаях преобразования речи в текст, пунктуационно небрежных записей в социальных сетях и работы с оцифровываемыми документами автоматическое определение знаков препинания стало одним из актуальных направлений в области обработки естественного языка (NLP). В данной статье глубоко анализируется проблема автоматического восстановления знаков препинания в текстах на узбекском языке, а также рассматриваются существующие методы. В частности, оценивается эффективность моделей на основе трансформеров, таких как BERT и ELECTRA, а также преимущества статистических и правилосообразных подходов. В исследовании проводились практические эксперименты на языке Python, сосредоточенные на

предложениях с участием модальных слов и вопросительных суффиксов, и предлагались передовые идеи по определению знаков препинания с учетом семантических и синтаксических особенностей языка в процессах NLP. Результаты данного подхода могут послужить прочной основой для будущих исследований.

Annotation. Automatic punctuation restoration plays a crucial role in ensuring accurate understanding of text content, identifying sentence boundaries, and simplifying the reading process. Particularly in speech-to-text conversion, punctuation-less writings on social media, and the digitization of documents, the automatic detection of punctuation marks has become one of the pressing directions in the field of natural language processing (NLP). This article deeply examines the issue of automatic punctuation restoration in Uzbek language texts and analyzes existing methods. Specifically, it evaluates the effectiveness of transformer-based models such as BERT and ELECTRA, as well as the advantages of statistical and rule-based approaches. The study includes practical experiments conducted in Python, focusing on sentences involving modal words and interrogative suffixes, and proposes advanced insights for identifying punctuation marks by considering the semantic and syntactic features of the language in NLP processes. The results of this approach can serve as a solid foundation for future research.

Kalit soʻzlar: *Tinish belgilari, tinish belgilarini avtomatik tarzda qoʻyish (APR), soʻroq belgisi, vergul, modal soʻzlar, soʻroq yuklamasi.*

Hozirgi axborot texnologiyalari rivojlangan davrda matnlarni avtomatik qayta ishlash tobora muhim ahamiyat kasb etmoqda. Ayniqsa, oʻzbek tili kabi milliy tillardagi matnlarni qayta ishlash, tahlil qilish va tuzatish tizimlarini yaratish dolzarb vazifalardan biri hisoblanadi. Matnlarni qayta ishlashning muhim bosqichlaridan biri tinish belgilarini avtomatik tiklash jarayonidir. Tinish belgilari matn maʼnosini toʻgʻri ifodalashda muhim rol oʻynaydi. Ular jumlar oʻrtasidagi chegaralarni belgilaydi, matn qismlarini mantiqan ajratadi va fikrning aniq ifoda etilishini taʼminlaydi.

Bugungi kunda ijtimoiy tarmoqlar, messenjerlar va boshqa raqamli platformalarda yozilgan koʻplab matnlar tinish belgilarisiz yoki notoʻgʻri tinish belgilari bilan yaratilmoqda. Bundan tashqari, ogʻzaki nutqni yozma koʻrinishga oʻtkazishda, tarixiy hujjatlarni raqamlashtirishda va katta hajmdagi matnlarni qayta ishlashda tinish belgilarini avtomatik tiklash muhim masalalardan hisoblanadi.

Tinish belgilari tabiiy tilni qayta ishlashda (NLP) muhim rol oʻynaydi, chunki ular matnning sintaktik va semantik tuzilishini belgilashga yordam beradi. Koʻpchilik NLP ilovalari, masalan, nutqni matnga aylantirish tizimlari, asl nutq tarkibida tinish belgilarini saqlamaydi. Shu sababli, matnni keyinchalik tushunish va qayta ishlash uchun tinish belgilari avtomatik tarzda tiklanishi kerak.

Ushbu maqolada o'zbek tili matnlarida tinish belgilarni avtomatik tiklash muammosining dolzarbligi, mavjud yechimlar va ularning samaradorligi, hamda kelajakda ushbu yo'nalishda olib borilishi kerak bo'lgan ilmiy-tadqiqot ishlari haqida so'z yuritiladi.

So'nggi yillarda chuqur o'rganish (Deep Learning) usullari yordamida Avtomatik Nutq Tanib Olish (ASR) tizimlarining aniqligi sezilarli darajada oshdi. O'zbek tilidagi matnlarda tinish belgilarni tiklash masalasi bugungi kunda zamonaviy texnologiyalar va sun'iy intellekt yordamida ko'plab tadqiqotlar obyektiga aylangan. Ushbu sohada xalqaro tajribalar, xususan, ingliz va venger tillari uchun ishlatilgan usullar orqali samarali natijalarga erishilgan. BERT modellari asosida avtomatik tinish belgilari tiklash usullari Attila Nagy va boshqa tadqiqotchilar tomonidan taqdim etilgan. Ularning tadqiqoti ingliz tili uchun Ted Talks[1] ma'lumotlar bazasi va venger tili uchun Szeged Treebank ma'lumotlar bazasida o'tkazilgan bo'lib, samaradorlik 79.8% (ingliz tili) va 82.2% (venger tili) ko'rsatkichlariga erishilgan. Courtland va boshqalar transfer arxitekturasini ishlatgan holda BERT modellari ko'p so'zli bashoratlar yordamida aniqlikni oshirish usullarini taklif qilgan [1,2]. Ushbu usullar o'zbek tilida ham grammatik qoidalarga asoslangan katta ma'lumotlar bazasini shakllantirish orqali qo'llanilishi mumkin.

Shuningdek, Viet Dac Lai va tadqiqotchilar tadqiqotlarida livestream videolari transkriptlarida tinish belgilari tiklashning o'ziga xos qiyinchiliklari ta'kidlangan[3]. Ular Behance platformasidan olingan videolar uchun yangi annotatsiyalangan korpus (BehancePR)ni taqdim etganlar, bu esa spontan nutq, ko'p sonli so'z xatolari va hissiy ifodalar (masalan, undov belgisi) kabi xususiyatlarni o'z ichiga oladi.

O'zbek tilidagi matnlarda tinish belgilari tiklash masalasini o'rganishda zamonaviy sun'iy intellekt texnologiyalari muhim rol o'ynaydi. Martin Polacek va tadqiqotchilarning 2023-yilda nashr etilgan ishida avtomatik tinish belgilari tiklash (APR– Automatic Punctuation Restoration) yondashuv taklif qilingan, bu esa faqat matnli kiritishdan foydalanadi va ELECTRA-Small modeli asosida ishlaydi. Ularning tadqiqoti real vaqt rejimida televidenya va radio transkriptlarida tinish belgilari (nuqta, vergul, so'roq belgisi) tiklashda o'rtacha 71.2% va 69.4% natijalarini ko'rsatgan [4]. Bu ish o'zbek tilida ham real vaqtli transkripsiyada tinish belgilari tiklash uchun samarali algoritmlarni ishlatish imkoniyatini ko'rib chiqishga asos berishi mumkin. Ushbu tadqiqot shuni ko'rsatadiki, transformator modellari (masalan, ELECTRA) va kichik kontekstli kiritish bilan past kechikish (uch so'z) va yuqori aniqlikni ta'minlash mumkin. Bu o'zbek tilidagi matnlarda tinish belgilari tiklashda yangi yechimlar izlash uchun muhim manba hisoblanadi.

Adebayo Mustapha Bakare va boshqa tadqiqotchilar (2023)[5] ijtimoiy tarmoqlardagi matnlarda tinish belgilarni tiklash uchun transformator modeliga

asoslangan algoritm taklif qilishgan. Ular Amazon (34,659 sharh) va Telekom Malaysia (5000 sharh) ma'lumotlaridan foydalanib, turli BERT modellarini sinovdan o'tkazishgan. Natijada, RoBERTa-Large modeli ikki yo'nalishli LSTM bilan birgalikda eng yuqori aniqlikni ko'rsatgan: Amazon uchun 97% va Telekom uchun 90%. Bu tadqiqot ijtimoiy tarmoqlardagi noto'g'ri punktatsiyali matnlarni tahlil qilishda tinish belgilarni tiklashning sentiment tahliliga foydasini ta'kidlaydi. O'zbek tilidagi ijtimoiy media matnlari uchun ham ushbu yondashuv moslashtirilishi mumkinligi ko'rsatiladi.

Tinish belgilarini avtomatik tiklash modellari.

Nutqni avtomatik matnga aylantirish (ASR – Automatic Speech Recognition) tizimlarida tinish belgilarini tiklash (punctuation restoration) – bugungi kunda tabiiy tilni qayta ishlash sohasining dolzarb va chuqur o'rganilayotgan masalalaridan biridir. Gap tuzilmasining aniqligini ta'minlash, matnni o'qish va tushunishni osonlashtirish va hatto keyingi bosqichdagi mashinali tarjima yoki matn tahlilini aniq bajarish uchun tinish belgilarining to'g'ri qo'yilishi muhim ahamiyatga ega.

Transformer modellar asosidagi ilg'or yondashuvlar. Tinish belgilarini tiklashda transformer arxitekturasiga asoslangan zamonaviy modellar – BERT va ELECTRA samaradorligi solishtirilib tahlil qilingan[4]. Taklif etilgan tizim, ayniqsa, real vaqt rejimidagi nutq transkripsiyasi uchun moslashtirilgan bo'lib, matn asosida ishlovchi (text-only) ELECTRA-Small modeli 11 millisekund kechikish bilan ishlashga muvaffaq bo'lgan. U model nuqta ($F1 = 73.8$), vergul ($F1 = 79.2$) va so'roq belgisi ($F1 = 42.1$)larni aniqlashda yuqori natijalarga erishgan [4]. Bunday real vaqt yondashuvlar ayniqsa jonli subtitrlar va translyatsiyalar uchun dolzarb hisoblanadi.

Statistika va n-gram modellarga asoslangan yondashuvlar. Gravano, Jansche va Bacchiani (2009) tomonidan olib borilgan tadqiqotda esa katta hajmdagi korpuslardan foydalangan holda statistik yondashuvlar orqali punktuatsiyani tiklashga urinishlar amalga oshirilgan. Xususan, 58 milliondan[6] ortiq tokenlar asosida qurilgan n-gram modellar yordamida vergul va nuqta singari belgilarni muvaffaqiyatli aniqlashga erishilgan. Biroq kam uchraydigan belgilar – so'roq belgisi, tire va nuqtali vergul singari tinish belgilar bu yondashuvlarda nisbatan past aniqlik bilan tiklangan.

O'zbek tilidagi matnlarda APRning qo'llanilishi.

Tinish belgilari o'zbek tilida yozma nutqning asosiy tarkibiy qismi bo'lib, ularning qo'llanilishi nutqning semantik, grammatik xususiyatlariga qat'iy bog'liq. Fitratning “O'zbek tili qoidalari to'g'risida bir tajriba” asariga asoslanib, tinish belgilari “turish belgilari” sifatida tasniflanadi va ularning ma'nosi ularning vazifasi, yozuvdagi o'rni, shuningdek, gapning sintaktik tuzilishi va ma'no yukiga bog'liq[7].

Masalan, nuqta jumlani tugatishni, vergul qo'shimcha ma'lumotlarni ajratishni, so'roq va undov belgilari esa gapning ohang va maqsadini ifodalaydi.

Grammatik jihatdan bir xil shakllangan va bir xil leksik vositalarga ega bo'lgan quyidagi besh gapni olib ko'raylik, bularda besh xil ma'no birligi tinish belgilaridan bilinadi. Masalan:

1. Biz fazoni o'rganmoqdamiz. (Xabar ma'nosi)
2. Biz fazoni o'rganmoqdamiz? (So'roq ma'nosi)
3. Biz fazoni o'rganmoqdamiz! (Undov ma'nosi)
4. Biz fazoni o'rganmoqdamiz... (fikrning bo'linishi)
5. Biz fazoni o'rganmoqdamiz!.. (Kuchli emotsiya – sevinch – fikrning bo'linishi ma'nosi)[7].

Yuqoridagi misollar shuni ko'rsatadiki, tinish belgilarini avtomatik tiklash nafaqat matnni grammatik jihatdan to'g'ri tuzishda, balki uning mazmunan to'liq anglanishida ham muhim rol o'ynaydi. Ayniqsa, o'zbek tilida bir xil grammatik tuzilishga ega gaplar turli tinish belgilar yordamida turlicha ma'nolarni ifodalashi mumkin. Bu esa avtomatik tizimlar uchun jiddiy muammo bo'lib, tinish belgilarini to'g'ri joylashtirish orqali matnning semantik va pragmatik ma'nosini tiklash zarurligini ko'rsatadi. Yuqorida keltirilgan misollar, ya'ni “Biz fazoni o'rganmoqdamiz” gapining har xil tinish belgilar bilan ifodalanishi (xabar, so'roq, undov, noaniqlik va emotsional kuchlanish) — bu jarayonning nozikligini va sun'iy intellektga asoslangan tizimlarda ehtiyotkorlik bilan yondashish lozimligini anglatadi. Shuning uchun ham avtomatik tinish belgilarini tiklash modellarini o'zbek tilining intonatsiya va sintaktik xususiyatlariga moslashtirish dolzarb masalalardan biri hisoblanadi.

Shu bilan bir qatorda matndagi xatoliklarni tuzatish muammolari orasida tinish belgilarni qo'yish ham alohida e'tiborni talab qiladi. Tinish belgilar bo'yicha ikki turdagi masala ko'riladi. Birinchisi, matnda mavjud tinish belgilarni to'g'ri qo'yilganligini tekshirish. Ikkinchisi esa, matnda tinish belgilari umuman mavjud emas vaziyatlarda ularni to'g'ri qo'yish [8].

Ushbu maqolada matnda mavjud bo'lmagan tinish belgilarni tiklash masalasiga alohida e'tibor qaratilgan. Xususan, modal so'zlar va so'roq yuklamalarining matndagi vazifasi, ularning gap ma'nosi va ohangiga ta'siri nuqtayi nazaridan tahlil qilingan. Ushbu birliklarning grammatik va semantik xususiyatlariga tayangan holda, tinsh belgilarini aniqlashda qanday xizmat qilishi amaliy jihatdan o'rganilgan.

Amaliy ish jarayonida modal so'zlar va so'roq yuklamalari ishtirokidagi gaplarni tahlil qilish hamda ularni tegishli tinish belgilar bilan to'ldirishga oid

tajribalar o'tkazilgan. Ushbu yondashuv tinish belgilarni avtomatik tiklashda lingvistik yondashuvning muhimligini ko'rsatishga xizmat qilish mumkin.

Quyidagi jadvalda **modal so'zlar va so'roq yuklamalari ishtirokidagi gaplar namunalarini jadvallar shaklida ko'rsatiladi**. Jadvalda ba'zi modal so'roq yuklamalarining matndagi vazifasi, unga mos tinish belgisi ko'rsatiladi (Qarang:1 va 2-jadvalga).

1-jadval. Modal so'zlar va ularning turlari.

Modal so'z turlari	Modal so'zlar	Misollar
Rostlik, tasdiq-ishonch, qat'iylikni bildiruvchi so'zlar	darhaqiqat, haqiqatan, albatta, shubhasiz, so'zsiz, shaksiz.	Darhaqiqat, bu holat ko'p kuzatiladi.
Taxmin, gumon kabi ma'nolarni ifodalovchi so'zlar	shekilli, chog'i, chamasi, ehtimol, balki, aftidan, taxminan.	Bu ishni tugatish uchun, taxminan, qancha vaqt ketadi?
Fikrning o'zaro munosabati	xullas, demak, binobarin, qisqasi, umuman, masalan, chunonchi, jumladan, aksincha, avvalo, avvalambor, nihoyat.	Men ko'plab mevalarni yaxshi ko'raman, masalan, olma, banan va uzum.
Ta'kid	xususan, ayniqsa.	Xususan, ingliz tili o'quvchilarga eng ko'p qiyinchilik tug'diradi.

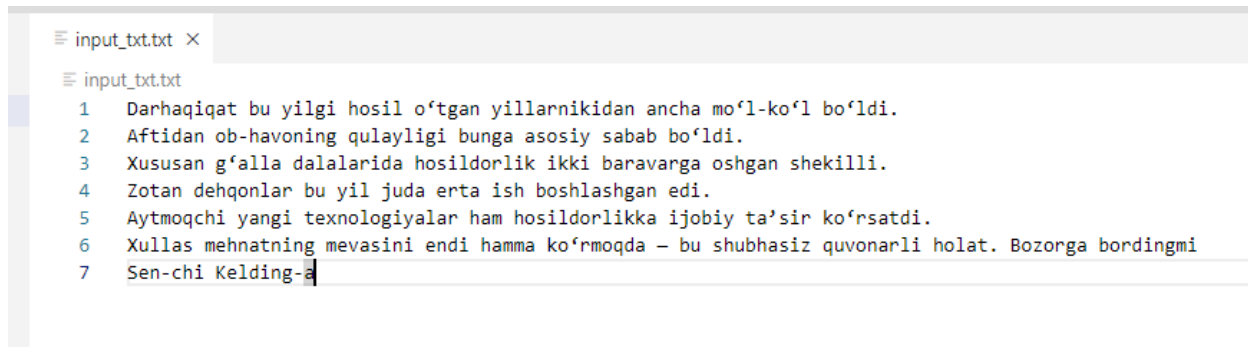
2-jadval. So'roq yuklamalari.

So'roq yuklamalari	Misollar
-mi	Ertaga biznikiga kelasanmi?
-chi	Men kitobni o'qib tugatdim, sen-chi?
-ya	Nahotki shu gaplar rost bo'lsa-ya?
-a	Bugunoq qaytasan-a?

Yuqoridagi jadval asosida tanlab olingan modal so'zlar va so'roq yuklamalari ishtirokidagi gaplar amaliy ish jarayonida dasturiy tahlil qilingan. Matnni tahlil qilish va tinish belgilarini avtomatik qo'shish jarayonida Python dasturlash tilidan va uning *nltk* (Natural Language Toolkit) kutubxonasidan foydalaniladi. Matnlar tokenlarga ajratish va so'zlarni aniqlashda, shuningdek, tinish belgilarini qo'yishda *re-kutubxonasi* yordamida muntazam ifodalardan (regular expressions) foydalaniladi. Bu yondashuv yordamida modal so'zlar va so'roq yuklamalarini aniqlash va ular bilan ishlash ko'zda tutilgan.

Amaliy tajriba quyidagi bosqichlarda amalga oshirildi:

1. Matn tayyorlandi va .txt formatida faylga joylashtirildi (qarang: 1-rasmga).



1-rasm. Dasturga matn kiritish jarayoni.

2. Dasturiy tahlil uchun Python kodi yozildi. Ushbu amaliy ishda matnlarni tahlil qilish va tinish belgilarini avtomatik qo'shish jarayonida Python dasturlash tilidan va uning *nltk* (Natural Language Toolkit) kutubxonasidan foydalanildi. Ushbu kodda matnlarni tokenlarga ajratish uchun *word_tokenize* va *sent_tokenize* funksiyalari hamda *uznatcorpora*, shuningdek, matnni birlashtirish va qayta tiklash uchun *TreebankWordDetokenizer* funksiyasidan foydalanildi. Tahlil jarayonida shuningdek, *re-kutubxonasi* yordamida maxsus tartibga solish va o'zgarishlar kiritish amallari bajarildi.

Dasturiy tahlil uchun Python kodi:

```
def modal_sozlarga_vergul_qosh(matn):  
    gaplar = sent_tokenize(matn)  
    yangi_gaplar = []  
  
    for gap in gaplar:  
        sozlar = sozlarni_ajrat(gap)  
        oxirgi_simvol = gap[-1]  
        yangi_sozlar = []  
        i = 0  
        while i < len(sozlar):  
            soz = sozlar[i]  
            soz_clean = re.sub(r'[.,!?', ' ', soz.lower()).strip()  
            is_modal = soz_clean in modal_sozlar  
            # print(soz_clean, is_modal)  
  
            if is_modal:  
                original_soz = soz  
                next_soz = sozlar[i + 1] if i < len(sozlar) - 1 else  
None  
                prev_soz = yangi_sozlar[-1] if yangi_sozlar else None  
  
            if i == 0:
```

```
yangi_sozlar.append(original_soz + ',')
i += 1
continue

elif i > 0 and i < len(sozlar) - 1 and next_soz not in
['.', '!', '?']:
    old_soz = yangi_sozlar.pop() if yangi_sozlar else
    ''

    if old_soz:
        yangi_sozlar.append(old_soz + ',')
        yangi_sozlar.append(original_soz + ',')
        i += 1
        continue

    else:
        if i > 0 and prev_soz and next_soz is None:
            old_soz = yangi_sozlar.pop()
            yangi_sozlar.append(old_soz + ', ' +
original_soz )

        else:
            yangi_sozlar.append(original_soz)
            i += 1
            break

    else:
        yangi_sozlar.append(soz)
        i += 1

yangi_gap = detokenizer.detokenize(yangi_sozlar)
if oxirgi_simvol not in ['.', ',', '?', '!', ]:
    oxirgi_simvol = ''
yangi_gaplar.append(yangi_gap + oxirgi_simvol)
return '\n'.join(yangi_gaplar)
```

3. Modal soʻzlar va soʻroq yuklamalari aniqlangach, ularning oʻrniga qarab quyidagi qoidalar boʻyicha mos tinish belgi tiklanadi:

– Modal soʻzlar gap boshida kelgan boʻlsa, vergul (,) belgisi: *Xullas, bu yilgi mavsumda oʻzgarishlar kutilmoqda, shuning uchun tayyor boʻlishimiz kerak;*

– Modal soʻzlar gap oʻrtasida kelsa, modal soʻzning ikkala tomonidan vergul (,) belgisi: *Ilmiy tadqiqotlar, jumladan, tajribalar ham bu jarayonni yaxshilashga yordam beradi;*

– Gap oxirida kelsa, modal soʻz oldidan vergul (,) belgisi qoʻyiladi: *Bu yilgi sayohat juda qiziqarli boʻladi, shubhasiz.*

4. Soʻroq yuklamalari gapni oxirida qoʻllanilsan soʻroq (?) belgisi qoʻyiladi.

Misol uchun, *Sen hali kelmadingmi?*

5. Natijalar.txt formatida fayl ko'rinishida chiqarib beriladi (qarang: 3-rasmga).

```
≡ output.txt
1 Darhaqiqat, bu yilgi hosil o'tgan yillarnikidan ancha mo'l ko'l bo'ldi.
2 Aftidan, ob-havoning qulayligi bunga asosiy sabab bo'ldi.
3 Xususan, g'alla dalalarida hosildorlik ikki baravarga oshgan, shekilli.
4 Zotan, dehqonlar bu yil juda erta ish boshlashgan edi.
5 Aytmoqchi, yangi texnologiyalar ham hosildorlikka ijobiy ta'sir ko'rsatdi.
6 Xullas, mehnatning mevasini endi hamma ko'rmoqda bu, shubhasiz, quvonarli holat.
7 Bozorga bordingmi? Sen-chi? Kelding-a?
```

3-rasm. Amaliy natijadan namuna.

Shu o'rinda ta'kidlash joizki, Python dasturlash tilida yozilgan kod faqat modal so'zlar va so'roq yuklamalari asosida ishlaydi. Bu yondashuv tinish belgilarini tiklashda ayrim istisno holatlar uchun yuqori samaradorlik ko'rsatkichni bermasligi mumkin. Biz bilamizki modal so'zning tinish belgisiz shakli ham to'g'ri deb qabul qilinishi mumkin. Masalan: “To'g'ri, yo'llar haydovchilar uchun muhim.”, “To'g'ri, yo'llarda ham qiyinchiliklar bo'ladi.” va “To'g'ri yo'llar haydovchilar uchun muhim.”, “To'g'ri, yo'llarda ham qiyinchiliklar bo'ladi.” kabi gaplarda faqat bitta vergul qo'yilishi natijasida gapning ma'nosi tubdan o'zgaradi. Bunday nozik holatlarni avtomatik tarzda to'g'ri aniqlash esa hali ham dolzarb muammo bo'lib qolmoqda. Bu esa mazkur sohada yanada chuqurroq lingvistik tahlil va zamonaviy algoritmlarni joriy qilish zaruratini ko'rsatadi.

Xulosa. Yuqoridagi tadqiqot shuni ko'rsatadiki, o'zbek tilida tinish belgilarini avtomatik tiklash masalasi sun'iy intellekt modellari, katta hajmdagi ma'lumotlar bazasi va domenlar o'rtasidagi farqlarni hisobga olgan holda yondashishni talab qiladi. Tinish belgilarining grammatik-sintaktik strukturadagi o'rni beqiyos bo'lib, ularning avtomatik aniqlanishi NLPning ko'plab sohalari uchun dolzarb masaladir. Chuqur o'qitish modellari va lingvistik qoidalarni uyg'unlashtirish bu boradagi samarali usullardan biri bo'lsa-da, matn konteksti noaniqligi, domenga xoslik va o'zbek tiliga xos grammatik xususiyatlar kelajakda maxsus korpuslar asosida hal etilishi mumkin. Ayniqsa, modal so'zlar va so'roq yuklamalarining kontekstga ko'ra o'zgaruvchanligi ularni alohida o'rganishni talab etadi va istiqbolli tadqiqot yo'nalishi sifatida qaralmoqda.

Foydalanilgan adabiyotlar:

1. Nagy A., Biall B., Ács J. *Automatic punctuation restoration with BERT models*. 2021. arXiv:2101.07343v1 [cs.CL] 18 Jan 2021.
2. Elov B., Alayev R., Abdullayev A., Aloyev N. Yuqori n-gram modellarini o'zbek tili matnlariga qo'llash. Raqamli Transformatsiya va Sun'iy Intellekt ilmiy jurnali. Volume 2, Issue 5, October 2024. – P. 152-162.

3. Lai V., Veyseh A., Dernoncourt F., Nguyen T. Punctuation Restoration. [Dublin, Ireland](#), 20-24 August 2023. – P. 446-450.
4. Poláček M., Cerva P., Zdansky J., Weingartova L. Online Punctuation Restoration using ELECTRA Model for streaming ASR Systems. In Interspeech. (2023).
5. Bakare A., Anbananthan K., Muthaiyah S., Krishnan J., Kannan S. Punctuation Restoration with Transformer Model on Social Media Data. Appl. Sci. 2023, 13, 1685
6. Gravano A., Jansche M., Bacchiani M. Restoring punctuation and capitalization in transcribed speech. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, 19-24 April 2009, Taipei, Taiwan.
7. Ishmurzayeva L. O'zbek tilshunosligida tinish belgilarining ilmiy tadqiqi. Journal of Innovations in Scientific and Educational Research, Volume-6, Issue-4, DOI: 10.5281/zenodo.7878232. (2023).
8. Ochilov M., Jurayev D., Narzullayev O. Matndagi tinish belgilarni tiklash muammolari va yechimlari. Kompyuter Lingvistikasi: Muammolar, Vechim, Istiqbollar, Vol. 1, No. 01, 2024.