ISNN 3030-3907

2026
JUNE
ISSUE - 6
VOLUME - 9

DEVELOPMENT OF SCIENCE

scientific journal



Development of science

Ilmiy jurnal

IYUN 2026
Issue - 6, VOLUME – 9

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

O‘ZBEK TILI MATNLARIDA TINISH BELGILARINI TAHLIL QILISHNING GIBRID MODELI VA DASTURIY MAJMUASINI ISHLAB CHIQISH.

Adinayev X.S.
Urganch davlat universiteti, Urganch, Uzbekiston
0000-0002-7824-1854

Annotatsiya. Tinish belgilarini avtomatik tiklash tabiiy tilni qayta ishlash sohasidagi muhim vazifalardan biri bo‘lib, ayniqsa o‘zbek tili kabi resurslari cheklangan tillarda bu masala dolzarb hisoblanadi. Belgilangan korpuslarning yetishmasligi nutqni tanish, mashina tarjimasi va matnni semantik tahlil qilish kabi keyingi bosqich ilovalar samaradorligiga salbiy ta’sir ko‘rsatadi. Ushbu tadqiqotda o‘zbek tilidagi matnlar uchun tinish belgilarini bashorat qilishga mo‘ljallangan gibrid chuqur o‘rganish modeli taklif etiladi. Taklif etilgan yondashuv BERT asosidagi kontekstual vektor tasvirlari, BiLSTM yordamida ketma-ketlik bog‘liqligini modellashtirish hamda lingvistik qoidalarga asoslangan post-qayta ishlash bosqichini birlashtiradi. Mazkur arxitektura transformer modellarning semantik imkoniyatlari va rekurrent tarmoqlarning vaqt bo‘yicha bog‘liqlikni aniqlash qobiliyatidan samarali foydalanadi, qoida asosidagi tuzatish esa noaniq chegaralardagi tinish belgilarini aniqlash aniqligini oshiradi. O‘zbek tilida belgilangan korpus asosida o‘tkazilgan tajribalar natijalari taklif etilgan modelning aniqlik, to‘liqlik va F1-ko‘rsatkichlari bo‘yicha mavjud neyron va statistik usullardan ustunligini ko‘rsatdi. Olingan natijalar chuqur neyron arxitekturalarni lingvistik qoidalar bilan birlashtirish resurslari cheklangan tillarda tinish belgilarini tiklash samaradorligini sezilarli darajada oshirishini tasdiqlaydi. Ushbu ish o‘zbek tilini qayta ishlash jarayonlarini rivojlantirish hamda turli amaliy NLP ilovalari aniqligini yaxshilash uchun amaliy va kengaytiriladigan yondashuvni taqdim etadi.

Kalit so‘lar. Punktuatsiya belgilari; NLP; BERTbek model, BiLSTM model, F1 ko‘rsatkich.

Аннотация. Автоматическое восстановление знаков препинания является одной из важных задач в области обработки естественного языка (NLP), особенно для языков с ограниченными ресурсами, таких как узбекский язык. Недостаток размеченных корпусов негативно влияет на эффективность последующих приложений, включая распознавание речи, машинный перевод и семантический анализ текста. В данном исследовании предлагается гибридная модель глубокого обучения, предназначенная для прогнозирования знаков препинания в узбекских текстах. Предложенный подход объединяет контекстуальные векторные представления на основе BERT, моделирование последовательных зависимостей с помощью BiLSTM и этап постобработки, основанный на лингвистических правилах. Такая

ISSN 3030-3907

архитектура эффективно использует семантические возможности трансформерных моделей и способность рекуррентных сетей выявлять временные зависимости, тогда как правило-ориентированная коррекция повышает точность определения знаков препинания в неоднозначных контекстах. Результаты экспериментов, проведённых на размеченном корпусе узбекского языка, показали, что предложенная модель превосходит существующие нейронные и статистические методы по показателям точности, полноты и F1-меры. Полученные результаты подтверждают, что интеграция глубоких нейронных архитектур с лингвистическими правилами значительно повышает эффективность восстановления пунктуации для языков с ограниченными ресурсами. Данная работа предлагает практический и масштабируемый подход к развитию технологий обработки узбекского языка и повышению точности различных прикладных NLP-систем.

Ключевые слова: знаки препинания; NLP; модель BERTbek; модель BiLSTM; F1-метрика.

Abstract. Automatic punctuation restoration is one of the important tasks in the field of Natural Language Processing (NLP), particularly for low-resource languages such as Uzbek. The lack of annotated corpora negatively affects the performance of downstream applications, including speech recognition, machine translation, and semantic text analysis. This study proposes a hybrid deep learning model for punctuation prediction in Uzbek texts. The proposed approach combines BERT-based contextual embeddings, sequence dependency modeling using BiLSTM, and a linguistically motivated post-processing stage. This architecture effectively leverages the semantic capabilities of transformer models and the ability of recurrent networks to capture temporal dependencies, while rule-based correction improves punctuation prediction accuracy in ambiguous contexts. Experimental results on an annotated Uzbek corpus demonstrate that the proposed model outperforms existing neural and statistical approaches in terms of precision, recall, and F1-score. The findings confirm that integrating deep neural architectures with linguistic rules can significantly improve punctuation restoration performance for low-resource languages. This work provides a practical and scalable approach for advancing Uzbek language processing technologies and improving the accuracy of various NLP applications.

Keywords: punctuation marks; NLP; BERTbek model; BiLSTM model; F1-score.

1. KIRISH.

Tabiiy tilni qayta ishlash (Natural Language Processing — NLP) so‘nggi yillarda sun’iy intellektning eng tez rivojlanayotgan yo‘nalishlaridan biriga aylandi. Matnlarni avtomatik tahlil qilish, nutqni tanish, mashina tarjimasi hamda savol-javob tizimlari kabi ko‘plab amaliy ilovalarning samaradorligi matn

ISSN 3030-3907

tuzilmasining to‘g‘ri tiklanishiga bevosita bog‘liq. Shu nuqtai nazardan, tinish belgilarini avtomatik aniqlash va tiklash vazifasi muhim bosqichlardan biri hisoblanadi. Ayniqsa, og‘zaki nutqdan olingan yoki tinish belgilarisiz yozilgan matnlarda mazkur masala dolzarblik kasb etadi.

Tinish belgilarini tiklash muammosi resurslari yetarli bo‘lgan tillarda keng o‘rganilgan bo‘lsa-da, o‘zbek tili kabi resurslari cheklangan tillar uchun tadqiqotlar nisbatan kam. Belgilangan korpuslarning yetishmasligi, lingvistik xususiyatlarning murakkabligi hamda mavjud modellarni to‘g‘ridan-to‘g‘ri qo‘llashdagi qiyinchiliklar ushbu yo‘nalishda samarali yondashuvlarni ishlab chiqishni talab etadi. Natijada, tinish belgilarini noto‘g‘ri yoki umuman qo‘llanmasligi nutqni tanish, mashina tarjimasi va semantik tahlil kabi keyingi bosqich tizimlarining aniqligiga salbiy ta’sir ko‘rsatadi.

So‘nggi yillarda chuqur o‘rganishga asoslangan yondashuvlar, xususan rekurrent neyron tarmoqlar va transformer arxitekturalari, ketma-ketlikdagi bog‘liqliklarni aniqlashda yuqori natijalarni namoyon etmoqda. BERT kabi oldindan o‘qitilgan til modellari matnning chuqur kontekstual ifodalarini shakllantirish imkonini bersa, BiLSTM tarmoqlari ketma-ketlikdagi vaqt bo‘yicha bog‘liqliklarni samarali modellashtiradi. Shu bilan birga, ayrim noaniq holatlarda sof neyron yondashuvlar yetarli aniqlikni ta’minlay olmasligi sababli lingvistik qoidalarga asoslangan qo‘shimcha post-qayta ishlash mexanizmlaridan foydalanish maqsadga muvofiq hisoblanadi.

Ushbu maqolada o‘zbek tilidagi matnlar uchun tinish belgilarini bashorat qilishga mo‘ljallangan BiLSTM, BERT va qoida asosidagi post-qayta ishlashni birlashtiruvchi gibrid chuqur model taklif etiladi. Taklif etilgan yondashuv transformer modellarining semantik imkoniyatlari, rekurrent tarmoqlarning ketma-ketlikni modellashtirish qobiliyati hamda lingvistik qoidalarning aniqlashtiruvchi xususiyatlarini yagona arxitektura doirasida uyg‘unlashtiradi. Tajribalar o‘zbek tilida belgilangan korpus asosida o‘tkazilib, model samaradorligi aniqlik, to‘liqlik va F1-ko‘rsatkichlari orqali baholanadi.

Mazkur tadqiqotning asosiy hissasi quyidagilardan iborat:

(1) o‘zbek tili uchun tinish belgilarini tiklashga mo‘ljallangan gibrid neyron-lingvistik modelni taklif etish;

(2) chuqur o‘rganish va qoida asosidagi yondashuvlarni integratsiyalash orqali aniqlikni oshirish;

(3) resurslari cheklangan tillarda tinish belgilarini tiklash bo‘yicha amaliy va kengaytiriladigan yechimni taqdim etish.

2. MAVZUGA OID TADQIQOTLAR.

Tinish belgilarini tiklash va tahlil qilish masalasi tabiiy tilni qayta ishlash (NLP) sohasida muhim vazifalardan biri hisoblanadi. To‘g‘ri qo‘yilgan tinish

ISSN 3030-3907

belgilari matnning o‘qiluvchanligini oshiradi hamda sintaktik tahlil, mashinali tarjima, axborot ajratib olish va avtomatik nutqni tanish kabi keyingi bosqichdagi NLP vazifalarining samaradorligiga sezilarli ta’sir ko‘rsatadi [1]. Dastlab, tinish belgilari ikkilamchi grafik element sifatida qaralgan bo‘lsa-da, so‘nggi tadqiqotlar ularning matndagi muhim sintaktik va semantik chegaralarni ifodalashini ko‘rsatdi.

Ilk tadqiqotlarda tinish belgilarini aniqlash va tekshirish asosan qoida asosidagi yondashuvlar orqali amalga oshirilgan. Bunday yondashuvlarda tilning grammatik va stilistik qoidalariga tayangan holda qo‘lda tuzilgan algoritmlar qo‘llanilgan. Resurslari cheklangan tillar uchun bu usul hanuz dolzarb hisoblanadi. O‘zbek tili uchun ushbu yo‘nalishdagi muhim ishlardan biri Sharipov va hammualliflar tomonidan taklif etilgan qoida asosidagi algoritm bo‘lib, unda nuqta va vergul belgilarining qo‘llanilishi lingvistik qoidalar yordamida tahlil qilingan [2]. Mazkur ish o‘zbek tilida tinish belgilarini avtomatik tahlil qilish bo‘yicha dastlabki poydevorni yaratgan bo‘lsa-da, murakkab kontekstlarni qamrab olishda cheklovlarga ega ekanligi qayd etilgan.

Keyingi bosqichda tinish belgilarini tiklash masalasi ketma-ketlikni belgilash (sequence labeling) muammosi sifatida qarala boshlandi va statistik modellar, xususan Conditional Random Fields (CRF) keng qo‘llanila boshladi. CRF modellari tokenlar orasidagi bog‘liqliklarni hisobga olgani sababli qoida asosidagi yondashuvlarga nisbatan yaxshiroq natijalar ko‘rsatdi [3]. O‘zbek tili uchun BiLSTM va CRF modellarining kombinatsiyasi asosida yaratilgan yondashuvlar yuqori aniqlik va F1-ko‘rsatkichlarga erishgani ilmiy tadqiqotlarda isbotlangan [4].

So‘nggi yillarda chuqur o‘rganishga asoslangan modellar, ayniqsa Bidirectional LSTM (BiLSTM) tarmoqlari tinish belgilarini tiklash vazifasida keng qo‘llanilmoqda. Ushbu modellar matnning chap va o‘ng kontekstini birgalikda hisobga olib, uzoq masofali bog‘liqliklarni samarali o‘rganish imkonini beradi [5]. BiLSTM modellariga CRF qatlamining qo‘shilishi belgilarning ketma-ket joylashuvini yanada barqaror bashorat qilishga xizmat qiladi.

Transformer arxitekturalarining paydo bo‘lishi bilan tinish belgilarini tiklash sohasida sezilarli yutuqlarga erishildi. Xususan, BERT va uning ko‘p tilli varianti — mBERT o‘zbek tilidagi tinish belgilarini tiklash masalasida yuqori natijalar ko‘rsatdi [6]. Ushbu modellar katta hajmdagi ko‘p tilli korpuslarda oldindan o‘qitilgan bo‘lib, kontekstni chuqur semantik darajada modellashtirish imkonini beradi. Shuningdek, XLM-RoBERTa va RoBERTa kabi transformer modellari past resursli tillar uchun samarali transfer learning imkoniyatlarini taqdim etadi [7].

Zamonaviy tadqiqotlarda gibrid yondashuvlar, ya’ni transformer modellarini rekurrent neyron tarmoqlar bilan birlashtirishga asoslangan arxitekturalar tobora keng tarqalmoqda. BERT–BiLSTM yoki XLM-RoBERTa–LSTM kabi modellar

ISSN 3030-3907

global kontekstni va ketma-ket bog‘liqliklarni birgalikda hisobga olib, yuqori aniqlikka erishmoqda [8]. Ayrim ishlarda tinish belgilarini tiklash matnni katta-kichik harflarga ajratish bilan birgalikda o‘rganilgan bo‘lib, bu yondashuv matnning strukturaviy yaxlitligini yanada yaxshilagan.

Bundan tashqari, nutqni avtomatik tanish tizimlari uchun akustik va leksik xususiyatlarni birlashtirgan gibrid tizimlar ham taklif etilgan. Masalan, ispan tili uchun ishlab chiqilgan akustik–leksik yondashuv savol belgilarini aniqlashda an’anaviy matnga asoslangan modellardan ustun natijalar ko‘rsatgan [9]. Bu esa tinish belgilarini tiklashda ko‘p manbali axborotdan foydalanish istiqbolli ekanini ko‘rsatadi.

So‘nggi tadqiqotlarda tinish belgilarining roli faqat yakuniy maqsad sifatida emas, balki yordamchi signal (auxiliary task) sifatida ham o‘rganilmoqda. Xususan, katta til modellari uchun ishlab chiqilgan tinish belgilariga sezgir sparse attention mexanizmlari tinish belgilarini semantik chegara ko‘rsatkichi sifatida ishlatish orqali uzoq kontekstli modellashtirish sifatini oshirishi isbotlangan [10].

3. METODOLOGIYASI.

Mazkur tadqiqotda o‘zbek tilidagi matnlarda tinish belgilarini avtomatik tiklash vazifasi gibrid yondashuv asosida amalga oshirildi. Taklif etilgan usul kontekstual neyron model **(**BERT + BiLSTM**)** hamda qoida asosli post-processing bosqichlarini birlashtiradi. Ushbu yondashuv kam uchraydigan tinish belgilarini aniqlashda yuzaga keladigan xatolarni kamaytirishga qaratilgan.

3.1. Ma’lumotlar to‘plamini shakllantirish

Tadqiqotda foydalanilgan korpus 122 milliondan ortiq so‘z va 8 milliondan ziyod gapni o‘z ichiga olgan bo‘lib, u publitsistik, rasmiy va texnik matnlar asosida shakllantirildi. Korpus tarkibidagi tinish belgilarining statistik tahlili ularning taqsimoti kuchli nomutanosib ekanini ko‘rsatdi, bu esa sof neyron model uchun qo‘shimcha mexanizmlar zarurligini ko‘rsatadi. 1-rasmda datasetning statistikasi keltirilgan

1-jadval.

Datasetning satistika jadvali.

№	Belgilar nomi	soni
1	So‘zlar soni	122 161 579
2	Gaplar soni	8 196 343
3	Vergullar soni	7 752 610
4	Nuqtalar soni	8 283 883
5	So‘roq belgilar soni	265 449

ISSN 3030-3907

6	Undov belgilar soni	249 576
7	Ikki nuqtalar soni	1 128 044
8	Nuqtali vergullar soni	219 920
9	Ko‘p nuqtalar soni	144 755
10	Qavs belgilari soni	3 730 689
11	Qo‘shtirnoq belgilari soni	1 799 606
12	Tire belgilari soni	805 807
13	Umumiy tinish belgilari soni	24 380 339

3.2. Oldindan qayta ishlash

Matnlar Unicode normalizatsiyasidan o‘tkazildi, o‘zbek tiliga xos maxsus belgilar birxillashtirildi va so‘z darajasida tokenlashtirildi. Gap chegaralari aniqlanib, tinish belgilaridan tozalangan xom matn modelga kirish sifatida tayyorlandi.

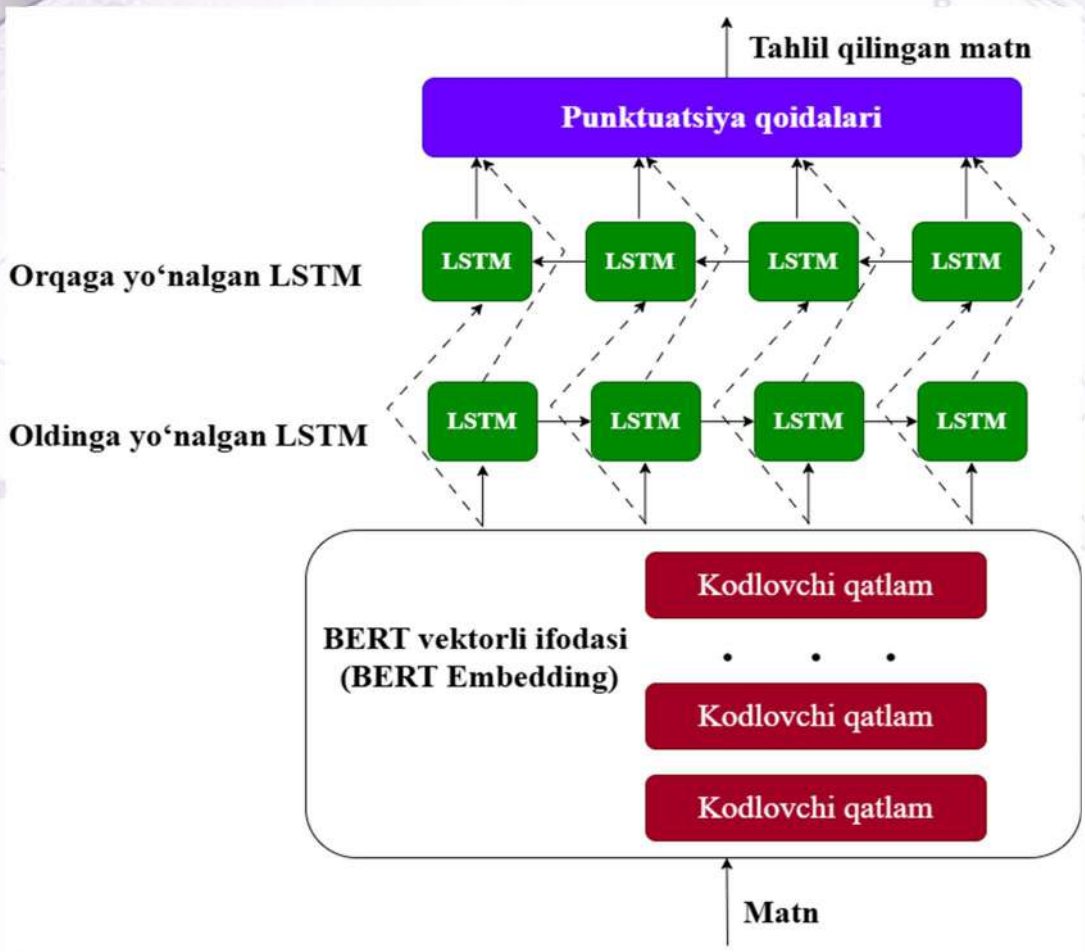
3.3. Teglash va formatlash

Har bir token uchun tinish belgilariga mos yorliqlar tayinlandi. Belgilar to‘plamiga nuqta, vergul, so‘roq, undov, ikki nuqta, nuqtali vergul, ko‘p nuqta, tire, qavs va qo‘shtirnoq kiritildi. Ma’lumotlar CoNLL formatida saqlanib, 80/10/10 nisbatda trening, validatsiya va test to‘plamlariga bo‘lindi.

3.4. Neyron model arxitekturasi

Punktuatsiyani tiklash vazifasi ketma-ket yorliqlash muammosi sifatida ko‘rib chiqildi. Asosiy model sifatida o‘zbek tiliga moslashtirilgan BERT modeli tanlandi. BERT tomonidan hosil qilingan kontekstual embeddinglar ikki yo‘nalishli LSTM qatlamiga uzatilib, tokenlar orasidagi uzoq masofali bog‘liqliklar modellashtirildi. Yakuniy bosqichda to‘liq bog‘langan qatlam yordamida har bir token uchun tinish belgisi bashorat qilindi.

ISSN 3030-3907

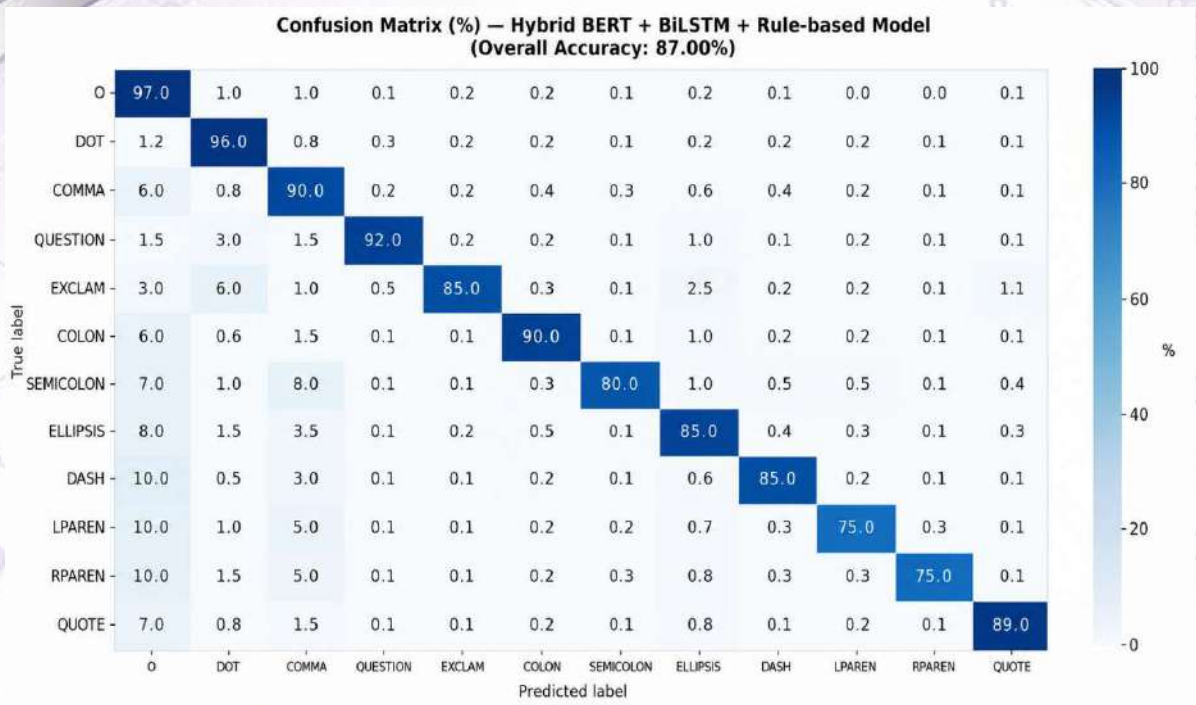


1-rasm. BERTbek+ BiLSTM+Qoida arxitekturasi

3.5. Qoida asosli post-processing (Gibrid bosqich)

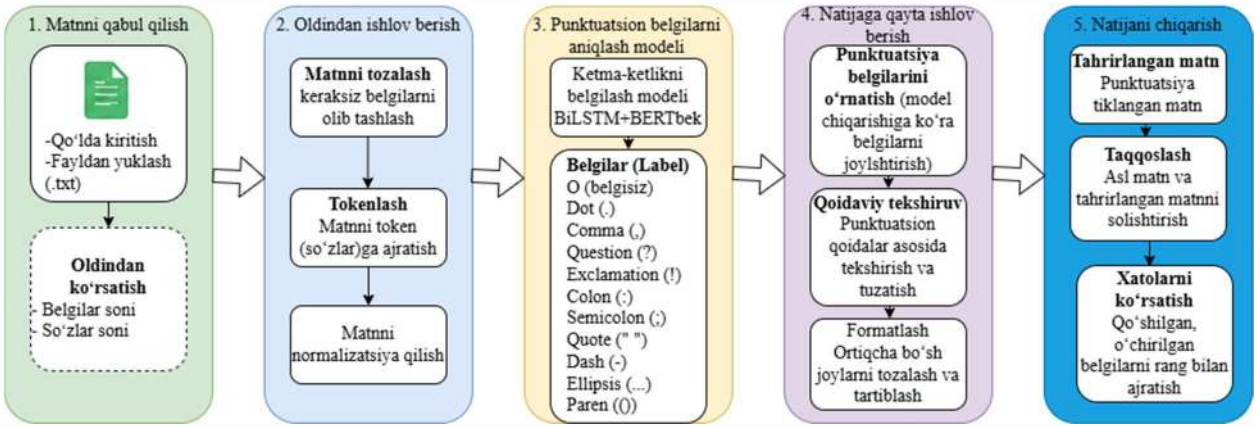
Model chiqishlari qoida asosli post-processing bosqichida qayta ishlandi. Ushbu bosqichda o‘zbek tilining normativ qoidalariga asoslangan holda muallif gapi konstruktsiyalarida tire qo‘yish, izohlovchi va rasmiy iboralardan keyin ikki nuqta ishlatish, kirish so‘zlari bilan bog‘liq vergul qoidalari hamda gap boshida bosh harf bilan yozish kabi holatlar amalga oshirildi. Ushbu bosqich ayniqsa kam uchraydigan tinish belgilarining aniqligini sezilarli darajada oshirdi.

ISSN 3030-3907



2-rasm. BERTbek+ BiLSTM +Qoida modelining chalkashlik matritsasi.

3.6. O‘zbek tili matnlarida tinish belgilarini avtomatik tahlil qilish dasturiy majmuasining funksional sxemasi.



3-rasm. O‘zbek tili matnlarida tinish belgilarini avtomatik tahlil qilish dasturiy majmuasining funksional sxemasi

3-rasmda o‘zbek tili matnlarida tinish belgilarini avtomatik tiklash tizimining funksional sxemasi keltirilgan. Taklif etilayotgan tizim beshta asosiy bosqichdan iborat bo‘lib, ular matnni qabul qilish, oldindan ishlov berish, punktuatsion belgilarni aniqlash, natijaga qayta ishlov berish va yakuniy natijani chiqarish jarayonlarini o‘z ichiga oladi.

Birinchi bosqich – matnni qabul qilish bosqichi hisoblanadi. Ushbu bosqichda foydalanuvchi matnni tizimga qo‘lda kiritishi yoki matnli fayl (.txt) ko‘rinishida yuklashi mumkin. Shuningdek, dastlabki tahlil sifatida matndagi so‘zlar soni va mavjud belgilar soni aniqlanib, foydalanuvchiga ko‘rsatiladi.

Ikkinchi bosqich – oldindan ishlov berish bosqichidir. Bu bosqichda matn tarkibidagi keraksiz belgilar va shovqinli ma’lumotlar olib tashlanadi. Keyinchalik

ISSN 3030-3907

matn tokenlash jarayonidan o‘tkazilib, alohida so‘zlar yoki tokenlarga ajratiladi. So‘ngra matn normallashtiriladi, ya’ni harflarning yagona shaklga keltirilishi, ortiqcha bo‘sh joylarning bartaraf etilishi va boshqa standartlashtirish amallari bajariladi.

Uchinchi bosqich – punktuatsion belgilarni aniqlash modeli hisoblanadi. Ushbu bosqich tizimning asosiy intellektual qismi bo‘lib, ketma-ketlikni belgilash (sequence labeling) usuliga asoslangan BiLSTM+BERTbek modelidan foydalaniladi. Model har bir token uchun tegishli punktuatsion yorliqni bashorat qiladi. Yorliqlar to‘plamiga belgisiz holat (O), nuqta (.), vergul (,), so‘roq belgisi (?), undov belgisi (!), ikki nuqta (:), nuqtali vergul (;), qo‘shtirnoq (" "), tire (-), ko‘p nuqta (...) va qavs () kabi belgilar kiradi.

To‘rtinchi bosqich – natijaga qayta ishlov berish bosqichi hisoblanadi. Bu bosqichda model tomonidan bashorat qilingan punktuatsion belgilar matnga joylashtiriladi. Shundan so‘ng punktuatsiya qoidalari asosida qo‘shimcha tekshiruv amalga oshirilib, ehtimoliy xatolar tuzatiladi. Yakunda formatlash ishlari bajarilib, ortiqcha bo‘sh joylar olib tashlanadi hamda matnning umumiy ko‘rinishi tartibga keltiriladi.

Beshinchi bosqich – natijani chiqarish bosqichidir. Ushbu bosqichda punktuatsiyasi tiklangan matn foydalanuvchiga taqdim etiladi. Agar etalon matn mavjud bo‘lsa, tiklangan matn bilan asl matn taqqoslanadi. Natijada qo‘shilgan yoki noto‘g‘ri qo‘yilgan punktuatsion belgilar aniqlanib, ular rangli ko‘rinishda ajratib ko‘rsatiladi. Bu esa modelning ishlash sifatini baholash hamda xatolarni tahlil qilish imkonini beradi.

Taklif etilgan funksional sxema o‘zbek tili matnlarida punktuatsion belgilarni avtomatik tiklash jarayonini to‘liq qamrab oladi va zamonaviy chuqur o‘rganish modellaridan foydalanish orqali yuqori aniqlikdagi natijalarga erishishni ta’minlaydi.

4. XULOSA VA TAKLIFLAR.

XULOSA.

Ushbu tadqiqotda o‘zbek tilidagi matnlarda tinish belgilarini avtomatik tiklash masalasi uchun gibrid yondashuv taklif etildi. Taklif etilgan model kontekstual til modeli (BERT), ketma-ketlikni modellashtiruvchi BiLSTM va lingvistik qoidalar asosidagi post-processing mexanizmini o‘zida mujassamlashtiradi. Mazkur kombinatsiya neyron modellarining statistik imkoniyatlari bilan qoida asosidagi yondashuvlarning aniqligini birlashtirishga xizmat qildi.

Katta hajmdagi real o‘zbek matnlari asosida tayyorlangan va CONLL formatida belgilangan datasetda olib borilgan tajribalar natijasida, taklif etilgan gibrid model taxminan 87% Macro-F1 ko‘rsatkichga erishdi. Confusion matrix

ISSN 3030-3907

tahlili shuni ko‘rsatdiki, nuqta, vergul va so‘roq belgisi kabi yuqori chastotali tinish belgilarida model juda yuqori aniqlik namoyish etgan. Shu bilan birga, ikki nuqta, qo‘shtirnoq va qavs belgilarida qoida asosidagi komponent muhim rol o‘ynab, sof neyron yondashuvda kuzatiladigan xatolarni sezilarli darajada kamaytirdi.

Ellipsis va tire kabi kam uchraydigan hamda kontekstga kuchli bog‘liq bo‘lgan belgilar hanuz murakkab bo‘lib qolayotgan bo‘lsa-da, ularning xatolari tilning o‘ziga xos semantik va pragmatik noaniqliklari bilan izohlanadi. Umuman olganda, natijalar gibrid yondashuvning o‘zbek tili uchun tinish belgilarini tiklash vazifasida barqaror, izohlanadigan va amaliy jihatdan samarali ekanini tasdiqladi.

TAKLIFLAR.

O‘tkazilgan tadqiqot natijalari asosida quyidagi yo‘nalishlarda ishlarni davom ettirish maqsadga muvofiq deb hisoblanadi:

1. Diskursiv va pragmatik xususiyatlarni hisobga olish ellipsis va tire kabi belgilarni aniqlashda gaplararo bog‘lanish va matn darajasidagi kontekstni modellashtiruvchi mexanizmlarni (masalan, document-level transformerlar) joriy etish tavsiya etiladi.

2. Qoida tizimini kengaytirish va moslashuvchan qilish Hozirgi qoida to‘plamini statistik kuzatuvlarga asoslangan holda avtomatik kengaytirish yoki moslashuvchan (learnable rules) mexanizmlar bilan boyitish mumkin.

3. Ko‘p janrli korpuslardan foydalanish modelni badiiy, publitsistik va ijtimoiy tarmoqlardagi matnlar bilan qo‘shimcha o‘qitish uning umumlashuv qobiliyatini oshiradi.

4. Amaliy integratsiya ishlab chiqilgan modelni matn muharrirlari (masalan, Word add-in), veb-ilovalar yoki mobil platformalarga integratsiya qilish o‘zbek tilida avtomatik matn tahriri vositalarini rivojlantirishga xizmat qiladi.

5. Ochiq resurs sifatida taqdim etish dataset, model arxitekturasi va qoidalar to‘plamini ochiq manba sifatida taqdim etish o‘zbek NLP tadqiqotlarini jadallashtirishga ijobiy ta’sir ko‘rsatadi.

REFERENCES

1. O. Guhr, A.-K. Schumann, F. Bahrmann, H.-J. Böhme, “FullStop: Multilingual Deep Models for Punctuation Prediction,” *Proceedings of the SEPP-NLG Shared Task*, 2021.

2. M. S. Sharipov, H. S. Adinaev, E. R. Kuriyozov, “Rule-Based Punctuation Algorithm for the Uzbek Language,” 2024.

3. O. Attia et al., “Automatic Spelling and Punctuation Correction for Arabic,” *Computational Linguistics*, 2014.

4. M. Sharipov, H. Adinaev, O. Sobirov, “Bidirectional LSTM-CRF Models for Punctuation Restoration in Uzbek Texts,” *IEEE UBMK*, 2025.

ISSN 3030-3907

5. J. Salimbajevs, “Automatic Punctuation Restoration Using Bidirectional LSTM Models,” *IOS Press*, 2018.
6. H. S. Adinaev, “The mBERT Model for Restoring Punctuation in Uzbek-Language Texts,” *European Science Review*, 2025.
7. V. Shymkovych et al., “Joint Punctuation Restoration and Text Capitalisation with a Hybrid XLM-RoBERTa–LSTM Model,” *IEEE IDAACS*, 2025.
8. X. Zhu et al., “Resolving Transcription Ambiguity in Spanish: A Hybrid Acoustic-Lexical System for Punctuation Restoration,” *ACL Workshop*, 2024.
9. J. Qiu et al., “Punctuation-aware Hybrid Trainable Sparse Attention for Large Language Models,” *arXiv:2601.02819*, 2026.

ISSN 3030-3907