KOMPYUTER LINGVISTIKASI; SUN'IY INTELLEKT VA TABIIY TILNI QAYTA ISHLASH

Madaminova Dildora Arabboy qizi
Andijon davlat chet tillari instituti
Filologiya va tillarni o‘qitish: koreys tili
yo‘nalishi talabasi
https://doi.org/10.5281/zenodo.23844728

Annotatsiya. Mazkur maqolada kompyuter lingvistikasining zamonaviy tilshunoslikdagi o‘rni, sun’iy intellekt texnologiyalari bilan o‘zaro aloqasi hamda tabiiy tilni qayta ishlashning nazariy va amaliy jihatlari yoritiladi. Tadqiqotning asosiy maqsadi kompyuter lingvistikasi doirasida sun’iy intellekt algoritmlarining tabiiy tilni tahlil qilish, qayta ishlash va avtomatik tushunish jarayonlaridagi imkoniyatlarini o‘rganishdan iborat. Maqolada tabiiy tilni qayta ishlashning morfologik, sintaktik va semantik tahlil, matnlarni tasniflash, mashina tarjimasi, nutqni tanib olish hamda matn yaratish kabi asosiy yo‘nalishlari tahlil qilinadi. Shuningdek, mashinaviy o‘rganish, chuqur o‘rganish va neyron tarmoqlarga asoslangan modellarni til ma’lumotlariga tatbiq etish masalalariga e’tibor qaratiladi.

Kalit so‘zlar: kompyuter lingvistikasi, sun’iy intellekt, tabiiy tilni qayta ishlash, mashinaviy o‘rganish, chuqur o‘rganish, neyron tarmoqlar, mashina tarjimasi, matn tahlili, o‘zbek tili, til texnologiyalari.

Abstract. This article examines the role of computational linguistics in modern linguistics, its relationship with artificial intelligence technologies, and the theoretical and practical aspects of natural language processing. The main objective of the study is to investigate the capabilities of artificial intelligence algorithms in analyzing, processing, and automatically understanding natural language within the framework of computational linguistics. The article discusses the principal areas of natural language processing, including morphological, syntactic, and semantic analysis, text classification, machine translation, speech recognition, and text generation. Particular attention is paid to the application of machine learning, deep learning, and neural network-based models to linguistic data.

Keywords: computational linguistics, artificial intelligence, natural language processing, machine learning, deep learning, neural networks, machine translation, text analysis, Uzbek language, language technologies.

Аннотация. В статье рассматриваются роль компьютерной лингвистики в современном языкознании, её взаимосвязь с технологиями искусственного интеллекта, а также теоретические и практические аспекты обработки естественного языка. Основная цель исследования заключается в изучении возможностей алгоритмов искусственного интеллекта при анализе, обработке и автоматическом понимании естественного языка в рамках компьютерной лингвистики. В статье анализируются основные направления обработки естественного языка, включая морфологический, синтаксический и семантический анализ, классификацию текстов, машинный перевод, распознавание речи и генерацию текста. Особое внимание уделяется применению машинного обучения, глубокого обучения и моделей на основе нейронных сетей для обработки языковых данных.

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

Ключевые слова: компьютерная лингвистика, искусственный интеллект, обработка естественного языка, машинное обучение, глубокое обучение, нейронные сети, машинный перевод, анализ текста, узбекский язык, языковые технологии.

Kirish. Bugungi globallashuv va raqamli transformatsiya sharoitida axborot-kommunikatsiya texnologiyalarining jadal rivojlanishi inson va kompyuter o‘rtasidagi muloqotni takomillashtirish, katta hajmdagi matnli ma’lumotlarni avtomatik qayta ishlash hamda tabiiy tillarni raqamli muhitga integratsiya qilish masalalarining dolzarbligini oshirmoqda. Ushbu jarayonlarda kompyuter lingvistikasi, sun’iy intellekt va tabiiy tilni qayta ishlash texnologiyalari muhim ilmiy-amaliy ahamiyat kasb etadi. Mazkur yo‘nalishlar tilshunoslik, informatika, matematika va kognitiv fanlar kesishmasida shakllanib, inson tilining qonuniyatlarini hisoblash modellari yordamida o‘rganish va tilga oid vazifalarni avtomatlashtirish imkonini beradi.

Kompyuter lingvistikasi tabiiy tilning fonetik, morfologik, sintaktik, semantik va pragmatik xususiyatlarini kompyuter yordamida modellashtirish, tahlil qilish hamda qayta ishlash tamoyillarini o‘rganuvchi fanlararo soha hisoblanadi. Uning asosiy vazifalari qatoriga elektron lug‘atlar va lingvistik korpuslarni yaratish, matnlarni avtomatik tahlil qilish, mashina tarjimasini amalga oshirish, nutqni tanib olish, matnlarni tasniflash hamda til birliklari o‘rtasidagi munosabatlarni aniqlash kiradi [1]. Mazkur vazifalarni samarali bajarishda sun’iy intellekt texnologiyalari, xususan, mashinaviy o‘rganish, chuqur o‘rganish va neyron tarmoqlarga asoslangan modellar keng imkoniyatlar yaratmoqda.

Sun’iy intellektning rivojlanishi tabiiy tilni qayta ishlash sohasida yangi bosqichni boshlab berdi. An’anaviy qoidalarga asoslangan tizimlardan statistik va neyron modellar tomon o‘tish til hodisalarini kontekst asosida tahlil qilish, so‘zlar o‘rtasidagi semantik munosabatlarni aniqlash va izchil matn yaratish imkoniyatlarini kengaytirdi. Xususan, transformator arxitekturasiga asoslangan modellar katta hajmdagi matnli ma’lumotlar yordamida o‘qitilib, tarjima, matnni umumlashtirish, savol-javob tizimlari va matn generatsiyasi kabi vazifalarda qo‘llanilmoqda. Biroq bunday modellarning samaradorligi o‘quv ma’lumotlarining sifati, hajmi, tilning grammatik xususiyatlari va hisoblash resurslariga bevosita bog‘liq.

Metodlar. Mazkur tadqiqot kompyuter lingvistikasi, sun’iy intellekt va tabiiy tilni qayta ishlash sohalariga oid ilmiy-nazariy qarashlar, mavjud yondashuvlar va texnologik yechimlarni o‘rganishga asoslangan. Tadqiqot metodologiyasi til hodisalarini hisoblash usullari yordamida tahlil qilish, tabiiy tilni qayta ishlash modellarining imkoniyatlarini o‘zaro taqqoslash hamda ularni o‘zbek tilining lingvistik xususiyatlari nuqtayi nazaridan baholash tamoyillarini qamrab oladi. Ushbu yondashuv kompyuter lingvistikasi va sun’iy intellekt o‘rtasidagi nazariy hamda amaliy munosabatlarni izchil o‘rganishga imkon beradi.

Tadqiqot jarayonida bir-birini to‘ldiruvchi bir nechta ilmiy metodlardan foydalanish nazarda tutiladi. Birinchi navbatda, **nazariy tahlil metodi** yordamida kompyuter lingvistikasi, sun’iy intellekt va tabiiy tilni qayta ishlashga oid ilmiy adabiyotlar, konsepsiyalar va mavjud yondashuvlar o‘rganiladi [2]. Ushbu metod asosiy tushunchalarning mazmunini aniqlash, sohaning rivojlanish yo‘nalishlarini tavsiflash va zamonaviy texnologiyalarning nazariy asoslarini yoritish uchun xizmat qiladi.

Ikkinchidan, **qiyosiy tahlil metodi** orqali tabiiy tilni qayta ishlashda qo‘llaniladigan qoidalarga asoslangan, statistik va neyron tarmoq modellarining xususiyatlari taqqoslanadi. Qoidalarga asoslangan yondashuvda lingvistik qoidalar va oldindan belgilangan algoritmlar muhim o‘rin tutsa, statistik usullar ma’lumotlardagi ehtimoliy qonuniyatlardan foydalanadi. Neyron tarmoqlarga asoslangan yondashuvlar esa murakkab til munosabatlarini katta hajmdagi ma’lumotlardan o‘rganishga imkon beradi. Ushbu

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

metod yordamida yondashuvlarning afzalliklari, cheklovlari va qo‘llanish sharoitlarini aniqlash ko‘zda tutiladi.

Uchinchidan, **tizimli tahlil metodi** kompyuter lingvistikasining alohida yo‘nalishlarini yagona texnologik jarayon doirasida ko‘rib chiqishga xizmat qiladi. Bunda matnni dastlabki qayta ishlash, tokenlarga ajratish, morfologik tahlil, sintaktik tuzilmani aniqlash, semantik talqin va yakuniy natijani shakllantirish kabi bosqichlar o‘zaro bog‘liq holda o‘rganiladi. Tizimli yondashuv tabiiy tilni qayta ishlashning bir bosqichida yuzaga keladigan xatolar keyingi bosqichlar natijasiga qanday ta’sir ko‘rsatishi mumkinligini tushuntirishga yordam beradi.

Shuni ta’kidlash lozimki, ushbu metodologik tavsif nazariy-tahliliy tadqiqot uchun mo‘ljallangan. Agar maqolada amaliy tajriba yoki dasturiy model ishlab chiqilgan bo‘lsa, tadqiqotning takrorlanuvchanligini ta’minlash maqsadida foydalanilgan ma’lumotlar to‘plami, namuna hajmi, dasturiy vositalar, model parametrlari, baholash mezonlari va tajriba natijalari alohida ko‘rsatilishi kerak. Agar empirik tajriba o‘tkazilmagan bo‘lsa, matnda tajriba natijalari olingani haqida da’vo qilinmasligi lozim.

Natijalar va tahlil. Ushbu tadqiqot doirasida kompyuter lingvistikasi va tabiiy tilni qayta ishlash texnologiyalarining asosiy yo‘nalishlari o‘zaro bog‘liq holda tahlil qilindi. Tahlil natijalari tabiiy tilni qayta ishlash tizimlarining rivojlanishini uchta asosiy yondashuv — qoidalarga asoslangan, statistik va neyron tarmoqlarga asoslangan modellar orqali tavsiflash mumkinligini ko‘rsatadi.

Tabiiy tilni qayta ishlash usullarining qiyosiy tahlili. Sun’iy intellekt texnologiyalari kompyuter lingvistikasining ko‘plab amaliy yo‘nalishlarini rivojlantirishga xizmat qiladi. Xususan, matnni avtomatik tasniflash, nomlangan obyektlarni aniqlash, mashina tarjimasi, nutqni tanib olish va matn yaratish jarayonlarida neyron modellar muhim o‘rin egallaydi.

Bunday tizimlarning samaradorligi faqat algoritmning murakkabligi bilan belgilanmaydi. O‘quv ma’lumotlarining sifati, ma’lumotlar to‘plamining mavzuviy xilma-xilligi, til birliklarining to‘g‘ri belgilanishi va natijalarni baholash mezonlari ham muhim ahamiyatga ega [3]. Shuning uchun muayyan vazifa uchun model tanlashda aniqlik bilan bir qatorda, uning izohlanishi, resurs talabi va amaliy qo‘llanish sharoitlarini ham hisobga olish lozim.

O‘zbek tilini avtomatik qayta ishlash muammolari. O‘zbek tilida tabiiy tilni qayta ishlash tizimlarini yaratishda morfologik tuzilish, so‘z shakllarining ko‘pligi, imlo variantlari va lingvistik resurslarning yetarliligi alohida e’tibor talab qiladi. Masalan, bir o‘zakka turli qo‘shimchalar qo‘shilishi natijasida hosil bo‘ladigan shakllarni avtomatik tahlil qilish uchun morfologik xususiyatlarni hisobga oluvchi algoritmlar zarur.

O‘zbek tilidagi nutqni tanib olishga oid ilmiy tadqiqotlarda ham til modeli va maxsus lingvistik resurslarning ahamiyati ko‘rsatilgan. Xususan, o‘zbek tili uchun UzLM til modelini ishlab chiqishga bag‘ishlangan tadqiqotda statistik va neyron yondashuvlar o‘rganilib, turli manbalardan yig‘ilgan katta hajmdagi matnli ma’lumotlardan foydalanilgan [4]. Bu holat o‘zbek tilidagi amaliy NLP tizimlarini rivojlantirishda tilga mos ma’lumotlar bazasi va til modellarini shakllantirish muhimligini ko‘rsatadi.

Munozara. Natijalar tahlili shuni ko‘rsatadiki, kompyuter lingvistikasi va sun’iy intellekt o‘rtasidagi hamkorlik tabiiy til bilan bog‘liq vazifalarni avtomatlashtirish uchun keng imkoniyatlar yaratadi. Biroq texnologik rivojlanishni faqat neyron tarmoqlar yoki yirik til modellarining paydo bo‘lishi bilan izohlash yetarli emas. Samarali tizim yaratish uchun lingvistik bilimlar, ma’lumotlar sifati, model arxitekturasi va baholash mexanizmlarini birgalikda hisobga olish talab etiladi.

Birinchi muhim masala — **lingvistik bilimlar va ma’lumotlarga asoslangan o‘rganish o‘rtasidagi munosabat**. Qoidalarga asoslangan modellar tilning aniq grammatik xususiyatlarini ifodalashda qulay bo‘lsa-da, tilning barcha istisnolari va kontekstual holatlarini qamrab olish qiyin. Neyron modellar esa katta hajmdagi ma’lumotlardan murakkab qonuniyatlarni o‘rganishi mumkin, biroq

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

ularning natijalari har doim ham grammatik yoki semantik jihatdan to‘g‘ri bo‘lavermaydi. Shu sababli ayrim vazifalarda lingvistik qoidalar va ma’lumotlarga asoslangan usullarni birlashtiruvchi gibrid yondashuvlar maqsadga muvofiq bo‘lishi mumkin.

Ikkinchi masala — **kontekstni tushunish va semantik aniqlik**. Transformator modellarining diqqat mexanizmi so‘zlar o‘rtasidagi kontekstual munosabatlarni o‘rganishga imkon beradi [5]. Biroq kontekstual bog‘lanishlarni hisobga olish tizimning inson kabi to‘liq tushunish qobiliyatiga ega ekanligini anglatmaydi. Rogers, Kovaleva va Rumshisky tomonidan BERT modeli bo‘yicha olib borilgan sharh tadqiqotida ushbu modellar tilga oid ko‘plab xususiyatlarni o‘zlashtirishi bilan birga, ularning ichki ishlash mexanizmlarini to‘liq tushuntirish murakkabligi ham qayd etilgan [4].

Uchinchi masala — **resurslari cheklangan tillar uchun texnologiyalar yaratish**. Ingliz tili kabi katta hajmdagi raqamli resurslarga ega tillarda ishlab chiqilgan modellarni o‘zbek tiliga bevosita tatbiq etish har doim ham kutilgan natijani bermaydi. O‘zbek tilining morfologik tuzilishi, so‘z yasalishi va qo‘shimchalar tizimi modelning tokenlashtirish, grammatik tahlil va matn yaratish jarayonlariga ta’sir ko‘rsatishi mumkin. Shu sababli ko‘p tilli modellardan foydalanish bilan bir qatorda, o‘zbek tiliga xos korpuslar va maxsus modellarni rivojlantirish ham muhimdir.

Xulosa. Mazkur maqolada kompyuter lingvistikasi, sun’iy intellekt va tabiiy tilni qayta ishlash texnologiyalarining o‘zaro aloqasi, nazariy asoslari hamda amaliy qo‘llanish imkoniyatlari tahlil qilindi. O‘rganilgan ilmiy yondashuvlarni umumlashtirish natijasida tabiiy tilni qayta ishlash zamonaviy tilshunoslikning muhim yo‘nalishlaridan biri bo‘lib, til hodisalarini avtomatik tahlil qilish, matnli axborotni qayta ishlash va inson-kompyuter muloqotini takomillashtirishda muhim ahamiyat kasb etishi aniqlandi.

Foydalanilgan adabiyotlar ro‘yxati

1. Abduraxmonova, N. Z. (2021). Kompyuter lingvistikasi. Darslik. Toshkent: «Nodirabegim» nashriyoti, 280 b.
2. Rahimov, A. (2011). Kompyuter lingvistikasi asoslari. Toshkent: «O‘zbekiston faylasuflari milliy jamiyati» nashriyoti, 160 b.
3. Pulatov, A. K., & Madvaliyev, A. (2018). O‘zbek tili milliy korpusi va kompyuter lingvistikasi masalalari. Toshkent: Fan.
4. Elov, A. R. (2022). O‘zbek tili uchun natural language processing (NLP) texnologiyalarini qo‘llash va avtomatik morfologik tahlil modellashtirish. Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi, Toshkent.
5. Rizayev, O., & Ziyayev, A. (2023). O‘zbek tili matnlarini avtomatik qayta ishlashda semantik va sintaktik tahlil usullari. O‘zbekiston ilmiy-amaliy jurnali, 4(2), 45-52.
6. Sharipov, M., & Mengliyev, B. (2021). O‘zbek tili va sun’iy intellekt: milliy korpus hamda til modellarini yaratish istiqbollari. Til va adabiyot ta’limi, (8), 12-16.

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