ISSN:3030-332X
Impact factor: 9,493
International journal of scientific researchers

KOMPYUTER LINGVISTIKASI: SUN'IY INTELLEKT VA TABIIY TILNI QAYTA ISHLASH (NLP)

Turamirzayeva Zilola
Andijon davlat chet tillari instituti
Xorijiy filologiya va turizm fakultetining
Filologiya va tillarni o'qitish:
koreys tili yo'nalishi 1-kurs talabasi
https://doi.org/10.5281/zenodo.23864162

Annotatsiya. Mazkur maqolada kompyuter lingvistikasi, sun'iy intellekt va tabiiy tilni qayta ishlash (Natural Language Processing — NLP) yo'nalishlarining o'zaro aloqasi, asosiy vazifalari hamda o'zbek tilini raqamli muhitda qayta ishlash bilan bog'liq masalalar yoritiladi. NLP doirasida matnni tokenlarga ajratish, morfologik va sintaktik tahlil, nutq qismlarini belgilash, nomlangan obyektlarni aniqlash, mashina tarjimasi va matn mazmunini avtomatik qayta ishlash kabi vazifalar ko'rib chiqiladi. Shuningdek, o'zbek tilining agglutinativ tuzilishi, imlo variantlari, belgilangan lingvistik ma'lumotlarning yetarli emasligi hamda sifatli korpuslarga ehtiyoj masalalari tahlil qilinadi. Maqolada o'zbek tili uchun NLP vositalarini rivojlantirishda korpus yaratish, lingvistik ekspertiza va texnologik yechimlarni birgalikda olib borish zarurligi asoslanadi.
Kalit so'zlar: kompyuter lingvistikasi, sun'iy intellekt, tabiiy tilni qayta ishlash, NLP, o'zbek tili, lingvistik korpus, morfologik tahlil, mashina tarjimasi, raqamli texnologiyalar.

Аннотация. В статье рассматриваются взаимосвязь компьютерной лингвистики, искусственного интеллекта и обработки естественного языка (Natural Language Processing — NLP), основные задачи данной области и проблемы цифровой обработки узбекского языка. Анализируются токенизация, морфологический и синтаксический анализ, определение частей речи и именованных сущностей, машинный перевод и автоматическая обработка содержания текста. Особое внимание уделяется агглютинативной структуре узбекского языка, вариантам написания, недостатку размеченных лингвистических данных и необходимости создания качественных корпусов. Обосновывается важность сочетания лингвистической экспертизы, корпусных ресурсов и современных технологических решений.
Ключевые слова: компьютерная лингвистика, искусственный интеллект, обработка естественного языка, NLP, узбекский язык, языковой корпус, морфологический анализ, машинный перевод.

Abstract. This article examines the relationship between computational linguistics, artificial intelligence, and Natural Language Processing (NLP), outlining the field's major tasks and the challenges of processing Uzbek in digital environments. It discusses tokenization, morphological and syntactic analysis, part-of-speech tagging, named entity recognition, machine translation, and automated text processing. Particular attention is paid to the agglutinative structure of Uzbek, spelling variation, the limited availability of annotated linguistic data, and the need for reliable corpora. The article argues that progress in Uzbek NLP depends on combining linguistic expertise, corpus development, and modern computational methods.

Keywords: computational linguistics, artificial intelligence, natural language processing, NLP, Uzbek language, linguistic corpus, morphological analysis, machine translation.

Kirish. Raqamli texnologiyalar kundalik hayot, ta'lim, ilm-fan va axborot almashinuvi jarayonlariga chuqur kirib kelayotgan bir paytda inson tilini kompyuter yordamida qayta ishlash masalasi alohida ahamiyat kasb etmoqda. Qidiruv tizimlari, ovozli yordamchilar, avtomatik tarjimonlar, matnni tahlil qilish dasturlari va suhbatlashuvchi tizimlarning ishlashi til birliklarini tanish, ularning o‘zaro munosabatini aniqlash hamda mazmunini hisoblash usullari orqali qayta ishlashga tayanadi. Ana shu vazifalar kompyuter lingvistikasi va tabiiy tilni qayta ishlash sohalarining asosiy tadqiqot obyektini tashkil etadi.

Kompyuter lingvistikasi til hodisalarini formal modellar, algoritmlar va hisoblash vositalari yordamida o‘rganadi. NLP esa ushbu bilimlarni matn va nutqni avtomatik tahlil qilish, tushunishga yaqin natijalar olish hamda til orqali muloqotni yengillashtirish uchun qo‘llaydi. Bu yo‘nalish sun'iy intellekt bilan chambarchas bog‘liq bo‘lsa-da, tilshunoslikning nazariy bilimlarini ham talab qiladi: so‘z shakllari, gap tuzilishi, ko‘pma'nolilik, kontekst va nutq maqsadini hisobga olmasdan turib, ishonchli natijaga erishish qiyin.

O‘zbek tilining raqamli imkoniyatlarini kengaytirish masalasi ham shu jarayonning muhim qismidir. O‘zbek tilida matnlar ko‘p bo‘lsa-da, ularni mashina o‘qiy oladigan, standart asosida belgilangan va turli vazifalarda foydalanishga tayyor shaklda to‘plash doimiy ilmiy-amaliy ishni talab qiladi. Shu sababli mazkur maqolaning maqsadi NLPning nazariy asoslarini yoritish, uning muhim vazifalarini tavsiflash va o‘zbek tili uchun dolzarb bo‘lgan muammolar hamda istiqbolli yo‘nalishlarni ko‘rsatishdan iborat.

Kompyuter lingvistikasi va NLP bo‘yicha ilmiy tadqiqotlar tilni hisoblash nuqtayi nazaridan modellashtirishga qaratilgan. Association for Computational Linguistics (ACL) hamjamiyati va ACL Anthology ilmiy nashrlar bazasida mashina tarjimasi, morfologik tahlil, nutqni qayta ishlash, matnni tasniflash va semantik tahlilga doir ko‘plab tadqiqotlar jamlangan. Ushbu ishlar sohada dastlab qo‘lda tuzilgan grammatik qoidalar ustuvor bo‘lganini, keyinchalik statistik usullar va katta hajmdagi ma'lumotlar asosida o‘rganuvchi neyron modellar keng qo‘llana boshlaganini ko‘rsatadi.

Zamonaviy NLP tizimlari ko‘pincha mashinaviy o‘rganish va chuqur o‘rganish usullaridan foydalanadi. Bunday yondashuvlarda model katta hajmdagi matnlardan til birliklari orasidagi qonuniyatlarni o‘rganadi. Biroq modelning natijasi faqat algoritmga emas, balki o‘rgatish ma'lumotlarining sifati, hajmi, mavzu doirasi va tilning xususiyatlari qanchalik hisobga olinganiga ham bog‘liq. Shu bois, ko‘p resursli tillar uchun yaratilgan yechimlarni boshqa tillarga o‘zgartirishsiz ko‘chirish har doim ham kutilgan samarani bermaydi.

O‘zbek tiliga bag‘ishlangan tadqiqotlarda ham aynan resurslar va lingvistik belgilash masalalari dolzarb ekani ta'kidlanadi. Mengliev va hammualliflarning 2024-yilda chop etilgan ishida o‘zbek tilidagi nomlangan obyektlarni aniqlash uchun belgilangan ma'lumotlar to‘plami va algoritmlar taqdim etilgan. Mualliflar o‘zbek tili uchun maxsus tayyorlangan ma'lumotlarning NLP tadqiqotlari hamda amaliy dasturlar yaratishdagi ahamiyatini ko‘rsatadi. O‘zbek tilidagi mashina tarjimasi haqida chop etilgan tadqiqotlarda esa parallel matnlar yetishmasligi, morfologik murakkablik va tarjima sifatini baholash masalalari muhim muammolar sifatida qayd etiladi.

Demak, mavjud adabiyotlar NLP taraqqiyoti texnologik modellar bilan bir qatorda, sifatli korpus, aniq lingvistik belgilash va tilga xos xususiyatlarni hisobga olishga ham tayanishini ko‘rsatadi.

Tadqiqot metodologiyasi. Maqolada tavsifiy, tahliliy va qiyosiy yondashuvlardan foydalanildi. Dastlab kompyuter lingvistikasi va NLPga oid asosiy tushunchalar umumlashtirildi. Keyin matnni avtomatik qayta ishlashning muhim bosqichlari o‘zbek tilining grammatik xususiyatlari bilan bog‘liq holda ko‘rib chiqildi. Shuningdek, ACL Anthology kabi ilmiy manbalar hamda o‘zbek tiliga doir ochiq

tadqiqotlarda keltirilgan natijalar tahlil qilindi. Maqola eksperimental model ishlab chiqishni emas, balki mavjud ilmiy qarashlar va amaliy muammolarni tizimlashtirishni maqsad qiladi.

Tahlil va natijalar.1.Kompyuter lingvistikasi va NLPning o'zaro munosabati.Kompyuter lingvistikasi tilning tovush, so'z, grammatik shakl, gap va ma'no kabi qatlamlarini kompyuterda ifodalash usullarini o'rganadi. NLP esa bu nazariy va amaliy bilimlarni matn hamda nutq bilan ishlaydigan tizimlarda qo'llaydi. Masalan, matn muharriridagi imlo tekshiruvi so'zning yozilishini lug'at bilan solishtirishi mumkin; murakkabroq tizim esa gapning tuzilishini tahlil qilib, so'zning ayni kontekstdagi vazifasini aniqlashga urinadi.

Sun'iy intellekt tilni qayta ishlashda bir necha vazifani bajaradi: ma'lumotlardan qonuniyatlarni o'rganish, yangi matn yaratish, savollarga javob berish, matnni qisqartirish, hissiy munosabatni aniqlash va turli tillar o'rtasida tarjima qilish. Shunga qaramay, sun'iy intellekt modelining ravon javob berishi uning barcha ma'lumotlari to'g'ri yoki har bir tilning madaniy xususiyatlarini to'liq anglaydi degani emas. Natijalarni tekshirish, manbani baholash va tilshunoslik mezonlarini qo'llash zarur.

2. Tabiiy tilni qayta ishlashning asosiy bosqichlari .NLP tizimida matn odatda bir necha bosqichda qayta ishlanadi. Birinchi bosqich — tokenizatsiya, ya'ni matnni so'zlar yoki kichik birliklarga ajratish. Keyingi bosqichda normalizatsiya orqali imlo va formatdagi farqlar tartibga keltiriladi. Lemmatizatsiya so'zning lug'aviy shaklini aniqlashga, morfologik tahlil esa so'z tarkibidagi grammatik ma'nolarni belgilashga yordam beradi. Masalan, "kitoblarimizdan" shaklini "kitob" leksemasi va ko'plik, egalik hamda chiqish kelishigi ma'nolari bilan bog'liq holda tahlil qilish mumkin.

Nutq qismlarini belgilash (part-of-speech tagging) so'zning ot, fe'l, sifat yoki boshqa turkumga mansubligini aniqlaydi. Sintaktik tahlil gapdagi birliklarning o'zaro munosabatini ko'rsatadi. Nomlangan obyektlarni aniqlash esa shaxs, tashkilot, joy, sana va boshqa maxsus nomlarni ajratib olishga qaratiladi. Ushbu jarayonlar axborot qidiruvi, elektron hujjatlarni saralash, matnlarni tahlil qilish va avtomatik tarjima uchun muhim asos hisoblanadi.

Semantik tahlilda so'z yoki gapning kontekstdagi ma'nosini aniqlashga harakat qilinadi. Chunki bitta so'z turli vaziyatlarda turlicha ma'no berishi mumkin. Masalan, "bosh" so'zi tana a'zosi, rahbar yoki biror narsaning boshlanish qismi ma'nolarida keladi. Bunday ko'pma'nolilikni to'g'ri talqin qilish uchun faqat so'zning o'zi emas, uning atrofidagi birliklar va matn mazmuni ham hisobga olinishi lozim.

3. O'zbek tilini avtomatik qayta ishlashdagi muammolar. O'zbek tili turkiy tillarga xos agglutinativ xususiyatga ega: grammatik ma'nolar ko'pincha so'z negiziga ketma-ket qo'shimchalar qo'shilishi orqali ifodalanadi. Natijada bitta negizdan ko'plab so'z shakllari hosil bo'lishi mumkin. Bu holat morfologik tahlil va lemmatizatsiya tizimlaridan qo'shimchalar ketma-ketligini, ularning vazifasini hamda ayrim shakllarning kontekstga bog'liqligini puxta hisobga olishni talab qiladi.

Ikkinchi muammo yozuv va imlo bilan bog'liq. O'zbek tilidagi apostrof belgilarining turli ko'rinishlarda yozilishi, lotin va kirill yozuvlarining parallel qo'llanishi, norasmiy matnlardagi qisqartmalar va imlo xatolari dastur uchun bir xil birlikni aniqlashni qiyinlashtiradi. Masalan, bir so'z turli klaviatura yoki matn manbalarida turlicha belgi bilan yozilishi mumkin. Normalizatsiya bosqichi bunday farqlarni imkon qadar bir xillashtirishi, lekin mazmunli farqlarni yo'qotmasligi kerak.

Uchinchi muammo — belgilangan ma'lumotlarning cheklanganligi. Modelni o'rgatish yoki baholash uchun shunchaki katta hajmdagi matn yetarli emas; matnlarda so'z turkumlari, sintaktik munosabatlar yoki nomlangan obyektlar ishonchli belgilangan bo'lishi ham muhim. O'zbek tili uchun turli uslub va sohalarni qamrab olgan, mualliflik huquqi va maxfiylik talablariga rioya qilgan holda tuzilgan korpuslar zarur. Bunday resurslar kam bo'lsa, model ayrim mavzu yoki matn turlarida yaxshi, boshqa vaziyatlarda esa sust ishlashi mumkin.

Yana bir masala — ma’no va madaniy kontekst. Maqol, kinoya, ibora, hurmat shakllari, shevaga xos so‘zlar va rasmiy-norasmiy nutq farqlari faqat grammatik qoidalar orqali to‘liq hal etilmaydi. Shu sababli o‘zbekcha NLP tizimlarini yaratishda tilshunoslar, dasturchilar, tarjimonlar va soha mutaxassislarining hamkorligi muhim.

4.NLPning amaliy qo’llanish sohalari.NLPning eng ko‘p uchraydigan qo‘llanishlaridan biri mashina tarjimasidir. Avtomatik tarjima foydalanuvchiga matn mazmunini tez anglashga yordam beradi, biroq so‘zma-so‘z tarjima ibora, kinoya yoki grammatik munosabatlarni noto‘g‘ri berishi mumkin. Shu bois, ayniqsa rasmiy hujjat, ilmiy matn va huquqiy materiallarda avtomatik tarjimani mutaxassis tekshiruvidan o‘tkazish maqsadga muvofiq.

Ta’limda NLP imlo va grammatikani tekshirish, matnni soddalashtirish, o‘quv materiallarini qidirish, avtomatik fikr-mulohaza berish va til o‘rganayotganlarga mashq tayyorlashda qo‘l keladi. Kutubxona va arxivlarda katta hajmdagi hujjatlarni mavzu bo‘yicha tasniflash, tashkilotlarda murojaatlarni yo‘nalishlarga ajratish, ommaviy axborot matnlarida mavzular yoki munosabatlarni tahlil qilish ham uning amaliy imkoniyatlaridandir.

Bunday vositalardan foydalanishda foydalanuvchi ma’lumotlarining maxfiyligi, mualliflik huquqi, natijadagi xatolar va algoritmik tarafkashlik masalalari e’tibordan chetda qolmasligi lozim. NLP tizimi inson qarorini to‘liq almashtiradigan vosita sifatida emas, balki zarur hollarda tekshiriladigan va nazorat qilinadigan yordamchi texnologiya sifatida qaralishi kerak.

O’zbek tili uchun rivojlanish istiqbollari. O‘zbek tilidagi NLP vositalarini rivojlantirish uchun, avvalo, turli janr va uslublardagi matnlardan iborat muvozanatli milliy korpuslarni kengaytirish zarur. Korpusga badiiy adabiyot, ilmiy maqolalar, yangiliklar, rasmiy hujjatlar, ta’limiy materiallar va og‘zaki nutq namunalari kiritilishi mumkin. Har bir matnning manbasi, sanasi, janri va boshqa tavsiflari metama’lumot sifatida saqlansa, undan ilmiy va amaliy foydalanish osonlashadi.

Ikkinchidan, morfologik va sintaktik jihatdan belgilangan ma’lumotlar to‘plamlarini yaratish hamda ularni yagona formatlarda tarqatish lozim. Bunda belgilash yo‘riqnomalari aniq bo‘lishi, bir xil birlik turli mutaxassislar tomonidan imkon qadar bir xil tahlil qilinishi va sifat nazorati yo‘lga qo‘yilishi kerak. Uchinchidan, tayyorlangan vositalarni faqat umumiy aniqlik ko‘rsatkichi bilan emas, balki turli janr, sheva, yozuv turi va murakkab grammatik shakllarda ham sinash zarur.

To‘rtinchidan, ochiq ilmiy hamkorlik muhim. Universitetlar, ilmiy markazlar va texnologik jamoalar korpus, lug‘at, morfologik analizator, imlo tekshirgich va baholash to‘plamlarini bir-biriga mos tarzda ishlab chiqsa, bir xil ishni takrorlash kamayadi. O‘zbek tilining raqamli muhitdagi mavqeini mustahkamlash uchun nafaqat yangi model yaratish, balki uni tekshirish, hujjatlashtirish va uzoq muddat foydalanishga yaroqli holatda saqlash ham zarur.

Xulosa va takliflar. Kompyuter lingvistikasi va tabiiy tilni qayta ishlash bugungi sun’iy intellekt texnologiyalarining muhim ilmiy-amaliy yo‘nalishlaridan biridir. Ushbu soha til birliklarini avtomatik tahlil qilish, matn va nutqni qayta ishlash, tarjima, axborot qidiruvi hamda ta’limiy vositalarni rivojlantirish imkonini beradi. Shu bilan birga, tizimlarning samaradorligi algoritmning murakkabligigagina emas, balki lingvistik bilimlar, ma’lumotlar sifati va sinov metodikasiga ham bog‘liq.

O‘zbek tilida NLPni rivojlantirishda agglutinativ morfologiya, imlo variantlari, ko‘pma’nolilik, kontekst va belgilangan korpuslar yetishmasligi kabi masalalar ustuvor hisoblanadi. Bu muammolarni bartaraf etish uchun quyidagi takliflarni ilgari surish mumkin: birinchidan, o‘zbek tilining turli janrlarini qamrab olgan ochiq va muvozanatli korpuslarni ko‘paytirish; ikkinchidan, morfologik, sintaktik va semantik belgilash mezonlarini standartlashtirish; uchinchidan, yaratilgan dasturlarni ona tili sohiblari va tilshunoslar ishtirokida sinash; to‘rtinchidan, natijalarni shaffof baholash va xatolarni muntazam tahlil

qilish; beshinchidan, ta’lim va ilmiy tadqiqotlarda NLP vositalaridan mas’uliyatli foydalanish madaniyatini shakllantirish.

Shunday qilib, o‘zbek tili uchun samarali sun’iy intellekt yechimlarini yaratish texnologik va lingvistik yondashuvlarning birligini talab etadi. Tilshunoslik asosida tuzilgan sifatli resurslar va izchil baholash tizimi mavjud bo‘lgandagina NLP vositalari o‘zbek tilining boy imkoniyatlarini raqamli muhitda to‘laqonli aks ettira oladi.

Foydalanilgan adabiyotlar ro‘yxati (References)

1. Association for Computational Linguistics. ACL Anthology: computational linguistics, natural language processing and speech technology bo‘yicha ilmiy nashrlar bazasi. https://aclanthology.org/
2. Association for Computational Linguistics. What is the ACL and what is Computational Linguistics? https://www.aclweb.org/portal/what-is-cl
3. Mengliev, D., Barakhnin, V., Abdurakhmonova, N., Eshkulov, M. (2024). Developing named entity recognition algorithms for Uzbek: Dataset insights and implementation. Data in Brief, 54, 110413. https://doi.org/10.1016/j.dib.2024.110413
4. Zokirova, Kh. (2024). O‘zbekistonda mashina tarjimasi: muammolar, yutuqlar va kelajak yo‘nalishlari. Xorijiy lingvistika va lingvodidaktika. https://doi.org/10.47689/2181-3701-vol2-iss4/S-pp203-208
5. Elov, B., Amirkulov, M., Suyunova, M. (2025). Metadata and formatting issues in the Uzbek-English parallel corpus and existing NLP tools for the Uzbek language. Topical Issues of Technical Sciences. https://doi.org/10.47390/ts-v3i9y2025no3
6. IEEE Technology Navigator. Computational Linguistics / Natural Language Processing: overview of core tasks and methods. https://technav.ieee.org/area/computational-linguistics/
7. Abjalova, M. O‘zbek tili multikorpusi – zamonaviy ta’limiy tizim. O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari. https://www.conferences.tsuull.uz/index.php/conferences/article/view/763