

ISSN 2181-922X

LANGUAGE & CULTURE

UZBEKISTAN O'ZBEKISTON

UZBEKISTAN

TIL VA MADANIYAT

KOMPYUTER
LINGVISTIKASI

2024 Vol. 1 (6)

www.compling.tsuull.uz

ISSN 2181-922X

O‘ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2024 Vol. 1 (6)

compling.tsuull.uz

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti

Bosh muharrir:

Botir Elov

Bosh muharrir o'rinbosari:

Shahlo Hamroyeva

Mas'ul kotib:

Oqila Abdullayeva

Tahrir kengashi

Shuhrat Sirojiddinov (O'zbekiston), Eshref Adali (Turkiya), [Viktor Zaxarov] (Rossiya), Vladimir Benko (Slovakiya), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat O'rxun (Turkiya), Suyun Karimov (O'zbekiston), Abduvali Qarshiyev (O'zbekiston), Muxammadjon Musayev (O'zbekiston), Kamoliddin Shukurov (O'zbekiston), O'tkir Hamdamov (O'zbekiston), Tal'at Zuparov (O'zbekiston), Bahodir Mo'minov (O'zbekiston), Faxriddin Nurullayev (O'zbekiston), Zulxumor Xolmanova (O'zbekiston), Muqaddas Abdurahmonova (O'zbekiston), Habibulla Madatov (O'zbekiston), Azizaxon Raxmanova (O'zbekiston), Ruhillo Alayev (O'zbekiston), Rasuljon Atamuratov (O'zbekiston), Malika Abdullayeva (O'zbekiston), Mannon Ochilov (O'zbekiston), Xolisa Axmedova (O'zbekiston), Zilola Xusainova (O'zbekiston).

Jurnal haqida ma'lumot

“O'zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi – Oliy attestatsiya komissiyasi ilmiy nashrlar ro'yxatidagi “O'zbekiston: til va madaniyat” akademikjurnalining ilovasi hisoblanib, unda professor-o'qituvchilar, doktorantlar, stajor-tadqiqotchilar, mustaqil izlanuvchilar, magistrantlarning kompyuter lingvistikasi, jumladan, tabiiy tilga ishlov berish (NLP), o'zbek tilining formal grammatikasi, korpus lingvistikasi, mashina tarjimai, nutqni qayta ishlash tizimlari, intellektual tizimlar, kompyuter leksikografiyasi hamda lingvistik ontologiyalar kabi sohalarga oid tadqiqotlari nashr qilinadi.

Jurnal ilovasi bir yilda to'rt marta chop etiladi.

O'zbek, turk, rus va ingliz tillarida yozilgan maqolalar qabul qilinadi.

Jurnalda kitoblarga yozilgan taqrizlar, adabiyotlar sharhi, konferensiyalar hisobotlari va tadqiqot loyihalari natijalari ham e'lon qilinadi.

Mualliflar fikri tahririyat nuqtayi nazaridan farq qilishi mumkin.

“O'zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi 2023-yildan chiqa boshlagan.

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti. O'zbekiston, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko'chasi, 103-uy.

E-mail: kompling@navoiy-uni.uz

Website: compling.tsuull.uz

Alisher Navo'i Tashkent State University of the Uzbek Language and Literature

Chief editor:	Botir Elov
Deputy editor-in-chief:	Shahlo Hamroyeva
Responsible secretary:	Oqila Abdullayeva

Editorial board

Shukhrat Sirojiddinov (Uzbekiston), Eshref Adali (Turkiye), [Viktor Zakharov] (Russia), Vladimir Benko (Slovakia), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat Orhun (Turkey), Suyun Karimov (Uzbekistan), Abduvali Karshiyev (Uzbekistan), Mukhammadjon Musayev (Uzbekistan), Kamoliddin Shukurov (Uzbekistan), O'tkir Hamdamov (Uzbekistan), Tal'at Zuparov (Uzbekistan), Bahadir Mo'minov (Uzbekistan), Fakhridin Nurullayev (Uzbekistan), Zulkhumor Kholmanova (Uzbekistan), Muqaddas Abdurakhmonova (Uzbekistan), Habibulla Madatov (Uzbekistan), Azizakhan Raxmanova (Uzbekiston), Ruhillo Alayev (Uzbekistan), Rasuljon Atamuratov (Uzbekistan), Malika Abdullayeva (Uzbekistan), Mannon Ochilov (Uzbekistan), Kholisa Akhmedova (Uzbekistan), Zilola Khusainova (Uzbekistan).

Information about the magazine

"Uzbekistan: language and culture. "Computer Linguistics" series is an appendix of the academic journal "Uzbekistan: Language and Culture" in the list of scientific publications of the Higher Attestation Commission, in which computer linguistics, including natural language processing (NLP) of professors-teachers, doctoral students, intern-researchers, independent researchers, master's students, researches related to formal grammar of the Uzbek language, corpus linguistics, machine translation, speech processing systems, intelligent systems, computer lexicography and linguistic ontologies are published.

The magazine supplement is published four times a year.

Articles written in Uzbek, Turkish, Russian and English languages are accepted.

The journal also publishes book reviews, literature reviews, conference reports, and research project results.

The opinion of the authors may differ from the editorial point of view.

"Uzbekistan: language and culture. "Computer Linguistics" series has been published since 2023.

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi. Yusuf Khos Hajib street, 103, Yakkasaray district, Tashkent, Uzbekistan.

E-mail: kompling@navoiy-uni.uz

Website: kompling.tsuull.uz

MUNDARIJA

Firuza Nurova

Jahon kompyuter lingvistikasida bir necha soʻz-shakldan iborat leksemalarga ishlov berish tajribasi haqida6

Iqbola Xolmonova

Oʻzbek-turk parallel korpusi uchun matnlar tokenizatsiyasi masalasi.....21

Botir Elov, Shahlo Hamroyeva, Marjona Hamroqulova

Nlpda semantik teglash usullari.....32

Oqila Abdullayeva

Oʻzbek tili matnlarida sintaktik teg va teglash masalasi.....46

Aziza Raxmanova

Modern methods of teaching the linguistic basics of the uzbek and english languages.....58

Nargiza Shamiyeva

The main principals of creating a bilingual thesaurus for the uzbek language.....66

Zarnigor Khayatova

Uzbek paraphrasing software: how your words get a makeover (without losing their meaning!).....78

CONTENT

Firuza Nurova

About the experience of processing lexemes consisting of several word forms in world computer linguistics.....19

Iqbola Xolmonova

The issue of text tokenization for the uzbek-turkish parallel corpus.....31

Botir Elov, Shahlo Hamroyeva, Marjona Hamroqulova

Semantic tagging methods in nlp.....44

Oqila Abdullayeva

The issue of syntactic tags and tagging in uzbek language texts.....56

Azizaxon Raxmanova

O'zbek va ingliz tillarining lingvistik asoslarini o'qitishning zamonaviy metodlari.....64

Nargiza Shamiyeva

O'zbek tili uchun bilingval tezaurus yaratishning asosiy tamoyillari.....76

Zarnigor Xayatova

O'zbekcha parafrazlash dasturi: sizning so'zlaringiz qanday o'zgaradi? (ma'noni saqlagan holda).....85

JAHON KOMPYUTER LINGVISTIKASIDA BIR NECHA SO'Z-SHAKLDAN IBORAT LEKSEMALAR-GA ISHLOV BERISH TAJRIBASI HAQIDA

Firuza Nurova¹

Annotatsiya. Mazkur maqolada jahon kompyuter lingvistikasida bir necha so'z-shakldan iborat leksemalarga ishlov berish tajribasi haqida so'z boradi. Bir necha so'z-shakldan iborat leksemalarni tarkibiy qismlarga ajratish, sintaktik va semantik tahlilni o'z ichiga olgan turli yondashuvlar va texnologiyalar yuzasidan olib borilgan tadqiqot yo'nalishlari tahlil qilindi. O'zbek tilshunosligida bir necha so'z-shakldan iborat analitik shakllarning matn tarjimasidagi muammolari hamda yechilishi kerak bo'lgan masalalari yoritilgan. Shuningdek, ko'p so'z shakllaridan tashkil topgan leksemalarni qayta ishlash mavzusi hisoblash tilshunosligining muhim tadqiqot sohasi bo'lib, turli davrlarda bu sohaga hissa qo'shgan olimlar haqida ham ma'lumot berilgan. Kompyuter tilshunosligida bir necha so'z shakllaridan tashkil topgan leksemalarni qayta ishlashning murakkabligi ularning noaniqligi, shakl va ma'nolarning xilma-xilligi bilan bog'liq bo'lib, bu kontekstni hisobga oladigan va to'g'ri so'z birikmalarini ajratishni amalga oshiradigan, iboralarning ma'nosini tushunadigan samarali tahlil usullarini ishlab chiqishni taqozo etishi xususidagi mulohazalar aytib o'tilgan.

Kalit so'zlar: *kompyuter lingvistikasi, morfologik tahlil algoritmi, semantik tahlil algoritmlari, semantik modellar, tabiiy til, morfoanalizator, fe'l-larning analitik shakllari, semantika, matn korpuslari, til texnologiyalari, Deep Learning, ko'p tilli tizimlar.*

Ko'p so'z shakllaridan tashkil topgan leksemalarni qayta ishlash mavzusi hisoblash tilshunosligining muhim tadqiqot sohasi bo'lib, turli davrlarda bu sohaga ko'plab olimlar hissa qo'shgan. Quyida ba'zi mashhur tadqiqotchilar va ularning ushbu mavzuga qo'shgan hissalarini keltirilgan:

1. Noam Xomskiy (Noam Chomskiy):– Noam Xomskiy kompyuter tilshunosligi va umuman nazariy tilshunoslik sohasidagi yetakchi tadqiqotchilardan biridir. Uning grammatika nazariyasi va tilning sintaktik tahliliga oid ishlari ko'p so'zli leksemalarni qayta ishlash usullarini ishlab chiqishga sezilarli ta'sir ko'rsatdi.

¹Nurova Firuza Kamolovna – Buxoro davlat pedagogika instituti O'zbek va rus tillari kafedrasida katta o'qituvchisi.

E-pochta: firuzaru870@gmail.com

ORCID: 0009-0005-3510-6483

2. Karen Spark Jonson (Karen Sparck Jones):- Karen Spark Jonson axborot qidirish va tabiiy tilni qayta ishlash sohasiga katta hissa qoʻshgan. Uning statistik matn tahlili va axborot olish usullari boʻyicha ishi koʻp soʻzli leksemalarni qayta ishlashda muhim ahamiyatga ega.

3. Jeyms Martin (James Martin):- Jeyms Martin hisoblash tilshunosligi va sunʼiy intellekt sohasidagi tadqiqotlari bilan mashhur. U koʻp soʻzli leksemalarni qayta ishlash algoritmlari va usullarini, jumladan, sintaktik va semantik tahlilni yaratishga katta hissa qoʻshgan.

4. Yoxan Bos (Johan Bos):- Yoxan Bos tabiiy tilning semantik tahlili va koʻp soʻzli leksemalarni qayta ishlashning semantik modellarini ishlab chiqishga ixtisoslashgan. Uning hisoblash tilshunosligi boʻyicha olib borgan tadqiqotlari kontekstdagi soʻz va iboralarning maʼnosini tushunishga muhim hissa qoʻshadi.

5. Jeyson Ismey (Jeyson Eysner):- Jeyson Ismey statistik modellashtirish va hisoblash tilshunosligida mashinalarni oʻrganish boʻyicha tadqiqotlari bilan mashhur. Uning ijodi koʻp soʻz leksemalarni qayta ishlashning turli jihatlariga, jumladan, morfologik va sintaktik tahlilga tegishli.

Bu kompyuter lingvistikasida koʻp soʻzli leksemalarni qayta ishlash sohasiga katta hissa qoʻshgan tadqiqotchilarning kichik roʻyxati. Boshqa koʻplab olimlar tabiiy tildagi matnlarni yanada toʻgʻri va samarali tahlil qilish usullari va texnologiyalarini yanada rivojlantirishga intilib, bu boradagi faoliyatini davom ettirmoqda.

Kompyuter tilshunosligi olamida kompyuter tizimlari tomonidan tabiiy tilni samarali tushunish va tahlil qilish maqsadida koʻp soʻzli leksemalarni qayta ishlash algoritmlari va usullari faol oʻrganilmoqda. Buning uchun tarkibiy qismlarga ajratish, sintaktik va semantik tahlilni oʻz ichiga olgan turli yondashuvlar va texnologiyalar qoʻllaniladi. Keling, ulardan baʼzilarini batafsil koʻrib chiqaylik:

1. Komponent qismlarga ajratish:

- Koʻp soʻzli leksemalarni qayta ishlash uchun koʻpincha tahlil qilish usuli yoki morfologik tahlildan foydalaniladi. Bu jarayon morfema va soʻz shakllarini ajratib olish, shuningdek, ularni gap boʻlaklari va boshqa grammatik belgilarga koʻra tasniflashni oʻz ichiga oladi.

- Morfologik tahlil algoritmlari grammatik xususiyatlarini aniqlash uchun soʻz shakllari va qoidalari lugʻatlaridan foydalanishni oʻz ichiga olishi mumkin.

2. Sintaktik tahlil:

- Ko'p so'zli leksemalar tarkibiy qismlarga ajratilgandan so'ng sintaktik tahlilga o'tkaziladi, bu ularning sintaktik tuzilishini, ifodadagi so'zlar o'rtasidagi munosabatlarni aniqlash imkonini beradi.

- Sintaktik tahlil qilish algoritmlari turli usullarga asoslaniishi mumkin, masalan, kontekstsiz grammatika, rekursiv algoritmlar, bog'liqlik usullari va boshqalar.

- Izohlash uchun ishlatilishi mumkin bo'lgan tahlil daraxti yoki gapning sintaktik tuzilishi keyingi sintaktik tahlilning natijasi bo'lib xizmat qiladi.

3. Semantik tahlil:

- Ko'p so'zli leksemalar sintaktik tahlildan so'ng ularning ma'nosi va semantik munosabatlarini aniqlashga qaratilgan semantik tahlilga o'tkaziladi.

- Semantik tahlil algoritmlari turli usullardan foydalanishi mumkin, jumladan semantik tarmoqlar, so'zlarning vektor tasvirlari, kontekst tahlili va boshqalar.

- Semantik tahlil natijasi ko'p so'zli leksemalarning ma'nosini izohlash bo'lib, ular yordamida turli vazifalarni, ya'ni mashina tarjimai, matnni tahlil qilish, savollarga javob berish va boshqalarni bajarish mumkin.

Ko'p so'zli leksemalarni qayta ishlash sohasidagi tadqiqotlar yangi algoritmlar va usullarni ishlab chiqish, shuningdek, ularni turli tillar va matnni qayta ishlash vazifalari uchun moslashtirish va optimallashtirishni o'z ichiga oladi. Ushbu yondashuv bizga keng ko'lamli ilovalarda, jumladan, qidiruv tizimlari, chatbotlar, matnlarni izohlash va boshqalarda qo'llanilishi mumkin bo'lgan aniqroq va samarali tabiiy tillarni qayta ishlash tizimlarini yaratishga imkon beradi.

Ko'p so'zli leksemalarni qayta ishlash usullari va algoritmlarini yanada takomillashtirish matnni avtomatik tahlil qilish va ularni kompyuter tizimlari tomonidan talqin qilishning aniqligi va sifatini oshirish kompyuter lingvistikasining asosiy tadqiqot yo'nalishi hisoblanadi. Bu tabiiy tilni samarali qayta ishlash va tushunish kontekstida muhim ahamiyatga ega, bu ko'plab ilovalar, jumladan, mashina tarjimai, avtomatik umumlashtirish, matnni chuqur tahlil qilish va boshqalar uchun asosdir.

Keyingi takomillashtirishning asosiy yo'nalishlari quyidagilardan iborat:

1. Morfologik tahlil algoritmlarini takomillashtirish:

- so'zlar va so'z shakllarining morfologik xususiyatlarini ajra-

tib ko'rsatishning aniqroq va samarali algoritmlarini ishlab chiqish, bu ularning grammatik xususiyatlarini va kontekstual ishlatilishini aniqroq tahlillash imkonini beradi.

2. Yuqori darajadagi sintaktik modellarni ishlab chiqish:

- nafaqat mahalliy grammatik qoidalarni, balki butun jumlarlar va matnlarning kontekstual xususiyatlarini hisobga olishga qodir bo'lgan yanada murakkab va moslashtirilgan sintaktik modellarni yaratish.

3. Semantik tadqiqot usullaridan foydalanish:

- so'z va iboralarning ma'nosini chuqurroq tahlil qilish, ularning turli kontekstlardagi semantik aloqalari va ma'nolarini hisobga olgan holda, so'zlarning vektorli tasviri va kontekstual tushunishga asoslangan semantik modellarni amalga oshirish.

4. Ma'nolarni ajratish usullarini ishlab chiqish:

- matndagi so'z va iboralarning ma'nolarini noaniqliklarni bartaraf etish va kontekst va semantik xususiyatlardan kelib chiqqan holda eng ehtimoliy ma'noni tanlash uchun yangi yondashuvlarni tadqiq qilish va ishlab chiqish.

5. Mashinani o'rganish va chuqur o'rganish integratsiyasi:

- Matndagi birikmalarni avtomatik ravishda ajratib olish va tahlil qilish uchun kompyuter lingvistikasini o'rganish va chuqur o'rganish usullaridan foydalanish, bu sizga ko'p so'zli leksemalarni qayta ishlash uchun yanada samarali va moslashuvchan modellarni yaratish imkonini beradi.

Ushbu yaxshilanishlarning maqsadi tabiiy tildagi matnlarni yuqori darajadagi aniqlik va kontekstni tushunish bilan avtomatik ravishda tahlil qila oladigan va talqin qila oladigan aniqroq va ishonchli kompyuter tizimlarini yaratishdir. Bu tabiiy tillarni qayta ishlash sohasida yangi va takomillashtirilgan ilovalarni ishlab chiqish va kompyuter tizimlari bilan aloqa va o'zaro aloqa sifatini oshirish imkonini beradi.

Bir necha so'z shakllaridan tashkil topgan leksemalarni qayta ishlashda morfologik, sintaktik va semantik modellardan foydalanish asosiy o'rin tutadi, chunki bu ularning tuzilishi va semantik munosabatlarini chuqurroq anglash imkonini beradi. Keling, ushbu modellarning har biri kompyuter tilshunosligida qanday qo'llanilishini batafsil ko'rib chiqaylik:

1. Morfologik modellar:

- Morfologik modellar so'zlarning tuzilishi va grammatik belgilarini, masalan, gap bo'lagi, hol, son, zamon va hokazolarni tahlil qilish imkonini beradi.

- Ko'p so'zli leksemalarni qayta ishlashda morfologik model-lar alohida so'z shakllarini aniqlash va ularning grammatik xususi-yatlarini aniqlashga yordam beradi, bu esa keyingi sintaktik va se-mantik tahlil uchun muhim ahamiyatga ega.

- Masalan, morfologik tahlil modeli "chiroyli" so'zining oddiy darajadagi sifat ekanligini aniqlashi mumkin.

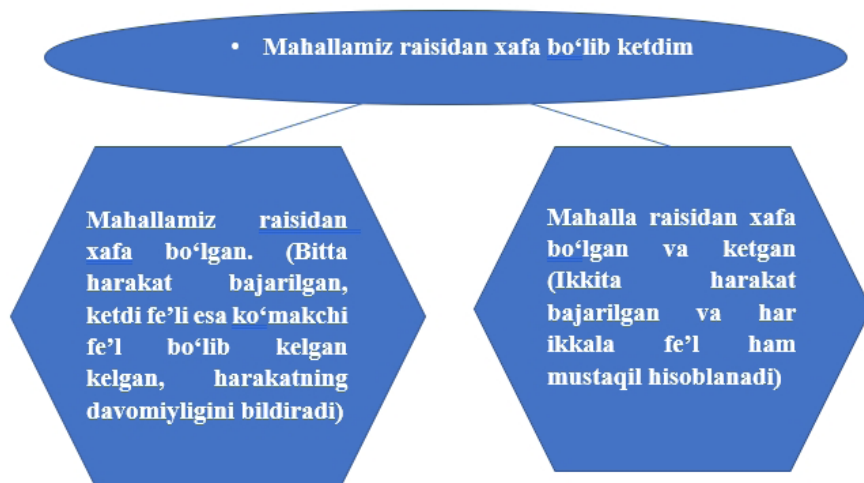
2. Sintaktik modellar:

- Sintaktik modellar gaplarning sintaktik tuzilishini tahlil qilishga, so'zlar orasidagi bog'lanishni va ularning gapdagi rollarini aniqlashga yordam beradi.

- Ko'p so'zli leksemalarni qayta ishlash sharoitida sintaktik modellar so'zlar orasidagi grammatik munosabatni aniqlashga, so'z birikmasini tarkibiy qismlarga ajratishga yordam beradi.

- Masalan, "chiroyli uy" iborasida "uy" so'zi ot, "chiroyli" esa uning sifatlovchisi ekanligini va ular "sifat+ot" sintaktik yasalishi-ni sintaktik model yordamida aniqlashi mumkin. Yoki "xafa bo'lib ketdi" birikmasi ikki xilda talqin qilinadi: Mahallamiz raisidan xafa bo'lib ketdim. (X.To'xtaboyev) gapi orqali bu birikmani tahlil qilamiz, ya'ni bu ko'p so'zli so'z-shaklni so'z birikmasi, ham fe'lning analitik shakli hisoblanmish ko'makchi fe'lli so'z qo'shilmasi bo'lib kela oladi

- Semantik modellar so'z va iboralarning ma'nosini tahlil qilish bilan birga ularning semantik munosabatlarini aniqlashga yor-dam beradi.



- Ko'p so'zli leksemalarni qayta ishlashda semantik modellar alohida so'zlarning ma'nolarini va ularning o'zaro munosabatlarini tushunishga yordam beradi, bu esa so'z birikmasini bir butun sifati-da to'g'ri talqin qilishda muhim ahamiyatga ega.

- Masalan, semantik model "chiroyli uy" iborasida "chiroyli"

so'zi "tashqi ko'rinishda idrok etilishiga ma'qul" xususiyatni, "uy" so'zi esa bu belgi qaysi obyektни bildirishini aniqlashi mumkin.

Ushbu modellarning barchasi bir-biriga o'zaro ta'sir qiladi va birgalikda kompyuter tizimlariga tabiiy tilni aniqroq tushunish va tahlil qilish imkonini beradi. Ko'p so'zli leksemalarni qayta ishlashda matn tahlilining yanada samarali va aniq usullarini yaratish uchun morfologik, sintaktik va semantik modellardan foydalanish zarur bo'lib, bu hisoblash tilshunosligining mashina tarjimai, matnни idrok qilish, axborotlashtirish kabi turli muammolarini hal qilishda muhim ahamiyatga ega.

O'zbek tili morfoanalizatori uchun fe'llarning analitik shakllarini modellashtirish, ayniqsa, bir necha so'z shakllaridan tashkil topgan leksemalar haqida gap ketganda, hisoblash tilshunosligi doirasidagi qiziqarli tadqiqot sohasidir. Shu nuqtayi nazardan, turli jihatlar, jumladan, o'zbek tilining morfologik murakkabligi, fe'llarni tahlil qilish va tegishli ishlov berish usullarini ishlab chiqish ko'rib chiqiladi.

1. O'zbek tilining morfologik murakkabligi:

- O'zbek tili turkiy tillar oilasiga mansub bo'lib, murakkab morfologiyaga, jumladan, fe'l shakllari va analitik konstruksiyalarning keng tizimiga ega. Bu morfologik tahlil va fe'llarni ajratishda qiyinchiliklar tug'diradi.

2. Fe'llarning analitik shakllarini tahlil qilish:

- O'zbek tilidagi fe'lning analitik shakllari fe'l o'zak va yordamchi fe'l yoki ko'makchi fe'llardan tashkil topgan, zamon va modal ma'nolarni ifodalash uchun xizmat qiluvchi konstruksiyalardir. Bunday yasashga misol qilib "o'qimoq" (o'qimoq) + "olmoq" (harakatni bajarishga imkoniyat holati)ni keltirish mumkin.

3. Modellashtirish va qayta ishlash:

- Fe'lning analitik shakllarini samarali qayta ishlash uchun bunday konstruksiyalarni to'g'ri tanib, tahlil qilishga qodir model-lar va morfoanaliz algoritmlarini ishlab chiqish bo'yicha izlanishlar olib borilmoqda. Bunga grammatik qoidalar va lug'atlarni yaratish, tizimni analitik shakllarni tan olishga o'rgatish uchun mashinani o'rganish usullaridan foydalanish kiradi.

4. Kontekst va semantikani hisobga olgan holda:

- Fe'llarning analitik shakllarini modellashtirishda gapning konteksti va semantikasini hisobga olish muhimdir. Masalan, konstruksiyalarning temporal ma'nolarini to'g'ri aniqlash uchun atrofda-gi so'zlarni va kontekst xususiyatlarini tahlil qilish kerak.

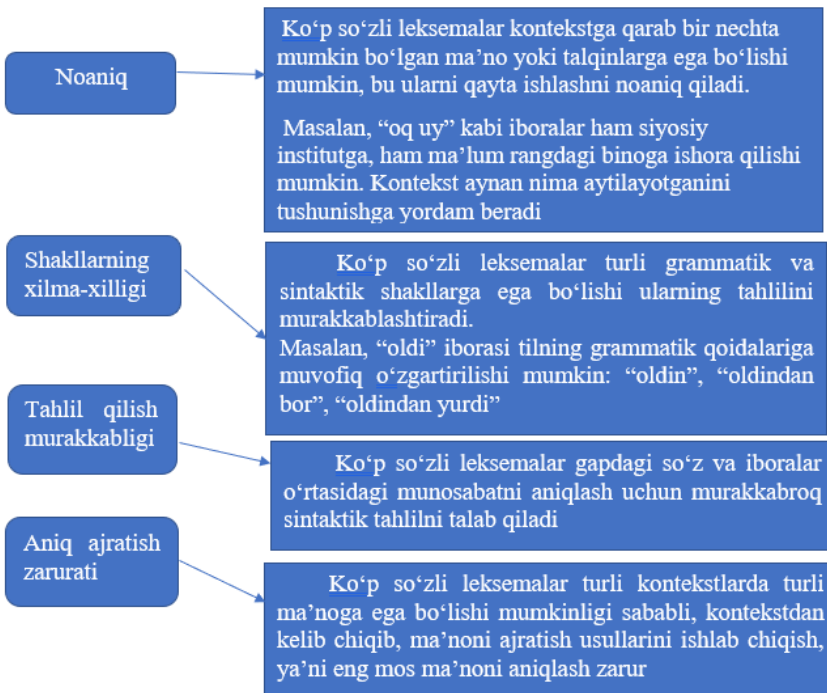
5. Tajribalar va sifatni baholash:

- Tadqiqotlar doirasida fe'llarning analitik shakllarini qayta

ishlash bo'yicha ishlab chiqilgan uslub va modellarning sifatini baholash bo'yicha tajribalar o'tkazilmoqda. Bu taklif etilayotgan yondashuvlarning samaradorligi va to'g'riligini aniqlash uchun turli matn korpuslarida test o'tkazish va natijalarni tahlil qilishni o'z ichiga oladi.

Demak, kompyuter tilshunosligi doirasida o'zbek tili morfoanalizatori uchun fe'llarning analitik shakllarini modellashtirish bo'yicha olib borilayotgan tadqiqotlar tilning morfologik tuzilishini tahlil qilishdan tortib, tegishli ishlov berish usullari va modellarini ishlab chiqish va sinab ko'rishgacha bo'lgan turli bosqichlarni o'z ichiga oladi. Bu tadqiqotlar til texnologiyalarini rivojlantirish, o'zbek tilidagi matnlarni yanada to'g'ri va to'liq tahlil qilishni ta'minlashda muhim ahamiyatga ega.

Bir nechta so'z shakllaridan tashkil topgan leksemalarni qayta ishlash bir nechta sabablarga ko'ra kompyuter tilshunosligida qiyin vazifadir:



Semantik tahlilning murakkabligi ham kompyuter tilshunosligida qiyin vazifa sanaladi:

- Ko'p so'zli leksemalarning ma'nosini tushunish uchun semantik tahlil, ya'ni ifodadagi so'zlar o'rtasidagi ma'no va munosabatlarni aniqlash talab etiladi.

- Masalan, "kompyuterda ishlash" iborasi uchun siz kompyuterdan foydalanish bilan bog'liq harakat haqida

gapirayotganimizni tushunishingiz kerak.

Umuman olganda, kompyuter tilshunosligida bir necha so'z shakllaridan tashkil topgan leksemalarni qayta ishlashning murakkabligi ularning noaniqligi, shakl va ma'nolarning xilma-xilligi bilan bog'liq bo'lib, bu kontekstni hisobga oladigan va to'g'ri so'z birikmalarini ajratishni amalga oshiradigan, iboralarning ma'nosini tushunadigan samarali tahlil usullarini ishlab chiqishni taqozo etadi.

Ko'p so'zli leksemalarni qayta ishlash global hisoblash tilshunosligida, ayniqsa, turkiy tillar haqida gap ketganda bir qator afzalliklarga ega. Asosiy afzalliklardan ba'zilari quyida keltirilgan:

1. Morfologik murakkablikni hisobga olgan holda:

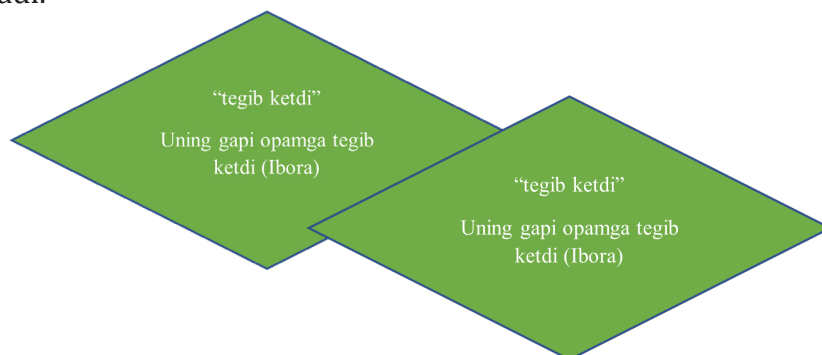
- Turkiy tillar boy morfologiyaga ega, ya'ni so'zlar kontekst va grammatik qoidalarga qarab turli shakllarga ega bo'lishi mumkin. Ko'p so'zli leksemalarni qayta ishlash ushbu morfologik murakkablikni hisobga olish va tahlil algoritmlarini muayyan kontekstga moslashtirish imkonini beradi.

2. Alohida hal qilish:

- Turkiy tillarda ko'pincha kontekstga qarab bir nechta ma'noga ega bo'lgan so'zlar mavjud. Ko'p so'zli leksemalarni qayta ishlash noaniqlikni hal qilish va so'zning eng mos ma'nosini tanlash uchun kontekstdan foydalanish imkonini beradi.

3. Mashina tarjimasini yaxshilandi:

- Ko'p so'zli leksemalarni qayta ishlash turkiy tillar uchun mashina tarjimasini sifatini oshirishga yordam beradi. Kontekstdagi iboralarni tushunish ushbu tillarning grammatikasi va semantikasini hisobga olgan holda aniqroq va tabiiy tarjimalarni yaratishga imkon beradi.



4. Kontekstga moslashish:

- Ko'p so'zli leksemalarni qayta ishlash turkiy tillardagi matnlarni tahlil qilishda kontekstni hisobga olish imkonini beradi. Bu sizga turli kontekst va vaziyatlarga adekvat javob bera oladigan yanada moslashuvchan tizimlarni yaratishga imkon beradi.

5. Resurslar va ilovalarni ishlab chiqish:

- Turkiy tillar uchun ko'p so'zli leksemalarni qayta ishlash usullarining ishlab chiqilishi turli ilovalar va resurslarni, jumladan, mashina tarjimasini tizimlari, matn tahlili, axborot qidirish va boshqa ko'plab dasturlarni yaratishga yordam beradi. Bu turkiy tillar jamiyatida til texnologiyalarini rivojlantirish uchun yangi imkoniyatlar ochadi.

Shunday qilib, ko'p so'zli leksemalarni qayta ishlash turkiy tillar uchun global hisoblash tilshunosligida muhim o'rin tutib, ushbu tillardagi matnlarni yanada aniqroq va moslashtirilgan tahlil qilishni ta'minlaydi, bu esa ushbu tillar jamoasida til texnologiyalari va muloqotining rivojlanishiga xizmat qiladi.

Ko'p so'z leksemalarda ikkilamchi va ko'p ma'nolilik muammosi shundan iboratki, bir ibora qo'llanilgan kontekstga qarab bir necha xil ma'no yoki izohlarga ega bo'lishi mumkin. Bu tabiiy tilni qayta ishlashning asosiy muammolaridan biri bo'lib, matnni tog'ri talqin qilish uchun maxsus kontekstli tahlil algoritmlarini ishlab chiqishni talab qiladi. Keling, ushbu muammoni batafsil ko'rib chiqaylik:

1. Ko'p so'zli leksemalarda ko'p ma'nolilik:

- Ko'p so'zli leksemalar ko'pincha ikkilanishli, ya'ni bir nechta mumkin bo'lgan ma'no yoki qarama-qarshi yoki ko'p ma'noli bo'lishi mumkin.

2. Ma'nolarning xilma-xilligi:

- Ko'p so'z leksemalar qo'llanilayotgan kontekstga, shuningdek, foydalanuvchi yoki matn muallifining shaxsiy xohishiga ko'ra turli ma'noga ega bo'lishi mumkin.

- Masalan, "oq uy" iborasi Vashingtondagi siyosiy muassasaga ham, ma'lum bir rangdagi binoga ham tegishli bo'lishi mumkin.

3. Kontekstli tahlilga ehtiyoj:

- Ko'p so'zli leksemalarni to'g'ri talqin qilish uchun ular qaysi kontekstda qo'llanganligini hisobga olish kerak. Kontekst qo'shni so'zlar, jumla, matn, shuningdek, muhokama mavzusi yoki muloqot maqsadi kabi tashqi kontekstni o'z ichiga olishi mumkin.

- Kontekstli tahlil ko'p so'z leksemaning atrofdagi kontekstdan kelib chiqib, eng ehtimolli ma'nosini aniqlash imkonini beradi.

4. Qiymatlarning noaniqligi:

- Ko'p so'zli leksemalarda ma'nolarning noaniqlik va xilma-xilligi muammosini hal qilish uchun kontekstdan kelib chiqib, eng mos ma'no yoki ma'noni tanlash imkonini beruvchi aniqlanish algoritmlari qo'llaniladi.

- Ma'noni ajratish turli usullar yordamida, masalan, statistik modellar, mashinani o'rganish, leksik va semantik ma'lumotlar bazalaridan foydalanish orqali amalga oshirilishi mumkin.

Umuman olganda, ko'p so'zli leksemalarda ma'nolarning noaniqlik va xilma-xilligi muammosi kompyuter tizimlari tomonidan matnni to'g'riroq tushunish va izohlashni ta'minlash maqsadida kontekstni hisobga oladigan va ma'nolarni ajrata oladigan kontekstual tahlil algoritmlarini ishlab chiqishni taqozo etadi. Bu turli xil tabiiy tillarni qayta ishlash dasturlari uchun muhim, masalan, mashina tarjimai, ma'lumot qidirish, matnni tahlil qilish va boshqalar.

Global hisoblash tilshunosligida ko'p so'zli leksemalarni qayta ishlashni o'rganish va amaliy qo'llashning hozirgi tendensiyalari tabiiy til matnlarini samarali tahlil qilish va tushunish muhimligini ko'rsatadi. Quyida asosiy tendensiyalardan ba'zilari keltirilgan:

1. Deep Learningdan foydalanish:

- Hozirgi tendensiyalardan biri ko'p so'zli leksemalarni qayta ishlashda chuqur o'rganishdan keng foydalanish hisoblanadi. Chuqur o'rganish usullari matnlardan xususiyatlarni avtomatik ravishda ajratib oladigan va katta hajmdagi ma'lumotlar bilan samarali ishlaydigan modellarni yaratadigan murakkab neyron tarmoqlarni yaratishga imkon beradi.

2. Semantik modellarni ishlab chiqish:

- Ko'p so'zli leksemalarni qayta ishlashda so'z vektorini ko'rsatish va kontekstni tushunishga asoslangan semantik modellar tobora ommalashib bormoqda. Bu modellar matndagi so'z va iboralar o'rtasidagi semantik munosabatlarni hisobga olish imkonini beradi, bu esa matnlarni tahlil qilish va izohlashning aniqligini oshiradi.

3. Ko'p tilli tizimlar:

- Xalqaro kommunikatsiyalar va jahon iqtisodiyotining rivojlanishi bilan ko'p so'zli leksemalarni qayta ishlash uchun ko'p tilli tizimlarga ehtiyoj ortib bormoqda. Tadqiqotchilar va ishlab chiquvchilar turli tillardagi matnlar bilan samarali ishlay oladigan va yuqori sifatli mashina tarjimai va tahlilini ta'minlaydigan tizimlarni yaratish ustida ishlamoqda.

4. Matnli ma'lumotlarni keng miqyosda qayta ishlash:

- Ijtimoiy tarmoqlar, yangiliklar maqolalari, ilmiy nashrlar va boshqalar kabi katta hajmdagi matnli ma'lumotlar paydo bo'lishi bilan bunday ma'lumotlarni keng miqyosda qayta ishlay oladigan tez va tezkor ta'minlovchi ko'p so'zli leksemalarni samarali qayta ishlash usullarini ishlab chiqish zarurati tug'iladi.

5. Boshqa sohalar bilan integratsiya:

- Ko'p so'zli leksemalarni qayta ishlashning hozirgi tendensiyalari kompyuter fanining tasvir, tovush va ma'lumotlarni qayta ishlash kabi boshqa sohalari bilan integratsiyani o'z ichiga oladi. Bu matnni tahlil qilish va tabiiy tilni talqin qilish uchun yanada to'liq va integratsiyalashgan tizimlarni yaratishga imkon beradi.

Ushbu tendensiyalar tadqiqotchilar va ishlab chiquvchilarning ko'p so'zli leksemalarni qayta ishlash uchun aniqroq, samarali va kengaytiriladigan tizimlarni yaratish istagini aks ettiradi, ular turli sohalarda, jumladan, mashina tarjimasini, ma'lumot qidirish, matn matn tahlil qilish va boshqalarda muvaffaqiyatli qo'llanilishi mumkin.

Natural Language Processing (NLP) texnologiyalari turli sohalarda, jumladan, mashina tarjimasini tizimlarini ishlab chiqish, matnlarni avtomatik tahlil qilish va murakkab ko'p so'zli leksemalarni samarali qayta ishlash zarur bo'lgan boshqa ilovalarda muhim rol o'ynaydi. Keling, asosiy NLP texnologiyalarini va ularning bunday tizimlarda qo'llanilishini ko'rib chiqaylik:

1. Morfologik tahlil:

- Morfologik tahlil so'zlarning tuzilishi va grammatik belgilarini, masalan, gap bo'lagi, hol, zamon va hokazolarni tahlil qilishni o'z ichiga oladi.

- Mashinali tarjima tizimlarida morfologik tahlil so'zlarning asosiy shakllarini aniqlash va to'g'ri tarjima qilish uchun ularning grammatik xususiyatlarini aniqlashga yordam beradi.

- matnni avtomatik tahlil qilishda morfologik tahlil kalit so'zlarni aniqlash va ularning matndagi chastotasini hisoblash uchun qo'llaniladi.

2. Tahlil:

- Sintaktik tahlil gapning sintaktik tuzilishini aniqlash va so'zlar o'rtasidagi munosabatlarni aniqlashga qaratilgan.

- Mashinali tarjima tizimlarida sintaktik tahlil gaplarning sintaktik daraxtlarini qurish va tarjima qilishda ularning tuzilishini hisobga olish imkonini beradi.

- Avtomatik matn tahlilida sintaktik tahlil so'zlar orasidagi bog'lanishni aniqlashga, asosiy ibora va gaplarni ajratib ko'rsatishga yordam beradi.

3. Semantik tahlil:

- Semantik tahlil kontekstdagi so'z va iboralarning ma'nosini tushunishga qaratilgan.

- Mashinali tarjima tizimlarida semantik tahlil tarjima qilishda so'zlarning ma'nolarini va ularning kontekstini hisobga olishga yordam beradi, bu esa so'zlarning noaniqligi tufayli noto'g'ri

tarjimalardan qochishga yordam beradi.

- Avtomatik matn tahlilida matn mavzusini aniqlash, so'zlar o'rtasidagi munosabatlarni aniqlash va matnlarni semantik toifalarga ajratish uchun semantik tahlildan foydalaniladi.

4. Aniqlik va kontekstual tahlil:

- Ma'nolarni ajratish va kontekstual tahlil matnni izohlashda kontekstni hisobga olish va noaniqliklarni bartaraf etish imkonini beradi.

- Mashina tarjimasi tizimlarida disambiguatsiya va kontekstual tahlil jumla kontekstidan kelib chiqqan holda eng mos tarjimalarni tanlashga yordam beradi.

- Matnni avtomatik tahlil qilishda bu texnologiyalar so'zlar orasidagi semantik aloqalarni tushunish va matnni tahlil qilishda ularni hisobga olish imkonini beradi.

5. Mashinani o'rganish va chuqur o'rganish:

- NLP tizimlarini ishlab chiqishda mashinani o'rganish va chuqur o'rganish usullari muhim rol o'ynaydi, chunki ular katta hajmdagi ma'lumotlardan naqshlarni avtomatik ravishda ajratib olish va matn tahlili sifatini yaxshilash imkonini beradi.

- Mashina tarjimasi tizimlari va avtomatik matn tahlili turli tillar va kontekstlarga moslasha oladigan modellarni yaratish uchun mashinani o'rganish va chuqur o'rganishdan foydalanadi.

Tabiiy tillarni qayta ishlash texnologiyalari mashina tarjimasi tizimlarini, avtomatik matn tahlilini va murakkab ko'p so'zli leksemalarni samarali qayta ishlash muhim bo'lgan boshqa ilovalarni ishlab chiqishda asosiy rol o'ynaydi. Ushbu texnologiyalardan foydalanish tabiiy tildagi matnni avtomatik tushuna oladigan va tahlil qila oladigan aniqroq va samarali tizimlarni yaratish imkonini beradi.

Foydalanilgan adabiyotlar

- Абдурахмонова Н. Машина таржимаси. –Тошкент, 2008.
Abduhamidova L. Tilshunoslikning yangi yo'nalishi: kompyuter lingvistikasi. –Toshkent, 2015
Аврорин В.А. Проблемы изучения функциональной стороны языка (к вопросу о предмете социолингвистики). –Л., 1975. –С. 177.
Базарова Б.Б. Введение в корпусную лингвистику. –Улан-Удэ, 2016. –Б.19
Very A Brief. Introduction to Computational Linguistics. Weiwei Su. Institute of Computer Science and Technology Peking

University. 2017.

D. Yu and L. Deng "Automatic Speech Recognition: A Deep Learning Approach" (Publisher: Springer), 2014.

Курбонова Ф. Компьютер луғатлари: тезаурус. –Тошкент, 2014.

Mengliyev B., Bobojonov S., Hamroyeva Sh. O'zbek tilining milliy korpusi. – Toshkent: Ma'rifat / 2018.

Марчук Ю.Н. Методы моделирования перевода. –М.: Наука, 1985.

ABOUT THE EXPERIENCE OF PROCESSING LEXEMES CONSISTING OF SEVERAL WORD FORMS IN WORLD COMPUTER LINGUISTICS

Firuz Nurova¹

Abstract. This article delves into the challenges associated with processing lexemes comprising multiple word forms within the realm of computer linguistics. It examines various methodologies and technologies, encompassing both syntactic and semantic analyses. Furthermore, it elucidates the issues encountered in Uzbek linguistics concerning the translation of analytical forms comprised of multiple word forms, along with the requisite solutions. Additionally, it underscores the significance of investigating lexemes with multiple word forms within computational linguistics, providing insights into pioneering researchers in this domain. The intricacy inherent in processing such lexemes in computer linguistics stems from their ambiguity, diverse forms, and meanings, necessitating consideration of context for accurate word combination parsing. The text also advocates for the development of effective analytical methods capable of grasping the intended meaning of expressions.

Key words: *computational linguistics, algorithm for morphological analysis, algorithms for semantic analysis, models for semantics, human language, morphological analyzer, verb analytical forms, complexity of morphology, meaning, textual collections, linguistic technologies, deep learning, systems for multiple languages.*

References

- Abdurahmanova N. Machine translation. -Tashkent, 2008.
- Abduhamidova L. A new direction of linguistics: computer linguistics. -Tashkent, 2015
- Avorin V.A. Problem study of the functional side of the language (questions about the subject of sociolinguistics). -L., 1975. -S. 177.
- Bazarova B.B. Introduction to corpus linguistics. -Ulan-Ude, 2016.- B.19
- Very A Brief. Introduction to Computational Linguistics. Weiwei Su. Institute of Computer Science and Technology Peking University. 2017.
- D. Yu and L. Deng "Automatic Speech Recognition: A Deep Learning

¹*Nurova Firuz Kamolovna* – senior teacher of the Department of Uzbek and Russian languages of the Bukhara State Pedagogical Institute.

E-mail: firuzaru870@gmail.com

ORCID: 0009-0005-3510-6483

Approach” (Publisher: Springer), 2014.

Kurbonova F. Computer dictionaries: thesaurus. -Tashkent, 2014.

Mengliyev B., Bobojonov S., Hamroyeva Sh. National corpus of the Uzbek language. - Tashkent: Marifat / 2018.

Marchuk Yu.N. Methody modelirovaniya perevoda. - M.: Nauka, 1985.

Jurnal 2017-yil 26-oktyabrda O'zbekiston Respublikasi Matbuot va axborot agentligi tomonidan 0936-raqam bilan ro'yxatdan o'tgan.

Jurnal O'zbekiston Respublikasi Oliy Attestatsiya Komissiyasi tomonidan filologiya fanlari bo'yicha falsafa doktori (PhD) va fan doktori (DSc) dissertatsiyalari asosiy ilmiy natijalari chop etilishi lozim bo'lgan ro'yxatga kiritilgan (30.10.2021. № 308/6).

Tahririyatga kelgan maqolalar mualliflarga qaytarilmaydi.

Manzil: Toshkent shahri, Yakkasaroy tumani, Yusuf Xos
Hojib ko'chasi 103-uy.
Telefonlar: +99871 281-45-11, +99871 281-41-93.
Website: compling.tsuull.uz
E-mail: kompling@navoiy-uni.uz

Bosishga 29.02.2024-yilda ruxsat etildi.
Bichimi 70x100 1/16, Ofset bosma. "Cambria" garniturasida.
Shartli b.t. 7,51. Nashr b.t. 7,62.

"O'zbekiston: til va madaniyat" jurnali tahririyatida tayyorlandi va sahifalandi.
"YASHNOBOD NASHR" bosmaxonasida chop etildi.
Adadi 300 nusxa. Buyurtma №2.
Bosmaxona manzili: Toshkent shahar Yashnobod tumani,
58-a harbiy shaharcha.