

RAVISH SO‘Z TURKUMINI GRAMMATIK POS TEGGLASH ALGORITMLARI: NAZARIY YONDASHUV, AMALIY TATBIQ VA KELGUSIDAGI YO‘NALISHLAR

Xudayberganov Nizomaddin Uktambay o‘g‘li,
o‘qituvchi

nizomaddin@navoiy-uni.uz

ToshDO‘TAU

Karimova Zilola Dilmurod qizi,

Kompyuter lingvistikasi mutaxassisligi magistranti

karimovazilola791@gmail.com

ToshDO‘TAU

Annotatsiya. Ushbu maqola o‘zbek tilidagi ravish so‘z turkumini aniqlash va grammatik POS(Part-of-Speech) teglash algoritmlarining nazariy asoslari, amaliy tatbiq usullari va eksperimental natijalarini chuqur tahlil qiladi. Tadqiqotda statistik modellar (HMM, CRF), qoidaviy yondashuv va morfologik tahlil usullari birgalikda o‘rganiladi. Jadval va blok-sxemalar yordamida algoritm strukturalari va samaradorlik ko‘rsatkichlari vizual tarzda ifodalanadi. Asosiy maqsad avtomatlashtirilgan tilni qayta ishlash tizimlarida ravish so‘zlarini, ularning ma‘noviy va tuzilishiga ko‘ra turlarini aniqlashning aniqligi va samaradorligini oshirishga qaratilgan yangi metodologiyalarni ishlab chiqishdir.

Abstract. This article provides an in-depth analysis of the theoretical foundations, practical implementations, and algorithms for identifying adverbial word classes models(HMM, CRF), rule-based approaches, and morphological analysis techniques. Through diagrams, tables, and performance metrics are visually presented. The objective is to develop novel methodologies to enhance the accuracy and efficiency of automated natural language processing systems in adverb identification.

Аннотация. В данной статье представлен детальный анализ теоретических основ, практической реализации и алгоритмов для идентификации наречных классов слов с использованием моделей (HMM, CRF), подходов на основе правил и методов морфологического анализа. Диаграммы, таблицы и показатели производительности визуально представлены. Цель исследования разработка новых методологий для повышения точности и эффективности автоматизированных систем обработки естественного языка при идентификации наречий.

Kalit so‘zlar: *ravish so‘z turkumi, POS teglash, POS algoritmlari, morfologik tahlil, statistik modellar, qoidaviy yondashuv*

Kirish. Soʻz turkumlarini (POS) belgilash – tabiiy tilni qayta ishlash (NLP) sohasida eng muhim mavzulardan biri boʻlib, asosiy qurilish bloklaridan va amaliyotlaridan biridir. Shu sababli, soʻz turkumlarini belgilash NLPda keng qoʻllaniladigan mavzu boʻlib, matndagi har bir soʻzga kontekstga muvofiq tegishli sintaktik yorliqni berishni maqsad qiladi. Soʻz turkumlarini avtomatik ravishda aniqlash POS tagging yoki grammatik belgilash deb ataladi. Bu jarayon feʼl, sifat, ravish, ot va boshqa soʻz turkumlariga tegishli boʻlgan soʻzlarni tasniflashni oʻz ichiga oladi. Ushbu texnologiya mashinada tarjima qilish, soʻz maʼnolarini ajratish, savollarga javob berish, sintaktik tahlil qilish va boshqa koʻplab NLP ilovalarida qoʻllaniladi. Soʻz turkumlarini aniqlash oddiy soʻzlarni maʼlum bir turkumga moslashtirishdan ancha murakkabroq. Chunki bir soʻz turli kontekstlarda har xil soʻz turkumiga ega boʻlishi mumkin. Shu sababli, POS belgilash uchun yagona universal qoidalar mavjud emas. Matn korpusidagi soʻz turkumlarini qoʻlda aniqlash juda qiyin. Doimiy ravishda yangi soʻzlar va kontekstlar paydo boʻlib turadi, shuning uchun POS belgilashni qoʻlda bajarish samarali emas. Shu sababli, biz POS belgilash jarayonida mashinani oʻrganishga asoslangan usullarga tayanamiz [1].

Jahon tilshunosligida ham bir qancha olimlar POS teglash boʻyicha samarali izlanishlar va tadqiqotlar olib borishgan. Ulardan S.K. Cameron POSning asosiy yetakchilaridan biridir va u ijobiy tashkilotlarni yaratish va rivojlantirish haqida koʻplab ilmiy maqolalar yozgan. Uning asarlari tashkilotlarda yuksak samaradorlik, jamoaviy ishlash va ijobiy psixologiyaning oʻrni haqida keng tahlillarni oʻz ichiga oladi. R.E. Quinn POSni rivojlantirishda muhim rol oʻynagan olimlardandir. U oʻz tadqiqotlarida organizatsion oʻzgarishlar va ijobiy liderlikni qanday amalga oshirish kerakligi haqida asosiy nazariyalarni ishlab chiqqan. Quinn oʻz tadqiqotlarida tashkilotlardagi ijobiy oʻzgarishlar va innovatsiyalarni qanday boshqarish mumkinligi haqida chuqur tahlil qilgan. U liderlik, ijobiy tuzilmalar va oʻzgarishlarni amalga oshirishga oid koʻplab ishlanmalarga ega. Jane Dutton POSning tadqiqotida katta hissa qoʻshgan olimlardan biridir. Uning asarlari, ayniqsa, tashkilotlarda rahbarlik va jamoaviy munosabatlar asosida oʻrganilgan. Dutton ijobiy psixologiyaning tashkilotlarda qanday ishlashini va rahbarlarning ishlash samaradorligini qanday oshirishi mumkinligini oʻrganadi.

Oʻzbek tilshunosligida kompyuter lingvistika sohasining oʻrganilishi XXI asrdan boshlandi. Bugunga qadar sohaga oid koʻplab tadqiqotlar olib borilmoqda. Oʻzbek tilida POS teglash oʻziga xos xususiyat bilan oʻzbek tilining boyligini chuqur va keng ekanini yaqqol ifoda eta oladi. Shuning uchun ham POS teglash masalasini maxsus oʻrganish va dasturiy taʼminotini ishlab chiqish katta ahamiyat kasb etadi. Oʻzbek tilining milliy korpusi ishlab chiqildi. Shunga qaramay, teglash jarajonida muammolar vujudga kelayotgani, bu jarayondagi kamchiliklar mavjudligi bu sohaning dolzarb ekanligini belgilaydi.

Leksik birliklarni grammatik POS teglash (Part-of-Speech tagging) – bu tilshunoslikda soʻzlarning grammatika jihatdan qanday rol oʻynashini aniqlash, yaʼni soʻz turkumini belgilash bilan bogʻliq muhim tadqiqot sohasidir[2]. Bu sohada Oʻzbekistonlik olimlar ham tadqiqotlar olib borishgan, lekin bu yoʻnalishning oʻzbek tilshunosligida nisbatan yangi va rivojlanayotgan sohaga kirishi mumkin. Leksik birliklar va grammatik POS teglashni oʻrganish asosan kompyuter lingvistikasi, sunʼiy intellekt va tabiiy tilga ishlov berish (NLP) texnologiyalariga asoslanadi.

Adabiyotlar tahlili va metodologiya

Oʻzbek tilshunosligida teglash masalasi boʻyicha amaliy va nazariy ishlar olib borilmoqda. Alisher Navoiy nomidagi Toshkent davlat oʻzbek tili va adabiyoti universiteti Kompyuter lingvistikasi kafedrası oʻqituvchilari tomonidan POS teglash boyicha bir qancha ishlar olib borilmoqda. Xususan, B.Elov, Sh.Xamroyeva, N.Xudoyberganov, O.Abdullayeva va boshqalarning ishlari diqqatga sazovordir. POS teglash – bu berilgan gapdagi har bir soʻz shakliga uning turkum (ot, feʼl, sifat, son, ravish yoki olmosh)ga mansubligini belgilash (teglash) vazifasidir. POS teglash tabiiy tilga ishlov berish (Natural Language Processing, NLP) sohasining asosiy vazifalaridan biri boʻlib, pipeline konveyerining muhim bosqichi hisoblanadi. POS teglash korpusdagi har bir soʻzga grammatik kategoriya belgilashni oʻz ichiga oladi. Yaʼni soʻzning morfologik xususiyatlarini aniqlashda ushbu teglashdan foydalanish samarali hisoblanadi. Ushbu teglashda otlar, feʼllar, sifatlar, ergash gaplar, olmoshlar, bogʻlovchilar, yuklamalar va boshqalarni teglashni kiritish mumkin. POS belgilari jumla tuzilmalarini tahlil qilish va lingvistik tadqiqotlar oʻtkazishda yordam beradi. Oʻzbek maxsus lingvistik korpuslarini tuzish va ularni teglash masalasi boʻyicha Sh.Hamroyeva, N.Abdurahmonova M.Abjalova A.Eshmoʻminov Sh.Gulyamova, O.Abdullayeva, D.Elova kabilarning tadqiqotlari alohida ahamiyatga molik.

Zamonaviy texnologiyalar, jumladan, kompyuter lingvistikasi sohasida ravish soʻz turkumini avtomatik aniqlash dolzarb masalalardan biri hisoblanadi. Bu jarayonda Part-of-Speech (POS) teglash algoritmlaridan foydalaniladi. POS teglash jarayonida tilning oʻziga xos morfologik va sintaktik xususiyatlari inobatga olinadi, bu esa lingvistik modellarni samarali qoʻllashni talab qiladi[3].

Soʻz turkumlari tilshunoslikda tilning grammatik tizimini tushunishga yordam beradigan asosiy konseptlardan biridir. Ravish (pronominal) soʻz turkumi, oʻzining grammatik funksiyasi va strukturasi jihatidan boshqa soʻz turkumlaridan farq qiladi[4]. Ravish soʻzlarning grammatik teglashini avtomatik ravishda amalga oshirish esa tilni tahlil qilishda muhim ahamiyatga ega, chunki bu jarayon tilning sintaktikva semantik tuzilishini anglash uchun zarur boʻlgan asosiy bosqichlardan biridir[5]. Maqolada ravish soʻz turkumini grammatik POS teglash algoritmlari, lingvistik modellari, ularning turlari va metodologiyalari haqida soʻz boradi.

Asosiy qism. O'zbek tilining murakkab morfologik tuzilishini hisobga olish muhimdir. Ushbu jarayonda quyidagi bosqichlar amalga oshiriladi:

- Asos va qo'shimchalarni ajratish: So'zning asosiy qismlari tahlil qilinadi.
- Blok-sxema: Quyidagi diagramma algoritmnining asosiy bosqichlarini ifodalaydi:

Quyidagi jadvalda amaliy misollar va ularning teglash natijalari keltirilgan:

1-jadval. Ravishlarni aniqlash va POS teglash bo'yicha amaliy natijalar

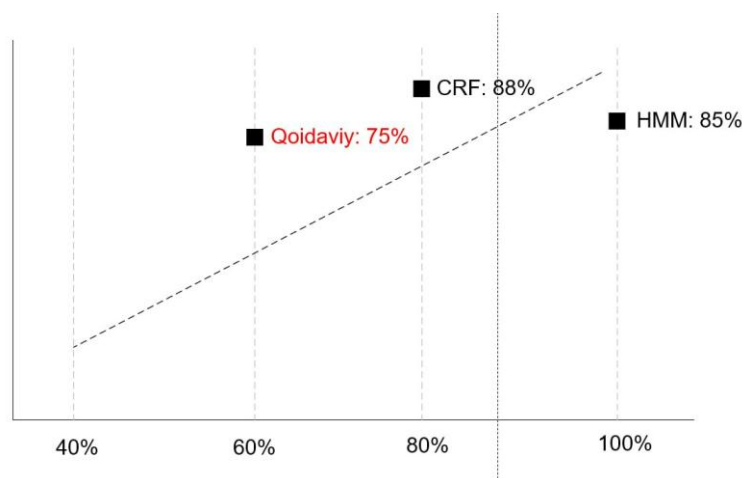
Matn namunasi	Aniqlangan ravish so'zlari	Algoritm natijasi (POS teglash)
Sardor juda sekin yugurdi	juda, sekin	Ravish: (juda, sekin) – matn kontekstiga mos ravishda
Ra'no tomonidan o'qilgan she'r o'zgacha tarzda eshitildi.	o'zgacha tarzda	Ravish: [o'zgacha, tarzda] – semantik izoh bilan
U o'g'rilikni juda tez payqagan edi.	Juda tez	Ravish: [juda] – qo'shimcha ibora sifatida
U nihoyatda puxta va tezjavob berdi	nihoyatda, puxta, tez	Ravish: [nihoyatda, puxta, tez] – umumiy tahlil

Bu misollar yordamida statistik va qoidaviyondashuvlarning samaradorligi, aniqligi va kamchiliklari tajriba asosida tahlil qilinadi.

Tadqiqot jarayonida quyidagi bosqichlar amalga oshirildi:

1. Matn yig'ish: O'zbek tilidagi akademik maqolalar, yangiliklar va adabiyotlardan iborat korpus shakllantirildi.
2. Avtomatik teglash: Dastlabki avtomatlashtirilgan teglash algoritmlari yordamida natijalar olinib, keyin qo'lda tahrir qilindi.
3. Model sinovlari: HMM, CRF va qoidaviy yondashuvlar yordamida modellar o'rgatilib, ularning aniqlikko'rsatkichlari taqqoslandi.

Quyidagi diagramma eksperimental natijalarning aniqlik ko'rsatkichlarini aks ettiradi:



2-rasm. Algoritm aniqlik ko'rsatkichlari diagrammasi

Natijalar shuni ko'rsatadiki, CRF modeli kontekstual bog'lanishlarni hisobga olib eng yuqori aniqlikka erishgan. Shuningdek, qoidaviy yondashuv barqaror natija bergan, ammo moslashuvchanlik jihatidan statistic modellarga nisbatan pastroq ekanligi aniqlangan.

Eksperimental jarayonda quyidagi xatoliklar va cheklovlar aniqlangan:

- *Morfologik noaniqliklar:* Ba'zi hollarda o'zak va qo'shimchalarni to'g'ri ajratib bo'lmaslik.
 - *Kontekstual cheklovlar:* Ravish so'zlarining kontekstini to'liq aks ettirishda qiyinchiliklar[6].
 - *Qoidaviy yondashuvning moslashuvchanligi:* Qo'lda yangilanish zarurati va keng korpusga moslashtirishda qiyinchiliklar[6].
- Kelgusida quyidagi takomillashtirishlar rejalashtirilmoqda:
- Korpus hajmini va xilma-xilligini oshirish;
 - Statistik va qoidaviy yondashuvlarni gibridlashtirish;
 - Chuqur o'rganish algoritmlari (LSTM, Transformer) integratsiyasini amalga oshirish.

Ushbu tadqiqotda quyidagi metodologik bosqichlar qo'llanildi:

- *Ma'lumot yig'ish:* O'zbekiston va xalqaro manbalardan olingan korpuslar asosida ma'lumotlar to'planildi.
- *Avtomatik teglash:* Dastlabki avtomatlashtirilgan algoritmlar yordamida tezkor natija olish, so'ngra inson nazorati ostida qayta tahrir qilish.
- *Sinov va taqqoslash:* Turli modellarning aniqlik va samaradorligini eksperimental tarzda sinab ko'rish[7].

Amaliy jarayonda quyidagilar bajarildi:

- *Modellarning o'rgatilishi:* HMM, CRF va qoidaviy yondashuvlar alohida sinovdan o'tkazildi.
- *Kombinatsion yondashuvlar:* Ensemble metodlari yordamida yakuniy natija sifatini oshirishga intilish.
- *Vizualizatsiya:* Diagrammalar va jadvallar yordamida natijalar va xatoliklar vizual tarzda aks ettirildi.
- *Neyravliy tarmoqlar:* LSTM va Transformer kabi chuqur o'rganish metodlari orqali POS teglashsamaradorligini oshirish.
- *Semantik vektor modellar:* Word2Vec, GloVe va BERT kabi metodlardan foydalanish orqali so'zlar orasidagi semantik bog'lanishni aniqlash.
- *Ensemble metodlar:* Turli modellarning natijalarini birlashtirish orqali umumiy aniqlikni oshirish[8].

Kelgusida quyidagi yo'nalishlar kengaytirilishi rejalashtirilmoqda:

- *Hibrid tizimlar*: Statistik, qoidaviy va chuqur o'rganish yondashuvlarini integratsiya qilish.
- *Korpusni kengaytirish*: Dialektlar, stilistik xususiyatlar va turli mavzularni inobatga olgan korpuslarni yaratish.
- *Real vaqt rejimi*: Interaktiv va mobil platformalarda POS teglash tizimlarini tatbiq etish.
- *Keng qamrovli sinovlar*: Aniqlik, tezlik va resurs samaradorligini o'lchash bo'yicha eksperimental sinovlarni kengaytirish.

Taqdim etilgan natijalar asosida quyidagi xulosalar chiqarildi:

- Statistik modellarning ustunligi: CRF va HMM modellari katta hajmdagi korpuslarda samaradorligini isbotladi, ayniqsa CRF modeli kontekstual bog'lanishlarga asoslanganligi sababli yuqori aniqlikka erishdi.
- Qoidaviy yondashuvning barqarorligi: Qoidaviy usullar barqaror va tushunarli natijalarni ta'minlasa-da, moslashuvchanlik jihatidan cheklanganligi aniqlangan.
- Hibrid metodlar istiqbollari: Statistik va qoidaviy yondashuvlarni chuqur o'rganish metodlari bilan birlashtirish orqali aniqlik va samaradorlikni yanada oshirish mumkinligi aniqlangan.

Xulosa.

Ushbu maqolada o'zbek tilidagi ravish so'z turkumini grammatik POS teglash algoritmlarining nazariy asoslari, metodologiyasi, amaliy tatbiq jarayoni va eksperimental natijalari chuqur tahlil qilindi. Natijalar shuni ko'rsatadiki, CRF modeli eng yuqori aniqlikni ta'minlasa-da, qoidaviy yondashuvlarning barqarorligi va tushunarli natijalari ham e'tiborga loyiqdir. Kelgusida hibrid yondashuvlar va chuqur o'rganish algoritmlarini integratsiya qilish orqali real vaqt rejimida ishlovchi, yuqori aniqlikka ega tizimlar ishlab chiqish istiqbollari mavjud.

Foydalanilgan adabiyotlar:

1. Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). Prentice Hall. P. 95.
2. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press. P. 133.
3. Brants, T. (2000). TnT – A Statistical Part-of-Speech Tagger. In *Proceedings of the Sixth Conference on Applied Natural Language Processing*. P. 224-231.
4. O'zbek Tili Morfologiyasi. (2018). Tashkent: Fan va Texnika. P. 23-24



5. Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003. P. 173-180.
6. Habash, N. (2010). Introduction to Arabic Natural Language Processing. Morgan & Claypool Publishers. P. 8-9.
7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In Advances in Neural Information Processing Systems P. 3111-3119.
8. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In Proceedings of the Empirical Methods in Natural Language Processing (EMNLP). P. 1532-1543.