

UDK: 811.512.133`42

O‘ZBEK TILIDAGI MATNLARDAN SIFAT, RAVISH, YORDAMCHI SO‘Z VA ORALIQ SO‘ZLAR TURKUMIGA OID SO‘ZLARNI ANIQLASH

Sharipov Maqsud Siddiqovich

UrDU t.f.n, AT kafedrası dotsenti

maqsbek72@gmail.com,

Jumaniyozova Zebiniso Xushnud qizi

UrDU magistranti

zebinisojumaniyozova0922@gmail.com

“Ma’mun” universiteti NTM o‘qituvchisi

Annotatsiya. O‘zbek tilshunosligida boshqa so‘z turkumlari singari sifat, ravish va yordamchi so‘z turkumi ham katta ahamiyat kasb etadi. Barcha turkiy tillar kabi o‘zbek tili ham agglutinatıv til bo‘lib so‘z o‘zagi yoki o‘zakga affiks (qo‘shimcha)lar qo‘shilishi bilan yasaladi. O‘zbek tilidagi matnlardan sifat, ravish, yordamchi so‘z va oraliq so‘zlar turkumiga oid so‘zlarni aniqlash uchun tub sifat va tub ravish so‘zlar bazasi yaratildi. Ushbu maqolada o‘zbek tilidagi matnlaridan sifat, ravish va yordamchi so‘z turkumiga tegishli so‘zlarni aniqlashning qoidaga asoslangan algoritm yaratish jarayoni ko‘rib chiqiladi.

Kalit so‘zlar: *ma’lumotlar bazasi, algoritm, affikslar, lemma, chekli avtomat (FSM), tokenlash, deterministik bo‘lmagan chekli avtomat (NFA).*

Abstract. In Uzbek linguistics, like other word groups, adjectives, adverbs and auxiliary words are of great importance. Like all Turkic languages, the Uzbek language is an agglutinative language and is formed by adding affixes to the stem or stem. A database of basic adjectives and basic adverbs was created to identify words related to adjectives, adverbs, auxiliary words, and intermediate words from Uzbek texts. In this article, the creation of a rule-based algorithm for identifying words belonging to adjectives, adverbs and auxiliary words from Uzbek language texts is considered.

Key words: *database, algorithm, affixes, lemma, finite state machine (FSM), tokenization, Non-Deterministic finite automata (NFA).*

Абстракт. В узбекском языкознании, как и в других группах слов, большое значение имеют прилагательные, наречия и вспомогательные слова. Как и все тюркские языки, узбекский язык является агглютинативным языком и образуется путем добавления аффиксов к основе или основе. База данных основных прилагательных и основных наречий была создана для выявления слов, связанных с прилагательными, наречиями, вспомогательными словами и промежуточными словами из узбекских текстов. В данной статье рассматривается создание основанного на правилах алгоритма выявления слов,

входящих в состав прилагательных, наречий и вспомогательных слов из текстов на узбекском языке.

Ключевые слова: база данных, алгоритм, аффиксы, лемма, конечный автомат (FSM), токенизация, недетерминированные конечные автоматы (NFA).

Tildagi barcha soʻzlar oʻziga xos belgi-xususiyatlari, soʻrogʻi va qanday umumlashgan maʼno ifodalashiga va gapda biror gap boʻlagi vazifasida kelishiga koʻra soʻz turkumlariga ajratiladi. Soʻz turkumlari maʼno va vazifasiga koʻra quyidagi guruhlariga boʻlinadi:

1. Mustaqil soʻzlar: Feʼl, ot, sifat, son, olmosh va ravish;
2. Yordamchi soʻzlar: Bogʻlovchi, koʻmakchi va yuklama;
3. Oraliq soʻzlar: Undov, taqlid va modal soʻzlar;

Hozirgi kunda zamonaviy texnologiyalar imkoniyatlari samaradorligini yaxshilash maqsadida matndan sifat, ravish, yordamchi va oraliq soʻz turkumiga tegishli soʻzlarni aniqlash zarurati tugʻiladi. Bu masalani tadqiqot maqsadi va vazifalari, oʻrganilganlik darajasini aniqlash maqsadida maʼlumotlar toʻplash ishiga kirishildi. Maʼlumotlar bazasini toʻplashda Sayfullayeva R., Mengliyev B. va boshqalarning “Hozirgi oʻzbek adabiy tili” va Erkaboyeva N. “Oʻzbek tilidan maʼruzalar toʻplami” darsliklari asos sifatida olindi. Bu maʼlumotlardan foydalangan holda oʻzbek tilining izohli lugʻati 1-2-3-4-5-jildlaridan sifat va ravish soʻz turkumiga doir soʻzlar ajratib olindi. Oʻzbek tili izohli lugʻatining 1-2-3-4-5-jildlaridan hammasi boʻlib, 1601 ta sifat va 877 ta ravish soʻz turkumiga doir soʻzlar aniqlandi va jadval koʻrinishida shakllantirildi. Ajratib olingan maʼlumotlar bazasi txt. fayl sifatida baza qilib kiritildi. Bazadan foydalangan holda sifat va ravish soʻz turkumiga oid soʻzlar tub va yasama soʻzlar guruhlariga ajratib olindi. Sifat va ravishning shakl yasovchi va soʻz yasovchi qoʻshimchalari bazasi shakllantirildi.

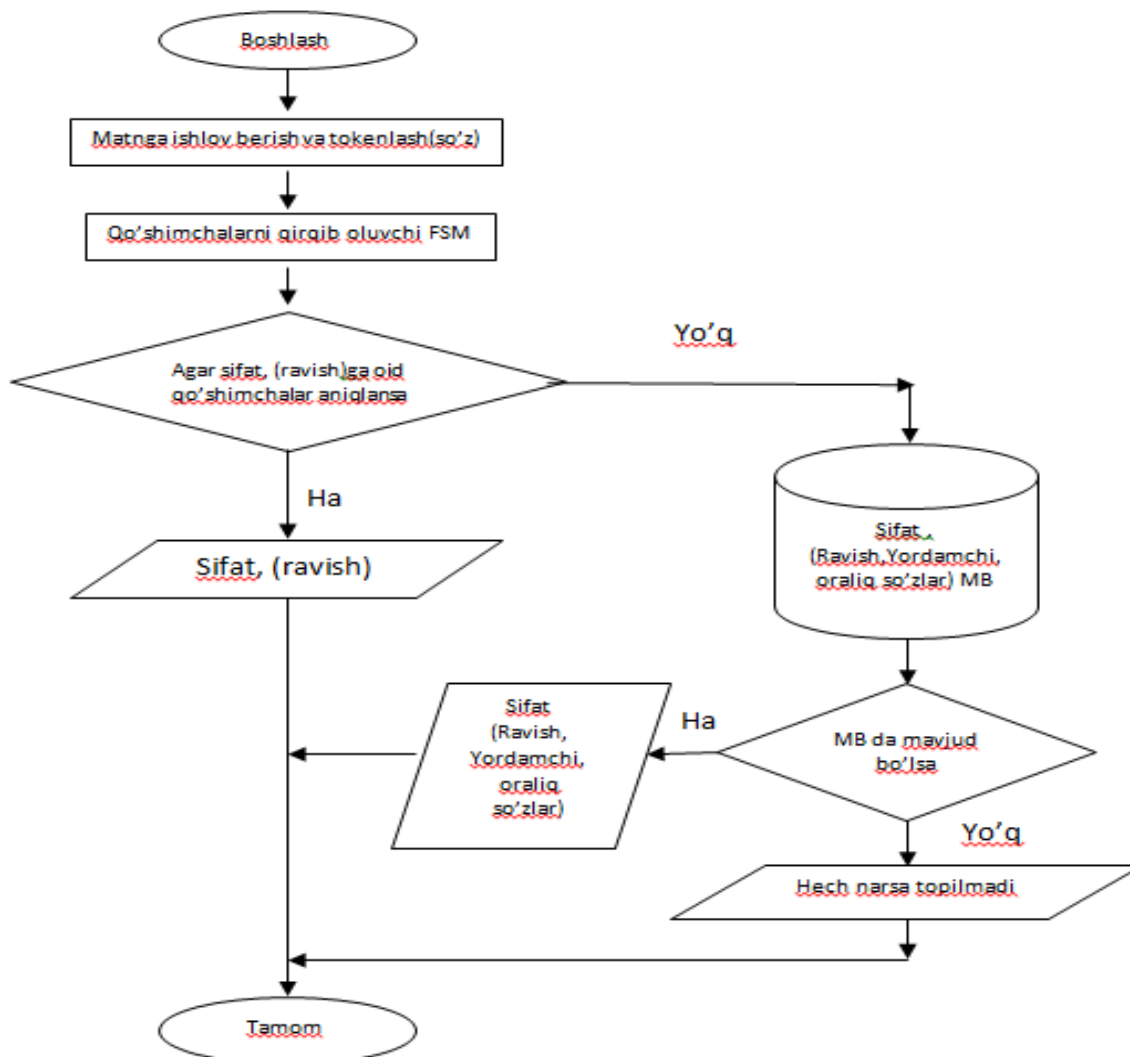
Soʻz turkumlari	Sifat lemma yasovchi affikslar	Ravish lemma yasovchi affikslar
Otdan	56	18
Feʼldan	40	16
Sifatdan	11	5
Ravishdan	3	4
Sondan	0	0
Olmoshdan	2	3
Taqlid soʻzlardan	3	0
Undov lemmadan	0	0
Modal lemmadan	2	2
Jami yasovchi affikslar	117	48

2-jadval: Boshqa so‘z turkumlaridan sifat va ravish lemma yasovchi affikslar miqdori

O‘zbek tilida sifat va ravish so‘z turkumiga tegishli so‘zlarni aniqlashda old qo‘shimchalardan ham foydalanish mumkin.

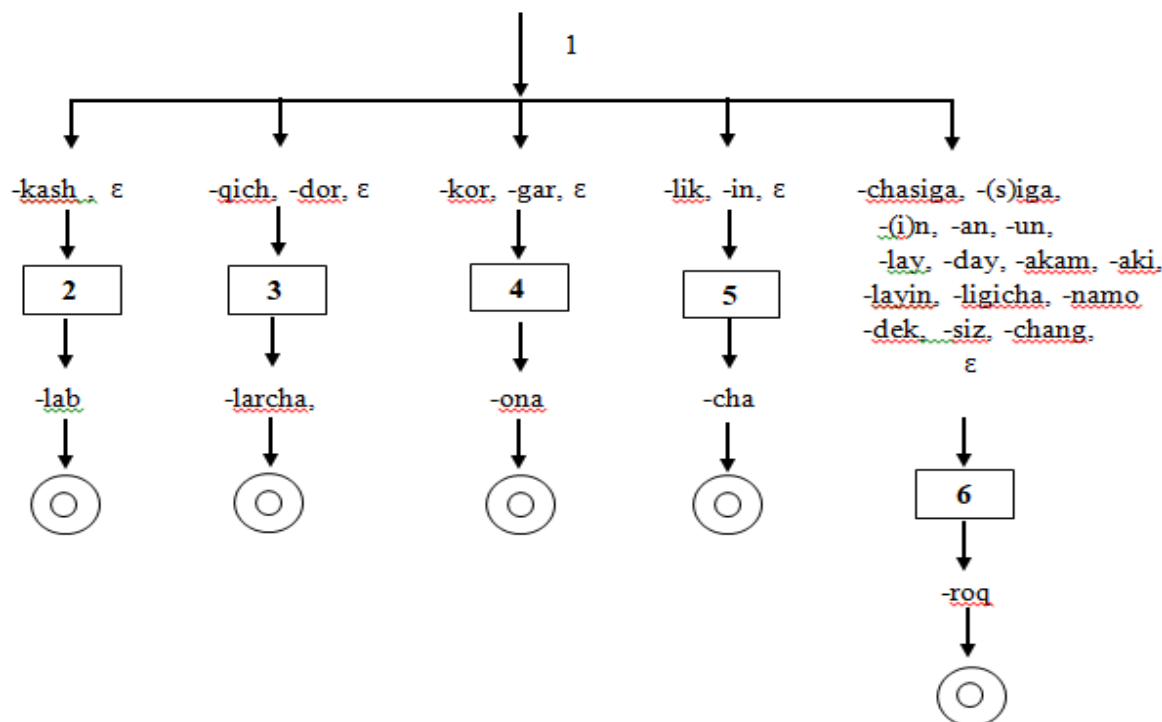
G‘ayri-	+ Sifat(Adj) = Sifat(Adj)	g‘ayri+tabiiy(Adj)=g‘ayritabiiy(Adj)
Ser-	}	ser+shox(N)=sershox(Adj)
Be-		be+baxt(N)=bebaxt(Adj)
Xush-		xush+bichim(N)=xushbichim(Adj)
Bad-		bad+axloq(N)=badaxloq(Adj)
Ba-	+ Ot(N) = Sifat(Adj)	ba+havo(N)=bahavo(Adj)
No-	}	no+umid(N)=noumid(Adj)
Bo-		bo+adab(N)=boadab(Adj)
Bar-		bar +kamol(N)=barkamol(Adj)
Ham-		ham+fikr(N)=hamfikr(Adj)
Bar-	}	bar+vaqt(N)=barvaqt(Adv)
	+ Ot(N) = Ravish(Adv)	
Be-		be+ixtiyor(N)=beixtiyor(Adv)

Bular yordamida o‘zbek tilidagi matnlardan sifat va ravish so‘z turkimini aniqlovchi algoritm ishlab chiqildi.



3-jadval: Matndan sifat va ravish so‘z turkumiga tegishli so‘zlarni aniqlovchi
algoritm

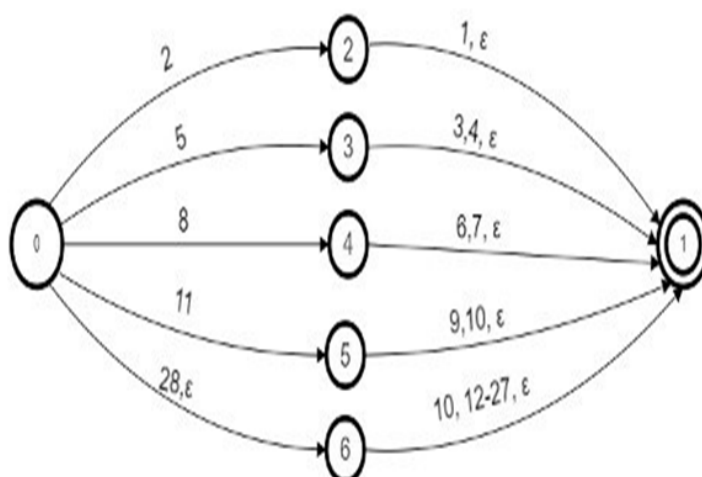
Algoritm asosan eng birinchi berilgan matnni tokenlab oladi, ya’ni matnni so‘zlarga ajratib olinadi. Tub sifatlarni va so‘z o‘zaklarini ma’lumotlar baza (MB)dan qidirib topadi. Yasama sifat va ravishlarni aniqlash uchun sifat va ravishning shakl yasovchi va so‘z yasovchi qo‘shimchalari uchun alohida-alohida chekli avtomat (FSM)larni qurib olindi.



4-jadval: Ravish soʻz turkumi yasovchi qoʻshimchalarni qidirish chekli avtomat (FSM)

Chekli Avtomat (FSM)larga asoslangan holda deterministik boʻlmagan chekli avtomat (NFA) tuzib olindi:

1. -kash
2. -lab
3. -qich
4. -dor
5. -larcha
6. -kor
7. -gar
8. -ona
9. -lik
10. -in
11. -cha
12. -chasiga
13. -siga
14. -iga
15. -n
16. -an
17. -un
18. -lav
19. -day
20. -akam
21. -aki
22. -lavin
23. -ligicha
24. -namo
25. -dek
26. -siz
27. -chang
28. -roq



5-jadval: Ravish soʻz turkumi yasovchi qoʻshimchalarni qidirish deterministik boʻlmagan chekli avtomat (NFA)

O'zbek tilida yordamchi va oraliq so'z turkumiga tegishli so'zlar qo'shimcha olmaganligi sababli bu so'z turkumiga tegishli so'zlarni ma'lumotlar bazasidan topiladi. O'rganishlar natijasida yordamchi so'zlar: bog'lovchilardan 33 tasi, ko'makchi 82 tasi, yuklama 30tasi va oraliq so'zlar: taqlid so'zlar 163tasi, undov so'zlar 82tasi va modal so'zlar 67 tasi aniqlandi va ma'lumotlar bazasiga kiritildi.

Xulosa qilib aytish mumkinki, ushbu ma'lumotlar bazasi bizga sifat, ravish, yordamchi va oraliq so'z turkumiga oid so'zlarni aniqlashda asos bo'lib xizmat qiladi. Ushbu ma'lumotlar bazasidan foydalangan holda Python dasturlash tilida algoritm tuzish maqsad qilindi.

Foydalanilgan adabiyotlar.

1. “O'zbekiston milliy ensiklopediyasi” Davlat ilmiy nashriyoti. -T.: 2020.
2. E. Kuriyozov, U. Salaev, S. Matlatipov, and G. Matlatipov, “Text classification dataset and analysis for Uzbek language,” *arXiv preprint arXiv:2302.14494*, 2023.
3. S. Matlatipov, H. Rahimboeva, J. Rajabov, and E. Kuriyozov, “Uzbek Sentiment Analysis Based on Local Restaurant Reviews,” in *CEUR Workshop Proceedings*, 2022, pp. 126–136. [Online]. Available: www.scopus.com
4. M. Sharipov, E. Kuriyozov, O. Yuldashev, and O. Sobirov, “UzbekTagger: The rule-based POS tagger for Uzbek language,” *arXiv preprint arXiv:2301.12711*, 2023.
5. M. Sharipov and O. Yuldashov, “UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language,” *arXiv preprint arXiv:2210.16011*, 2022.
6. X. Madatov, M. Sharipov, and S. Bekchanov, “O'ZBEK TILI MATNLARIDAGI NOMUHIM SO'ZLAR,” *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, vol. 1, no. 1, 2021.
7. K. Madatov, S. Bekchanov, and J. Vičič, “Accuracy of the Uzbek stop words detection: a case study on "School corpus",” *arXiv preprint arXiv:2209.07053*, 2022.
8. K. Madatov, S. Bekchanov, and J. Vičič, “Dataset of stopwords extracted from Uzbek texts,” *Data Brief*, vol. 43, p. 108351, 2022.
9. M. Sharipov and O. Sobirov, “Development of a rule-based lemmatization algorithm through Finite State Machine for Uzbek language,” *arXiv preprint arXiv:2210.16006*, 2022.