

O'ZBEK TILIDAGI MATNLARDAN FE'L SO'Z TURKUMIGA OID SO'ZLARNI QOIDAGA ASOSLANGAN HOLDA QO'SHIMCHALAR ASOSIDA ANIQLASH

Sharipov Maqsud Siddiqovich

t.f.n, dots. Urganch Davlat Universiteti

maqsbek72@gmail.com

Yakubova Dilafruz Kadamboyevna

yaqubovadilafruz@gmail.com

Urganch Davlat Universiteti magistranti

Yuldashov Ollabergan Ergash o'g'li

ollaberganyuldashov@gmail.com

Urganch Davlat Universiteti o'qituvchisi

Annotatsiya. Bu ishda biz o'zbek tili matnlaridan fe'llarni aniqlashning qoidaga asoslangan algoritm yaratish jarayonini ko'ramiz. Fe'llarni matnlardan aniqlash va morfologik tahlil qilish o'zbek tili uchun tabiiy tillarni qayta ishlashning bir qator masalalarini (POS tagging, mashinali tarjimon va h.k.) yechishda foydali bo'ladi. Barcha turkiy tillar kabi o'zbek tili ham agglutinatив til bo'lib so'z o'zagi yoki negiziga qo'shimchalar qo'shilishi bilan yasaladi. Bu masalani yechish uchun tub fe'llar bazasi yaratildi, fe'lning so'z yasovchi va shakl yasovchi qo'shimchalari bazasi shakllantirildi. Bular yordamida o'zbek tilidagi matnlardan fe'l so'z turkumini aniqlovchi algoritm ishlab chiqildi.

Kalit so'zlar: *FSM, NFA, O'zbek tili, fe'l, FST*

Abstract. In this work, we will see the creation of a rule-based algorithm for identifying verbs from Uzbek language texts. Identifying verbs from texts and morphological analysis will be useful for solving a number of natural language processing issues (POS tagging, machine translation, etc.) for the Uzbek language. Like all Turkic languages, the Uzbek language is an agglutinative language and is formed by adding suffixes to the root or base of the word. To solve this problem, a base of basic verbs was created, and a base of word-forming and form-forming suffixes of the verb was formed. With the help of these, an algorithm was developed for identifying verbs from Uzbek texts.

Key words: *FSM, NFA, Uzbek language, verbs, FST.*

Абстракт. В этой работе мы увидим создание основанного на правилах алгоритма идентификации глаголов из текстов на узбекском языке. Выявление глаголов из текстов и морфологический анализ будут полезны для решения ряда задач обработки естественного языка (POS-теги, машинный перевод и т. д.) для узбекского языка. Как и все тюркские языки, узбекский язык является агглютинативным языком и образуется путем добавления суффиксов к корню

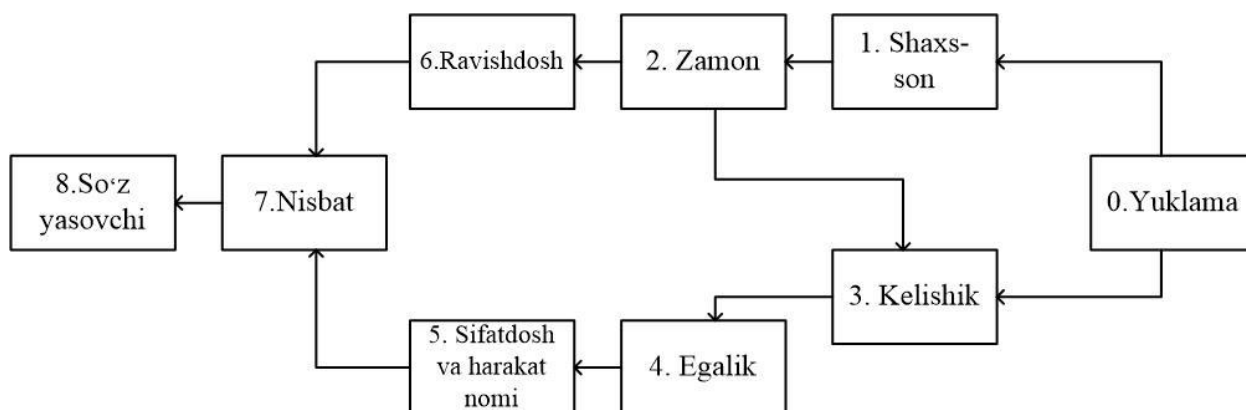
или основе слова. Для решения этой задачи была создана база основных глаголов, а также сформирована база словообразовательных и формообразующих суффиксов глагола. С их помощью был разработан алгоритм определения глаголов из узбекских текстов.

Ключевые слова: *FSM, NFA, Узбекский язык, глагол, FST.*

Tabiiy tillarni aniqlash (NLP: Natural Language Processing) sohasi boshqa sohalarga qaraganda o'zining dolzarbligi jihatidan ancha oldinlab ketdi va bu sohada barcha rivojlangan tillar qatori, o'zbek tili uchun ham bir qancha til resurslari va dasturiy vositalari ishlab chiqila boshlandi [1]–[3]. NLP sohasida lingvistik ma'lumotlar muhim rol o'ynaydi, shu jumladan, gap bo'laklari ham [4], [5]. O'zbek tili va boshqa ko'plab tillarda fe'l so'z turkumi juda katta rol o'ynab, ko'p hollarda gapning ildizi sifatida keladi. Shuning uchun bu so'z turkumi ustida izlanish olib borish muhim ahamiyat kasb etadi.

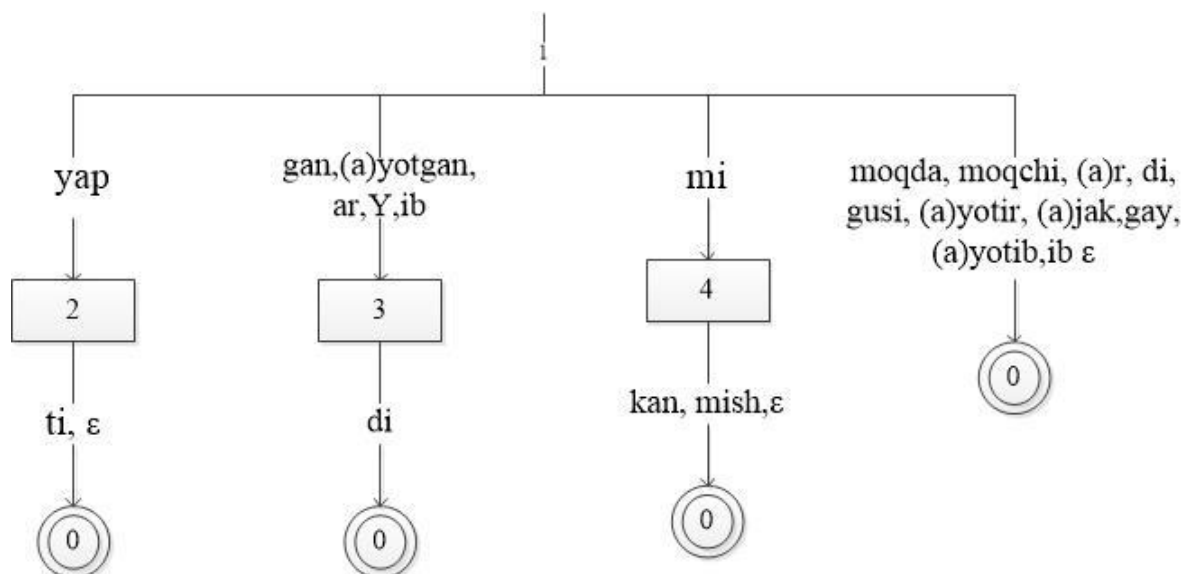
Ushbu maqolada o'zbek tilidagi matnlardan fe'llarni aniqlash algoritmi yaratish masalasi ko'riladi. O'zbek tilidagi fe'llar murakkab bo'lganligi uchun biz o'zbek tilidagi fe'llar uchun maxsus ma'lumotlar bazasini yaratdik. Fe'llar o'zbek tilida eng muhim va murakkab qismlardan biridir. Chunki fe'l butun jumlagga ta'sir qiladi va boshqaradi. Ushbu ish o'zbek tili uchun yaratilayotgan NLP modellar va dasturiy vositalar qatorida yangi ilmiy izlanishlar uchun imkon yaratadi. Shu paytgacha NLP sohasida o'zbek tili uchun bir qancha ilmiy ishlar qilingan, shular qatoriga o'zbek tili uchun yaratilgan nomuhim so'zlar dataseti [6], [7], lemmatizatsiya modellari [8], hamda sentiment tahlili datasetlari [9] kiradi.

O'zbek tili matnlardan fe'l so'z turkumiga oid so'zlarni aniqlashda biz tub fe'llar bazasini tuzdik, chunki tub fe'l qo'shimchalar bilan yasalmaganligi uchun uni FSMlar yordamida aniqlab bo'lmaydi. Yasama fe'llarni aniqlash uchun biz yuklamalar, shaxs-son, zamon, ravishdosh, sifat-dosh va harakat nomi, egalik, kelishik, nisbat, so'z yasovchi qo'shimchalar uchun alohida-alohida chekli avtomat (FSM)larni qurib chiqdik.



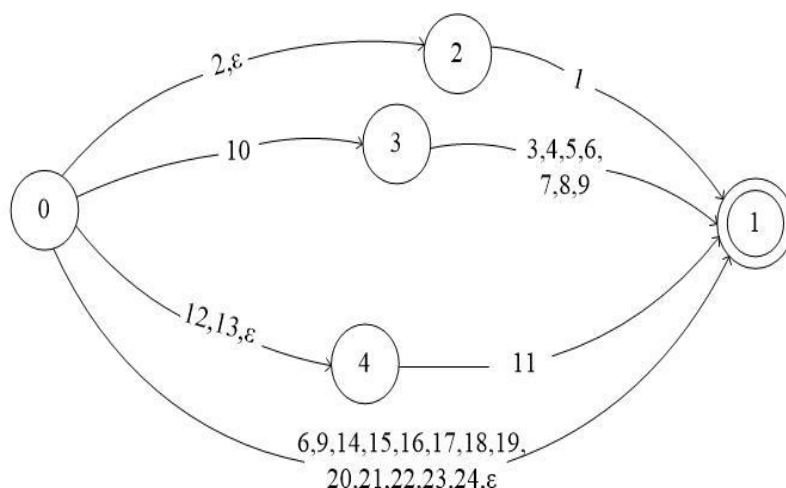
1-chizma. Fe'lni qo'shimchalar orqali aniqlash uchun FSM

Fe'lni qo'shimchalar orqali aniqlash 1-chizmada keltirilgan bo'lib bu yerda 9 ta Chekli avtomat (FSM)lar mavjud. Shulardan zamon (2)ning FSMini ko'rib chiqamiz.

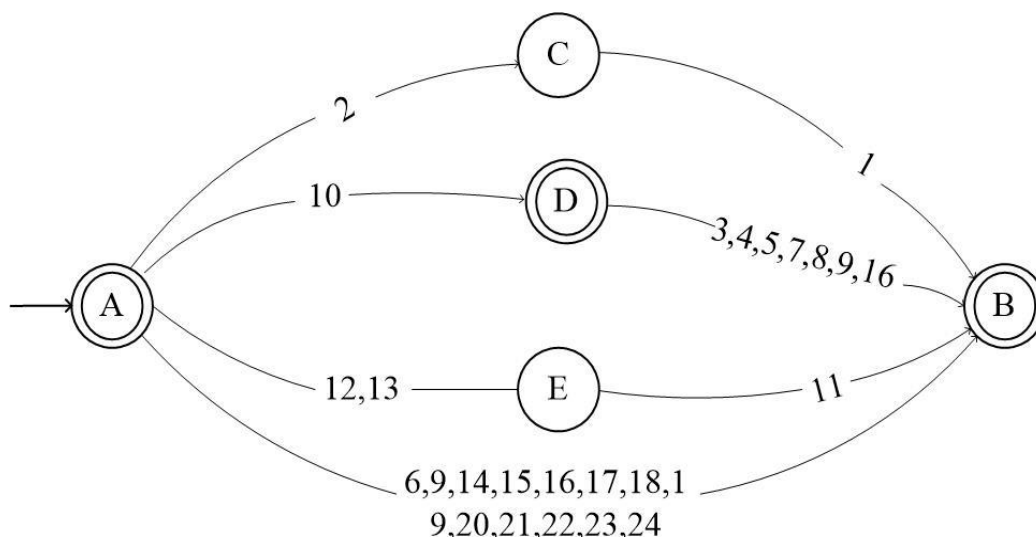


2-chizma: Zamon qo'shimchalarini qidirish FSM (chapdan o'ngga)

- | | |
|-------------|-------------|
| 1. -yap | 13. -mish |
| 2. -ti | 14. -moqda |
| 3. -gan | 15. -moqchi |
| 4. -ayotgan | 16. -gusi |
| 5. -yotgan | 17. -r |
| 6. -ar | 18. -ayotir |
| 7. -a | 19. -yotir |
| 8. -y | 20. -ajak |
| 9. -ib | 21. -jak |
| 10. -di | 22. -gay |
| 11. -mi | 23. -ayotib |
| 12. -kan | 24. -yotib |



3-chizma: Zamon qo‘shimchalarini qidirish deterministik bo‘lmagan chekli avtomat (NFA)



4-chizma: Zamon qo‘shimchalarini qidirish FSM (o‘ngdan chapga)

Shunday istisnolarni aniqladikki, fe’l so‘z turkumiga oid bo‘lmagan so‘zlar ham bu FSMda qo‘shimchalarini topishi mumkin. Chunki fe’lning vazifadosh shakllaridan sifatdosh va harakat nomi ot so‘z tukumiga hoslangan shakli bo‘lganligi sababli otning sintaktik va lug‘aviy qo‘shimchalarini qabul qiladi. Misol uchun ot so‘z turkumiga oid quyidagi *kitob-imiz(4)-ni(3)-mi(0)*, *mi*-qo‘shimchasi 0-FSMda, *ni*- qo‘shimchasi 3-FSMda, *imiz*- qo‘shimchasi 4-FSMda topilganligi sababli bizning FSM bu so‘zni fe’l deb chiqaradi. Quyida yana shunday istisnolar mavjud:

1. *qalam-imiz-ni-mi* (0,3,4)
2. *Xiva-dan-mi-siz-lar* (1,1,3,4)
3. *Xiva-dan-mi-siz-lar-a* (0,1,1,3,4)
4. *qizil-mi* (0)

5. *maktab-ga* (3)
6. *bir-ga-miz* (1,3)
7. *bir-ga-miz-a* (0,1,3)
8. *o'yin-i-ni* (3,4)

Soʻzlar faqat bitta *-miz*, *-ing*, *-imiz*, *-siz*, *-san*, *-man* qoʻshimchalari bilan tugagan boʻlsa u holda qoʻshimchani qirqib, bazaga murojat qilamiz.

Xulosa sifatida shuni aytish mumkinki, ushbu chekli avtomatlar bizga feʼl soʻz turkumiga oid soʻzlarni aniqlashda dastur tezligini oshirish uchun xizmat qiladi. Ushbu chizmalardan foydalangan holda Python dasturlash tilida dasturini tuzish maqsad qilindi.

Foydalanilgan adabiyotlar.

1. E. Kuriyozov, U. Salaev, S. Matlatipov, and G. Matlatipov, “Text classification dataset and analysis for Uzbek language,” *arXiv preprint arXiv:2302.14494*, 2023.
2. 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
3. M. Sharipov, E. Kuriyozov, O. Yuldashev, and O. Sobirov, “UzbekTagger: The rule-based POS tagger for Uzbek language,” *arXiv preprint arXiv:2301.12711*, 2023.
4. M. Sharipov and O. Yuldashov, “UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language,” *arXiv preprint arXiv:2210.16011*, 2022.
5. X. Madatov, M. Sharipov, and S. Bekchanov, “OʻZBEK TILI MATNLARIDAGI NOMUHIM SOʻZLAR,” *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, vol. 1, no. 1, 2021.
6. 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.
7. K. Madatov, S. Bekchanov, and J. Vičič, “Dataset of stopwords extracted from Uzbek texts,” *Data Brief*, vol. 43, p. 108351, 2022.
8. 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.
9. E. Kuriyozov, S. Matlatipov, M. A. Alonso, and C. Gómez-Rodríguez, “Construction and evaluation of sentiment datasets for low-resource languages: The



case of Uzbek,” in *Human Language Technology. Challenges for Computer Science and Linguistics: 9th Language and Technology Conference, LTC 2019, Poznan, Poland, May 17–19, 2019, Revised Selected Papers*, 2022, pp. 232–243