

O'ZBEK TILIDAGI MATNLARDA OT VA OLMOSH SO'Z TURKUMLARIGA TEGISHLI SO'ZLARNI ANIQLOVCHI ALGORITM YARATISH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, PhD

maqsbek72@gmail.com

UrDU dotsenti

Mavlonova Mahliyo Mavlon qizi

mmm399197@gmail.com

UrDU magistranti

Sobirov Og'abek Ollayor o'g'li

sobirovogabek0409@gmail.com

UrDU magistranti

Annotatsiya. Ushbu maqolada o'zbek tili matnlarida ot va olmosh so'z turkumlarini aniqlashning qoidaga asoslangan algoritmini yaratish ko'rib chiqilgan. Ot va olmoshlarni matnlarda aniqlash va morfologik tahlil qilish o'zbek tili uchun tabiiy til jarayonlarining (TTJ) bir qator masalalarini (*POS tagging, mashinali tarjimon va h.k.*) yechishda foydali bo'ladi. Barcha turkiy tillar kabi o'zbek tili ham agglyutinativ til bo'lib so'z o'zak yoki negiziga qo'shimchalar qo'shilishi orqali yasaladi. Bu masalani yechish uchun tub bo'lgan ot va olmosh so'zlar bazasi yaratildi, bundan tashqari ot va olmoshning so'z yasovchi va shakl yasovchi qo'shimchalar bazasi shakllantirildi.

Kalit so'zlar. *Tabiiy til jarayonlari (TTJ), teglash, so'z turkumlari, ot, olmosh, asos, qo'shimchalar, lemma, lemmatizatsiya, Chekli holatlar mashinasi (ChHM).*

Annotation. This article deals with the creation of a rule-based algorithm for identifying noun and pronoun word classes in Uzbek language texts. Identification and morphological analysis of nouns and pronouns in texts will be useful in solving a number of issues (*POS tagging, machine translation, etc.*) of natural language processes (NLP) for the Uzbek language. Like all Turkic languages, the Uzbek language is an agglutinative language, formed by adding suffixes to the root or base of the word. To solve this problem, a database of basic nouns and pronouns was created, in addition, the database of word-forming (derivational affixes) and form-forming (inflectional affixes) of nouns and pronouns was created.

Key words. *Natural Language Processing (NLP), POS-tagging, word classes, noun, pronoun, stem, affixes, lemma, lemmatization, Finite State Machine (FSM).*

Аннотация. В данной статье речь идет о создании основанного на правилах алгоритма определения классов имен существительных и местоимений в текстах на узбекском языке. Выявление и морфологический анализ существительных и местоимений в текстах будет полезен при решении

ряда вопросов (POS-тегирование, машинный перевод и др.) обработка естественного языка (ОЕЯ) для узбекского языка. Как и все тюркские языки, узбекский язык является агглютинативным языком, образованным путем добавления суффиксов к корню или основе слова. Для решения этой задачи была создана база данных основных существительных и местоимений, кроме того, создана база данных словообразовательных (словообразовательных аффиксов) и формообразующих (флективных аффиксов) существительных и местоимений.

Ключевые слова. Обработка Естественного Языка, тегирование части речи, классы слов, существительное, местоимение, основа, аффиксы, лемма, лемматизация, *Конечный Автомат (КА)*.

Kirish. Matnlarda soʻzlarni qaysi soʻz turkumiga tegishli ekanini aniqlovchi algoritmi ishlab chiqish tabiiy til jarayonlarining (TTJ) asosiy jarayonlaridan biridir. Ushbu maqolada oʻzbek tili matnlarida ot va olmosh soʻz turkumlarini aniqlashning qoidaga asoslangan algoritmini yaratish koʻrib chiqilgan. Bunda ushbu masalani yechish uchun otning *soʻz yasovchi* va *shakl yasovchi* qoʻshimchalar bazasi yaratildi, bundan tashqari tub boʻlgan *olmosh* soʻzlar bazasi shakllantirildi va shular asosida qoidaga asoslangan Chekli Holatlar Mashinasi (ChHM) yordamida algoritm ishlab chiqildi.

Tabiiy til jarayonlarining bir qancha sohalarida koʻplab ishlar olib borilgan, biz bu boradagi bir nechta ishlarni koʻrib chiqdik. Misol uchun: matnlarni teglash qoidalarga asoslangan tizimlardan ishlab chiqiladi (Kupiec, 1992), sintaktik tahlil qilingan datasetning yaratilishi (Sharipov et al., 2022), gap boʻlaklarini ajratib beruvchi dataset va dasturiy vositasi (Sharipov et al., 2023), soʻzlar oʻxshashlik va bogʻliqlik darajalarini oʻz ichiga olgan semantik tahlil dataseti (Salayev et al., 2022), soʻz lemmasini topish algoritmi (Sharipov & Sobirov, 2022). Ushbu ishlar qatorida matnlarda soʻzlarni qaysi soʻz turkimiga tegishli ekanini aniqlash tabiiy til jarayonlarining yana bir muhim masalasi ekanini koʻrishimiz mumkin.

Metodologiya. Tadqiqot maqsadi va vazifalari, oʻrganilganlik darajasini aniqlash maqsadida ilmiy yoʻnalishlarda erishilgan yutuqlar, shu bilan birga ushbu mavzu yuzasidan maʼlumotlar toʻplashdan ishiga kirishildi. Maʼlumotlar bazasi “Hozirgi oʻzbek adabiy tili”(SAYFULLAYEVA et al., 2009) va “Oʻzbek tilidan maʼruzalar toʻplami”(Erkaboeva, 2016) darsliklariga murojaat qilindi. Otlar haqida maʼlumotlar toʻplash ishlari tugagach, Oʻzbek tilining izohli lugʻati 1-2-3-4-5-jildlaridan ot va olmosh turkumiga doir soʻzlar ajratib olindi. Oʻzbek tili izohli lugʻatining 1-2-3-4-5-jildlaridan umumiy hammasi boʻlib, 14154 ta otning izohi bilan jadvashtirildi. Ajratilgan maʼlumotlar bazasi txt. fayl sifatida baza qilib kiritildi. Tadqiqotning keyingi bosqichida ot va olmosh soʻz turkumlariga doir oʻziga xos xususiyatlar kiritib chiqildi. Yaʼni ot va olmosh soʻz turkumlari asos va soʻz yasovchi, lugʻaviy shakl yasovchi va sintaktik shakl yasovchi qoʻshimchalarga ajratib chiqildi. Oʻzbek tilida jami soʻz yasovchi qoʻshimchalar, taxminan, 150 dan



	qo'shi m- chalari							Birlilik	Ko'plik		
Ot	-kor -chi -zor -dosh	- n ik i	- da gi	- gac ha	- da y - de k	- cha(k) - (a)lo q - chOq (- chiq) - (b)oy -bek -bibi - Xon(-jon) -gina - posh sha	- l a r	- (i) m - (i) ng - (s) i	- (i)miz - (i)ngi z -lari	-ning -ni -ga -da -dan	Paxtakor niki Gulchidek Gulzorlri mizning Sinf doshd ekmiz

1-jadval. Ot so'z turkumining qo'shimchalari.

Kishilik	Ko'rsatish	Belgilash (har bilan bog'liq barcha so'zlar)	Bo'lishsizlik (hech bilan bog'liq barcha so'zlar)	So'roq	O'zlik	Gumon
men	u	hamma	hech nima	kim	o'z	kimdir
sen	bu	barcha	hech bir	nima	o'zimga	nimadir
u	shu	bari	hech kim	qayer	o'zimda	qandaydir
biz	o'sha	jami	hech qanday	qanday	o'zimdanda	qayerdir
siz	ana	jamiyki	hech qayer	qanaqa	o'zimniki	qaysidir
ular	mana			qaysi	o'zimnikiga	nechtadir

2-jadval. Olmosh so'z turkumining tub so'zlari.

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

Natija. Ushbu ish orqali ishlab chiqilgan algoritmning ChHM chizmasi maqola oxirgi betida keltirilgan.

Xulosa. Tabiiy til jarayonlari masalalarida so'zlarni qaysi so'z turkumiga tegishli ekanini aniqlash masalasi asosida o'zbek tili matnlarida ot va olmosh so'z turkumlarini aniqlash masalasi ko'rib chiqildi, kelajakda ushbu masala dasturini tuzish reja qilindi.

Foydalanilgan adabiyotlar:

1. Erkaboeva, N. (2016). O'zbek tilidan ma'ruzalar to'plami. Akademnashr.
2. Kupiec, J. (1992). Robust part-of-speech tagging using a hidden Markov model. *Computer Speech & Language*, 6(3), 225–242.
3. Salaev, U., Kuriyozov, E., & Gómez-Rodríguez, C. (2022). SimRelUz: Similarity and Relatedness scores as a Semantic Evaluation Dataset for Uzbek Language. 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, SIGUL 2022 - Held in Conjunction with the International Conference on Language Resources and Evaluation, LREC 2022 - Proceedings, 199–206. www.scopus.com
4. Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova Z., Abuzalova M. Hozirgi o'zbek adabiy tili. -Toshkent, 2009.
5. Sharipov, M., Kuriyozov, E., Yuldashev, O., Sobirov, O. (2023). UzbekTagger: The rule-based POS tagger for Uzbek language. *ArXiv Preprint ArXiv:2301.12711*.
6. Sharipov, M., Mattiev, J., Sobirov, J., & Baltayev, R. (2022). Creating a morphological and syntactic tagged corpus for the Uzbek language. *ArXiv Preprint ArXiv:2210.15234*.
7. Sharipov, M., & Sobirov, O. (2022). Development of a rule-based lemmatization algorithm through Finite State Machine for Uzbek language. *ArXiv Preprint ArXiv:2210.16006*