

O'ZBEKISTON RESPUBLIKASI FANLAR AKADEMIYASI

O'ZBEKISTON RESPUBLIKASI AXBOROT TEXNOLOGIYALARI
VA KOMMUNIKATSIYALARINI RIVOJLANTIRISH VAZIRLIGI

**INFORMATIKA VA ENERGETIKA
MUAMMOLARI**
O'zbekiston jurnali

Узбекский журнал
**ПРОБЛЕМЫ
ИНФОРМАТИКИ И ЭНЕРГЕТИКИ**

Uzbek Journal

**OF THE PROBLEMS OF
INFORMATICS AND ENERGETICS**

1

2023

FAN VA TEXNOLOGIYA

O'ZBEKISTON RESPUBLIKASI FANLAR AKADEMIYASI
АКАДЕМИЯ НАУК РЕСПУБЛИКИ УЗБЕКИСТАН
O'ZBEKISTON RESPUBLIKASI RAQAMLI TEXNOLOGIYALAR
VAZIRLIGI
МИНИСТЕРСТВО ЦИФРОВЫХ ТЕХНОЛОГИЙ РЕСПУБЛИКИ
УЗБЕКИСТАН

ИНФОРМАТИКА
VA ENERGETIKA
MUAMMOLARI

1-2023

ПРОБЛЕМЫ
ИНФОРМАТИКИ
И ЭНЕРГЕТИКИ

*Журнал под таким названием издается с января 1992 г.
по 6 номеров в год*

Fan va texnologiyalar nashriyot-matbaa uyi

ТАШКЕНТ– 2023

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Ш.Х. ФАЗЫЛОВ (*главный редактор*)
Х.М. МУРАТОВ (*зам. главного редактора*)
Б.М. АЗИМОВ (*ответственный секретарь*)
О.З.ТОИРОВ (*ответственный секретарь*)
С.Ф.АМИРОВ
Р.А.ЗАХИДОВ
Х.З.ИГАМБЕРДИЕВ
М.А. ИСМАИЛОВ
О.Х.ИШНАЗАРОВ
М.М. КАМИЛОВ
М.М.КАРИМОВ
М.М.МУСАЕВ
Т.Х.НАСЫРОВ
Н. РАВШАНОВ
М.А.РАХМАТУЛЛАЕВ
Л.Ф.СУЛЮКОВА
Р.Х.ХАМДАМОВ

Адрес редакции:

100173, г.Ташкент, ул.Фозилтепа, 226

Телефоны +998 90 927 72 13, +998 99 365 43 28

e-mail: informatika-energetika@mail.ru

*Редактор: Ш. Културбога
Технический редактор: А. Муйитов
Компьютерная верстка: Ш. Мирзиёева*

Журнал зарегистрирован Агентством по печати и информации Республики Узбекистан 22.12.2006 г.
Регистрационный №0046. Изд. лист. А1 №149, 14.08.09. Изд. №3-75. Сдано в набор 02.08.2023. Подписано в печать,
11.10.2023. Формат 60×84 1/8.
Гарнитура литературная. Печать офсетная. Бумага офсетная. Усл.печ. л. - 11.8. Уч.-изд. л.7.8. Тираж: 150. Заказ №25
Цена договорная.

Издательство «Fan va texnologiyalar nashriyot-matbaa uyi»: 100173, г.Ташкент, ул.Фозилтепа, 226.
Отпечатано в типографии Научно-исследовательского института развития цифровых технологий и
искусственного интеллекта при Министерстве цифровых технологий
Республики Узбекистан: 100125, Ташкент ш., Буз-2 майдон, 17А уй.

«Fan va texnologiyalar nashriyot-matbaa uyi»

ТАШКЕНТ – 2023

СОДЕРЖАНИЕ

Информатика и управление

| | |
|--|----|
| Б.М.Азимов, А.Р.Рузикулов, Ш.З.Ихсанова. Моделирование и оптимальное управление движением хлопкоуборочной машины при различных условиях неровности полевых дорог | 3 |
| Б.Элов, Р.Алаев, З.Хусанова. Алгоритм UZB Stemming на основе разделения аффиксов..... | 14 |
| А.Х.Ишмухамедов. Решения задач надежности систем передачи данных..... | 28 |
| Э.К.Самандаров. Анализ знаний школьников с помощью простого алгоритма Байесовской классификации..... | 37 |

Энергетика

| | |
|---|----|
| О.З. Тоиров, С.С.Халиков. Функции показателей надежности насосных установок насосных станций машинного водоподъема..... | 45 |
| Н.М.Арипов,З.Ф.Мирзарахмедов,Ш.Б.Джаббаров,Б.Б. Рахмонов. Математическое описание контактных передатчиков и моделирование микроэлектронных передатчиков для кода "KJ" и разработка микроэлектронных передатчиков..... | 52 |
| П.Ж. Алланазарова, О.Х. Ишпазаров. Математическая модель системы «Асинхронный двигатель – центробежный насос»..... | 59 |
| И.К.Хужаев, Х.А.Мамадалиев, Ш.Хожиккулов, Ш.Равшанов.Применение метода Фурье для исследования распространения волны уплотнения в трубопроводах с демпфером..... | 65 |

Информационные и телекоммуникационные технологии

| | |
|--|----|
| Ю.А.Буланова, С.С.Садыков, И.Р.Самандаров. Исследование методов фильтрации шума на маммографических снимках..... | 76 |
| Л.Ф. Сулюкова, Б.А. Ёркулов. Комплекс программных продуктов для оценки качества образовательных информационных систем..... | 87 |

МУНДАРИЖА
Информатика ва бошқарув

| | |
|---|----|
| Б.М.Азимов, А.Р.Рўзикулов, Ш.З.Ихсанова. Даланинг турли потекис йўллари шароитларида пахта териш машинаси ҳаракатини моделлаштириш ва оптимал бошқариш..... | 3 |
| B.Elov, R.Alayev, Z.Xusainova. Affikslarni ajratishga asoslangan uzb stemming algoritmi..... | 14 |
| A.X.Ishmuxamedov. Ma'lumotlarni uzatish tizimlarining ishonchligi masalarini echis..... | 28 |
| E.K. Samandarov. sodda Bayes algoritmi yordamida klassifikatsiya masalasini yechish..... | 37 |

Энергетика

| | |
|---|----|
| O.З. Тоиров, С.С. Халиков. Машинали сув кўтариш насос станциялари насос қурilmаларининг ишончлилик кўрсаткичларини функциялари..... | 45 |
| N.M. Arifov, Z.F. Mirzaraxmedov, Sh. B. Djabbarov, B. B. Rahmonov. Kontaktli transmitterlarning matematik ta'rifi va mikroelektron uzatgichlarni "kj" kodi uchun modellashtirish va mikroelektron uzatgichlarni ishlab chiqish..... | 52 |
| P. Ж. Алланазарова, О. Х. Ишназаров. «Асинхрон мотор – марказдан қочма насос» тизимининг математик модели..... | 59 |
| И.К.Хўжаев, Х.А.Мамәдалиев, Ш.Хожикулов, Ш.Равшанов. Демп-ферли қувурларда зичланиш тўлқини тарқалишини тадқиқ этиш учун фурье усулининг татбиқи..... | 65 |

Ахборотли ва телекоммуникацияли технологиялар

| | |
|---|----|
| Yu. A. Bulanova, S. S. Sodiqov, I. R. Samandarov. Mammografiya tasvirlarida shovqinni filtrlash usullarini tadqiq qilish..... | 76 |
| L. F. Sulyukova, B. A. Yorkulov. Ta'lim axborot tizimlari sifatini baholash uchun dasturiy mahsulotlar majmuasi..... | 87 |

- различных условиях движения / Решение об официальной регистрации программы для электронно-вычислительных машин № DGU 20201456, Ташкент., 2020. –49 с.
7. Афанасьев В.Н. Теория оптимального управления непрерывными динамическими системами: М.: Изд-во физического факультета МГУ, 2011. –170 с.
 8. Васильев Ф. П. Численные методы решения экстремальных задач. М.: Наука, 1988. С. 421–485.
 9. Гуськов В.В. и др. Тракторы: Теория. М.: Машиностроение, 1988, С.39-40, 210–212.
 10. Динамика машин и управление машинами: Справочник / Астапов В.К., Бабицкий В.И., Вульфсон И.И. и др. Под ред. Крейнина Г.В. М.: Машиностроение, 1988. С.39–40.
 11. Смирнов Г.А. Теория движения колесных машин. М.: Машиностроение, 1990. С.89–94.

Научно-исследовательский институт
“Развитие цифровых технологий и
искусственного интеллекта при
Министерстве цифровых технологий
Республики Узбекистан”

Дата поступления
10.04.2022

UDK:811.512.133:004.4

B.ELOV, R.ALAYEV, Z.XUSAINOVA

AFFIKSLARNI AJRATISHGA ASOSLANGAN UZB STEMMING ALGORITMI

Bugungi kunda ba'zi tabiiy tilni qayta ishlash (NLP) vazifalarni hal qilishda so'z shakllarini ularning o'zakkacha qisqartirishga to'g'ri keladi. So'zdan barcha flektiv affikslarni olib tashlash va so'zning qolgan qismini lemmatizatsiya qilish NLPning muhim vazifalaridan biri hisoblanib, ushbu jarayon stemming deb yuritiladi. Stemming – NLPda matnga ishlov berishning dastlabki bosqichi bo'lib, ko'plab NLP vazifalarida amalga oshirilishi talab etiladigan jarayon. O'zbek tilining aglutinativ tuzilishga ega ekanligi sababli, lingvistik qoidalarga asoslangan stemming algortimlarini ishlab chiqish maqsadga muvofiq. Ushbu maqolada o'zbek tili so'zlari affikslarini ajratishga asoslangan stemming algoritmi keltiriladi.

Kalit so'zlar: tabiiy tilni qayta ishlash, NLP, o'zak, stem, tovush o'zgarishi, onomim, affikslarni ajratish, NER, neologizm, POS teglash, Stemming algortimi.

Б. Элов, Р. Алаев, З.Хусайнова

Алгоритм UZB Stemming на основе разделения аффиксов

Обосновано, что при решении некоторых задач обработки естественного языка (NLP) приходится сужать словоформы до их сути. Показано, что снятие всех флексивных аффиксов из слова и лемматизация остальной части слова считаются одной из важных задач NLP – процесса stemming. Объяснено, что Stemming – это начальный этап обработки текста в NLP, процесс, который требуется выполнять во многих задачах NLP. Поскольку узбекский язык имеет агглютинативную структуру, целесообразно разработать алгоритмы stemming, основанные на лингвистических правилах. В данной статье представлен алгоритм Stemming, основанный на разделении аффиксов узбекских слов.

Ключевые слова: обработка естественного языка, NLP, ядро, STEM, изменение звука, омоним, разделение аффиксов, NER, неологизм, теги POS, алгоритм Stemming.

B. Elov, R. Alaev, Z. Khusainova

UZB Stemming algorithm based on affix differences

Today, when solving some problems of natural language processing (NLP), it is necessary to narrow word forms to their essence. Removing all inflectional affixes from a Word and lemmatizing the rest of the word is considered one of the important tasks of NLP, and this process is called stemming. Stemming is the initial stage of text processing in NLP, a process that needs to be performed in many NLP tasks. Since the Uzbek language has an agglutinative structure, it is advisable to develop stemming algorithms based on linguistic rules. This article presents a Stemming algorithm based on the separation of affixes of Uzbek words.

Keywords: natural language processing, NLP, kernel, STEM, sound change, homonym, separation of affixes, NER, neologism, POS tags, stemming algorithm.

Kirish. Soʻzlarning tarkibiy qismlari bilan shugʻullanadigan morfologik tahlil NLPning asosiy yoʻnalishlaridan biri hisoblanadi. Bugungi kunda morfologik tahlilni yanada samaraliroq amalga oshirish uchun turli usullar va algortimlar ishlab chiqilgan hamda joriy qilingan [1,2,3].

Bir qarashda, lugʻatda barcha soʻzning flektiv shakllarini saqlash va hech qanday morfologik tahlilsiz tilga ishlov berish mumkindek tuyuladi. Bu yondashuv morfologik jihatdan sodda tillar uchun mos boʻlishi mumkin. Ammo affikslar birikkanidan keyin soʻz koʻplab turli shakllarni olishi mumkin boʻlgan agglyutinativ tillarga ushbu yondashuvni qoʻllash samarali natijani bermaydi. Agglyutinativ tillar uchun pos teglash va stemmingni amalga oshirish uchun zarur tushuncha va terminlar izohi [4] tadqiqotda keltirilgan.

Oʻzakni aniqlash, odatda, soʻzdan oldqoʻshimcha va oʻzakdan keyingi affikslarni olib tashlash orqali amalga oshiriladi [5,6]. Stemming jarayoni soʻzlarning affikslarini shunchaki kesib tashlaydigan "qoʻpol" evristik jarayon sifatida qaralishi mumkin. Mualliflarning fikriga koʻra, lemmatizatsiyadan farqli oʻlaroq, stemming jarayonida lugʻatdan yoki morfologik tahlidan foydalanilmaydi [7]. Stemming natijasida hosil qilingan oʻzak *haqiqiy soʻz* yoki uning morfologik ildiziga oʻxshash boʻlmashligi mumkin. NLPda soʻzning turli morfologik variantlarini ularning umumiy shakl (oʻzak, stem)ni aniqlaydigan metod stemming algoritmi deyiladi [8,9].

Stemming jarayonining maqsadi – oʻxshash soʻzlarni bir xil oʻzakgacha qisqartirish. Stemming jarayoni soʻzlarning yasalishi bilan bogʻliq: u oʻzak va yasalmaning oʻzgarishiga asoslanadi. Stemming natijasida soʻzshakllar konflyatsiya¹ga olib kelinadi. Oʻzbek tilidagi soʻzlar prefiks va suffiks kabi affikslar bilan yasalishi mumkin. Affikslar *soʻz yasovchi (derivational)* va *shakl yasovchi (inflectional)*ga ajratiladi. *Shakl yasovchi qoʻshimchalar* soʻzga birlik/koʻplik, hozirgi/oʻtgan/kelasi zamon kabi grammatik maʼnolarni yuklaydi. Lekin yangi soʻz hosil qilmaydi (leksik maʼno qismini oʻzgartirmaydi). *Soʻz yasovchi qoʻshimchalar:* oʻzakdagi maʼno bilan bogʻliq, ammo yangi soʻz yasaydi (yangi leksik maʼno hosil boʻladi).

Amalga oshirilgan ilmiy tadqiqotlar. Hozirgi kunda juda koʻplab NLP tadqiqotchilari tomonidan stemming algoritmlari ishlab chiqilgan. Zamonaviy stemming algoritmlari odatda uchta sinfga boʻlinadi: *qoidaga asoslangan, statistik va gibridd* algoritmlar.

Qoidalarga asoslangan stemmerlar avtomatik boʻlmagan qoidalardan foydalangan holda stemmlarni aniqlashga qaratilgan. *Qoidalarga asoslangan ommabop stemmerlar* sifatida Lovins [10], Porter [11,12] va Krovets [13]larni keltirish mumkin. Qoidalarga asoslangan stemmlash algoritmlari odatda nazorat qilinadi.

¹ Konflatsiya (conflation) – soʻzshakllarining umumiy oʻzagini aniqlash.

Statistik stemmlash algoritmlari stemmlarni o'rganish uchun statistik usullardan foydalanadi. Xu va Croft [14], Porter stemmerining [11] kamchiliklarini bartaraf etish uchun tasodifiy yuzaga keladigan statistik so'zdan foydalanadigan usulni taqdim etishgan. Tasodifiy statistik ma'lumotlariga asoslanib, ular Porter stemmeri tomonidan yaratilgan sinflar sonini kamaytirish uchun *grafni qismlarga ajratish algoritmini* amalda qo'llashgan [11].

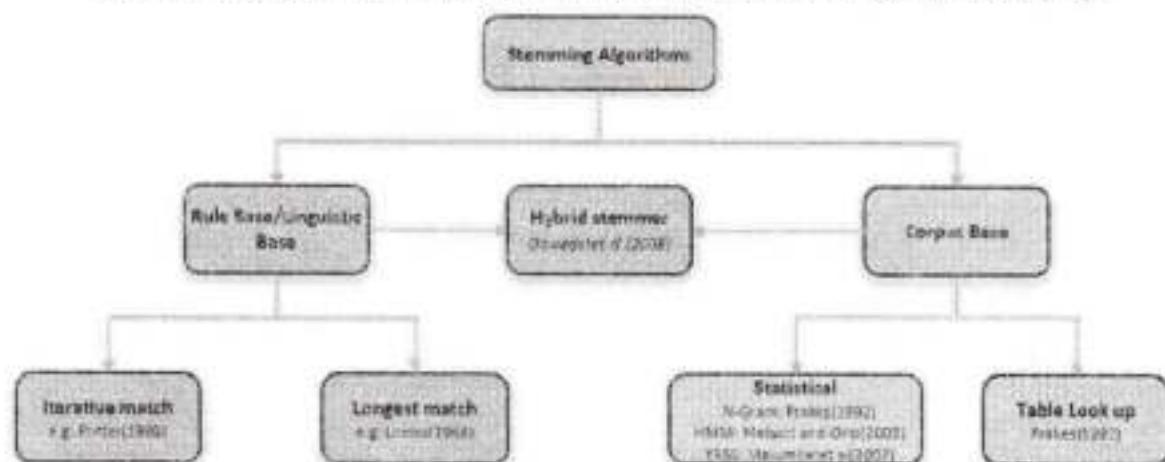
Gibrid stemmlash algoritmlari qoidalarga asoslangan va statistik usullarni yagona tizimga birlashtiradi. Ba'zi gibrid stemmlash algoritmlari Shrivastava [15], Goweder [16] va Adam [17] tomonlaridan ishlab chiqilgan.

Graflarga asoslangan stemming algoritmi Bacchin tomonidan taklif qilingan [18]. Algoritm satr ostilar to'plamini aniqlash uchun birinchi bosqichda har bir so'zni barcha mumkin bo'lgan bo'linish nuqtalariga ajratadi. Ikkinchi bosqichda satr ostilar to'plamidan foydalangan holda yo'naltirilgan graf hosil qilinadi. Nihoyat grafdagi satr ostidagilar chastotasiga qarab prefiks va suffiks ballari hisoblanib, stem aniqlanadi.

Majumder tomonidan YASS (Yet Another Suffix Striper) [19] deb nomlangan stemmlash algoritmi taqdim etilgan bo'lib, u satrlar orasidagi masofa o'lchovidan foydalanadigan klasterlash algoritmiga asoslangan. Satrlar orasidagi masofa o'lchovi so'zlar orasidagi morfologik o'xshashlikni baholash uchun ishlatilgan.

So'zlar orasidagi o'xshashlikni aniqlashga asoslangan stemmer Peng tomonidan ishlab chiqilgan [20]. Ushbu stemmer qidiruv tizimi natijalarini yaxshilash uchun IR vazifalari uchun qo'llanilgan va yuqori samaradorlikni bergan.

A.Jabbar va S.Iqbal stemming algoritmlarini quyidagi tasniflagan (1-rasm) [21].



1-rasm. Stemming algoritmlar tasnifiga A.Jabbar va S.Iqbal yondashuvi

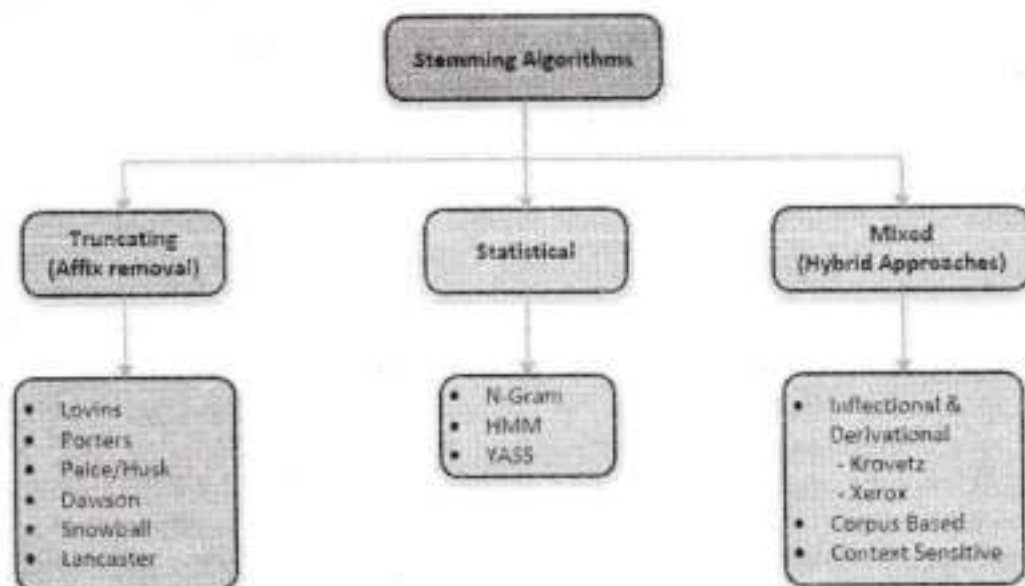
Hind olimi G.Anjali stemming algoritmlarini kesish (truncating), statistik (statistical) va aralash (mixed) usullar kabi uch guruhga ajratgan: [9]. Ushbu guruhlarning har biriga mos algortimlar o'ziga xos yondashuvlar asosida so'z variantlariga mos o'zaklarni aniqlaydi. Ushbu algortim va usullar tasnifi quyidagi 2-rasmda keltirilgan.

Ushbu maqolada o'zbek tilidagi so'zlarni affikslarni ajratishga asoslangan usul orqali stendlash algortimini ishlab chiqish nazarda tutilgan. Shu sababli, ushbu tadqiqot ishida Truncating (Affix removal) yondashuviga asoslangan ilmiy ishlar sharhi keltiriladi. Zamonaviy stemming algortimlari haqida batafsil ma'lumot mualliflarning [22] tadqiqot ishida yoritilgan.

Affikslarni ajratishga asoslangan usullar

Ushbu usullar so'zning qo'shimchalari yoki prefiks(affikslar)ini olib tashlash (truncating)ga asoslanadi. Bu turdagi eng oddiy stemmer Tuncate(n) stemmeri bo'lib, u so'zni n-belgisigacha qisqartadi. Bunda, so'zning birinchi n ta harfni saqlagan holda, qolganlarini olib tashlaydi [23,24]. Bu usulda uzunligi n dan qisqa so'zlar o'z shaklida

saqlanadi. Soʻz uzunligi kichik boʻlgan holda ortiqcha (keraksiz) amallar soni ortadi. Affiksni olib tashlashga asoslangan yana bir yondashuv S-stemmer algoritmidagi ingliz tilidagi otlarning birlik va koʻplik shakllarini birlashtiradi. Bu algoritm Donna Xarman tomonidan taklif qilingan [25]. S-stemmer algoritmidagi koʻplikdagi qoʻshimchalarni birlik shakliga aylantirish qoidalarini ishlab chiqilgan [23].



2-rasm. Stemming algoritmlarining tasnifi

Brute force/Table Lookup usulida oʻzgartirilgan yoki asos soʻzlar oʻzaklarga moslashtiriladi. Agglyutinativ shakllarga asoslangan tillar (turk, oʻzbek) uchun bu usulda katta hajmdagi jadvallar hosil qilinadi [26]. Hozirgi kunda *soʻzdan affikslarni grammatik qoidalar asosida ajratish* yoki *soʻz morfemasidan, jumladan, POS teglash orqali oʻzakni aniqlash* algortimlaridan foydalanilmoqda. Ushbu yondashuvlarning barchasi tabiiy tilni bilishni talab qiladi.

Baʼzi yondashuvlarda POS teglash amali gapdagi soʻzlarni ketma-ketlikdagi belgilash/teglash muammosi sifatida qaraladi. Bunday turdagi yondashuvlar asosidagi algortimlarda koʻp hollarda Yashirin Markov Modeli (Hidden Markov Models, HMMs)dan foydalaniladi [27]. Merialdo uch sinfdan iborat Markov modelini taqdim etgan boʻlib, tadqiqotda korpusdagi oʻquv maʼlumotlarining turli oʻlchamlari uchun turli parametrlarni baholash usullari taqqoslangan [28]. Korpus maʼlumotlari uchun nisbiy chastotali trening ishlatilgan va tegsiz maʼlumotlar uchun Maksimal ehtimollik (Maximum Likelihood) usuli qoʻllanilgan.

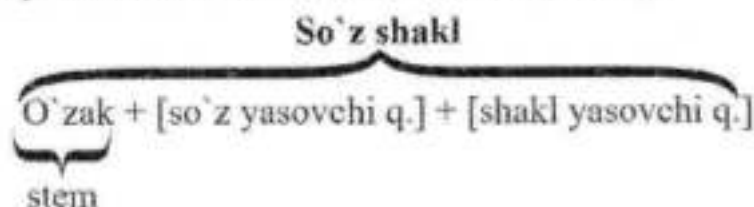
Banko va Moore har bir soʻzni faqat joriy soʻz “teg”idan emas, balki oldingi va keyingi soʻz “teg”lari (valentlik)ni oʻz ichiga olgan uchta qoʻshni teg asosida kontekstli HMM teggerini taqdim etadilar [29]. Ushbu model asosiy HMM bilan solishtirganda koʻproq kontekstli maʼlumotlarni oʻz ichiga olgan va samarali natija qaytargan.

Jonson HMMga asoslangan POS teglashda ishlatiladigan turli parametrlarni solishtirgan. Shu maqsadda Expectation Maximization (EM), Variatsion Bayes va Gibbs namunalaridan foydalangan [30]. Tadqiqot Gibbs namunasi va variatsion Bayes baholovchisi bilan solishtirganda EM algortimining past samaradorligini koʻrsatib bergan.

Metodologiya. Oʻzbek tilida gaplar alohida soʻzlardan tashkil topadi. Morfologik jihatdan soʻzlar oʻzakka baʼzi qoʻshimchalar qoʻshish orqali hosil qilinadi. Bu jarayonda soʻzda fonetik oʻzgarishlar (phonetic harmony) yuzaga kelishi mumkin va bu bevosita matnda oʻz aksini topadi. Oʻzakning oʻzi ham soʻzning oʻziga xos

ma'nosini ifodalovchi so'z bo'lishi mumkin. Affikslar gapda muhim rol o'ynasa-da, mustaqil ma'noga ega emas.

Affikslar *so'z yasovchi (derivational suffixes)* va *shakl yasovchi (inflectional suffixes)* turga ajratiladi [31]. Shakl yasovchi qo'shimchalar so'zning faqat grammatik vazifasini o'zgartiradi. So'z yasovchi qo'shimchalarni o'zakka qo'shish orqali so'zda semantik o'zgarish yuzaga kelishi mumkin. Shakl yasovchi qo'shimchalar so'zda sintaktik o'zgarishlarni keltirib chiqaradi. O'zakka avval so'z yasovchi qo'shimchalar, so'ngra shakl yasovchi qo'shimchalar biriktiriladi. Biroq o'zakka to'g'ridan-to'g'ri shakl yasovchi qo'shimchalar biriktirilishi ham mumkin (3-rasm).



3-rasm. O'zbek tilidagi so'zning umumiy morfologik tuzilishi

Morfologik tahlil sohasida hozirgacha bajarilgan ishlarni ikki sinfga birlashtirish mumkin: stemga va affikslarni olib tashlashga asoslangan. Stemga asoslangan morfologik tahlilni amalga oshirish uchun dastlab so'zning stemi, so'ngra affiksiari aniqlanadi. Affikslarni olib tashlashga asoslangan yondashuvda dastlab so'zdagi affikslar aniqlanadi. Affikslar olib tashlangandan so'ng, so'zning qolgan qismini **stem** deb hisoblash mumkin yoki bu taxminni tasdiqlash uchun lug'atdan foydalanish mumkin.

So'zga qo'shila oladigan qo'shimchalar soni va ularning ko'p sonli birikmalari agglyutinativ tillarda o'zakni aniqlash jarayonini murakkab muammoga aylantiradi. Chunki ko'pchilik agglyutinativ tillarda qo'shimchalar kombinatsiyasi murakkab so'z shakllarini hosil qiladi. Yangi so'z yasovchi yoki so'z shakllarini hosil qiluvchi qo'shimchalar soni bo'yicha ko'rsatkichlar ham har xil hisoblanadi. O'zbek tilida **228** ta so'z yasovchi, **69** ta lug'aviy shakl hosil qiluvchi va **41** ta sintaktik shakl hosil qiluvchi qo'shimchalar mavjudligi darsliklarda ko'rsatiladi. Agar ularning variantlarini ham qo'shsak, aslida bu son yana oshib ketadi².

O'zbek tilidagi so'zlarning stemini aniqlash uchun o'zak va unga birikadigan barcha turdagi qo'shimchalar aniqlab chiqiladi. An'anaviy stemming algoritmlari qo'shimchalar va ba'zi morfologik qoidalarga asoslangan bo'lib, stemming jarayoni natijasida stemdagi noaniqlik yuzaga kelishi mumkin. Stemming jarayonida odatda so'zdagi barcha turdagi qo'shimchalar olib tashlanadi. Lekin bu usulda stemming amalga oshirilganda, ba'zi hollarda noto'g'ri natija olinishi mumkin. Masalan, "olmani" so'zining stemi "olma", "olmadi" so'zining stemi esa "ol". Ya'ni so'z shaklining gapdagi so'z turkumiga qarab stem ham o'zgarishi mumkin. O'zbek tilidagi so'zlarni stemini aniqlashda quyidagi muammolar yuzaga kelishi mumkin:

- o'zak va qo'shimchani bitta o'zak bilan omonim bo'lishi;
- so'zning tovush o'zgarishiga uchrashi;
- neologizm va NERlarni stemmlash.

O'zak va qo'shimchani bitta o'zak bilan omonim bo'lishi. Ko'p ma'noli o'zakni aniqlash ancha murakkab jarayon bo'lib, stemming jarayonida gap darajasidagi semantik ma'lumotlar e'tiborga olinmaydi. Ba'zida so'zning POS tegi uning o'zagining POS tegi bilan bir xil bo'lmasligi mumkin.

POS teglash – bu berilgan gapdagi har bir so'zshaklga uning turkum (*ot, fe'l, sifat, son, ravish* yoki *olmosh*)ga mansubligini belgilash (teglash) vazifasidir. POS

² <https://github.com/KhZilola/Python-Codes>

teglash lug'at asosida yoki lug'atsiz amalga oshirilishi mumkin. POS teglash bo'yicha amalga oshirilgan ilmiy tadqiqotlarning aksariyatida [32,33] so'zlarga asoslangan (word-based) yondashuv qo'llanilgan. Agglyutinativ tillarning ba'zilarida POS teglash jarayonini amalga oshirish uchun so'zlarning stemlaridan foydalaniladi [34]. Quyidagi 1-jadvalda o'zbek tilidagi so'zning lemmasi, stemi va POS tegiga namunalar keltirilgan [4,35].

1-jadval

O'zbek tilidagi so'zning lemmasi, stemi va POS tegi

| № | so'z | Lemma | POS | Stem | POS | Root | POS |
|---|-------------------|-----------------|-----|----------|-----|---------|-----|
| 1 | muzladi | muzlamoq | VB | muz | N | muz | N |
| 2 | issiqroq | issiq | JJ | isi | VB | isi | VB |
| 3 | soddalashtiriladi | soddalashtirmoq | VB | sodda | JJ | sodda | JJ |
| 4 | ixtiyoriy | ixtiyoriy | JJ | ixtiyor | N | ixtiyor | N |
| 5 | qo'llaniladigan | qo'llamoq | VB | qo'l | N | qo'l | N |
| 6 | yo'lakda | yo'lak | N | yo'l | N | yo'l | N |
| 7 | qishlog'im | qishloq | N | qishlog' | ? | qishloq | N |

O'zbek tilida so'z yasovchi va shakl hosil qiluvchi qo'shimchalar o'rtasida omonimiya hodisasi uchraydi [36,37,38]. Bu esa stemming jarayonida muammoli vaziyatni yuzaga keltiradi. Quyidagi 2-jadvalda omonim qo'shimchalar ro'yxatidan namuna keltirilgan[4]:

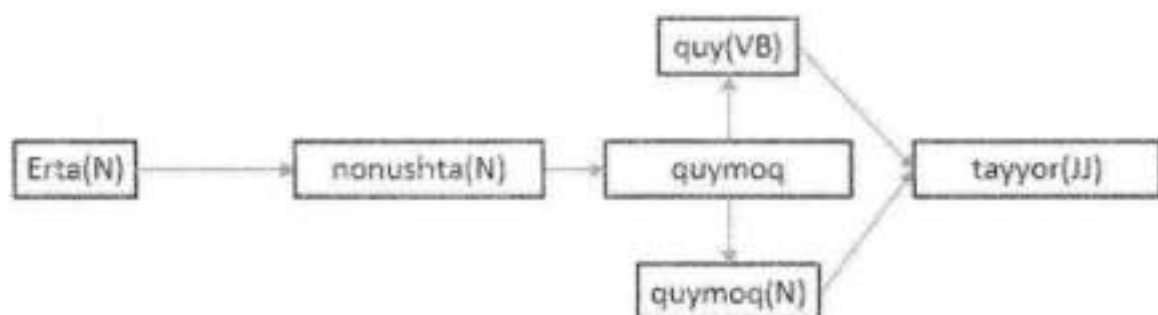
2-jadval

So'z yasovchi va shakl hosil qiluvchi qo'shimchalar o'rtasida omonimiya

| | Shakl yasovchi qo'shimcha | So'z yasovchi qo'shimcha |
|---|-----------------------------------|---|
| - | av (lug'aviy shakl yas.) boray | kuchay (fe'l) |
| - | gi (lug'aviy shakl yas.) borgim | supurgi (ot), yozgi (sifat) |
| - | da (sintaktik shakl yas.) uyda | undamoq (fe'l) |
| - | i (sintaktik shakl yas.) do'sti | jannati (sifat), boyi (fe'l) |
| - | in (lug'aviy shakl yas.) ko'rin | ekin (ot), sog'in (sifat) |
| - | im (sintaktik shakl yas.) uyim | bilim (ot), ayrim (sifat) |
| - | ir (lug'aviy shakl yas.) o'chir | gapir (fe'l) |
| - | iq (lug'aviy shakl yas.) siniqmoq | yo'liq (fe'l), ochiq (sifat), chiziq (ot) |
| - | y (lug'aviy shakl yas.) o'qiy | qoray (sifat) |

O'zbek tilidagi gaplarda stemdagi noaniqlikni quyidagi 4-rasmda ko'rish mumkin:

Ertalab nonushtaga **quymoq** tayyorlandi.



4-rasm. O'zbek tilidagi gaplarda stemdagi noaniqlik

So'zning tovush o'zgarishiga uchrashi. Shakl yasovchi qo'shimchalarni o'zakning oxirgi harflariga qo'shish najjasida ba'zi hollarda so'zda fonetik o'zgarishlar (insertion, deletion, phonetic harmony, and assimilation) yuzaga kelishi mumkin [4].

O'zbek tilidagi so'zlarda *tovush ortishi*, *tushishi* va *almashinishi* (weaking, assimilation) kabi uch xil fonetik o'zgarishlar amalga oshirilishi mumkin (3-jadval).

3-jadval

O'zbek tilidagi so'zlarda tovush ortishi, tushishi va almashinishi

| So'z | O'zak | stem | Turi |
|------------|---------|----------|------|
| shahrim | shahar | shahr | 1 |
| qishlog'im | qishloq | qishlog' | 2 |
| huquqim | huquq | huquq | 0 |
| ayblov | aybla | ayblo | 2 |
| aldov | alda | aldo | 2 |
| angla | ong | ang | 2 |
| anov | ana | ano | 2 |

Turi: 0-istisno, 1-tushishi, 2-almashinishi, 3-ortishi.

Stemni aniqlashdagi tovush o'zgarishiga uchrashi muammosini hal qilish uchun, birinchi bosqichda o'zak va qo'shimchalarning chegaralari aniqlanadi, ikkinchi bosqichda esa lemmatizatsiya amalga oshiriladi. Lemmatizatsiya natijasida xato hosil qilingan stemlar lug'atda mavjud root (o'zak)ga o'zgartiriladi.

Neologizm va NERlarni stemmlash.

NER (Named-entity recognition) – matndagi asosiy obyektlarni ajratib olish va ularni oldindan belgilangan toifalarga ajratishi mumkin bo'lgan NLP usuli. Matndan *shaxs ismlari*, *joylashuv nomlari*, *kompaniya nomlari* va shu kabi lug'atda mavjud bo'lmagan nomlangan obyektlarni aniqlash jarayoni ko'plab NLP vazifalarini hal qilishda muhim bosqich sanaladi. NLPda nomlangan obyektning tanib olish, odatda, *obyektning identifikatsiyalash*, *obyektning ajratib olish* yoki *obyektning bo'laklarga ajratish* deb ham yuritiladi [38]. NER obyektlari va neologizmlar lug'atda mavjud bo'lmaganligi sababli ularni tanib olish juda murakkab masala hisoblanadi.

NERlarning stemini topish muammosi yuzaga kelganda shakl yasovchilar kesiladi, so'z yasovchi shaklidagi qo'shimcha yoki so'zning qismi qoldiriladi, shu qism NER sanaladi: *O'zbekistondan* so'zshaklining stemi *O'zbekiston*.

Shunday qo'shimchalar borki, ular so'z yasovchi va shakl yasovchi vazifasida keladi (4-jadval).

4-jadval

So'z yasovchi va shakl yasovchi omonim qo'shimchalar

| Shakl yasovchi va so'z yasovchi qo'shimchalar | | |
|---|------|-------|
| -ay | -k | -chak |
| -gi | -ka | -chiq |
| -da | -kin | -choq |
| -i | -la | -qa |
| -in | -lab | -qin |
| -im | -m | -sa |
| -ir | -ma | -siz |
| -iq | -moq | -xon |
| -y | -cha | |

Bunday qo'shimchalar bosh harf bilan yozilgan so'zlarning tarkibida kelganda, shakl yasovchilar va so'z yasovchilar tarkibida bo'lsa, so'z shakl tarkibida qoldiriladi va shu shaklida stem deb olinadi. Masalan, *Jon Kennedi* so'zning tarkibida *-i* harfi bor. Dastur o'zakni bilmaganligi sababli, ya'ni mazkur so'z o'zbek tili lug'atida mavjud emasligi sababli o'zakni ajratolmay qoladi, natijada *-i* qo'shimchasini kesib, *Kenned* so'zini o'zak deb olishi mumkin. Bunday holatdan qochish maqsadida shakl yasovchi va so'z yasovchilar orasida omonimiya hosil qiluvchi qo'shimcha bilan shakldosh bo'lgan har qanday birlik so'zshakl tarkibida qoldiriladi.

Neologizmlarni stemplash muammolari. Neologizmlarning yangiligi dastlab paydo bo'lgan vaqtlardagina sezilib turadi: vaqt o'tgach, ular "yangilik" xususiyatini yo'qotib, odatda, faol so'zlar qatoriga o'tadi³. Neologizmning shakliy neologizm, semantik neologizm, funksional neologizm, ijtimoiy neologizm, texnologik neologizm, stilistik neologizm kabi turlari mavjud.

Neologizmlarning paydo bo'lish yo'llari xilma-xil bo'lib, ular tilning mavjud lug'aviy tarkibi va grammatik qonun-qoidalari asosida yangi so'z yasash yo'li, shuningdek, mavjud so'zning lug'aviy ma'nolaridan birini yangi ma'noda qo'llash yo'li bilan va boshqa tildan so'z qabul qilish orqali hosil qilinadi.

Neologizmlar tarkibida -izm (neologizm), -ik (daltonik), -la (gugilash) kabi qo'shimchalar uchraydi.

Neologizmlar lug'atda mavjud bo'lmaganligi sababli ularni stemplashda muammolar yuzaga chiqadi. Ularning tarkibidagi qo'shimchalar, so'zning bir qismining qo'shimchaga o'xshab qolishi muammolari shular shumlasidan. Bunday holatda shakl yasovchi qo'shimchalar bazasida mavjud qo'shimchalar kesiladi. Qolgan qism stemga teng keladi. O'zbek tilidagi neologizmlar va NERlarga mos stem yuqoridagi 3-rasmda keltirilgan.

Neologizmlar keng jamoatchilik tomonidan ma'lum vaqt oralig'ida faol qo'llanib og'zaki va yozma nutqqa ko'chganda tilning leksik boyligi sifatida rasman e'tirof etilishi, ya'ni lug'atga kiritilishi mumkin. Bu jarayondan so'ng ularni stemplash lug'atdagi so'zlarni stemplash qoidasi asosida amalga oshiriladi.

Taklif qilinayotgan yechim

Affikslarni olib tashlashga asoslangan o'zbek tili stemmerini ishlab chiqish uchun *til korpusi*, *affikslar bazasi*, *o'zbek tili morfoleksikoni*, *root (o'zaklar bazasidan foydalanildi* [4]. So'z shaklining so'z turkumi (POS tegi)ni inobatga olib, stemni aniqlash uchun quyidagi 2-bosqichdagi qadamlar bajarildi.

1-bosqich. Ma'lumotlar (dataset)ni tayyorlash:

1. O'zakka qo'shilishi mumkin bo'lgan **prefiks va suffiks qo'shimchalar** aniqlanadi (5,6-jadvallar).

5-jadval

O'zbek tilidagi prefiks qo'shimchalari

| Prefix qo'shimchalar | | |
|----------------------|---------|-------|
| ba- | be- | ham- |
| bo- | no- | bar- |
| bad- | ser- | be- |
| bar- | g'ayri- | xush- |

6-jadval

O'zbek tilidagi suffiks qo'shimchalar (POS bo'yicha)

| POS | so'z yasovchi (derevatsion) | shakl yasovchi (flektiv) |
|------------|-----------------------------|--------------------------|
| Ot (N) | 51 | 17 |
| Son (Num) | 1 | 5 |
| Fe'l (Vb) | 30 | 22+10 (nisbat) |
| Sifat (JJ) | 97 | 9 |
| Olmosh (P) | - | - |
| Ravish (R) | 18 | 5 |

2. Har bir so'z turkumiga oid bir nechtadan (faol) so'z shakllari olinib, ularning stemmini so'z shaklidan o'chirib tashlash orqali, berilgan so'z turkumidagi stemplarga qo'shilishi mumkin bo'lgan prefiks va suffiks qo'shimchalar satrlar to'plami aniqlandi.

³ <https://uz.wikipedia.org/wiki/Neologizmlar>

So'z shakllarni prefiks, stem va suffiks qo'shimchalar satriga ajratish

| So'z shakl | prefiks | stem | suffiks qo'shimchalar satri |
|---------------------|---------|--------|-----------------------------|
| <i>kitobni</i> | | kitob | ning |
| <i>kitoblarning</i> | | kitob | larning |
| <i>kitob</i> | | kitob | |
| <i>kitobda</i> | | kitob | da |
| <i>kitoblariga</i> | | kitob | lariga |
| <i>olmadi</i> | | ol | di |
| <i>olmaysan</i> | | ol | ysan |
| <i>olmaga</i> | | olma | ga |
| <i>olmalari</i> | | olma | lari |
| <i>shahriga</i> | | shahr | iga |
| <i>shahringdan</i> | | shahr | ingdan |
| <i>shaharda</i> | | shahar | da |
| <i>badavlat</i> | ba | davlat | |
| <i>berahm</i> | be | rahm | |

7. $p + s' + q = w$ shartni qanoatlantiradigan, barcha berilgan har bir so'z shakli uchun $g = \langle p, s, q \rangle$ qiymatlar to'plami aniqlandi. Bunda,

w – so'z shakli;

s' – stem;

$p \in P$;

P – prefiks qo'shimchalar satri to'plami;

$s \in S, S$ – stem so'z turkumlari to'plami;

$q \in Q, Q$ – suffiks qo'shimchalar satri to'plami;

$g \in G, G$ – har bir elementi $\langle p, s, q \rangle$ uchta parametrdan iborat to'plam.

8. Aniqlangan $\langle p, s, q \rangle$ qiymatlariga misol 9-jadvalda keltirilgan:

Prefiks, stem va suffiks qo'shimchalar

| Prefiks (p) | Stem so'z turkumi (s) | Suffiks qo'shimchalar satri (q) |
|-------------|-----------------------|---------------------------------|
| | ot | ning |
| | ot | larning |
| | ot | da |
| | ot | lariga |
| | fe'l | di |
| | fe'l | ysan |
| | ot | ga |
| | ot | lari |
| | ot | iga |
| | ot | ingdan |
| | ot | da |
| ba | ot | |
| be | ot | |

9-jadvaldagi qiymatlar aniqlanganidan so'ng quyida keltirilgan **UzbStemming algoritmi** yordamida so'z shaklining stemini aniqlash mumkin.

II-bosqich. UzbStemming algoritmi.

Algoritmda quyidagi belgilashlardan foydalaniladi:

R – o'zaklar to'plami;

- G – $\langle p, s, q \rangle$ qiymatlar to'plami;
 T – G to'plam elementlarining p va q qiymatlari uzunliklari bo'yicha kamayish tartibida tartiblangan to'plam;
 V – unli harflar to'plami;
 C – undosh harflar qiymatlar to'plami;
 w – so'z shakli;
replacelastletter (x, y) – x satrning oxirgi harfini y harfga almashtirish funksiyasi;
removelastvowel (x) – x satrning oxiridan boshlab 2-o'rinda turgan unli harfini o'chirish funksiyasi;
right (x, n) – x satrning oxiridan boshlab n -harfini aniqlash funksiyasi;
left (x, n) – x satrning boshidan boshlab n -harfini aniqlash funksiyasi;
RPOS (r) – r o'zakning so'z turkumlari ro'yxatini aniqlash funksiyasi. O'zaklar va ularning so'z turkumlari bazasi asosida aniqlanadi;
QPOS (p, q) – p va q qo'shimchalar satrining qo'shilishi mumkin bo'lgan so'z turkumlari ro'yxatini aniqlash funksiyasi. 9-jadval asosida aniqlanadi.
1. w satr kiritilsin.
 2. T to'plamning har bir elementi uchun.
 - 2.1. Agar w satri p bilan boshlanib q bilan tugasa, **2.2-qadamga** o'tilsin, aks holda **2.6-qadamga** o'tilsin.
 - 2.2. w satrning p va q qismlari kesib yuborilsin, qolgan qismini r' bilan ifodalaymiz. Bunda, $p + r' + q = w$ tenglik o'rinli.
 - 2.3. r' satrida kamida bitta unli mavjud bo'lsa, **2.4-qadamga** o'tilsin, aks holda **2.6-qadamga** o'tilsin.
 - 2.4. Agar $r' \in R$ bo'lsa, **4-qadamga** o'tilsin, aks holda **2.5-qadamga** o'tilsin.
 - 2.5. R to'plamning har bir elementi r uchun
 - 2.5.1 Agar $RPOS(r) \cap QPOS(p, q) = \emptyset$ bo'lsa, **2.5.19-qadamga** o'tilsin, aks holda **q1 = left(q, 1)**, **2.5.2-qadamga** o'tilsin.
 - 2.5.2 Agar $right(r, 1) = "k"$ va $q1 = "i"$ bo'lsa, **2.5.3-qadamga** o'tilsin, aks holda **2.5.5-qadamga** o'tilsin.
 - 2.5.3 $z = replacelastletter(r, "g")$.
 - 2.5.4 Agar $r' = z$ bo'lsa, **4-qadamga** o'tilsin.
 - 2.5.5 Agar $right(r, 1) = "q"$ va $q1 = "i"$ bo'lsa, **2.5.6-qadamga** o'tilsin, aks holda **2.5.8-qadamga** o'tilsin.
 - 2.5.6 $z = replacelastletter(r, "g'")$.
 - 2.5.7 Agar $r' = z$ bo'lsa, **4-qadamga** o'tilsin.
 - 2.5.8 $l = length(r)$.
 - 2.5.9 Agar $q1 = "i"$ va $l > 4$ va $right(r, 1) \in C$ va $right(r, 2) \in V$ bo'lsa, **2.5.10-qadamga** o'tilsin, aks holda **2.5.12-qadamga** o'tilsin.
 - 2.5.10 $z = removelastvowel(r)$.
 - 2.5.11 Agar $r' = z$ bo'lsa, **4-qadamga** o'tilsin.
 - 2.5.12 Agar $"Verb" \in RPOS(r)$ bo'lsa, **2.5.13-qadamga** o'tilsin, aks holda **2.5.19-qadamga** o'tilsin.
 - 2.5.13 Agar $right(r, 1) = "a"$ bo'lsa, **2.5.14-qadamga** o'tilsin, aks holda **2.5.16-qadamga** o'tilsin.
 - 2.5.14 $z = replacelastletter(r, "o")$.
 - 2.5.15 Agar $q1 \in {"v", "q"}$ va $r' = z$ bo'lsa, **4-qadamga** o'tilsin.
 - 2.5.16 Agar $right(r, 1) = "i"$ bo'lsa, **2.5.17-qadamga** o'tilsin, aks holda **2.5.19-qadamga** o'tilsin.
 - 2.5.17 $z = replacelastletter(r, "u")$.
 - 2.5.18 Agar $q1 \in {"v", "q"}$ va $r' = z$ bo'lsa, **4-qadamga** o'tilsin.

- 2.5.19 *R* to'plamning keyingi elementi olinsin, **2.5.1-qadamga** o'tilsin.
- 2.6. *T* to'plamning keyingi elementi olinsin va **2.1-qadamga** o'tilsin.
3. "**Stem topilmadi**". **5-qadamga** o'tilsin.
4. Stem *r'*.
5. Tamom.

Izoh: Takroriy so'zlar uchun stem "-" bilan ajratilgan faqat bitta qismi uchun keltirilgan algoritm yordamida aniqlanadi. Juft so'zlar uchun stem "-" bilan ajratilgan har bir qismi uchun alohida-alohida keltirilgan algoritm yordamida aniqlanishga urinib ko'riladi. Qaysi qismlar uchun stem aniqlansa, o'sha qismlarning stemlari birgalikda natija sifatida olinadi. Agar juft so'zning "-" bilan ajratilgan bironta qismi uchun stem topilmasa, stemni butun juft so'z uchun keltirilgan algoritm yordamida aniqlanadi.

Qo'shma so'zlar uchun stem so'zning har bir qismi uchun alohida-alohida keltirilgan algoritm yordamida aniqlanadi.

Affikslarni olib tashlashga asoslangan yuqorida keltirilgan UzbStemmerini o'zbek tili ta'limiy korpusi (<http://uzschoolcorpara.uz/uz/Search>)dagi 100000 dan iborat gaplarga qo'llash natijasida **97.5%** aniqlikdagi natijani berdi. UzbStemmer algoritmi samaradorligini oshirish maqsadida, algoritmning I bosqich (ma'lumotlarni tayyorlash)ning 3-qadamida aniqlangan affikslar guruhida mavjud bo'lmagan qo'shimchalar kombinatsiyasini til qoidalariga asoslangan gibrid yondashuvdan foydalanish maqsadga muvofiq. Shuningdek, til korpusi hajmini oshirish, NER obyektlari va neologizmlarni samarali aniqlash usullarini ishlab chiqish, omonimiya muammosini bartaraf qilish orqali stemming algoritmi aniqligini oshirish imkoniyati mavjud.

Xulosa. Lug'at orqali POS teglash va stemmingni amalga oshirish tabiiy tilni qayta ishlashning ko'plab vazifalariga qiyinchilik tug'diradi. POS teglash va stemlash uchun til korpusidan foydalanish orqali lug'at bilan bo'ladigan muammolar hal qilinadi. Til korpusi ustida o'tkazilgan turli tajribalar shuni ko'rsatadiki, o'zak ma'lumotlarini sintaktik vazifa bilan birlashtirish morfologik jihatdan boy til uchun POS teglash natijasini yaxshilaydi, bu esa NLP vazifasining hal qilish samaradorligini oshirishga xizmat qiladi. Maqolada korpusda turli xil bog'liqliklarni qabul qiladigan bir nechta turli xil qo'shma modellar taqdim etildi. Umumiy eksperimental natijalar *til qoidalari* va *statistik yondashuvlarga* asoslangan, *HMM modeli yordamida POS teglash* imkoniyatiga ega gibrid UzbStemmeri orqali o'zbek tili matnlarini stemmingni amalga oshirish yuqori ko'rsatkichlarni taqdim etadi. Natijalar shuni ko'rsatadiki, semantik ma'lumotlardan foydalanish orqali stemlash va POS teglash algoritmi aniqligini sezilarli darajada yaxshilaydi. Mualliflar bugungi kunda N-grammlar va neyron tarmoqlariga asoslangan o'zbek stemmerini ishlab chiqarish ustida tadqiqot olib bormoqdalar. Ushbu maqolada keltirilgan yondashuv neyron tarmoqlarga asoslangan stemmerni ishlab chiqishga asos bo'lib xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. Böllüçü N., & Can B. Unsupervised joint PoS tagging and stemming for agglutinative languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, (2019). №18(3). <https://doi.org/10.1145/3292398>.
2. Kışla T., & Karaoğlu B. A hybrid Statistical Approach to Stemming in Turkish: An Agglutinative Language // *Anadolu University Journal of Science and Technology-A Applied Sciences and Engineering*, (2016). №17(2). <https://doi.org/10.18038/btda.31812>.
3. Elov B. Hamroyeva Sh., Axmedova X. Methods for creating a morphological analyzer // *14th International Conference on Intelligent Human Computer*

- Interaction, IHCI Tashkent.2022, 19–23 October. https://link.springer.com/chapter/10.1007/978-3-031-27199-1_4.
4. Elov B., Hamroyeva Sh., Abdullayeva O., Xusainova Z., Xudayberganov N. Agglyutinativ tillar uchun POS teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida). *Til va madaniyat, Kompyuter lingvistikasi*. Vol.1(6). (2023) B. 35–50.
 5. Eiman Tamah Al-Shammari Towards An ErrorFree Stemming in // *Proceedings of ADIS European Conference Data Mining, 2008*. P. 160–163.
 6. Paice C. D. Another Stemmer. *ACM SIGIR Forum*, (1990) №24(3). <https://doi.org/10.1145/101306.101310>.
 7. Elov B.B., Khamroeva M.Sh., Xusainova Z.Y. NLP (tabiiy tilga ishlov berish)ning pipeline konveyeri. *MUHAMMAD AL-XORAZMIY AVLODLAR 2023*. № 1 (23), mart 181–191-b.
 8. Paice C. D. An evaluation method for stemming algorithms. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1994*. https://doi.org/10.1007/978-1-4471-2099-5_5.
 9. Anjali M., & Jivani G. A Comparative Study of Stemming Algorithms. *Int. J. Comp. Tech. Appl.*, (2011) №2(6).
 10. Lovins J. B. Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics*. (1968) №11 June.
 11. Porter M. F. An algorithm for suffix stripping. *Program*. (2006). №40(3). <https://doi.org/10.1108/00330330610681286>.
 12. Porter M.F. *Snowball: A language for stemming algorithms*, 2001.
 13. Krovetz R. Viewing morphology as an inference process. *Artificial Intelligence*, (2000). №118(1–2). [https://doi.org/10.1016/S0004-3702\(99\)00101-0](https://doi.org/10.1016/S0004-3702(99)00101-0).
 14. Xu J., & Croft W. B. Corpus-based stemming using cooccurrence of word variants. *ACM Transactions on Information Systems*. (1998). №16(1). <https://doi.org/10.1145/267954.267957>.
 15. Manish Sh., Nitin A., Bibhuti M. Morphology based natural language processing tools for indian languages. In *Proceedings of the 4th Annual Inter Research Institute Student Seminar in Computer Science, IIT, Kanpur, India, April, 2011*. Citeseer. 81–92-p.
 16. Goweder A., Alhami H., Rashed T, Al-Musrati A. A hybrid method for stemming Arabic text // *Journal of computer Science*, 2010. V. 1-7-p.
 17. Adam G., Asimakis K., Bouras C., & Pouloupoulos V. An efficient mechanism for stemming and tagging: The case of Greek language. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2010). 6278 LNAI(PART 3). https://doi.org/10.1007/978-3-642-15393-8_44.
 18. Bacchin M., Ferro N., & Melucci M. The effectiveness of a graph-based algorithm for stemming. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2002). 2555. https://doi.org/10.1007/3-540-36227-4_12.
 19. Majumder P., Mitra M., Parui S. K., Kole G., Mitra P., & Datta, K. YASS: Yet another suffix stripper. *ACM Transactions on Information Systems*, (2007). 25(4). <https://doi.org/10.1145/1281485.1281489>.
 20. Peng F., Ahmed N., Li X., & Lu Y. Context sensitive stemming for web search. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'(2007)*. 07. <https://doi.org/10.1145/1277741.1277851>.

21. Jabbar A., Iqbal S., Khan M. U. G., & Hussain S. A survey on Urdu and Urdu like language stemmers and stemming techniques. *Artificial Intelligence Review*. (2018). 49(3). <https://doi.org/10.1007/s10462-016-9527-1>.
22. Xusainova Z., Samatboyeva M. Tabiiy tilni qayta ishlash (NLP)da stemming jarayoni tavsifi. "KOMPYUTER LINGVISTIKASI: MUAMMOLAR, YECHIM, ISTIQBOLLAR" Xalqaro ilmiy-amaliy konferensiya, Vol.1. № 01.2023. 69-75-b.
23. Frakes W. B., & Fox C. J. Strength and similarity of affix removal stemming algorithms. *ACM SIGIR Forum*, (2003). 37(1). <https://doi.org/10.1145/945546.945548>.
24. Sirsat S. R., Chavan V., & Mahalle H. S. Strength and Accuracy Analysis of Affix Removal Stemming Algorithms. *International Journal of Computer Science and Information Technologies*, 2013. № 4(2).
25. Harman D. How effective is suffixing? *Journal of the American Society for Information Science*, 1991. № 42(1). [https://doi.org/10.1002/\(SICI\)1097-4571\(199101\)42:1<7::AID-ASIS2>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-4571(199101)42:1<7::AID-ASIS2>3.0.CO;2-P).
26. Sharma A., Kumar R., & Mansotra V. Proposed Stemming Algorithm for Hindi Information Retrieval. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, 2016. 3297(6). <https://doi.org/10.15680/IJRCCE.2016>.
27. Gao J., & Johnson M. A comparison of Bayesian estimators for unsupervised Hidden Markov Model POS taggers. *EMNLP 2008. Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL*, 2008. <https://doi.org/10.3115/1613715.1613761>.
28. Merialdo B. Tagging English text with a probabilistic model. *Computational Linguistics*, (1994. № 20(2).
29. Banko M., & Moore R.C. Part of speech tagging in context. *COLING 2004 - Proceedings of the 20th International Conference on Computational Linguistics*, 2004. <https://doi.org/10.3115/1220355.1220435>.
30. Haghghi A., & Klein D. Prototype-driven learning for sequence models. *HLT-NAACL 2006 - Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings of the Main Conference*. <https://doi.org/2006./10.3115/1220835.1220876>.
31. Ҳожиев А. Ўзбек тилида сўз ясалishi. Тошкент, 2005.
32. Goldwater S., & Griffiths T. L. A fully Bayesian approach to unsupervised part-of-speech tagging. *ACL 2007 - Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 2007.
33. Gael J., Vlachos A., & Ghahramani Z. The infinite HMM for unsupervised PoS tagging. *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*, <https://doi.org/10.3115/1699571.1699601>.
34. Taner Dinçer B., & Karaođlan B. Stemming in agglutinative languages: A probabilistic stemmer for Turkish. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2003). 2869. https://doi.org/10.1007/978-3-540-39737-3_31
35. Elov B., Hamroyeva Sh., Abdullayeva O., Uzoqova M. O'zbek tilida pos tegging masalasi: muammo va takliflar. *Til va madaniyat (Amaliy filologiya)*. Vol.5(4). 2022. B. 45-63.
36. Elov B., Akhmedova K. A. Mathematical Model That Semantically Analyzes Polysemantic Words // *Journal of Pedagogical Inventions and Practices*. Vol.3. 2021. P. 119–122.

37. Elov B.&, Akhmedova X. Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in The Uzbek Language. Proceedings - 7th International Conference on Computer Science and Engineering, UBMK, 2022. <https://doi.org/10.1109/UBMK55850.2022.9919453>.
38. Elov B., Samatboyeva M. Identifying NER (Named Entity Recognition) objects in Uzbek language texts // Science and Innovation International Scientific Journal. Vol. 2. Issue 4. April, 2023. UIF-2022: 8.2 | ISSN: 2181-3337. P.44-57.

Toshkent davlat o'zbek tili va adabiyoti
universiteti
Mirzo Ulug'bek nomidagi O'zbekiston
Milliy universiteti

28.07.2022-y.
qabul qilingan