

MUHAMMAD AL-XORAZMIY  
AVLODLARI  
ILMIY-AMALIY VA AXBOROT-  
TAHLILY JURNAL

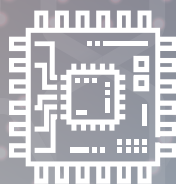
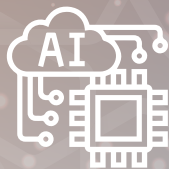
DESCENDANTS OF MUHAMMAD  
AL-KHWARIZMI  
SCIENTIFIC-PRACTICAL AND  
INFORMATION-ANALYTICAL JOURNAL



1(23)/2023

ISSN-2181-9211

MUHAMMAD AL-XORAZMIY NOMIDAGI  
TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI



# MUHAMMAD AL-XORAZMIY AVLODLARI

Ilmiy-amaliy va axborot-tahliliy jurnal 2017 yilda  
ta'sis etilgan

1(23)/2023

## Tahririyat kengashi a'zolari

Maxkamov B.SH. – Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti (TATU) rektori, Tahririyat kengashi raisi

Sultanov Dj.B. – Tahririyat kengashi raisi o'rinbosari

Tashev K.A. – Tahrir kengashi raisi o'rinbosari

Raximov B.N. – t.f.d., prof. bosh muharrir

Nosirov X.X. – PhD, dots. bosh muharrir o'rinbosari

## Muharrirlar:

Kamilov M.M. – t.f.d., prof., akademik.

Musayev M.M. – t.f.d., prof.

Abduraxmonov K.P. – f.-m.f.d., prof.

Jumanov J.X. – t.f.d., prof.

Muxamediyeva D.T. – t.f.d., prof.

Isayev R.I. – t.f.n., prof.

Yusupov A. – f.-m.f.d., prof.

Yakubova M.Z. – t.f.d., prof. (Qozog'iston)

Xalikov A.A. – t.f.d., prof. (TTYTMI)

Nazarov A.M. – t.f.d., prof. (TDTU)

Jmud V.A. – professor (Rossiya)

Miroslav Skoric – professor (Avstriya)

Dzhurakhalov A. – professor (Belgiya)

Abrarov S.M. – professor (Kanada)

Kyamakya K. – professor (Avstriya)

Chedjou J.Ch. – professor (Avstriya)

Davronbekov D.A. – t.f.d., prof.

Anarova Sh.A. – t.f.d., prof.

Pisetskiy Y.V. – t.f.d., prof.

Nishonov A.X. – t.f.d., dots.

Muminov B.B. – t.f.d., prof.

Raximov N.O. – t.f.d., dots.

Amirsaidov U.B. – t.f.d., dots.

Kerimov K.F. – t.f.d., dots.

Ganiyev A.A. – t.f.n., dots.

Gavrilov I.A. – t.f.n., dots.

Gubenko V.A. – t.f.n., dots.

Pulatov Sh.U. – t.f.n., dots.

Muradova A.A. – PhD, dots

Shaxobiddinov A.SH. – PhD

Madaminov X.X. – PhD, dots

Xudaybergenov T.A. – PhD, dots

Ro'ziboyev O.B. – PhD, dots

Yaxshibayev D.S. – PhD, dots.

Mirsagdiyev O.A. – PhD, dots.

Puziy A.N. – PhD

Berdiyev A.A. – PhD, bosh muharrir yordamchisi

Arabboyev M.M. – PhD, texnik muxarrir

Begmatov Sh.A. – PhD, texnik muxarrir

Xudaybergenov J.D. – texnik muxarrir

## MUNDARIJA

### DASTURIY VA KOMPYUTER INJINIRING TEXNOLOGIYALARINING ZAMONAVIY MUAMMOLARI

- Botirov F.B.** Axborot xavfsizligi insidenti: axborotni himoyalashga qaratilgan yondashuvlar 3
- Siddiqov I.Kh., Fayzullayeva B.B.** Simulation of continuous-discrete systems with multiple synchronous interruptions 7
- Babadjanov E.S., To'liyev X.I., Kenjayev X.B.** Sog'in sigirlarning sut mahsuldorlik tahlili va laktatsiya egri chizig'ini qurish usuli 11
- Xudoyqulov Z.T., Boyquziyev I.M., Allanov O., Murodov M.M.** GOST R 34.12.2015 shifrlash standarti algoritmlarini kriptotahlil usullariga baholash natijalari 15
- Nishanov A.X., Beglerbekov P.J., Kudiyarov B.C., Muxsinov Sh.Sh., Mengturaev F.Z.** Номинал белгилри тимсолларни таснифлашда информатив белгиларни танлаш ва кластерлаш алгоритми 19
- Tashev K.A., Yuldashева H.C.** Шахсни идентификацияловчи биометрик тизимлар 23
- Otaxонова B.I., Oлимжонов O.O., Рахимов И.И.** Матнли маълумотларда учрайдиган орфографик хатоликларни бартараф этиш усуллари 28
- Rahimov M.F., To'rayeva M.Sh.** Mashinani o'qitish usulidan foydalangan holda yurak kasalliklarini tasniflash 32
- Xudaybergenov T.A., Hamrayeva S.I., Jumaniyazova M.K.** SK6812 RGB LED asosida taqiladigan yorug'lik identifikatorli bilaguzukni IEEE802.15.7 8-CSK Symbol Point asosida xaritalashni loyihalashtirish 36
- Irgasheva D.Ya., Agzamova M.Sh.** Анализ биометрических технологий аутентификации и современных систем защиты информации 40
- Abdirazakov F.B.** Nutq signallarga raqamli ishlov berish usullari tahlili 49

### OPTIK ALOQA TIZIMLARI, TELEKOMMUNIKATSIYA TARMOQLARI VA KOMMUTATSIYA TIZIMLARINING RIVOJLANISH TAMOIYILLARI

- Iсаев P., Турсимуратов С.С.** Об одном методе оценки показателей надежности волоконно-оптических линии связи 56
- Рахимов Б.Н., Кенгесбаев С.К., Тураев Б.Э.** Создание оптоэлектронного преобразователя фотоколориметра 62
- Maxmudov S.O., Temirova D.X.** "Aqlli shahar" tizimi infrastrukturasi modelini yaratish 66
- Maxmudov I.A., Tillaboyev M., Saparbayev R.K.** Multiservis tarmog'i resurs va xizmatlarini taqdim etish jarayonlarini intellektuallashtirish masalalari model va usuli 72
- Matqurbonov D.M.** Модель системы управления маршрутизацией с использованием нечетких множеств 76
- Тураев Б.Э.** Оптоэлектронные методы определения содержания веществ в жидких и газообразных средах 82

### RAQAMLI TELEVIDENIYE VA RADIOESHITTIRISH, SIMSIZ TEXNOLOGIYALAR VA RADIOTEXNIKANI RIVOJLANTIRISH ISTIQBOLLARI

- Иманкул М.Н., Давронбеков Д.А., Манбетова Ж.Д.** Интегрированные системы связи для служб быстрого реагирования 86

**Muassis:**

*Muhammad al-Xorazmiy nomidagi  
Toshkent axborot texnologiyalari  
universiteti*

**Manzil:**

*100084, O'zbekiston, Toshkent sh., Amir  
Temur ko'chasi, 108*

**Telefon:** 71 238-64-38;

**e-mail:** [alxorazmiy@tuit.uz](mailto:alxorazmiy@tuit.uz)

**Jurnal sayti:** <http://alxorazmiy.uz>

**Bosishga ruxsat etildi:**

*Qog'oz bichimi 60x84 1/8*

*Bosma tabog'i 15,5. Adadi 100 nusxa*

*Buyurtma raqami №195 "Fan va*

*texnologiyalar Markazining*

*bosmaxonasi" da chop etildi*

*Toshkent shahri Oltazor ko'chasi, 171.*

*Jurnal O'zbekiston Matbuot va*

*axborot agentligida 2017 yil*

*22 iyunda 0921 raqami bilan ro'yxatdan*

*o'tgan.*

*Jurnal yilda 4 marotaba*

*(har chorakda) chop etiladi.*

Автоматизация процесса определения параметров реле железнодорожной автоматики и телемеханики	90
<b>Mukhammedova D.B., Gavrilov I.A., Puziy A.N., Alkhamov R.R.</b> The hybrid wavelet video encoder with interframe coding on a contour selection basis and its efficiency estimation	94
<b>Mallayev O.U., Ruzimov O.N.</b> Elektr energiya iste'molini ONLINE monitoring qilish qurilmasi va uning strukturasi	103
<b>Raximov B.N., Berdiyev A.A., Xudayberganov J.D.</b> Telekommunikatsiya tizimlarida akustik aks-sado shovqinlarini bartaraf qilishda zamonaviy usullarining qiyosiy tahlili	109
<b>Djumanov J. X., Anorboyev E. A., Jamolov X.M.</b> Kuzatuv quduqlaridan yer osti suvlari sathi va tok o'tkazuvchanligini aniqlovchi masofaviy ma'lumot uzatish qurilmasini yaratish	113
<b>Raximov B.N., Ibragimov D.B.</b> Формирование шумовых сигналов с использованием деаутификации через отражающую антенную решетку ВАН-АТТА и антенну крeосан направленных против радиоуправляемых взрывных устройств	118
<b>Халиков А.А., Хуррамов А.Ш., Ўроков О.Х.</b> IP- тармок асосидаги тезкор технологик радио алоқа тармоғи ишончилигини ҳисоблаш методикаси	126
<b>Abdullaev A.X.</b> Magnit maydondagi tok o'tkazuvchi halqasimon plastinkaning chiziqli bo'lmagan deformatsiyasi	133
<b>Мадаминов Х.Х., Алиев У.Т., Худайбергенов Ж.Д.</b> Ер усти LPWAN тармоқлари технологиялари ва уларни сунъий йўлдош тизимга мослаштириш имкониятларини таҳлил қилиш	137
<b>Sultonova M.O.</b> Atmosfera optik aloqa tizimlari asosidagi transport tarmoqlarining ishonchligini baholash usuli	144
<b>Rahimov B. N., Berdiyev A.A., Babajanova A.T., Alimuhamedova M.E.</b> Ko'p kanalli sensorlardagi ma'lumotlarni avtomatlashtirilgan tahlilini olib boruvchi apparat-dasturiy majmuasining arxitekturasini ishlab chiqish	147

**ILMIY AXBOROTLAR**

<b>Boltayev S.T., Valiyev S.I., Mirsagdiyev O.A., Kasimova Q.A.</b> Turli toifadagi poyezdlar uchun yo'l keshimlariga poyezdlarni yaqinlashish uslubiyatini ishlab chiqish	152
<b>Кодиров Ф.М.</b> Телекоммуникация электр таъминоти тизимларини таҳлил қилишнинг тизимли ёндашуви	160
<b>Sultanov R.R.</b> Research and design structure and formal model of information system for managing educational system and e-content design based on elements of artificial intelligence	167
<b>Mukhtarov F.M., Muminov K.Z.</b> Goals and objectives of information security	174
<b>Beknazarova S.S., Kucharova Sh.S.</b> 3D modellashtirish va misollar bilan animatsiya	177
<b>Elov B.B., Khamroeva Sh.M., Xusainova Z.Y.</b> NLP (tabiiy tilga ishlov berish) ning Pipeline konveyeri	181
<b>Orinbaev A.B.</b> Elektr energiyasini uzatishda sarflanadigan xarajatlarni optimallashtirishning deterministik modeli	192
<b>Сейтназаров К.К., Базарбаева А.К., Туремуратова Б.К.</b> Критерии оценивания студентов в высших учебных заведениях и их анализ	197
<b>Yusupov Ya.T., Aripova M.X., Ibragimova B.B.</b> "Radioelektron qurilmalar va tizimlar" ta'lim yo'nalishidagi mutaxassislik fanlarining "Oliy matematika" fani bilan bog'liqligi tahlili	201
<b>Igamberdiev K.A.</b> Gidravlik tizimlarini tebranma harakatlarini turg'unlik va stabilashtirish jarayonining matematik modeli	205
<b>Mўминов Б.Б., Даулетов А.Ю.</b> Ташкилотларда электрон ҳужжат айланиш тизимларининг мантиқий - функционал жараёнларини лойиҳалаш ва вақтини ҳисоблаш	208
<b>Meliqo'ziyev R.Sh.</b> Harbiy bo'lnma jangovar imkoniyatlarini baholash mezonini va algoritimi	215



Elov B.B., Khamroeva Sh.M., Xusainova Z.Y.

## NLP (tabiiy tilga ishlov berish) ning Pipeline konveyeri

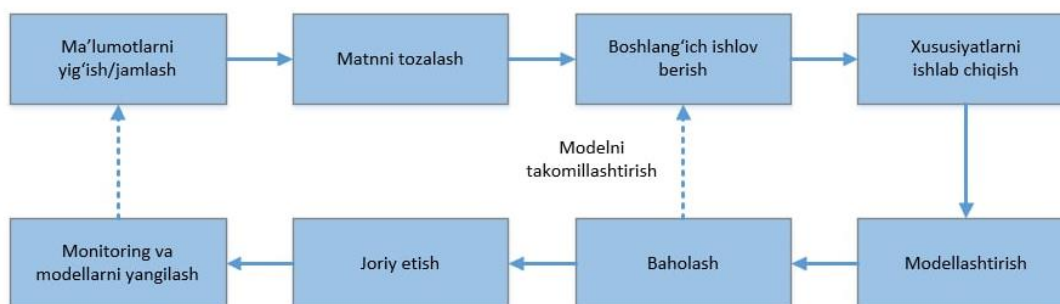
NLP masalasi bir necha kichik qismlarga ajratilib, bosqichma-bosqich hal qilinishi lozim. Ushbu maqolada NLP masalasini hal qilishning har bir bosqichida zarur bo'lgan matnni qayta ishlashning barcha shakllari ko'rib chiqiladi. NLPda matnni bosqichma-bosqich qayta ishlash pipeline jarayoni (konveyeri) deb yuritiladi. Har qanday NLP modelini yaratishda pipeline jarayoni amalga oshirilishi lozim bo'lgan qadamlar ketma-ketligidir. Matnni qayta ishlash konveyerini rejalashtirish va ishlab chiqish har qanday NLP loyihasini yaratishning boshlang'ich nuqtasi sifatida qaraladi. Ushbu maqolada pipeline jarayonini amalga oshirish uchun bajariladigan qadamlar va ularning NLP vazifalarini hal qilishdagi o'rni haqida mulohaza yuritiladi. Ushbu maqolada NLP pipeline konveyeridagi eng keng tarqalgan dastlabki ishlov berish bosqichlari tahlil qilindi. Qayta ishlash bosqichlarining barchasi turli NLP kutubxonalarida oldindan o'qitilgan, foydalanish mumkin bo'lgan modellar sifatida aniqlangan. Zarurayot bo'lganda, berilgan masala shartiga qarab qo'shimcha, moslashtirilgan oldindan ishlov berish qadamlari ishlab chiqilishi mumkin. Muayyan dastlabki ishlov berish bosqichi berilgan NLP muammosiga qanday xizmat qilishini faqat ko'plab tajribalar bilan aniqlash mumkin.

**Kalit so'zlar:** NLP, pipeline jarayoni (konveyeri), tokenizatsiya, lemmatizatsiya, stemming, POS-tagging, morfologik tahlil.

### Kirish

Odatda, NLP masalasi bir necha kichik qismlarga ajratilib, bosqichma-bosqich hal qilinishi lozim. Ushbu maqolada NLP masalasini hal qilishning har bir bosqichida zarur bo'lgan matnni qayta ishlashning barcha shakllari ko'rib chiqiladi. NLPda matnni bosqichma-bosqich qayta ishlash **pipeline jarayoni** (konveyeri) deb yuritiladi [1-4]. Har qanday NLP modelini yaratishda pipeline jarayoni amalga oshirilishi

lozim bo'lgan qadamlar ketma-ketligidir. Matnni qayta ishlash *konveyerini rejalashtirish* va *ishlab chiqish* har qanday NLP loyihasini yaratishning boshlang'ich nuqtasi sifatida qaraladi. Ushbu maqolada pipeline jarayonini amalga oshirish uchun bajariladigan qadamlar va ularning NLP vazifalarini hal qilishdagi o'rni haqida mulohaza yuritiladi. Quyidagi 1-rasmda zamonaviy NLP tizimini ishlab chiqish uchun umumiy pipeline konveyerining asosiy komponentlari ko'rsatilgan [5].



1-rasm. Pipeline konveyerining asosiy bosqichlari

Pipeline konveyerining asosiy bosqichlari quyidagilardan iborat:

1. *Ma'lumotlarni yig'ish.*
2. *Matnni tozalash.*
3. *Boshlang'ich ishlov berish.*
4. *Xususiyatlarni ishlab chiqish.*
5. *Modellashtirish.*
6. *Baholash.*
7. *Joriy etish.*
8. *Monitoring va modelni yangilash.*

### Asosiy qism

Har qanday NLP tizimini ishlab chiqish jarayonida birinchi qadam berilgan vazifaga tegishli ma'lumotlarni yig'ishdir. Qoidalarga asoslangan NLP tizimini ishlab chiqish uchun *qoidalarni loyihalash* va *sinab ko'rish* uchun training (test) ma'lumotlariga ega bo'lish kerak. Biz olgan ma'lumotlar kamdan-kam hollarda "toza" bo'ladi va keyingi bosqichda *matnni tozalash* amali bajariladi. Matn tozalanganidan so'ng, matn

ma'lumotlarini *kanonik shaklga aylantirish* kerak. Ushbu amal *dastlabki ishlov berish* bosqichida amalga oshiriladi [6-10]. Shundan so'ng NLP vazifasini bajarish uchun zarur bo'lgan *xususiyatlarni ishlab chiqish* lozim. Ushbu xususiyatlar *modellashtirish algoritmlari* orqali tushunarli formatga aylantiriladi. Keyingi bosqichda *modellashtirish* va *baholash* amalga oshiriladi. Bu bosqichda bir yoki bir nechta til modellari yaratiladi va ular tegishli baholash ko'rsatkichlari (metrikalari) yordamida taqqoslanadi. Ishlab chiqilgan modellar orasidan eng yaxshisi tanlab olingandan so'ng ushbu modelni joriy etish lozim. Va nihoyat, doimiy tarzda *modelning ishlashini kuzatib boramiz* va, agar zaruriyat bo'lsa, uning ishlashini takomillashtirishimiz, uni yangilashimiz zarur. Dastlabki bosqichlarda ko'p vaqt *xususiyatlarni ishlab chiqish, modellashtirish* va *baholashda* sarflanadi [6,7,11-15].

### Dastlabki bosqichlar



Yuqorida aytilganidek, NLP dasturi, odatda, matnni soʻzlarga (tokenlarga) va gaplarga boʻlish orqali tahlil qiladi. Shunday qilib, har qanday NLP pipeline konveyeri **matnni jumlalarga ajratish** (*gaplar segmentatsiyasi*) va keyinchalik **gaplarni soʻzlarga ajratish** (*soʻzlarni tokenizatsiyalash*)ni toʻgʻri amalga oshiruvchi tizimni ishlab chiqish zarur.

*Gap segmentatsiyasi.* Oddiy qoidaga koʻra, nuqta va soʻroq belgilari paydo boʻlganda matnni gaplarga ajratish orqali gaplarni segmentatsiyalashimiz mumkin. Ammo oddiy qoidani buzishi mumkin boʻlgan *qisqartmalar, manzil shakllari* (*sh.k – shunga koʻra, va b.-va boshqalar*) yoki koʻp nuqta (...) boʻlishi mumkin. Python tilida mavjud NLP kutubxonalaridagi standart metodlar orqali gap va soʻzlarni ajratish imkoniyati taqdim etilgan. Quyidagi Python tilidagi kodda **Natural Language Tool Kit (NLTK)** kutubxonasidan gap va soʻz ajratuvchidan qanday foydalanish koʻrsatilgan:

```
from nltk.tokenize import sent_tokenize, word_tokenize
mytext = "Odatda NLP masalasini bir nechta kichik qismlarga ajratib bosqichma-bosqich hal qilish lozim. Ushbu maqolada NLP masalasini hal qilishning har bir bosqichda zarur boʻlgan matnni qayta ishlashning barcha shakllarini koʻrib chiqiladi."
```

```
my_sentences = sent_tokenize(mytext)
```

*Soʻzlarni tokenizatsiyalash.* Tilning asosiy birligi uning soʻz boyligi – leksikasidir. Ammo lugʻat (leksika) oʻz ichiga nimani oladi? **“Soʻz”** atamasi keng maʼnoli boʻlib, ilmiy uslubda qoʻllash uchun ushbu atamani aniqlashtirib olish lozim. Agar matndan ajratuvchilar (*oralig boʻshliqlar, baʼzi tinish belgilari va boshqalar*) orqali soʻzlarni ajratib olsak, matnda koʻplab **tokenlar** hosil qilinadi [16,17]. **Token** matndan chegarasi bilan ajralib turadigan har qanday birlikdir. *“Daraxt bir yerda koʻkaradi”* gapida 4 ta token mavjud: *daraxt, bir, yerda, koʻkaradi*. *“Halol mehnat yerda qolmas”* jumlasida 4 ta token mavjud: *halol, mehnat, yerda, qolmas*.

Gapni tokenizatsiya qilish singari gaplarni soʻzlarga aylantirish uchun tinish belgilarining mavjudligiga qarab matnni soʻzlarga boʻlishning oddiy qoidasidan boshlashimiz mumkin. NLTK kutubxonasi buni amalga oshirishga imkon beradi:

```
for sentence in my_sentences:
    print(sentence)
    print(word_tokenize(sentence))
```

Mavjud boʻlgan yechimlar (tokenizatorlar)ning koʻpchiligi berilgan matnni 100% toʻgʻri tokenlarga ajratmasligini va ushbu algoritmlar mukammal emasligini yodda tutish kerak. Masalan, quyidagi gapni koʻrib chiqamiz:

*“Sumbula sapchib oʻrnidan turdi-da: «Xayriyat, tushim ekan», – dedi hansirab.”*

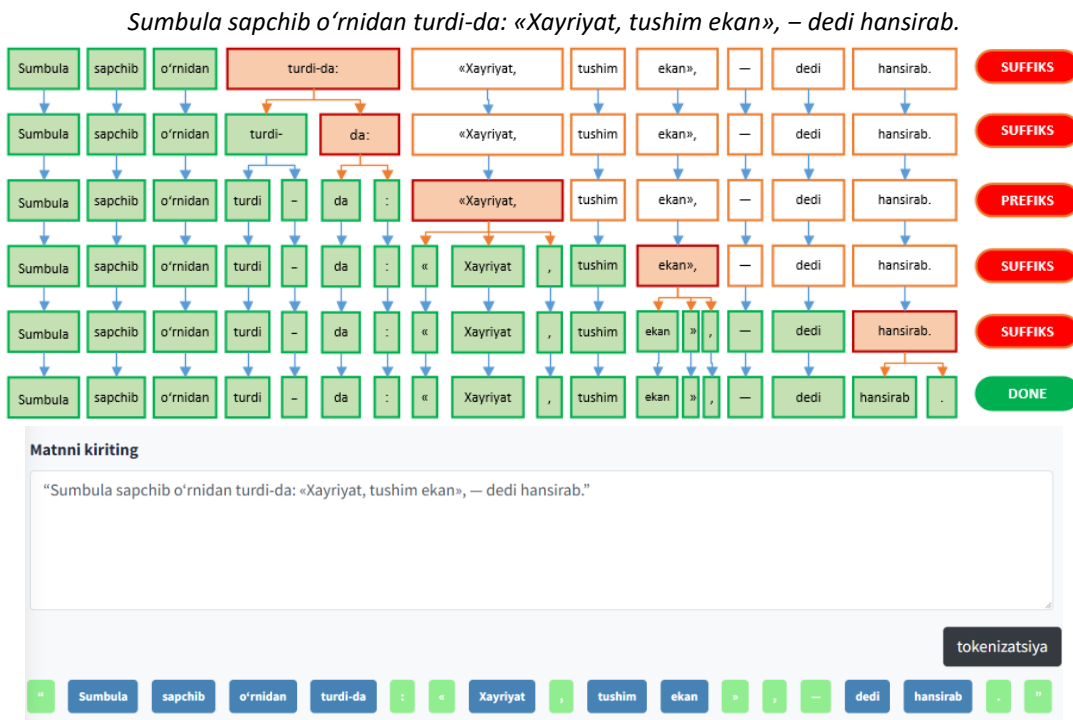
Agar biz yuqoridagi gapni NLTK tokenizer orqali ishlatsak, *[oʻrnidan] [.] [turdi] [-] [da] [:]* alohida tokenlar sifatida aniqlanadi. Shuningdek, agar tvitlarni tokenlamoqchi boʻlsak, tokenizator *heshtegni* ikkita tokenga ajratadi: *“#”* belgisi va *undan keyingi qator*. Bunday hollarda, bizning maqsadimiz uchun (oʻzimiz tomonimizdan) yaratiladigan maxsus tokenizatoridan foydalanishimiz kerak boʻlishi mumkin. Shuning uchun tokenizator, baʼzan **grafematik tahlilning** baʼzi elementlari bajarilishi lozim:

- *harflar ketma-ketligi;*
- *sonlar;*
- *tinish belgilari;*
- *ierogliflar;*
- *ajratuvchilar;*
- *turli grafik belgilar.*

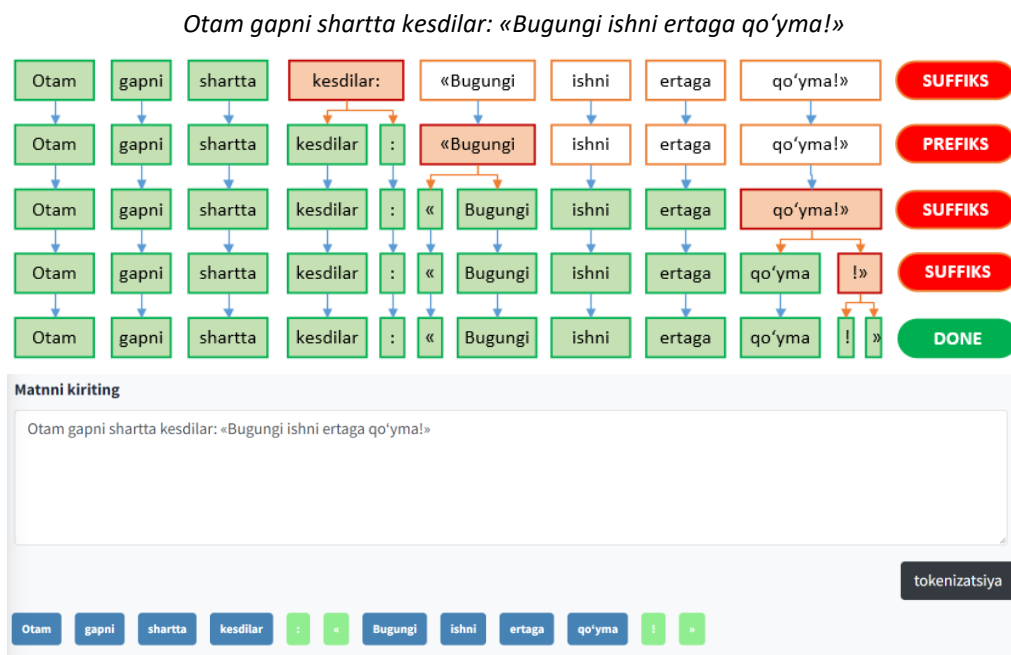
Quyida oʻzbek tilidagi 2 ta gapga mos tokenizatsiya jarayoni va <http://uznatcorpara.uz/uz/Tokenizer> dasturidagi varianti keltirilgan:

**Muntazam bajarildigan qadamlar:** NLP pipeline konveyerida muntazam bajariladigan boshlangʻich qayta ishlash amallarini koʻrib chiqaylik. Aytaylik, biz yangiliklar saytidagi maqolaning toifasini *siyosat, sport, biznes* va boshqa mavzulardan biri sifatida belgilaydigan dasturiy taʼminotni ishlab chiqmoqdamiz. Bizda gapni segment va tokenlarga ajratuvchi dastur mavjud boʻlsin. Bu holda yangiliklarni guruhlariga ajratish vositasini ishlab chiqish uchun qanday maʼlumotlar foydali ekanligi haqida oʻylashni boshlashimiz kerak.

Oʻzbek tilidagi tez-tez ishlatiladigan soʻzlarning baʼzilari, masalan, *albatta, ammo, asosan, aynan, balki, barcha, bilan, biroq* kabi soʻzlar ushbu vazifa uchun zarur hisoblanmaydi. Chunki ular toʻrtta toifani ajratish uchun oʻz-oʻzidan hech qanday mazmunni anglatmaydi. Bunday soʻzlar **nomuhim soʻzlar** deb ataladi va, odatda, keyingi tahlildan olib tashlanadi. Biroq oʻzbek tili uchun nomuhim soʻzlarning standart roʻyxati mavjud emas. Mualliflar tomonidan ishlab chiqilgan oʻzbek tilidagi nomuhim soʻzlarning baʼzilari quyidagi jadvalda keltirilgan:



2-rasm. Tokenizatsiya jarayoniga namuna



3-rasm. Tokenizatsiya jarayoniga namuna

1-jadval

O'zbek tilidagi umumiy nomuhim so'zlar

afsuski	beri	eng	lla	mening	orqada	shak-shubhasiz	ustida
aftidan	bilan	esa	iloyim	Misli	orqaga	shekilli	ustidan
agar	binoan	essiz	ishonamanki	misoli	orqali	shu	ustiga
aksincha	biroq	evaziga	ishqilib	mobaynida	o'sha	shubhasiz	va
albatta	biroz	faqat	jihatdan	mobodo	ostida	shunchaki	vaqtda
allaqachon	biz	gar	joiz	modomiki	oxir	shunday	xayriyat
alqissa	bizning	garchand	juda	mos	o'zi	shunga	xo'sh
ammo	bo'yi	garchi	jumladan	na, na	o'zim	shuning uchun	xolos

<b>ammo-lekin</b>	bog'liq	go'yo	Kabi	nachora	o'zimiz	shuningdek	xuddi
<b>aqalli</b>	bois	go'yoki	kerak	nafaqat	o'zingiz	singari	xullas
<b>arafasida</b>	boshqa	goh	keyin	nafar	o'zlari	siz	xususan
<b>aro</b>	bu	chunonchi	keyingi	natijada	pastda	sizniki	ya'ni
<b>aslida</b>	bular	ham	kim	negaki	pastga	sizning	yana
<b>aslo</b>	bundan	hamda	kimdir	nimagaki	payida	tabiiyki	yanada
<b>asosan</b>	bunday	hamma	kimga	o'rniga	qachonki	tag'in	yaxshi
<b>asosiy</b>	butun	hammasi	ko'p	o'rtasida	qadar	tahminan	yaxshiyam
<b>avvalambor</b>	chamasi	haqiqatda	ko'plab	o'sha	qarab	tashqari	yo
<b>avvalgi</b>	chog'i	haqiqatdan	ko'proq	o'z	qarata	to'g'risi	yo'qsa
<b>avvalo</b>	chunki	har	ko'ra	o'zi	qaratilgan	toki	yo'q-yo'q
<b>axir</b>	chunonchi	har holda	ko'proq	o'zim	qayta	tomon	yo'sinda
<b>aynan</b>	darhaqiqat	har qalay	lekin	o'zimiz	qaytanga	tufayli	yoki
<b>ayni</b>	darhol	hatto	lozim	o'zingiz	qisqasi	turli	yonida
<b>ayniqsa</b>	darkor	hech	ma'lum	o'zini	qo'yingki	u	yonidan
<b>aytaylik</b>	dastavval	hokazo	mabodo	o'zlari	quyida	uchun	yoniga
<b>aytganicha</b>	davomida	holda	mana	o'z-o'zidan	quyidagi	ular	yo'q
<b>aytganday</b>	demak	hozir	masalan	ochig'i	ravishda	ularni	yoxud
<b>ba'zi</b>	deyarli	hozirda	mayli	oid	rostdan	ularning	yuqorida
<b>balki</b>	deylik	ichida	mazkur	olaylik	rosti	unda	yuqoriga
<b>baravarida</b>	doim	ichidan	mazmuni	oldiga	sana	unga	yuzasidan
<b>barcha</b>	doir	ichiga	men	oldin	sari	uni	zero
<b>barchasi</b>	doirasida	ichkarida	meni	orasida	sayin	uning	zeroki
<b>baribir</b>	ehtimol	ichra	menimcha	orasiga	seningcha	ushbu	zotan

Ba'zi NLP paketlarida nomuhim so'zlar ro'yxatlari (ba'zi chet tillar uchun) mavjud bo'lib, ko'p hollarda, berilgan masalaga ko'ra ular o'zaro farq qilishi mumkin. *Tinish belgilarini va/yoki raqamlarni olib tashlash* matnni tasniflash, ma'lumot olish va ijtimoiy media tahlili kabi ko'plab NLP muammolari uchun umumiy qadam hisoblanadi. Quyidagi dastur kodida ingliz tilidagi berilgan matnlar to'plamidan *nomuhim so'zlar, raqamlar, tinish belgilari va kichik harflarga o'tkazish* qanday olib tashlanishi ko'rsatilgan:

```
from uzcorpus import stopwords
from string import punctuation
def preprocess_corpus(texts):
mystopwords=set(stopwords.words("uzbek"))
def remove_stops_digits(tokens):
return [token.lower()for token in tokens
if token not in mystopwords not
token.isdigit() and token not in
punctuation]
return
remove_stops_digits(word_tokenize(text))
for text in texts]
```

Shuni ta'kidlash kerakki, ushbu to'rtta jarayon barcha NLP masalalari uchun *majburiy* yoki *ketma-ket* bajarilishi talab etilmaydi. Yuqoridagi funktsiya NLP loyihamizga matnni qayta ishlash bosqichlarini qanday amalga oshirishni ko'rsatadi. Ushbu funktsiya orqali matn

ma'lumotlariga dastlabki ishlov berilgan. So'z darajasidagi xususiyatlarni hisobga oladigan muntazam bajariladigan dastlabki ishlov berish bosqichlari – stemming va lemmatizatsiya hisoblanadi.

#### Natijalar va muhokama

*Stemming va lemmatizatsiya.* Stemming jarayonida so'zdagi *qo'shimchalar olib tashlanadi* va so'zni qandaydir **asosiy shaklga qisqartirish** (so'zshaklni qo'shimchasiz holga keltirish) amalga oshiriladi.

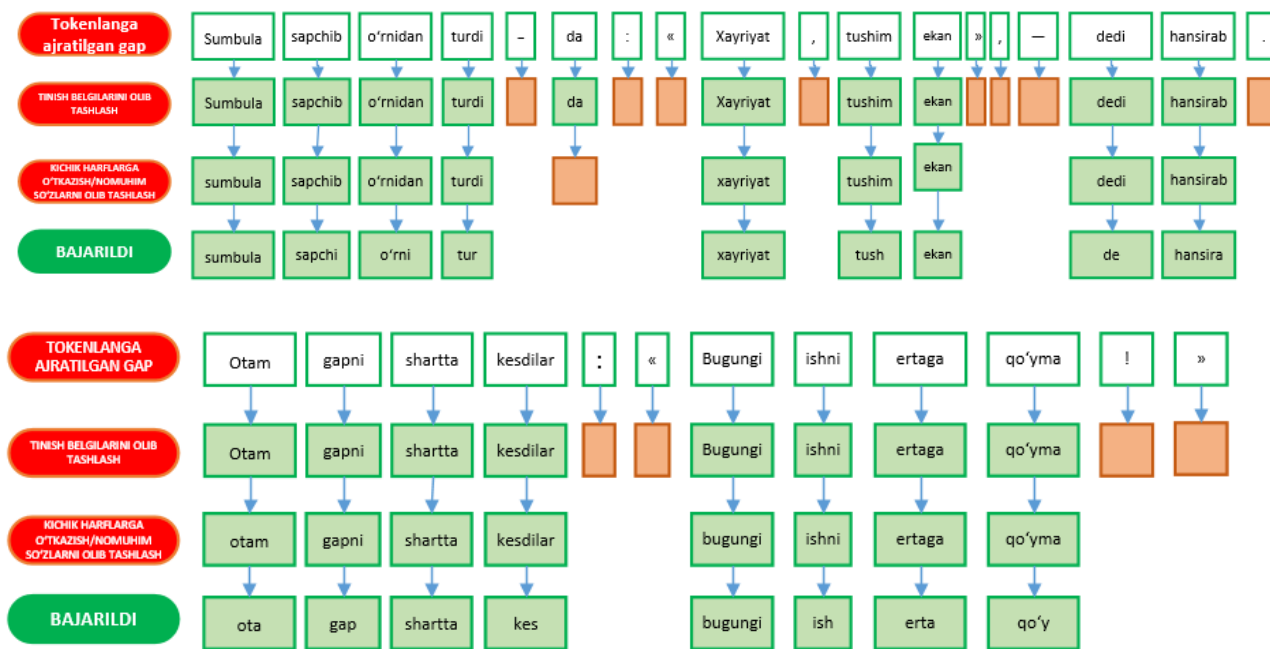
Har bir token uchun uning (*lemma deb ataluvchi*) boshlang'ich (*yoki normal*) shakli mavjud. Nutq(matn)da ushbu boshlang'ich shakl turli grammatik shakllarni olgan holda (fleksiya sodir bo'lishi ham mumkin) ishlatiladi. Quyidagi rasmda ikkita gapning stemming bosqichlari ko'rsatilgan.

Bu o'rinda **so'zshakl** atamasiga ham aniqlik kiritishimiz kerak. So'zshakl lemma (leksema)ning nutqda qo'llanishi shaklidir. **“Daraxt”** lemmasi ot so'z turkumiga oid bo'lgan turli grammatik shakllarni olib, ko'plab so'zshakllarni hosil qilishi mumkin:

- *daraxt, daraxtni, daraxtning, daraxtga, daraxtdan;*
- *daraxtlar, daraxtlarni, daraxtlarning, daraxtlarga, daraxtlardan;*
- *daraxtim, daraxting, daraxti, daraxtimiz, daraxtingiz, daraxtlari.*



STEMMING



4-rasm. Stemming jarayoniga namunalar

Shu tartibda boshqa lemmalar ham nutqda qo'llanish vaqtida so'zshakl ko'rinishga keladi. O'zbek tilida so'z tarkibining odatiy strukturasi quyidagi ko'rinishda bo'ladi:

**“O‘zak + so‘z yasovchi + lug‘aviy shakl + sintaktik shakl”**

Grammatik vositaning joylashishidagi tartib, izchillik uning ma'no va grammatik xususiyati bilan bog'liq bo'lib, quyidagi ketma-ketlikda shakllantiriladi:

- 1) yangi lug‘aviy ma’no hosil qiluvchi;
- 2) lug‘aviy ma’noga ta’sir qiluvchi;
- 3) lug‘aviy ma’noga ta’sir qilmaydigan, lekin so’zni bog‘laydigan vosita qo‘shiladi.

Qo‘shimchanning joylashuvidagi me’yoriy holat ba’zan buziladi:

*opa-lar-im / opa-m-lar, ayt-di-ng-lar / ayt-di-lar-ing.*

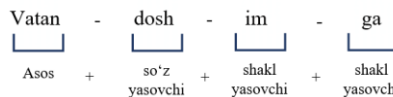
O‘zbek tilida **qo‘shimcha morfemalar** o‘zakdan keyin ketma-ket qo‘shiladi. Morfotaktika qoidalari so‘zshaklda morfema va allomorflarning qanday ketma-ketlikda joylashuvini aniqlaydi. Shularga tayanib, o‘zbek tilidagi ayrim morfemalarning bog‘lanish imkoniyatini quyidagi jadvalda kuzatamiz:

Yuqoridagi jadvaldan *-lar, -im, -ing, -imiz, -si, -ni* kabi qo‘shimchalar o‘zakka to‘g‘ridan-to‘g‘ri qo‘shilishini ko‘rish mumkin. Boshqa affiksial morfemaning esa chap tomoniga qo‘shiladi. *-man, -san, -miz, -siz* qo‘shimchalari esa o‘zakka to‘g‘ridan-to‘g‘ri qo‘shila olmaydi. Chunki ulardan oldin o‘zakka avval boshqa morfema (masalan, zamon qo‘shimchasi) qo‘shilib, undan keyingi pozitsiyani bu qo‘shimchalar guruhi egallaydi, ular affiksial morfemalardan o‘ng tomonda joylashish xususiyatiga ega.

3-jadval. Chap tomonda joylashuvchi morfemalar

morfemalar	chap tomonda joylashish xususiyati (+/-)	misol
<b>-lar</b>	+	bola+lar+dan
<b>-man</b>	-	bor+a+man
<b>-san</b>	-	o‘qi+y+san
<b>-miz</b>	-	yoz+a+miz
<b>-siz</b>	-	ayt+ib+siz
<b>-ing</b>	+	kel+ing
<b>-im</b>	+	kitob+im
<b>-imiz</b>	+	daftar+imiz
<b>-si</b>	+	uka+si
<b>-ni</b>	-	ashula+ni

Demak, o‘zbek tilida o‘zak va qo‘shimchalarning joylashuvchi minimal tarzda quyidagi ko‘rinishda bo‘ladi:

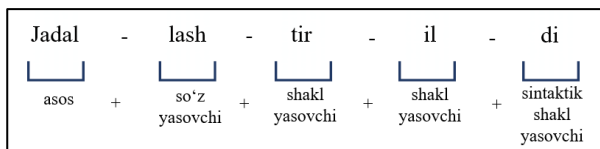


Yangi so‘z yoki so‘zshakllarini tilning eng kichik birligi

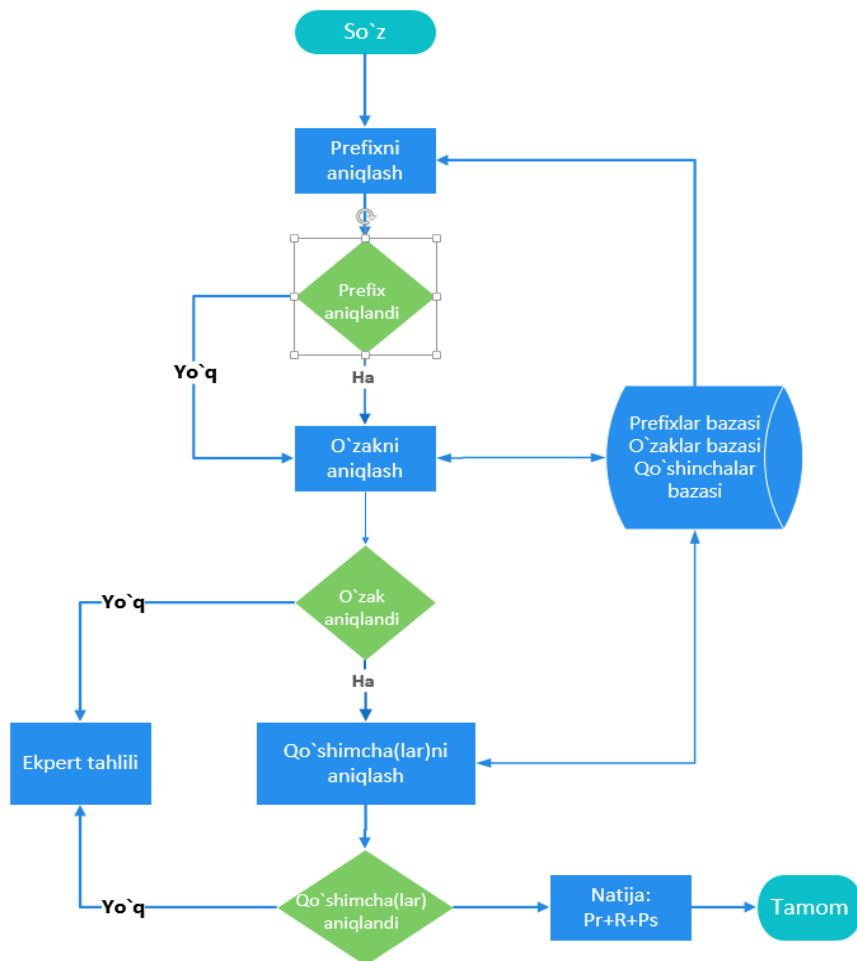
– **morflarni** birlashtirish orqali hosil qilinadi. Morflar *o‘zak* (so‘z asosi) va *qo‘shimchalarga* bo‘linadi:

- *oldqo‘shimcha* (so‘z yasovchi);
- *so‘zdan keyin qo‘shiluvchi so‘z yasovchi*;
- *sintaktik shakl yasovchi*;
- *lug‘aviy shakl yasovchi qo‘shimchalar*.

So‘zning asosiy lug‘aviy ma’nosini o‘zak anglatadi. Qo‘shimchalar esa so‘zga qo‘shimcha ma’no beradi. So‘zni morfemalarga ajratish **morfemik tahlil** deyiladi (5-rasm). Masalan:



Quyidagi 5-rasmda soʻz tarkibini aniqlash algoritmini keltiramiz:



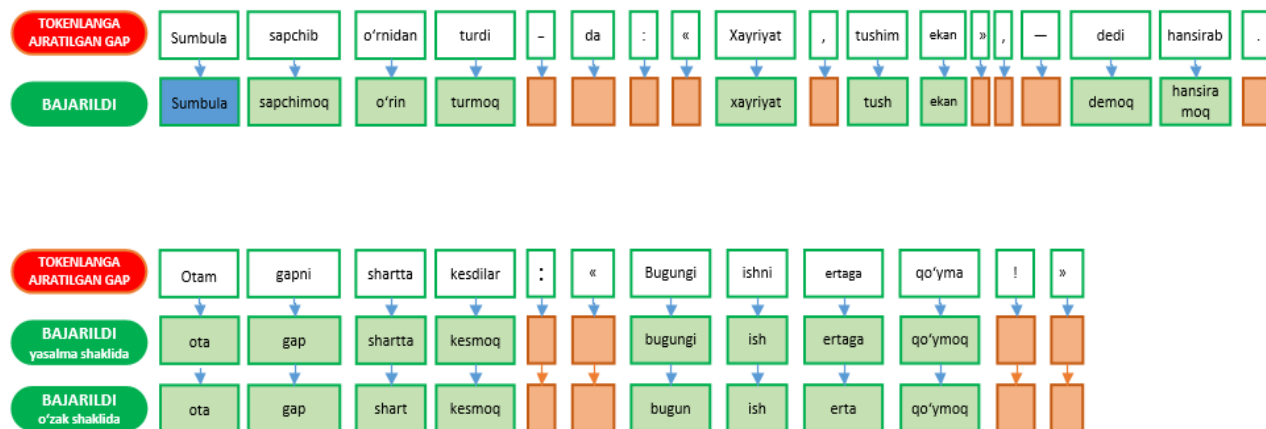
5-rasm. Oʻzbek tilidagi soʻz tarkibini aniqlash algoritmi

*Lemmatizatsiya* – bu soʻzning barcha turli soʻzshakllarini **asosiy soʻz** yoki **lemma** bilan taqqoslash jarayoni. Bu stemming jarayoniga oʻxshash koʻrinadi, ular, aslida, boshqacha. Quyida lemmatizatsiya jarayoni koʻrsatiladi: Eʼtibor berilishi lozim jihat shuki, stemming va lemmatizatsiyani farqlash lozim. Chunki koʻpincha oʻzbek tilida stemming va lemmatizatsiya natijasi bir xildek koʻrinadi, ammo ular batamom farqli jarayonlar sanaladi: stemming oʻzakdan qoʻshimchani kesish, lemmatizatsiya esa oʻzakning lugʻatdagi variantini

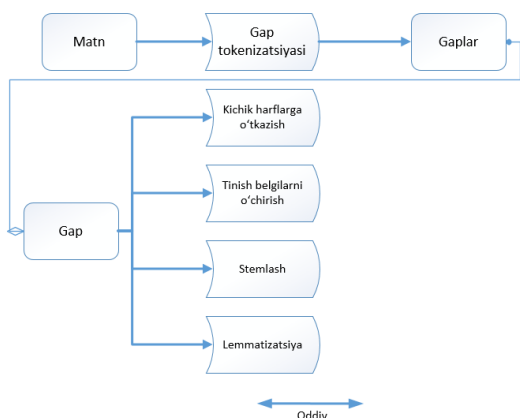
aniqlash jarayonidir. Quyida stemming va lemmatizatsiya jarayoni tahliliga misollar keltiramiz (7-rasm).

Lemmatizatsiya koʻproq lingvistik bilimlarni talab qiladi, zero *oʻzbek tili matnlari uchun samarali natija qaytaruvchi lemmatizatorlarni modellashtirish va ishlab chiqish* hozir ham NLP tadqiqotlarida ochiq muammo boʻlib qolmoqda.

Berilgan matndan *nomuhim soʻzlar, raqamlar, tinish belgilari va kichik harflarni* olib tashlash.



6-rasm. Lemmatizatsiya jarayoniga namunalar



7-rasm. Stemming va lemmatizatsiya o'rtasidagi farq qadamlarining hammasi ham doim ham zaruriy qadamlar hisoblanmaydi. Misol uchun, agar biz matndagi raqamlar va tinish belgilarni olib tashlasak, birinchi navbatda, olib tashlangan narsalar unchalik muhim bo'lmisligi mumkin. Biroq biz, odatda, matnda stemming jarayonini bajarishdan oldin bosh harflarni kichik harf bilan almashtiramiz.

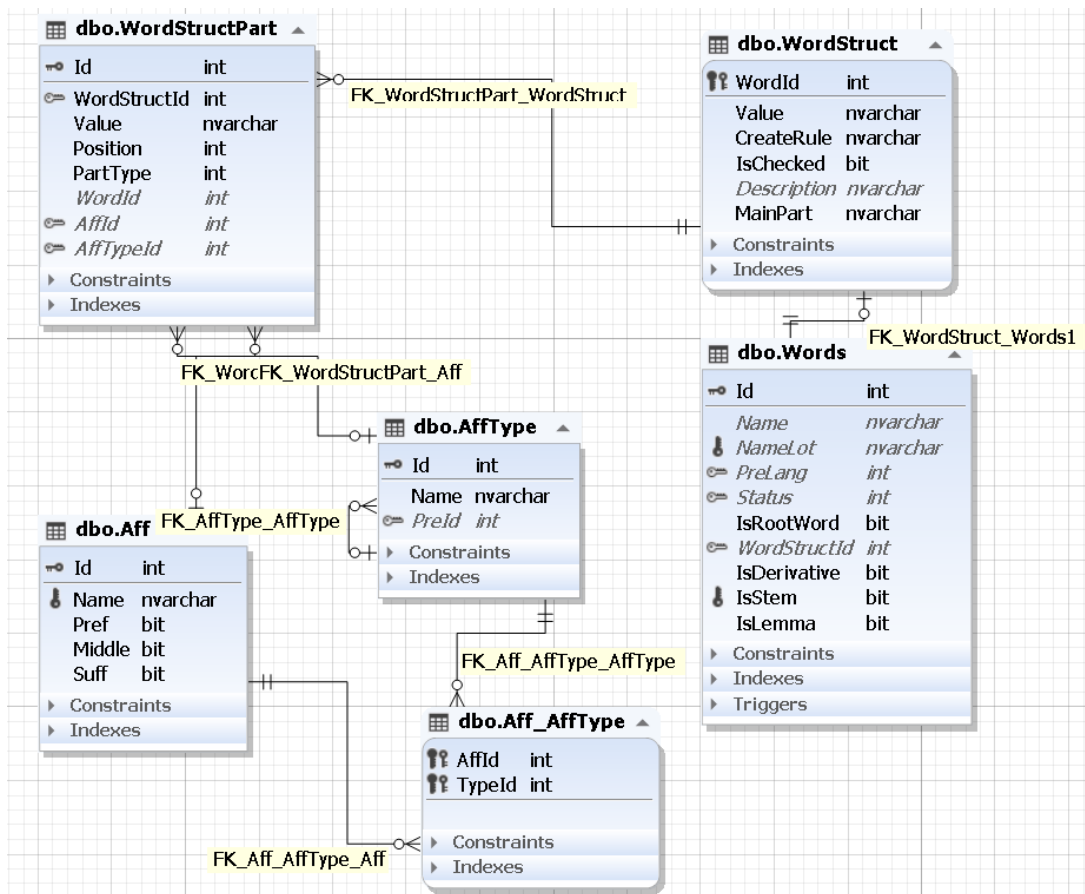
Lemmatizatsiya so'z va uning kontekstining ma'lum miqdordagi lingvistik tahlilni o'z ichiga olganligi sababli unga stemming jarayonidan ko'ra ko'proq vaqt talab etiladi va u, odatda, faqat zarurat tug'ilganda qo'llanadi. NLP pipeline konveyerining dastlabki ishlov berish bosqichlari uchun qaysi lemmatizator yoki stemmerdan foydalanish berilgan masalaning shartiga mos holda tanlanadi.

Lemmatizatsiya jarayonidan oldin biz matndan leksema va kichik harflarni olib tashlamaymiz. Chunki lemmani olish uchun so'zning turkumini (part of speech) bilishimiz kerak va bu gapdagi barcha tokenlarning buzilmasligini talab qiladi. Ma'lumotlarimizni qanday qayta ishlashni aniq tushunganingizdan so'ng bajarilishi kerak bo'lgan dastlabki ishlov berish amallarining ketma-ket ro'yxatini tayyorlash maqsadga muvofiq. Mualliflar tomonidan ishlab chiqilgan o'zbek tili morfoanalizatoridan (<http://uznatcorpara.uz/>) NLP pipeline konveyeridagi tokenizatsiya, stemming va lemmatizatsiya boshqichlariga namunalar va ma'lumotlar bazasidagi ER model quyida keltirilgan:

STEMMING	LEMMATIZATSIYA
qishlog'im → qishlog'	qishlog'im → qishloq
etagi → etag	etagi → etak
shunaqa → shun	shunaqa → shu
bunday → bun	bunday → bu
ikkovimiz → ikk	ikkovimiz → ikki
ayblo → ayblo	ayblo → aybla
taroq → taro	taroq → tara
shahrimiz → shahr	shahrimiz → shahar
singlim → singl	singlim → singil
o'rning → o'rn	o'rning → o'rin
nonni → non	nonni → non
shuhratim → shuhrat	shuhratim → shuhrat
maktabdan → maktab	maktabdan → maktab
olaning → olam	olaning → olam

8-rasm. Matn fragmenti uchun umumiy dastlabki ishlov berish bosqichlari





9-rasm. NLP pipeline konveyerining ER modeli

Matnni kiriting

"Sumbula sapchib o'rnidan turdi-da: «Xayriyat, tushim ekan», — dedi hansirab."

Analiz

№	So'z	Lemma	So'z turkumi	O'zak va qo'shimchalar
1	Sumbula	sumbula	Ot	{sumbula}
2	sapchib	sapchimoq	Fe'l	{sapchi}-b
3	o'rnidan	o'rin	Ot	{o'rin}-i-dan
4	turdi-da	turdi-da		{turdi-da}
5	Xayriyat	xayriyat	Modal	{xayriyat}
6	tushim	tushim	Ot	{tushim}
7	ekan	ekan	Fe'l	{ekan}
8	dedi	demoq	Fe'l	{de}-di
9	hansirab	hansiramoq	Fe'l	{hansira}-b

10-rasm. O‘zbek tili morfoanalizatori (<http://uznatcorpara.uz/>)

**Dastlabki ishlov berishning qo‘shimcha bosqichlari:** Yuqorida NLP pipeline konveyerida bir nechta umumiy dastlabki ishlov berish bosqichlari ko‘rib chiqildi. Bunda matnlarning mohiyati aniq ko‘rsatilmagan bo‘lsa-da, oddiy inglizcha matn bilan ishlash taxmin qilindi. Quyida yana bir nechta dastlabki ishlov berishning qo‘shimcha bosqichlarini keltiramiz.

**Matnni normallashtirish.** Ijtimoiy media postlaridagi yangiliklarni aniqlash masalasini ko‘rib chiqaylik. Ijtimoiy media matni gazetalarda foydalaniladgan tildan juda farq qiladi. So‘zlar turli yo‘llar bilan yozilishi mumkin, masalan, *qisqartirilgan shakllarda, telefon raqami turli formatlarda yozilishi mumkin, ismlar ba‘zan kichik harflar bilan yoziladi* va hokazo. Bunday ma‘lumotlar bilan ishlash uchun NLP vositalarini ishlab

chiqishda matnning barcha o'zgarishlarni qamrab oladigan **kanonik shaklini** hosil qilishimiz lozim. Ushbu jarayon **matnni normallashtirish** deb nomlanadi. Matnni normallashtirishning ba'zi umumiy qadamlari matnni barcha *kichik yoki katta harflarga aylantirish, raqamlarni matnga aylantirish* (masalan, 9 - to'qqizga), *qisqartmalarni kengaytirish* va hokazo. Matnni normallashtirishning oddiy usuli Spacy paketida taqdim etilgan.

*Tilni aniqlash.* Ko'pgina veb-kontent ingliz tilidan tashqari tillarda yoziladi. Misol uchun, bizdan mahsulotimiz haqidagi barcha sharhlarni internetda to'plashimiz so'ralgan deylik. Turli elektron tijorat veb-saytlarini tahlil qilish natijasida mahsulotimizga tegishli sahifalarni skanerlashni boshlaganimizda, biz bir nechta inglizcha bo'lmagan sharhlar paydo bo'lishiga guvoh bo'lamiz. NLP pipeline konveyerining asosiy qismi tilga xos vositalar bilan qurilganligi sababli inglizcha matnni kutayotgan NLP quvurimizda qayday o'zgarishlarni amalga oshirish lozim? Bunday hollarda **tilni aniqlash** NLP pipeline konveyerining birinchi qadami sifatida

amalga oshiriladi. Tilni aniqlash uchun Polyglot kabi NLP paketlaridan foydalanishimiz mumkin. Ushbu amal bajarilgandan so'ng NLP pipeline konveyerining keyingi qadamlari tilga xosligini kuzatish mumkin.

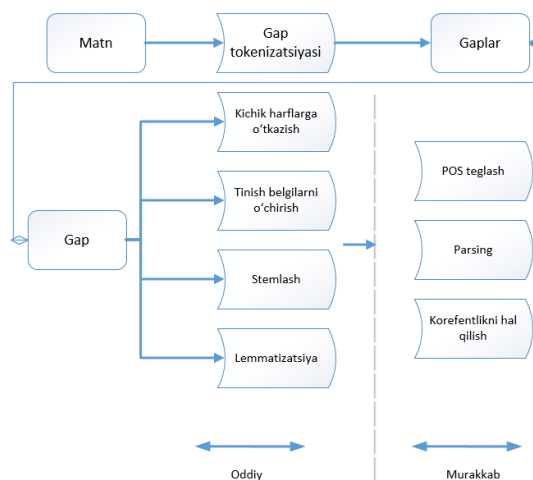
Dunyo bo'ylab ko'p odamlar kundalik hayotlarida bir nechta tillarda gaplashadilar. Shunday qilib, ijtimoiy tarmoqlardagi postlarda bir nechta tillardan foydalanadilar. Kodlarni aralashirishga misol sifatida 5-rasmdagi LDCdagi **Singlish** (Singapur jargoni + Ingliz tili) iborasini ko'rishimiz mumkin. Bitta iborada *tamil, ingliz, malay* va uchta *xitoy tilidagi* so'zlar mavjud. Kodlarni aralashirish tillar o'rtasida almashinishning ushbu hodisasiga ishora qiladi. Odamlar o'z yozuvlarida bir nechta tillardan foydalanganda, ular, ko'pincha, bu tillarda so'zlarni **lotin yozuvida, inglizcha imlo bilan** yozadilar. Demak, ingliz tilidagi matn bilan birga boshqa tildagi so'zlar ham yoziladi. Bu jarayon **transliteratsiya** sifatida tanilgan. Ushbu ikkala hodisa ham ko'p tilli jamoalarda keng tarqalgan va matnni boshlang'ich qayta ishlash jarayonida ularni inobatga olish kerak [5].



11-rasm. Singlish tilidagi post

Yuqorida umumiy boshlang'ich ishlov berish bosqichlari muhokama qilindi. Ushbu ro'yxat to'liq bo'lmasa-da, ma'lumotlar to'plamining tabiatiga qarab talab qilinishi mumkin bo'lgan turli xil boshlang'ich ishlov berish usullari haqida tasavvur beradi deb umid qilamiz.

**Kengaytirilgan ishlov berish:** Kompaniyadagi bir million hujjat to'plamidagi *shaxs va tashkilot nomlarini* aniqlash tizimini ishlab chiqish masalasini ko'rib chiqamiz. Biz ilgari muhokama qilgan matnni qayta ishlashning umumiy bosqichlari bu kontekstga mos kelmasligi mumkin. Ismlarni identifikatsiyalash bizdan POS teglarini amalga oshirishni talab qiladi. Chunki tegishli atoqli otlarni aniqlash shaxs va tashkilot nomlarini aniqlashda foydali bo'lishi mumkin. Loyihaning dastlabki ishlov berish bosqichida POS teglarini qanday qilamiz? Oldindan o'qitilgan va oson foydalanish mumkin bo'lgan POS teggerlari NLTK, spaCy va Parsey McParseface Tagger kabi NLP kutubxonalarida qo'llaniladi. Biz, odatda, o'zimizning POS-teglash yechimlarimizni ishlab chiqishimiz shart emas. Shuni ta'kidlash kerakki, bir xil dastlabki ishlov berish bosqichi uchun turli NLP kutubxonalaridagi natijalar o'rtasida farqlar bo'lishi mumkin.

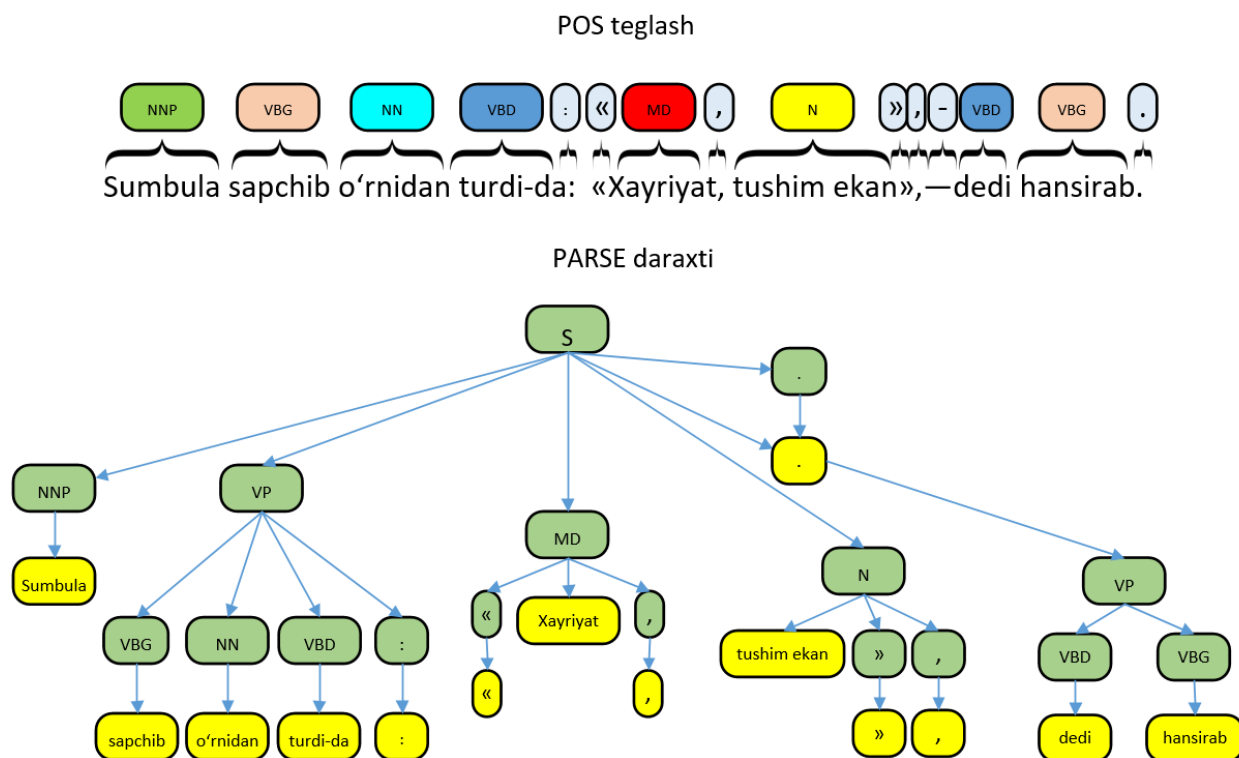


12-rasm. Matn fragmenti uchun kengaytirilgan ishlov berish bosqichlari

Bu amalga oshirishdagi farqlar va turli kutubxonalar o'rtasidagi algoritmlar bilan bog'liq. NLP loyihasida qaysi kutubxonadan (yoki kutubxonalardan) foydalanish – berilgan masala shartiga bog'liq. Keling, biroz boshqacha muammoni ko'rib chiqaylik: kompaniyamizning millionlab hujjatlar to'plamidagi shaxs va tashkilot nomlarini aniqlash bilan bir qatorda, bizdan ma'lum bir shaxs va tashkilot qandaydir tarzda bir-biriga bog'liqligini aniqlashimiz so'raladi. Buning uchun bizga gapdagi ikkita shaxs o'rtasidagi "munosabat"ni ko'rsatadigan shablonlarni aniqlash

usulini ishlab chiqish kerak. Bu bizdan gapning sintaktik ifodalanishining qandaydir shakliga ega bo'lishni talab qiladi. Bundan tashqari, biz obyektning bir necha

eslatmalarini aniqlash va bog'lash usulini aniqlashimiz lozim.



13-rasm. NLP pipeline konveyerida POS teglash va gap strukturasi namuna

### Xulosa.

Ushbu maqolada NLP pipeline konveyeridagi eng keng tarqalgan dastlabki ishlov berish bosqichlari tahlil qilindi. Qayta ishlash bosqichlarining barchasi turli NLP kutubxonalarida oldindan o'qitilgan, foydalanish mumkin bo'lgan modellar sifatida aniqlangan. Zarurayyat bo'lganda, berilgan masala shartiga qarab qo'shimcha, moslashtirilgan oldindan ishlov berish qadamlari ishlab chiqilishi mumkin. Muayyan dastlabki ishlov berish bosqichi berilgan NLP muammosiga qanday xizmat qilishini faqat ko'plab tajribalar bilan aniqlash mumkin. Alisher Navoiy nomidagi Toshkent Davlat o'zbek tili va adabiyoti universitet olimlari tomonidan ishlab chiqilgan o'zbek tili morfoanalizatorida (<http://uznatcorpara.uz/>) NLP pipeline konveyerining tokenizatsiya, stemming va lemmatizatsiya boshqichlari ishlab chiqilgan hamda amaliyotga tadbiiq etilgan.

#### Foydalanilgan adabiyotlar

1. Saloot, M. A., & Pham, D. N. (2021). Real-time Text Stream Processing: A Dynamic and Distributed NLP Pipeline. ACM International Conference Proceeding Series. <https://doi.org/10.1145/3459104.3459198>
2. Becquin, G. (2020). End-to-end NLP Pipelines in Rust. <https://doi.org/10.18653/v1/2020.nlpss-1.4>
3. Peng, N., Ferraro, F., Yu, M., Andrews, N., DeYoung, J., Thomas, M., Gormley, M. R., Wolfe, T.,

Harman, C., van Durme, B., & Dredze, M. (2015). A concrete Chinese NLP pipeline. NAACL-HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Demonstrations, Proceedings. <https://doi.org/10.3115/v1/n15-3018>

4. Noji, H., & Miyao, Y. (2016). Jigg: A framework for an easy natural language processing pipeline. 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - System Demonstrations. <https://doi.org/10.18653/v1/p16-4018>

5. S.Vajjala, B.Majumder, A.Gupta, H.Surana. Practical Natural Language Processing. A Comprehensive Guide to Building Real-World NLP Systems. 2020. 455 p.

6. Botir E.B., Axmedova X.I. Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in Uzbek Language // International conference on information science and communications technologies application, trends and opportunities (IEEE - UBMK - VII. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı). Ankara, 2022.

7. B. Elov, Sh. Hamraeva, X. Axmedova Methods for creating a morphological analyze // 14th International Conference on Intelligent Human Computer Interaction. 19-23 October 2022, Tashkent.



8. B. Elov, Sh. Hamroyeva, D. Elova. Morfologik analizatorni yaratish usullari // O'zbekiston: til va madaniyat. Amaliy filologiya masalalari. 2022-yil 5 (1) son. – B. 67-87.

9. Менглиев Б.Р., Хамроева Ш.М. Структура и единицы морфоанализатора узбекского языка // Компьютерная лингвистика и вычислительные онтологии. Выпуск 5 (Труды XXIV Международной объединенной научной конференции «Интернет и современное общество», IMS-2021, Санкт-Петербург, 24 – 26 июня 2021 г. Сборник научных трудов). – СПб.: Университет ИТМО, 2021. – 82 с

10. Elov B.B. N-gramm til modellari vositasida o'zbek tilida matn generatsiya qilish // Kompyuter lingvistikasi: muammolar, yechim va istiqbollari / Xalqaro ilmiy-amaliy konferensiya to'plami. Elektron nashr / ebook. – Toshkent: 2022.

11. Soysal, E., Wang, J., Jiang, M., Wu, Y., Pakhomov, S., Liu, H., & Xu, H. (2018). CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. Journal of the American Medical Informatics Association, 25(3). <https://doi.org/10.1093/jamia/ocx132>

12. Tenney, I., Das, D., & Pavlick, E. (2020). BERT rediscovers the classical NLP pipeline. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference. <https://doi.org/10.18653/v1/p19-1452>

Attardi, G. (2015). DeepNL: A deep learning NLP pipeline. 1st Workshop on Vector Space Modeling

13. for Natural Language Processing, VS 2015 at the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2015. <https://doi.org/10.3115/v1/w15-1515>

14. Koeva, S., Obreshkov, N., & Yalamov, M. (2020). Natural language processing pipeline to annotate bulgarian legislative data. LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings.

15. de Vries, W., van Cranenburgh, A., & Nissim, M. (2020). What's so special about BERT's layers? A closer look at the NLP pipeline in monolingual and multilingual models. Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020. <https://doi.org/10.18653/v1/2020.findings-emnlp.389>

16. Elov B. Tabiiy tilni qayta ishlash (nlp)da spacy modulidan foydalanish. Science and innovative development. 4/2022. Tashkent. 41-55 p.

17. Xusainova Z.Y. NLP: tokenizatsiya, stemming, lemmatizatsiya va nutq qismlarini teglash. O'zbek amaliy filologiyasi istiqbollari / Respublika ilmiy-amaliy konferensiya to'plami. Elektron nashr / ebook. – Toshkent: ToshDO'TAU, 26.10.2022. 159-163 b.

**Elov Botir Boltayevich** Alisher Navoiy nomidagi Toshkent Davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini mudiri t.f.f.d, dotsent.

E-mail: [elov@navoiy-uni.uz](mailto:elov@navoiy-uni.uz)

**Hamroyeva Shahlo Mirjanovna** Alisher Navoiy nomidagi Toshkent Davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar dotsenti f.f.d.

E-mail: [shaxlo.xamrayeva@navoiy-uni.uz](mailto:shaxlo.xamrayeva@navoiy-uni.uz)

**Xusainova Zilola Yuldashevna** Alisher Navoiy nomidagi Toshkent Davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini tayanch doktoranti.

E-mail: [xusainovazilola@navoiy-uni.uz](mailto:xusainovazilola@navoiy-uni.uz)

#### **B.B. Elov, Sh.M. Khamroeva, Z.Y. Xusainova NLP (TABIIY TILGA ISHLOV BERISH) NING PIPELINE KONVEYERI**

The problem of NLP should be divided into several small parts and solved step by step. This article examines all forms of text processing required at each stage of NLP problem solving. In NLP, the step-by-step processing of text is called a pipeline process (conveyor). A pipeline is a sequence of steps that must be followed when building any NLP model. Planning and developing a text processing in the NLP pipeline pipeline. All processing steps are defined as pre-trained, usable models in various NLP libraries. If necessary, additional, customized preprocessing steps can be developed depending on the given problem condition. How a particular preprocessing step works for a pipeline is considered the starting point for any NLP project. This article discusses the steps involved in implementing a pipeline process and their role in solving NLP tasks. This article analyzes the most common preprocessing steps given NLP problem can only be determined with a lot of experimentation.

**Keywords:** NLP, pipeline process (conveyor), tokenization, lemmatization, stemming, POS-tagging, morphological analysis.