

KORPUS LINGVISTIKASI

O'ZBEK TILIDAGI SO'ZLARNI MORFOLOGIK TAHLIL QILUVCHI DASTUR

Qarshiyev Abduvali Berkinovich

Fizika-matematika fanlari nomzodi

TATU Samarqand filiali professori

Karimov Suyun Amirovich,

Filologiya fanlari doktori

SamDU professori

Tursunov Muhammadsolih Sa'din o'g'li,

muhammadsolih927@gmail.com

TATU Samarqand filiali katta o'qituvchisi

Annotatsiya. Ushbu maqolada o'zbek tilidagi so'zlarni morfologik tahlil qiluvchi dasturdan foydalanish ko'rsatib o'tilgan. Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida o'zbek tilining elektron morfologik lug'atini yaratish masalasi qo'yiladi. Unga ko'ra, har bir so'zning qaysi so'z turkumiga xosligining aniq statistikasi ko'rsatiladi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. Undan keyin leksemalar so'z turkumlari bo'yicha to'plamlarga birlashadi. To'plamlarga ajratishda leksemalarning umumiy semantik, qolaversa, formal jihatlarini inobatga olinadi. To'plamlarga ajratilgan har bir so'zga qo'shilishi mumkin bo'lgan qo'shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo'shimchalar funksional belgilariga ko'ra guruhlariga bo'linib, ularning o'zbek tilidagi barcha mumkin bo'lgan mantiqiy bazasi shakllantiriladi.

Abstract. This article shows the use of the program for the morphological analysis of words in the Uzbek language. The creation of an electronic morphological dictionary of the Uzbek language is the first step towards creating an automatic text analyzer. It displays precise statistics on which word group each word belongs to. The tables reflect individual morphological features of each lexeme. After this, lexemes are combined into sets of word groups. When dividing into sets, the general semantic and formal aspects of lexemes are taken into account. A combination of suffixes is determined that can be added to each word divided into sets. Previously, all suffixes are divided into groups according to their functional characteristics and all possible logical bases for them in the Uzbek language are formed.

Аннотация. В данной статье показано использование программы для морфологического анализа слов узбекского языка. Создание электронного морфологического словаря узбекского языка является первым шагом на пути создания автоматического анализатора текста. В соответствии с ним отображается точная статистика того, к какой группе слов принадлежит

каждое слово. В таблицах отражены отдельные морфологические особенности каждой лексемы. После этого лексемы объединяются в наборы по группам слов. При разделении на наборы учитываются общие семантические и формальные аспекты лексем. Определяется комбинация суффиксов, которую можно добавить к каждому слову, разделенному на множества. Предварительно все суффиксы разделяются на группы по их функциональным признакам и формируются все возможные логические основы их в узбекском языке.

Kalit so‘zlar: *o‘zbek tili, so‘z turkumi, morfologik tahlil, so‘zning asosi, ma‘lumotlar bazasi, lug‘at, affikslar, dasturiy ta‘minot.*

So‘z turkumini aniqlovchi morfoanalizator uchun dastlabki algoritmlar ikki tarkibli tuzilishga asoslangan. Birinchi bosqich har bir so‘zning turkumini aniqlash bo‘lsa, ikkinchi bosqich ajratilgan har bir so‘zning grammatik qoidalariga tuzatishlar kiritish hisoblangan [Mahlov, Piotrowski, 2009]. O‘zbek tili uchun ma‘lumotlar bazasi juda muhim sanaladi. Shuni unutmash kerakki, morfoanaliz faqat shakl bilan cheklanib qolmasligi kerak. Kompyuter lingvistikasida chuqur tahlil qilish metodi semantik tahlil deb ataladi. Morfoanalizator quyidagi ikki bosqichdan iborat bo‘ladi:

- so‘zlarning asosini aniqlash;
- so‘z turkumini aniqlash;
- asosga ko‘ra har bir affiksni tahlil qilish. Bu jarayon chapdan o‘ngga yoki aksincha bo‘lishi mumkin;
- har bir qo‘shimchani turini belgilash.

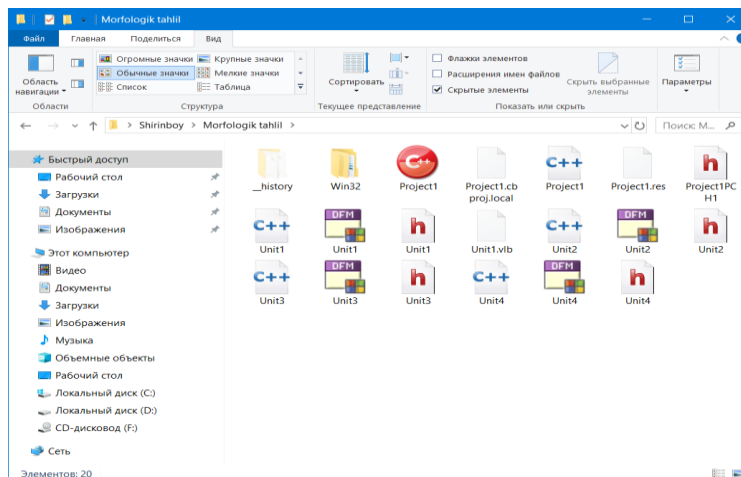
So‘zlarni ma‘lumotlar bazasidan o‘qib olish uchun ularning simvollar ketma-ketligiga amal qilinadi. Belgilash o‘sha so‘z yaxlit birlik sifatida aniqlanmagunicha bu holat takrorlanaveradi. Morfologik tahlil uchun komponent tahlil metodidan ham foydalanish mumkin. Unga ko‘ra so‘zning uzunligi aniqlanadi. Shu tarkibdagi simvollar soniga qarab ma‘lumotlar bazasidan qidiriladi. O‘sha so‘z bazadan aniqlanmagunicha tekshirish davom etaveradi.

O‘zbek tilida grammatik qo‘shimchalarning ketma-ketligi aniq belgilab qo‘yilgan, biroq qoidadan tashqari holatlar ham uchraydi. Chunonchi, ot va fe‘l so‘z turkumlarida quyidagicha:

OT	bola+jon+lar+im+dagi+lar+niki+mas+mi
FE‘L	yoz+dir+tir+il+ma+yotgan+lar+ga+mi yoz+dir+tir+il+ma+di

Morfotaktik qoidalar morfemalarning o‘zaro kombinatsiyalarini boshqaradi. Shu jihatdan, matnlarni tahlil qilganda so‘zlarning asos va affikslarni aniqlashda morfotaktik tamoyillarga tayaniladi [Baqayev, 2021].

Ish stolning  papkasida dasturiy majmuaning modullari va ularning Embarcadero XE3 C++ Builder dasturlash muhitida tuzilgan kodlari joylangan (1 – rasm):



1 – rasm. Dasturni ishga tushirish

1 – rasmdagi papkaning ichida  va  papkalariga ketma-ket kiriladi va dasturiy modulning  fayli ishga tushiriladi. Natijada ekranda quyidagi bosh oyna hosil bo'ladi (2 – rasm).



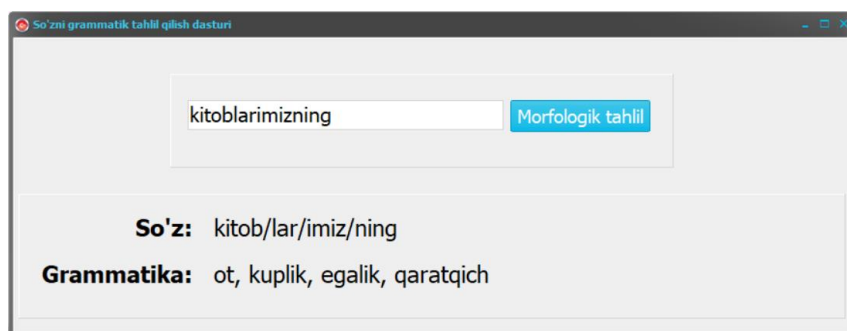
2 – rasm. Dasturning asosiy oynasi

2 – rasmdagi oynada dastur ikkita rejimda ishlashi ko'rinib turibdi. Bitta so'zni grammatik tahlil qilish va matnni grammatik tahlil qilish dasturlaridan iborat. “So'zni Grammatik tahlil qilish” tugmasi bosilsa 3 – oyna hosil bo'ladi. So'z kiritish maydoniga biror so'z kiritilsa, masalan “kitoblarimizning” so'zi kiritilsa, so'zdan birinchi navbatda asosni topish zarur bo'ladi. Bunda so'zdan asos sifatida “kitob” so'zini ajratib oladi. Sababi dasturning asoslar lug'atida kitob so'zi mavjud. Undan so'ng so'zning asosidan keyingi qo'shimchalarni aniqlaydi va qo'shimchalarga mos ravishda uning ma'nolarini foydalanuvchiga taqdim etadi (3-rasm).

Dastur “kitoblarimizning” so'zini quyidagicha tahlil qiladi:

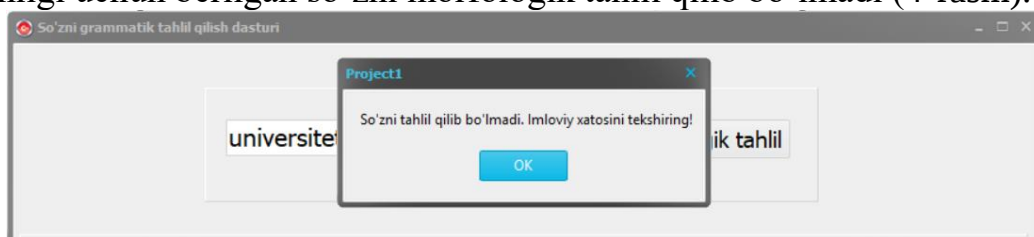
Asos: kitob – ot;

Qo'shimcha: -lar – ko'plik, -imiz – egalik, -ning – qaratqich.



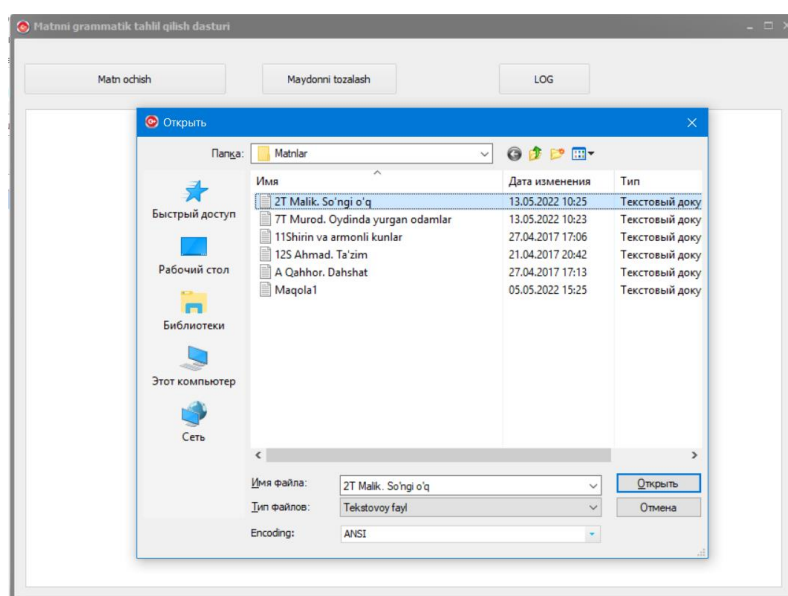
3 – rasm. So'zni grammatik tahlil qilish dasturi oynasi

Endi 3-rasmdagi dastur oynasiga “universitetning” degan so'zni kiritib, uni morfologik tahlil qilinsin. Bunda dastur asoslar lug'atida yuqoridagi so'zning asosi bo'lmaganligi uchun berilgan so'zni morfologik tahlil qilib bo'lmadi (4-rasm).



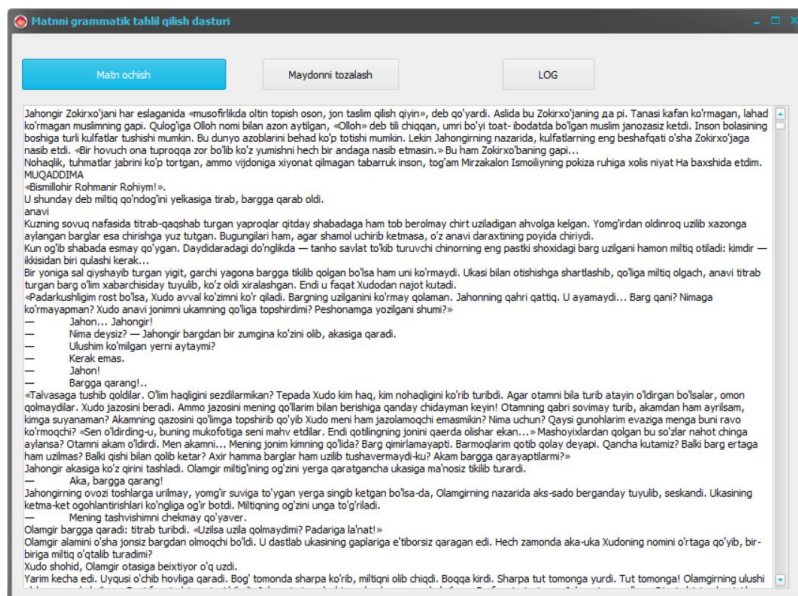
4 – rasm. So'zni Grammatik tahlil qilishda xatolik

Matnni grammatik tahlil qilish uchun 2-rasmdagi “Matnni grammatik tahlil qilish” tugmasi bosiladi. Keyingi qadamda “matnni ochish” tugmasi bosilib, dasturga matn yuklanadi (5-rasm).



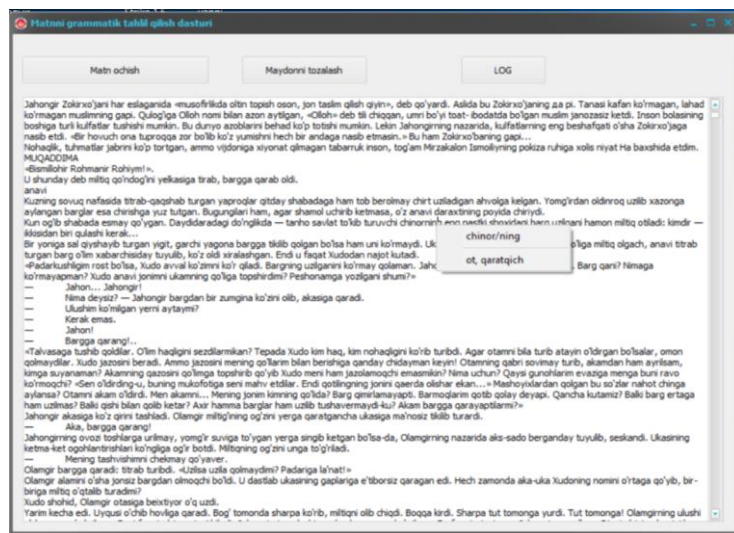
5 – rasm. Matnni dasturga yuklash

Matnni dasturga yuklagandan so‘ng matn dastur ishchi oynasida hosil bo‘ladi (6-rasm).

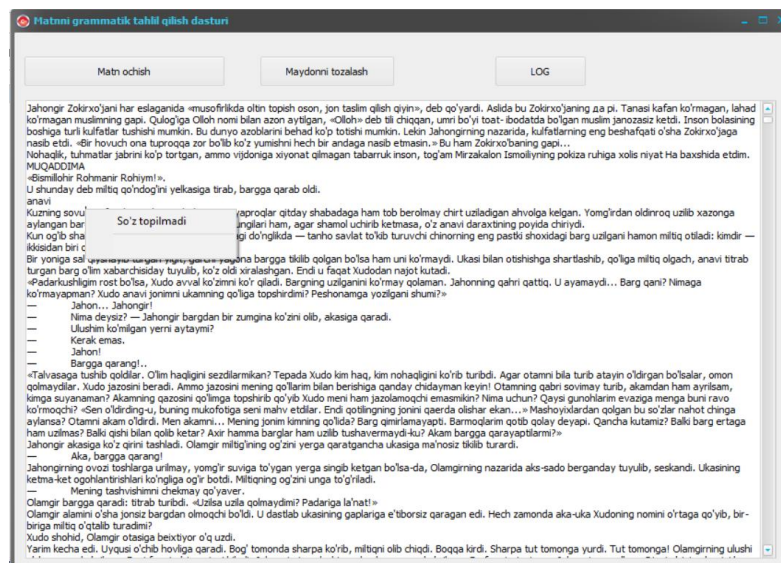


6 – rasm. Matnni dasturga yuklangan holati

Dasturga matn yuklangandan so‘ng, matnning ichida ixtiyoriy so‘zning, masalan “chinorning” so‘zini ustidan sichqonchaning o‘ng tugmasi bosilsa, tanlangan so‘zning grammatik xarakteristikalarini aniqlab beradi (7-rasm). Agar tanlangan so‘z bosilsa va bu so‘zning asosi lug‘atda bo‘lmasa, u holda “So‘z topilmadi” degan xabar paydo bo‘ladi. Masalan “sovuq” so‘zini morfologik tahlil qilinsin (8-rasm).



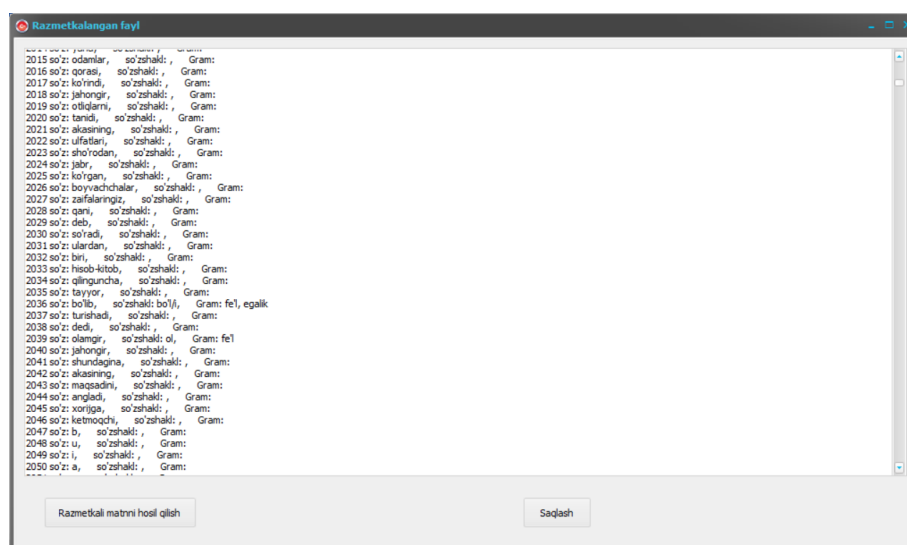
7 – rasm. Matndagi so‘zning morfologik tahlili



8 – rasm. Matndagi so‘zni morfologik tahlil qilib bo‘lmagan holat

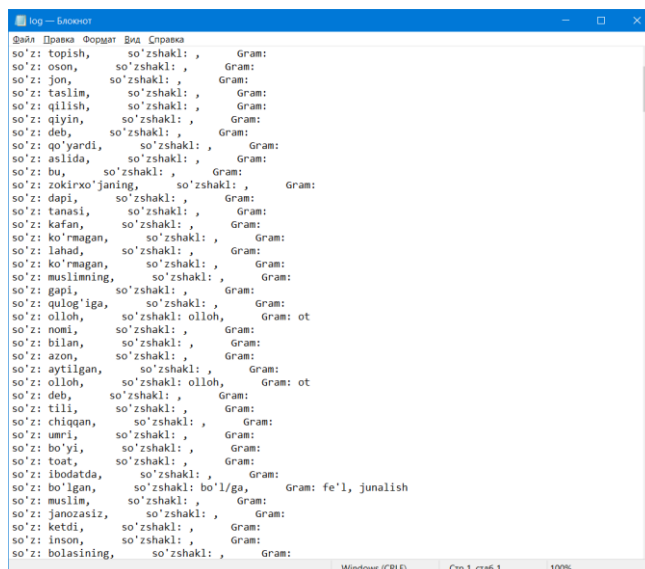
Dasturga yuklangan matnni to‘liq tahlil tahlil qilib, kompyuter xotirasiga saqlash lozim bo‘lsin. Bunda “LOG” tugmasi bosiladi va 9-rasm ochiladi.

9-rasmda “Razmetkali matnni hosil qilish” tugmasi bosiladi va tanlangan matndagi hamma so‘zlar razmetkalanib, foydalanuvchiga taqdim etiladi (10-rasm).



10 – rasm. Matnni razmetkalash

Matnni to‘liq razmetkalanganda ishchi oynaga tartib raqam, so‘z, so‘z shakl va grammatikasi chiqadi. So‘z shaklida so‘zning asosi va qo‘shimchalari ajratib ko‘rsatiladi. Hosil bo‘lgan razmetkali matnni kompyuter xotirasida saqlash uchun “saqlash” tugmasi bosiladi va “LOG.txt” nomli faylga saqlanadi.



11 – rasm. Faylda saqlangan razmetkali matnning ko‘rinishi

Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda dastlabki bosqich sifatida o‘zbek tilining elektron morfologik lug‘atini yaratish masalasi qo‘yiladi. Unga ko‘ra, har bir so‘zning qaysi so‘z turkumiga xosligining aniq statistikasi ko‘rsatiladi. Har bir leksemaning muayyan morfologik xususiyatlari jadvallarda aks etadi. Undan keyin leksemalar so‘z turkumlari bo‘yicha to‘plamlarga birlashadi. To‘plamlarga ajratishda leksemalarning umumiy semantik, qolaversa, formal jihatlarini inobatga olinadi. [Tursunov, Qarshiyev, 2019] To‘plamlarga ajratilgan har bir so‘zga qo‘shilishi mumkin bo‘lgan qo‘shimchalar kombinatsiyasi aniqlanadi. Undan oldin barcha qo‘shimchalar funksional belgilariga ko‘ra guruhlariga bo‘linib, ularning o‘zbek tilidagi barcha mumkin bo‘lgan mantiqiy bazasi shakllantiriladi.

Foydalanilgan adabiyotlar:

1. Mahlov C., Piotrowski M., JSLIM – “Computational Morphology in the Framework of the SLIM, Theory Language” / State of Art in Computational Morphology, Zurich, 2009, 15-p.
2. Baqayev I.I., “O‘zbek tili so‘z shakllarini morfologik tahlil qilish modellari va algoritmlari” / avtoreferat, Toshkent: TATU, 2021, 11-bet.
3. Tursunov M.S., Qarshiyev A.B., “O‘zbekcha matnlarni morfologik va leksik tahlil qilish algoritmlari va dasturlari” / International scientific conference “Actual problems of applied mathematics and information technologies”. – Toshkent, 14-15.11.2019, 290-291 betlar.