INTERNATIONAL INNOVATION AND RESEARCHES
International scientific journal Volume: 2 Issue: 3

O‘ZBEK TILIDAGI MATNLARDAGI ORFOGRAFIK XATOLARNI ANIQLOVCHI ALGORITM VA DASTURIY MAJMUA ISHLAB CHIQISH

Qo'ziyev Shahzod Ulug‘bek o'g'li
Ismailov Shixnazar Rashid o'g'li
Abidova Shaxnoza Bahodirovna
Abduvosiqov Oybek
Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti magistri
E-mail: shahzod4122@gmail.com

Annotatsiya. *Ushbu maqolada o‘zbek tilidagi matnlarda orfografik xatolarni aniqlash uchun aralash algoritm (statistik va qoida asosladi usullar) va dasturiy majmua ishlab chiqildi. Morfologik generator yordamida so‘z shakllari tahlil qilinadi, real-vaqt rejimida tekshirish imkoniyati yaratildi. Dastur ta’lim, matbuot va hujjatlar tayyorlashda matn sifatini oshirishga xizmat qiladi. Kelajakda dasturiy majmua milliy til texnologiyalarini rivojlantirish va jamiyatda til madaniyatini mustahkamlashda muhim rol o‘ynaydi.*

Kalit so‘zlar: *orfografik xatolar, o‘zbek tili, morfologik tahlil, modellar tahlili, dasturiy majmua.*

Аннотация. *В данной статье разработаны смешанный алгоритм (статистические и основанные на правилах методы) и программный пакет для обнаружения орфографических ошибок в узбекских текстах. Словоформы анализируются с помощью морфологического генератора, возможна проверка в режиме реального времени. Программа служит для улучшения качества текста в образовании, прессе и подготовке документов. В перспективе программный комплекс будет играть важную роль в развитии национальных языковых технологий и укреплении языковой культуры в обществе.*

Ключевые слова: *орфографические ошибки, узбекский язык, морфологический анализ, модельный анализ, программный пакет.*

Annotation. *In this article, a mixed algorithm (statistical and rule-based methods) and a software package have been developed to detect spelling errors in Uzbek texts. Word forms are analyzed using a morphological generator, and real-time checking is possible. The program will serve to improve the quality of texts in education, the press, and document preparation. In the future, the software package will play an important role in the development of national language technologies and strengthening language culture in society.*

Key words: *spelling errors, Uzbek language, morphological analysis, model analysis, software package.*

KIRISH

Zamonaviy axborot texnologiyalari rivojlanganiga qaramay, matnlardagi imlo xatolarini aniqlash muammosi, ayniqsa o‘zbek tilida, hali ham dolzarb hisoblanadi. O‘zbek tilining morfologik jihatdan boy va murakkab tuzilishi, so‘zlar ko‘plab shakl almashishlariga ega bo‘lishi, shuningdek, turli dialektik va fonetik o‘zgarishlar tufayli

aniq va samarali imlo tekshiruv algoritmlarini ishlab chiqish murakkab vazifaga aylanmoqda.

Bugungi kunda dunyoning ko‘plab tillari uchun ishlab chiqilgan imlo tekshirish dasturlari (masalan, Grammarly, LanguageTool) yuqori aniqlikda xatolarni topib, ularni tuzatish bo‘yicha tavsiyalar beradi. Biroq, o‘zbek tiliga mo‘ljallangan shunday tizimlar hali yetarlicha ishlanmagan. Buning asosiy sabablari orasida o‘zbek tiliga xos lingvistik modellarning etishmasligi, lug‘atlar bazasining to‘liq shakllanmaganligi va morfologik tahlil qilish algoritmlarining murakkabligi kabilar kiradi.

Ushbu tadqiqotda o‘zbek matnlaridagi orfografik xatolarni aniqlash uchun takomillashtirilgan lingvo-matematik modellar asosida algoritm va dasturiy ta’minot to‘plamini ishlab chiqish nazarda tutilgan. Ishlatiladigan yondashuvlar orasida statistik usullar (N-gram analizi, chastotali lug‘atlar), Levenshtein masofasi kabi klassik algoritmlar, shuningdek, mashina o‘rganish va sun’iy intellekt texnologiyalari (neytron tarmoqlar, transformer modellar) aralashmasidan foydalaniladi.

Imlo xatolarini avtomatik tuzatish tizimlari nafaqat matnlar sifatini oshirish, balki ma’lumotlarni qayta ishlash samaradorligini oshirish, ta’lim jarayonlarini yaxshilash va til xizmatlarini rivojlantirishda muhim rol o‘ynaydi. Loyiha natijasida ishlab chiqiladigan dasturiy mahsulot turli sohalarda – jumladan, matbuot, ta’lim, adabiyot va rasmiy hujjatlar tayyorlashda keng qo‘llanilishi mumkin.

Shu bilan birga, ishlab chiqilayotgan tizim o‘zbek tilining rivojlanishiga hissa qo‘shib, kelajakda tilni qayta ishlash (NLP) sohasidagi boshqa loyihalar uchun ham asos bo‘lib xizmat qiladi. Tadqiqot jarayonida yig‘ilgan lingvistik ma’lumotlar va ishlab chiqilgan algoritmlar o‘zbek tilidagi boshqa sun’iy intellekt qo‘llanmalari uchun ham foydali bo‘lishi kutilmoqda.

Tadqiqotning dolzarbligi: O‘zbek tilida imlo xatolarini aniqlash va tuzatishni avtomatlashtirish masalasi zamonaviy axborot jamiyatining talablariga mos keladigan eng dolzarb muammolardan biridir.

Tadqiqot maqsadi: O‘zbek tilidagi matnlardagi xatolarni aniqlovchi takomillashtirilgan modellar asosida algoritm va dasturiy majmua ishlab chiqish.

Tadqiqot vazifalari:

- O‘zbek tilidagi orfografik xatolarni tahlil qilish.
- Xatolarni aniqlashga oid mavjud lingvo-matematik modellarni tahlil qilish.
- Xatolarni aniqlashga oid mavjud lingvo-matematik modellar asosida o‘zbek tilidagi matnlarda orfografik xatolarni aniqlovchi algoritm ishlab chiqish.
- Ishlab chiqilgan algoritm asosida dasturiy majmua yaratish.

Imlo xatolarini aniqlash bo‘yicha mavjud dasturiy yechimlarni solishtirish va tahlil qilish quyidagi asosiy maqsadlar bilan amalga oshiriladi:

Dunyoda qo‘llanilayotgan imlo tekshirish dasturlari (masalan, Grammarly, LanguageTool) turli tillar uchun ishlab chiqilgan algoritmlarga asoslanadi. Ularning o‘zbek tiliga mosligini tekshirish orqali ularning afzalliklari (tezlik, keng lug‘at bazasi) va kamchiliklari (morfologik murakkablikni hisobga olmaslik, o‘zbekcha so‘z shakllarini tanimay olmaslik) aniqlangan.

Tahlil jarayoni o‘zbek tiliga xos lingvistik qoidalar (masalan, so‘zning shakl almashishi) bilan mavjud algoritmlarning “mos kelmasligini” ko‘rsatadi. Tahlil jarayoni orqali qaysi algoritmlar (statistik/metodlar, qoida asosladi tizimlar, AI) o‘zbek tilida yuqori natija berishi mumkinligi aniqlanadi.

1-Jadval. Mavjud orfografik xatoliklarni aniqlovchi dasturlar solishtirma tahlillari.

№	Dastur nomi	Pullik	Ko‘p tillarni qo’llab quvvatlash	Xatoliklarni aniqlash
1.	LanguageTool	+	+	85%-Grammatik va uslubiy xatolarni yaxshi aniqlaydi, O‘zbek tili uchun cheklangan.
2.	Microsoft muharriri	-	-	80%- Inglizcha uchun kuchli, lekin O‘zbek tilida cheklangan.
3.	Reverso	-	-	75%- Asosan tarjima va grammatika uchun.
4.	Orfo	+	+	70%-Rus tili uchun mo‘ljallangan.
5.	Text.ru	+	-	65%- Ko‘proq plagiat tekshiruviga mo‘ljallangan, grammatika ikkinchi darajada.
6.	Advego	-	+	68%- Uslubiy xatolarni aniqlaydi, ammo grammatik aniqlash o‘rtacha.
7.	Google Docs	-	+	78%- Ingliz va rus tillarida yaxshi ishlaydi, o‘zbek tili cheklangan.
8.	Glavred	-	-	60%- Uslubiy tahlilga ixtisoslashgan, grammatika kamroq aniqlanadi.
9.	Yurida-Orfo	+	-	75%- Huquqiy matnlar uchun mo‘ljallangan, maxsus xatoliklar aniqlanadi.
10.	Grammarly	+	+	Eng kuchlilaridan biri, ammo o‘zbek tilini qo‘llab-quvvatlamaydi.
11.	Orfocheck	-	-	87%- O‘zbek tiliga ixtisoslashgan, grammatik, punktuatsion va orfografik xatolarni aniqlashda yuqori aniqlik.

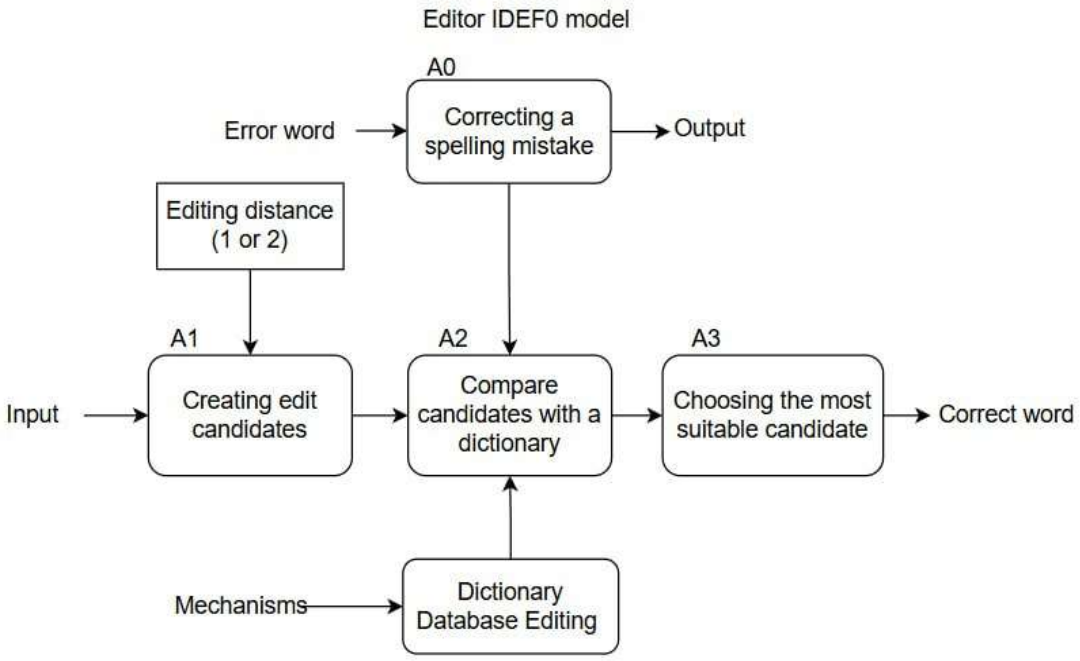
Ushbu ro‘yxatdagi dasturiy vositalar majmualar o‘zbek tilidagi matnlarni tahlil qilish va ortografik, grammatik yoki uslubiy xatolarni aniqlash uchun materiallarni o‘z ichiga oladi. Mavjud dasturiy yechimlarning tahlili shuni ko‘rsatadiki, o‘zbek tilidagi imlo xatolarini aniqlash muammosi hal etilmagan. Global dasturlar tilning

o‘ziga xos xususiyatlarini hisobga olmaydi, mahalliy yechimlar esa texnologik jihatdan qoloq. Innovatsion yechim sifatida lingvistik qoidalar, statistik usullar va AI texnologiyalarini uyg‘unlashtirgan algoritm taklif etiladi. Bu yondashuv nafaqat xatolarni aniqlash aniqlik darajasini oshiradi, balki o‘zbek tilini raqamli maydonda rivojlantirishga ham hissa qo‘shadi.

O‘zbek tilidagi matnlarni tahlil qilish va xatolarni aniqlash uchun dastur qidirasiz, Jadvaldan ko‘rinib turibdiki Orfocheck yaxshi tanlov bo‘ladi. Huquqiy matnlar bilan ishlasangiz, pullik versiyada Yurida-Orfo ham foydali bo‘lishi mumkin. Boshqa dasturlar asosida boshqa tillar (ingliz, rus) uchun mo'ljallangan va o‘zbek tilini qo‘llab-quvvatlamaydi. O‘zbek tilining orfografik tizimi talaffuzga yaqinligi bilan boshqa tillardan farq qiladi. O‘zbek tili punktuatsiyasini o‘ziga xos so‘z tuzilish va talaffuz qoidalariga mos bo‘lib, boshqa tillarga nisbatan murakkabroq va maxsus belgilardan (apostrof kabi) iboratdir.

Modellar tahlili

IDEF0 (Integration Definition for Function Modeling) — bu kompleks jarayonlarni tizimli ravishda tahlil qilish va vizual tasvirlash uchun ishlatiladigan funktsional modeldir.

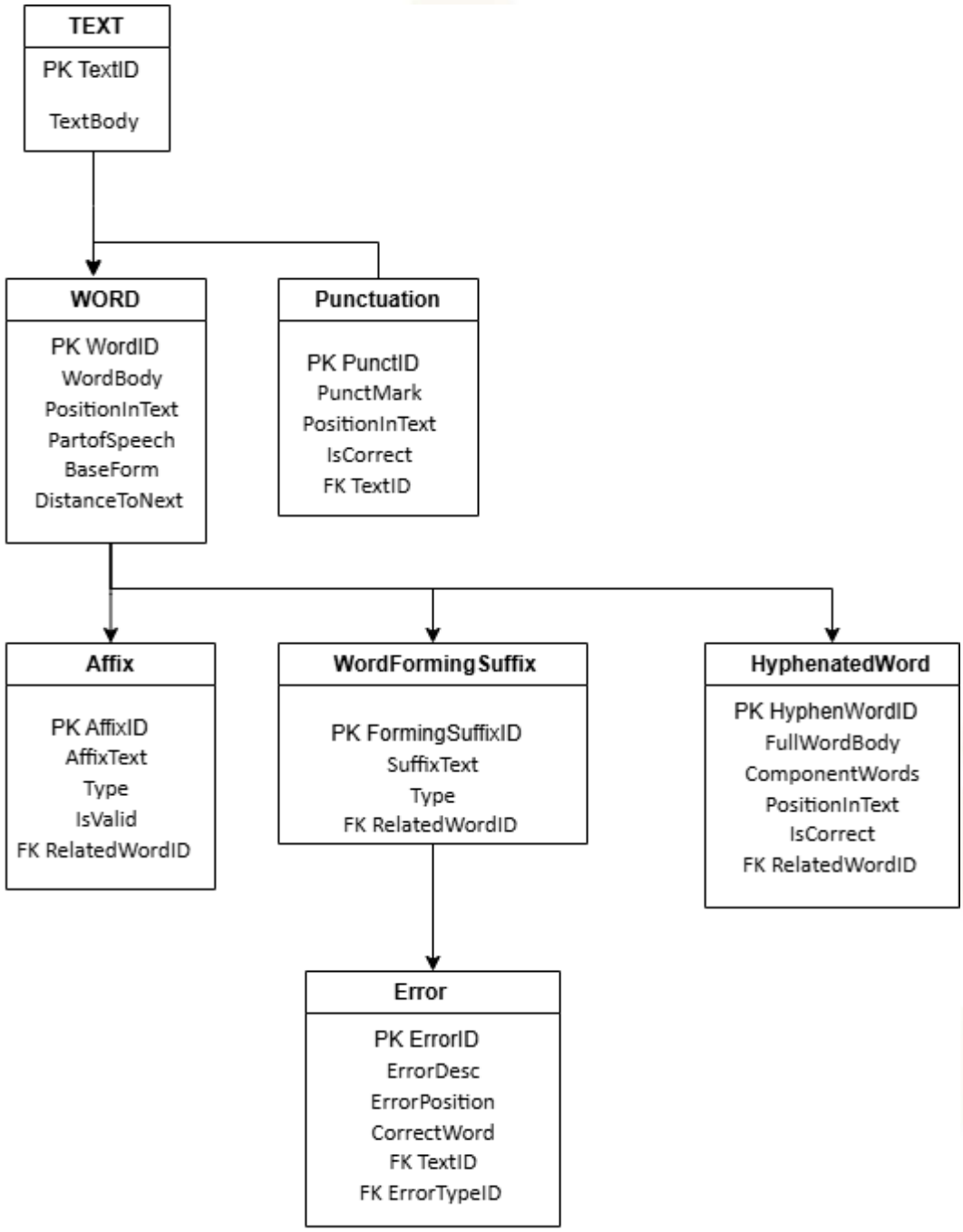


1-rasm. Tahrirlovchi IDEF0 modeli.

Yuqoridagi rasmda tahrirlovchi dastur berilgan model va algoritm asosida matndagi orfografik xatolarni tahrirlash jarayoni tasvirlangan.

Bu IDEF0 modeli - u jarayonlarni tizimli tahlil qilish va modellashtirish uchun jihozlash modellashtirish usuli. Uch modelda jarayonlarga kirish (Input), chiqish (Output), boshqaruv (Controls) va mexanizmlar (Mechanisms) sifatida tasvirlangan.

Bu model o‘zbek tilidagi matnni tahlil qilib, orfografik xatolarni tuzatish jarayonini amalga oshiradi. U sintaktik va morfologik, so‘zlarni muqobil shakllarini keltirib xatolarni tuzatish kabi bosqichlarni o‘z ichiga oladi. Bu modelni yanada kengroq yoritib berish maqsadida dastur jarayonining ketma-ketligini aks ettiruvchi - IDEF1x modeli yaratildi.



2-rasm. IDEF1x modeli

Yuqoridagi IDEF1x modeli o’zbek tilidagi matnda orfografik xatolarni aniqlash jarayonining ketma-ketligi keltirilgan. IDEF1x modeli - bu ma'lumotlar bazasi modellashtirish uchun grafik metodologiya. Asosan, ma'lumotlar tuzilmasi (yoki “ma'lumotlar modeli”) ni aniq va qat’iy shaklda ifodalash uchun belgi hisoblanadi.

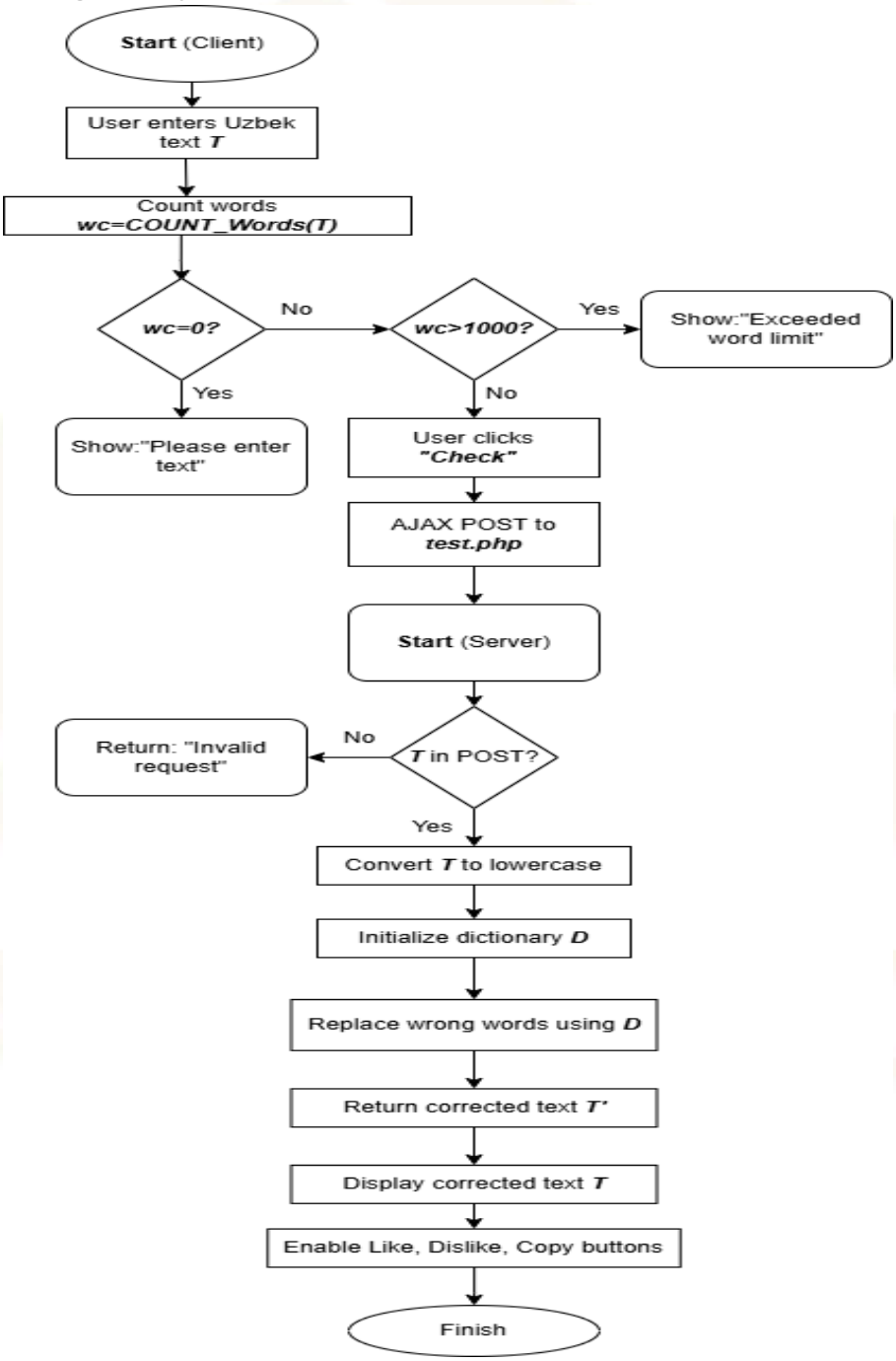
Uning asosiy maqsadi — ma’lumotlar va ular o‘rtasidagi munosabatlarni aniqlik bilan ko‘rish va normalizatsiya jarayonini osonlashtirishdan iboratdir. IDEF1x modeliga nimalarga e'tibor beradi:

• Entity (obyektlar, ya'ni ma'lumotlar haqida gapiradigan narsa yoki hodisa).

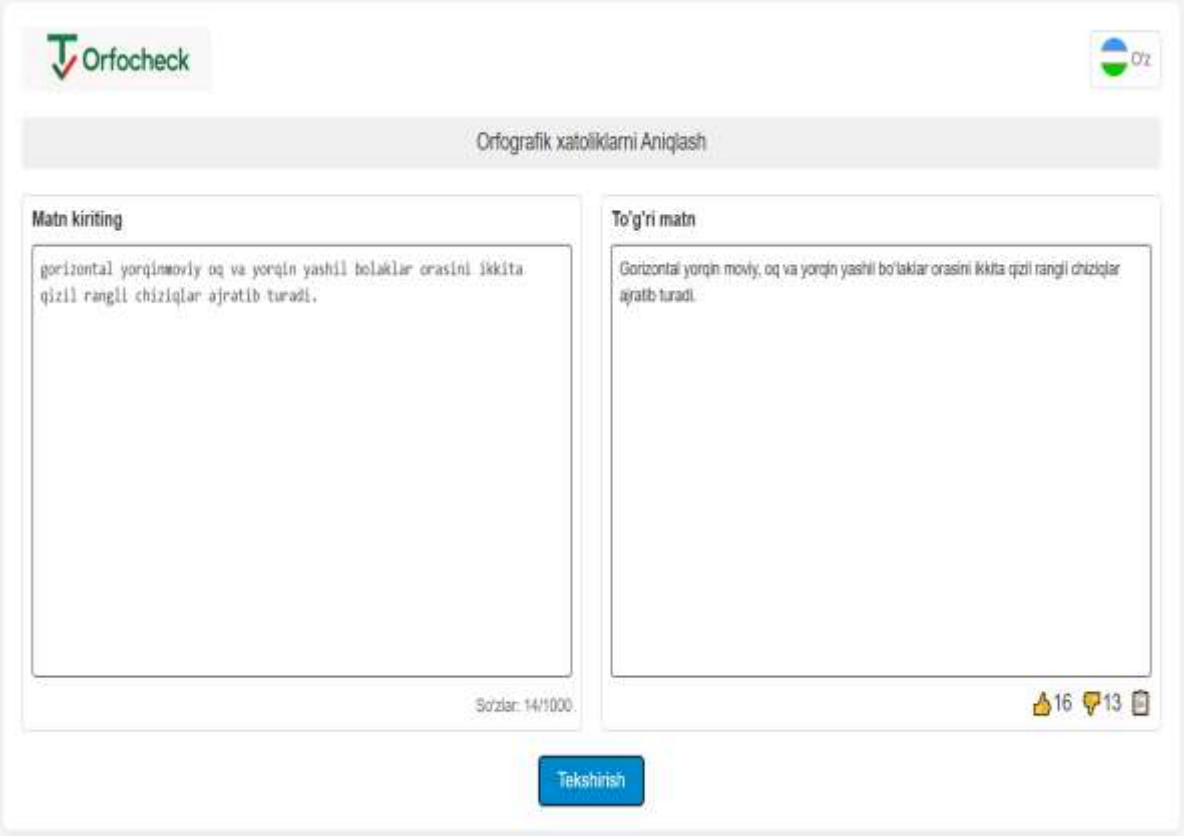
• Atribut (ob’ektlarning xossalari yoki atributlari).

• Munosabatlar (obyektlarning o'zaro bog'lanishlar).

Yuqoridagi jarayonlarni aks ettirgan takomillashtirilgan IDEF0 va IDEF1x modellari asosida algoritm yaratildi.



3-rasm. O’zbek tilida orfografik xatoliklarni aniqlash uchun dasturiy majmua algoritmi yaratildi.



4-rasm. Orfografik xatoliklarni aniqlash uchun Orfocheck dasturiy majmua interfeysi.

Rasmda "Orfocheck" dasturining o‘zbek tilidagi imlo xatolarini aniqlash va tuzatishdagi samaradorligi ko‘rsatilgan. Dastur matnni grammatik jihatdan to‘g‘rilash bilan birga, foydalanuvchiga vizual qo‘llab-quvvatlash orqali ishonchli yechim taklif etadi. Orfocheck Dasturining Ishlash Prinsipi: Xatolarni avtomatik aniqlash va dastur so‘zlarni lug‘at (ma’lumotlar bazasi) va morfologik qoidalar asosida solishtiradi.

Keltirilgan rasmda "Orfocheck" dasturiy majmuasining o‘zbek tilidagi imlo xatolarini aniqlash va tuzatishdagi samaradorligi ko‘rsatilgan. Foydalanuvchilarga qulay bo‘lishi uchun o‘zbek tilida kiritilgan matnni to‘g‘ridan-to‘g‘ri xató yozilgan matnni to‘g‘ri matn ko‘rinishida dasturiy majmua oynasida chiqarib beradi. Dastur matnni grammatik jihatdan to‘g‘rilash bilan birga, foydalanuvchiga vizual qo‘llab-quvvatlash orqali ishonchli yechim taklif etadi.

Dasturiy majmua quyidagi qismlardan iborat:

- Matn kiritish Maydoni.
- To‘g‘ri matn Maydoni.
- Tekshirish tugmasi.
- Vizual Belgilar.

Dasturning samaradorligi: Bir necha xatolarni bir vaqtning o‘zida aniqlash va tuzatish qobiliyatiga ega.

XULOSA

O‘zbek tilidagi matnlarda orfografik xatolarni aniqlovchi algoritm va dasturiy majmua ishlab chiqish bo‘yicha olib borilgan tadqiqot natijasidaasosiy yutuqlarga erishildi. O‘zbek tilini raqamli muhitda qo‘llash imkoniyatini kengaytirdi.Ta’lim, matbuot, rasmiy hujjatlar tayyorlash kabi sohalarda matnlar sifatini oshirishga xizmat qiladi. Kelajakda NLP sohasidagi loyihalar uchun lingvistik ma’lumotlar bazasini yaratdi. Ishlab chiqilgan algoritm va dasturiy majmua nafaqat imlo xatolarini tuzatish, balki o‘zbek tilining raqamli resurslarini boyitish, milliy til texnologiyalarini rivojlantirish va jamiyatda til madaniyatini mustahkamlashda muhim rol o‘ynaydi. Loyiha kelajakda sun’iy intellekt asosidagi yanada murakkab til qo‘llanmalari uchun poydevor bo‘lishi mumkin.

Foydalanilgan adabiyotlar ro‘yxati:

1. Dadaboyev, H., Hamidov, Z., & Xolmanova, Z. (2007). O ‘zbek adabiy tili leksikasi tarixi. Toshkent: Fan, 97-103.
2. Israilova, S. (2023). Tarixiy manbalarda ranglar talqini. Mirzo ulug’bek nomidagi 0 ‘zbekiston milliy universiteti ilmiy jurnali.
3. Zulhumor, X. O. L. M. A. N. O. V. A. (2017). Babür’ün Özbek Dilbilimini Geliştirmedeki Büyük Hizmetleri. Journal of Modern Turkish Studies, 14(1), 85.
4. Bilveren, T. (2015). Babürnâmedeki bazi kelimelerin tarihî ve etimolojik tahlili. Kilis 7 Aralık Üniversitesi Sosyal Bilimler Dergisi, 5(9), 114-118.
5. Israilova, S. (2023). The pragmatic occurrence and mental properties of color component units in context. Interpretation and researches, 1(5).
6. Saodat, I., & Ma’suda, E. (2023). Bilingvizm turlari va ikkinchi til leksik bazasi. In international scientific conferences with higher educational institutions (Vol. 1, No. 14.04, pp. 185-188).
7. Bakhronova, D., Alavutdinova, N., Israilova, S., & Vilchis, V. V. Color Lexemes in Context: Cognitive and Cultural Explorations.
8. Zulxumor, X. (2023). Bobur–semantikaning sohir bilimdoni. Роль наследия Захириддина Мухаммада Бабура в развитии восточной государственности и культуры, 1(1).
9. Saodat, I. Ko ‘k” leksemasining etimologik tavsifi. O ‘zbekiston, 75.
10. Israilova, S. (2022). “Ko ‘k” leksemasi semantikasini yoritishda tezaurus modelidan foydalanish. Computer linguistics: problems, solutions, prospects, 1(1).
11. Turapovna, I. S. (2021). Semantics of the lexeme" green". Academicia: An International Multidisciplinary Research Journal, 11(9), 440-448.
12. Israilova, S., & Alavutdinova, N. (2023). Intercultural Color Interpretation. Science and innovation, 2(C12), 111-115.
13. Xolmanova, Z., Saidaxmedova, O., & Nurullayeva, O. (2018). Lingvokulturologiyaga oid tushunchalar tadqiqi. Toshkent: Navro ‘z, 41-49.