

TAHRIRLOVCHI DASTUR (SPELLCHECKER) TEXNOLOGIYASI

Sobirova Nazira G'anijon qizi

nazirasobirova@gmail.com

Alisher Navoiy nomidagi Toshkent davlat o'zbek
tili va adabiyoti universiteti tayanch doktoranti

Annotatsiya. Kompyuter lingvistikasiga doir qator izlanishlar olib borish natijasida mashina tarjimasi, korpus texnologiyasi, kompyuter leksikografiyasi hamda dasturiy ilovalarning lingvistik asoslarini ishlab chiqish, leksikografik matnlarga ishlov berish va lingvostatistik tahlil etish borasida muayyan darajada natijalarga erishib kelinmoqda. Ayniqsa, bugungi shiddatkor davr uchun matnlarni turli jihatdan avtomatik tahrirlash va tahlil qilish, matnlarning avtomatik tarzda morfologik, sintaktik va semantik tahlilini (inglizcha parsing) ta'minlovchi tizimlar yaratish, mashina tarjimasi uchun mo'ljallangan dasturlarni ishlab chiqish muhimlik kasb etadi. Mazkur maqolada avtomatik tahrir hamda matnni avtomatik qayta ishlaydigan dasturlarning avtomatik tarjima sifatini yaxshilashdagi roli, xususan, spellchecker dasturlar, ularning avtomatik tarjimadagi ahamiyati, ushbu dasturlarning ishlab chiqish masalalari muhokama qilingan.

Abstract. As a result of carrying out a number of researches on computer linguistics, some results have been achieved in the field of machine translation, corpus technology, computer lexicography and the development of linguistic bases of software applications, lexicographic text processing and linguostatistical analysis. Especially for today's intense era, it is important to automatically edit and analyze text in various ways, to create systems that provide automatic morphological, syntactic and semantic analysis of text (English parsing), to develop software for machine translation.

This article discusses the role of automatic editing and automatic text processing programs in improving the quality of automatic translation, in particular, spellchecker programs, their importance in automatic translation, and development issues of these programs.

Аннотация. В результате проведения ряда исследований по компьютерной лингвистике достигнуты некоторые результаты в области машинного перевода, корпусных технологий, компьютерной лексикографии и разработки лингвистических основ программных приложений, лексикографической обработки текста и лингвостатистического анализа. Особенно в современную напряженную эпоху важно автоматически редактировать и анализировать текст различными способами, создавать системы, обеспечивающие автоматический морфологический, синтаксический и семантический анализ текста (синтаксический анализ английского языка), разрабатывать программное обеспечение для машинного перевода. В данной статье рассматривается роль программ автоматического

редактирования и автоматической обработки текста в повышении качества автоматического перевода, в частности программ проверки орфографии, их значение в автоматическом переводе, а также вопросы развития этих программ.

Kalit soʻzlar: *avtomatik tahrir, orfografiya, sunʼiy intellekt, korrektura, imlo qoidalari, tahlil metodlari, tokenizatsiya, lemmatizatsiya, stemmizatsiya, spellchecker.*

Dunyo tilshunosligida muayyan tillarda yaratilgan ogʻzaki va yozma manbalarni raqamlashtirish hamda ulardan tabiiy tillarni qayta ishlashda foydalanish muhim masalalardan biriga aylandi. Prezidentimiz tomonidan 2020-yil 20-oktabrdagi “Mamlakatimizda oʻzbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari toʻgʻrisida”gi Farmoni bilan tasdiqlangan “2020-2030-yillarda oʻzbek tilini rivojlantirish va til siyosatini takomillashtirish Konsepsiyasi”da belgilangan davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini taʼminlash maqsadida oʻzbek tilini Internet jahon axborot tarmogʻida ommalashtirish, unda munosib oʻrin egallashini taʼminlash, dasturiy mahsulotlarning oʻzbekcha ilovalarini yaratish, oʻzbek tilini oʻrgatuvchi kompyuter dasturlarini keng miqyosda amaliyotga tatbiq qilish hamda oʻzbek tilidagi matnlar tahririga moʻljallangan kompyuter dasturlarini yaratish kabi vazifalarining koʻrsatib berilishi bugungi kun tilshunos-dasturchi mutaxassislariga katta masʼuliyat yukladi. Shu jihatdan, oʻzbek tiliga zamonaviy axborot texnologiyalari vositalarida ishlov berilishiga erishish, buning uchun til korpuslari, elektron tarjimon, tezaurus, orfokorrektor, ularga dastlabki avtomatik ishlov berish vosita (lingvistik analizator)lari hamda ularning lingvistik taʼminotini yaratish singari masalalar dolzarblik kasb etmoqda. [Hamroyeva, Mengliyev, Gʻulomova, 2022].

Avtomatik tahrir - kompyuter lingvistikasining asosiy yoʻnalishlaridan biri boʻlib, u kompyuterdagi matn muharrirlari (Microsoft Word, Exsel, Wordpad, Lexicon) rivojlanishi bilan bogʻliqdir. Avtomatik tahrir deganda, kompyuterga kiritilayotgan matnning mexanik xatolari avtomatik tarzda toʻgʻrilanishi va xato ekanligi haqidagi signallarning foydalanuvchiga taklif etilishi nazarda tutiladi.

Xorijda kompyuter lingvistikasi sohasida matnlarni avtomatik tahlil va tahrir qilish masalasi koʻplab olimlar tomonidan oʻrganilgan [Peterson, James, 1980.] J.Peterson xatolarni tekshiruvchi va ularni tuzatuvchi dasturiy taʼminot (DT)lar borasida soʻz yuritgan. Matnning avtomatik tahlil jarayonini lingvistik aspektda oʻrganish D.Jurafskiy nomi bilan bogʻliq.

Avtomatik tahrirlash yoʻnalishida avtomatik tahrir qiluvchi dasturlar ishlab chiqilgan, ular ruscha, inglizcha, nemischa, fransuzcha, italyancha matnlarni tahrirlashga yoʻnaltirilgan. Hatto tatar, boshqird, qozoq, qirgʻiz, ozarbayjon kabi turkiy tillarga mansub matnlar uchun morfologik va sintaktik analizatorlar yaratilgan, MS WORD sistemasiga avtotahrir paketlari joylashtirilgan. Bunday



tahrirlash dasturlari, asosan, imloni tekshirishga mo'ljallangan. Rus tilini avtomatik tahlil qilish yo'nalishida R.G.Kotov, V.E.Berzon, V.G.Britvin, I.A.Melchuk, A.B.Kuznetsov, L.I.Belyayeva, V.A.Chijakovskiy, G.G.Belonogov, I.S.Duganova, L.L.Iomdin, E.K.Lavoshnikova, A.V.Sokirko, I.M.Nojoy, E.I.Bolshakova, D.V.Klishinskiy, N.N.Leonteva va boshqalar samarali tadqiqot olib borishgan.

Turkiy tillarga mansub matnlarni statistik metodlar asosida lingvistik tahlil qilish sohasida R.G.Piotrovskiy, M.A.Mahmudov, A.A.Piotrovskaya, T.Sadikov ishlari diqqatga sazovor. O'zbek tili materiallari bo'yicha KLga oid dastlabki tadqiqotlarni olib borgan olimlar H.Arziqulov, S.Rizayev va S.Muhamedovlardir. Ular, asosan, KLning statistik tahlil yo'nalishlari bo'yicha nazariy va amaliy ishlarni amalga oshirgan. H.Arziqulov o'zining “Informatika i pererabotka teksta sredstvami vychislitel'noy tekhniki” (“Информатика и переработка текста средствами вычислительной техники”) nomli monografiyasida kompyuter yordamida katta hajmdagi matnlarga ishlov berishning avtomatik tizimi asoslarini tahlil qilgan. S.Muhamedov va R.Piotrovskiy hammuallifligida yozilgan “Injenernaya lingvistika i opit sistemnostatisticheskogo issledovaniya uzbekskix tekstov” (“Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов”) nomli kitob, shuningdek, R.G.Piotrovskiy, T.Sadikovlarning ilmiy izlanishlarida lingvistik modellar, modellash va uning umumiy tamoyillari haqida fikr yuritilgan, o'zbekcha matnlarning kvantitativ modellari keltirilgan. S.Muhamedov va R.Piotrovskiy hammuallifligidagi asarda sun'iy intellekt yaratish va injener lingvistikasi metodlari bilan o'zbekcha nutqni avtomatik qayta ishlash jarayonida foydalanish uchun zarur bo'lgan o'zbekcha matnlarning leksik-morfologik tuzilishi statistik tahlili natijalari keltirilgan. S.Rizayevning “Kibernetika va tilshunoslik” asarida esa tilshunoslikda aniq metodlarning qo'llanilishi, bunda kompyuterdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari, matnlarni avtomatik qayta ishlash, mashina tarjimai muammolari, shuningdek, harflarning qo'llanilish chastotasini aniqlashda kompyuterning o'rni borasidagi ma'lumotlar keltirilgan. M.K.Ayimbetov statistik metodlardan foydalangan holda turkiy tillar – o'zbek, qozoq va qoraqalpoq tillari morfologiyasini modellash orqali ularning qarindosh tillar ekanligini isbotlab, ushbu tillarning kvantitativ tipologiyasini ishlab chiqqan bo'lsa, X.F.Isxakova turkiy tillarning formal morfologiyasini tadqiq qilish bo'yicha samarali izlanish olib borgan. XXI asr boshiga kelib o'zbek va ingliz tillarida bir qator izlanishlar qilgan, o'zbek KL rivojiga salmoqli ulush qo'shgan A.Po'latov boshchiligidagi o'zbek tilshunosligiga oid masalalarni kompyuter vositasida hal etishga bag'ishlangan bitiruv malakaviy ishlari, magistrlik dissertatsiyalari himoya qilindi, o'nlab ilmiy maqolalar chop etildi, o'quv qo'llanmalar yaratildi va kitoblar nashrdan chiqarildi. Ayniqsa, fe'l turkumini formallashtirish, ularning modelini ishlab chiqish va fe'llar asosida lingvistik ta'min yaratishga bag'ishlangan talay ishlar dunyo yuzini ko'rdi.

So'nggi yillarda M.X.Xakimov rahbarligida “Tarjimon LMX” sistemasi uchun bajarilgan tadqiqot ishlarida o'zbek va rus tillarining leksik-sintaktik tahlili

berilgan. Sh.Xamroyeva, N.Abdurahmonovalar ishlarida aniq metodlarning qo'llanilishi, bunda kompyuterdan foydalanish, til va nutq hodisalariga statistik yondashishning sabablari, matnlarni avtomatik qayta ishlash, mashina tarjimasini muammolari, shuningdek, harflarning qo'llanilish chastotasini aniqlashda kompyuterning o'rni borasidagi ma'lumotlar keltirilgan. M.K.Ayimbetov statistik metodlardan foydalangan holda turkiy tillar – o'zbek, qozoq va qoraqalpoq tillari morfologiyasini modellashtirish orqali ularning qarindosh tillar ekanligini isbotlab, ushbu tillarning kvantitativ tipologiyasini ishlab chiqqan bo'lsa, X.F.Isxakova turkiy tillarning formal morfologiyasini tadqiq qilish bo'yicha samarali izlanish olib borgan. XXI asr boshiga kelib o'zbek va ingliz tillarida bir qator izlanishlar qilgan, o'zbek KL rivojiga salmoqli ulush qo'shgan A.Po'latov boshchiligida o'zbek tilshunosligiga oid masalalarni kompyuter vositasida hal etishga bag'ishlangan bitiruv malakaviy ishlari, magistrlik dissertatsiyalari himoya qilindi, o'nlab ilmiy maqolalar chop etildi, o'quv qo'llanmalar yaratildi va kitoblar nashrdan chiqarildi. Ayniqsa, fe'l turkumini formallashtirish, ularning modelini ishlab chiqish va fe'llar asosida lingvistik ta'min yaratishga bag'ishlangan talay ishlar dunyo yuzini ko'rdi. So'nggi yillarda M.X.Xakimov rahbarligida “Tarjimon LMX” sistemasi uchun bajarilgan tadqiqot ishlarida o'zbek va rus tillarining leksik-sintaktik tahlili berilgan. Sh.Xamroyeva, N.Abdurahmonova va D.O'rinboyevalar izlanishlarida matnlarni tahlil qilish bo'yicha muayyan fikrlar berilgan, bu jihat lingvistik dasturlarni yaratishda lingvistik modullarni ishlab chiqishning muhim ahamiyatga egaligini belgilab bergan Jahon tilshunosligida XX asrning 50-yillaridan til va matn muammolari bilan bog'liq masalalar kompyuter texnologiyalari yordamida hal qilinmoqda. [Abjalova, 2020].

Kompyuter yordamida matnni tahrirlash bosqichlariga quyidagilar kiradi: **fonetik, grafematik, morfologik, sintaktik, semantik**. Bulardan matnni tahrirlashning morfologik bosqichida matn me'yorlashtiriladi. U

- *Tokenizatsiya* (matndagi so'z shakllarini aniqlaydi);
- *Lemmatizatsiya* (so'zlarning lug'atdagi ifodasini aniqlaydi);
- *Stemmizatsiya* (derivativ so'zlarning o'zagini aniqlash) kabi bosqichlarni o'z ichiga oladi.

Morfologik bosqich sintaktik tahlil bosqichi uchun juda muhim sanaladi.

Avtomatik tahrir va tahlil jarayonini quyidagicha tavsiflash mumkin:

Avtomatik tahrir va tahlil modullari	Texnologiya nomi	Dasturiy ta'minot nomi
Orfografik tahrir	Spelling	Spell checker
Grafematik tahrir (tahlil)	Tokenizatsiya/ segmentlash	Tokenizator
Morfologik tahlil	Stemming Lemmatizatsiya	Stemmer morfoanalizator
Sintaktik tahlil	Parsing	Parser

Matnlarni tahrirlash inson tomonidan (***muharirlik ishi***) yoki kompyuter vositasida (***avtomatik tahrir***) amalga oshirilishi mumkin. Shuni ta'kidlash lozimki, har ikki holatda ham mutaxassisning faoliyati (qobiliyat, temperament, xarakter, irodaviy sifatlar, ijtimoiy holat) muhim ahamiyatga ega. Muharir matnlarni bevosita tahrir va tahlil qiladi, avtomatik tahlilda esa tilshunos bilvosita qatnashadi, ya'ni dasturning lingvistik bazasini yaratib beradi. [Abjalova, Toshimov, 2011].

Avtomatik tahrirning asosiy uch vazifasi:

* *Orfografiyani* tekshirish – bu vazifa morfologik tahlil yordamida amalga oshiriladi, bunda etalon lug'atdagi baza asosida tekshiriladi.

* *Sintaksisni* tekshirish orfografiyani tekshirishdan ko'ra murakkabroq. Chunki orfografiyani tekshiruvchi dasturlar yaratilgan, lekin norasmiy matnlarning sintaktik jihatdan avtomatik tahriri masalasi haligacha hal etilgani yo'q. Rasmiy matnlarning avtomatik tahriri bo'yicha muhim natijalarga erishilgan.

* *Semantik* tahrirlash inson tafakkurini modellashtirishga asoslangan ***sun'iy intellekt*** tizimi bilan bog'liq ravishda o'rganilmoqda. (Sun'iy intellekt muayyan vazifalarni kompyuter orqali boshqariladigan algoritmdir.)

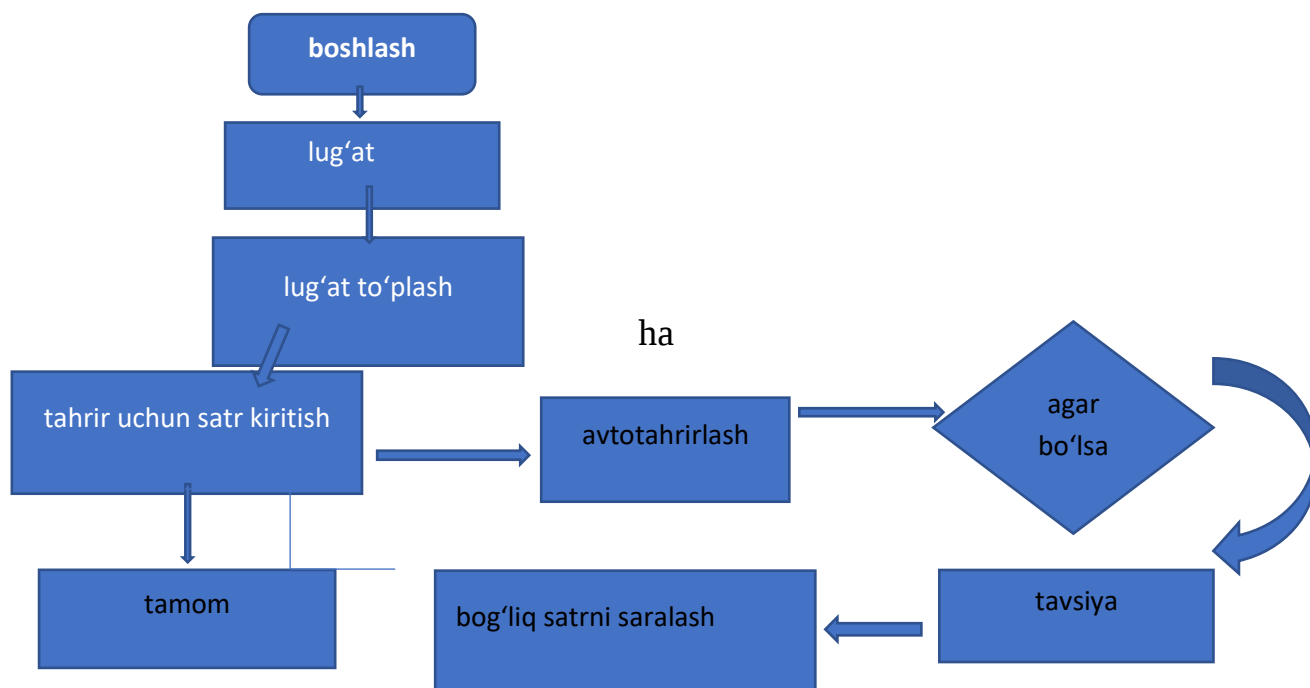
Ixtiyoriy tahrirlovchi dasturning algoritmi quyidagi bosqichlardan iborat bo'ladi:

1. Kompyuter matn tarkibidagi so'zlarni o'qib oladi;
2. Har bir so'z lug'atdagi imlosi bilan solishtiriladi. Leksik va grammatik atributlariga ko'ra so'zlarning lug'ati tuziladi;
3. Ajratilgan segment birliklar morfologik va sintaktik me'yorlarga ko'ra tekshiriladi.

Tahrirlovchi dasturlarning onlayn tizimlari ham yaratilgan, ularning imkoniyatlari bir-biridan farq qiladi, matnni har jihatdan tekshirishga qaratilgan. Hozirda ularning: *Grammarly, Reverso, Ginger software, white Smoke, language Tool, Paper Rater, Pro Writing Aid, Online Correction.com, Hemingway Editor, Spell Check plus, Grammar Slammer, Virtual Writing Tutor, Microsost Word, Google Docs, Slick Write, Grammar Check, WordPerfect Office X8, SentenceChecker.org, After the Deadline, Textly.ai* kabi turlari mavjud.

Matnlarni tahrirlovchi dasturlar matn tarkibida tushirib qoldirilgan yoki ortiqcha qo'llanilgan so'z yoki iborani aniqlash, imloviy va uslubiy xato qo'llanilgan so'zlarni aniqlash, matnning sintaktik strukturasi xatolarni aniqlash, grammatik xato qo'llanilgan shakllarni hamda so'zlarning sinonimlarini taqdim etish orqali semantik imkoniyatini oshirish singari bir qator funksional imkoniyatlarga ega bo'ladi.

Tahrirlovchi dastur algoritmining blok-sxemasini quyidagicha ifodalash mumkin:



Lingvistik tahrirlash (korrektura) – turli ko‘rinishdagi (ilmiy, badiiy, publitsistik va rasmiy uslublardagi) matnlarning orfografik, grammatik, stilistik hamda mantiqiy qurilishdagi xatolarni to‘g‘rilash demakdir.

MC Word dasturida orfografiyani tekshirish imkoniyati cheklanganligi bois hozirgi kunda maxsus **SpellChecker** dasturlar yaratilmoqda. Bunday dasturlar sirasiga quyidagilarni kiritish mumkin: *SpellLink Multilingual Pro* (ingliz, ispan, italyan, nemis, fransuz, portugal, ukrain tillaridagi so‘zlarning to‘g‘ri yozilishini tekshiradi. (*SpellLink -Enlis’h*, *SpellLink French*, *SpellLink Italian*, *SpellLink Spanish*, *Ruta 1.0* rus va ukrain tillaridagi so‘zlarning imlosini tekshirishga yordam beradi), *Asutyp*, **Spellex**, **Spell atche** kabi. Mazkur dasturlar Word dasturiga qo‘shimcha tarzda o‘rnatiladi.

Imloni tekshirish tizimi (spell-cheker ingl. spell checker) – kompyuter dasturi bo‘lib, kiritilgan matnning orfografik tahririni amalga oshiradi. Aniqlangan imlo xatolari maxsus belgilanadi, ya‘ni xato yozilgan leksema tagiga chiziladi. Ko‘p hollarda matn teruvchiga imloviy xatolarga ishora qilishdan tashqari dastur maxsus eslatmasi sifatida so‘zning to‘g‘ri yozilish variantlarini ham taklif qiladi. Shuningdek, matnga qanday tuzatish kiritish mumkinligiga izohlar ham beriladi.

Matnlarni **orfografik tahrirlovchi dastur** hozirda faqat offis dasturlari uchun emas, balki matn bilan bog‘liq boshqa tizimlarda ham keng qo‘llanilmoqda.

Spellcheckerlar odatda quyidagi vazifalarni bajaradi:

Imlo tekshirish: Spellcheckerlar matndagi yozuvning imlo xatolarini aniqlaydi. Bu xatolar o‘rtacha yozuvchilar tomonidan qilinadigan xatolar, harf tartibidagi xatolar, nomlarda yoki kichik harflarda xatolar, yozuv imlo qoidalari va boshqalar bo‘lishi mumkin.

Tavsiyalar berish: Spellcheckerlar imlo xatolarini aniqlagan holda, foydalanuvchiga to‘g‘ri variantlar yoki o‘zgartirishlar taklif etadi. Bu, to‘g‘ri

yozuvni tanlashga yordam beradi va yozuvchilarning yozuvlarini to'g'ri qilishga yordam beradi.

So'zlarni saqlash: Spellcheckerlar odatda o'z ichiga bir nechta so'zlar ketma-ketligini saqlaydi. Bu so'zlarning to'g'ri yozilganligini tekshirishda va foydalanuvchining foydalanishini qo'llab-quvvatlashda yordam beradi.

Foydalanuvchi lug'atini yaratish: Spellcheckerlar foydalanuvchilar o'zlarining so'zlarini qo'shishlari uchun imkon beruvchi o'zlarining lug'atlarini yaratish imkonini beradi. Bu, foydalanuvchi maxsus so'zlari, nomlarni yoki yuridik terminlarni qo'shish imkoniyatini beradi.

Spellcheckerlar uchun qo'llanadigan texnologiyalar va algoritmlar tez-tez rivojlanib borayotgan so'zlar, lug'atlar, imlo qoidalari va texnologiyalarga asoslanadi. Bu, yozuvchilarga yuqori sifatli imlo va yozuvlarni ta'minlashda yordam beradi.

Bugungi kunda jahon bo'ylab spellcheckerlar yuqori darajada ommalashtirilgan vositalar hisoblanadi. Bu vositalar matnlar ustida imlo xatolarini aniqlash va to'g'rilashda yordam beradi. Imlo tekshiruvchilari va spellcheckerlar yozuvning sifatini oshirishga yordam beradigan muhim vositalardir. Ular matnlar ustida yozuv xatolarini aniqlashda va to'g'rilashda yordam beradi va yozuvning o'qilishi va tushunishini osonlashtiradi. Imlo qoidalari va grammatika qoidalari spellcheckerlar va yozuvchilar uchun juda muhim vositalardir. Ular matndagi imlo xatolarini tez fursatda to'g'rilashda ko'maklashadi, shuningdek, yozuvchilar uchun to'g'ri yozuvni ta'minlashda ham yordam beradi.

Matnlarning leksik, morfologik va sintaktik sathlarda tahrir va tahlilini amalga oshiruvchi spell-cheker, orfografik tekshiruv (Speller), orfokorrektorlar, morfoanalizator, parser kabi sistemalarga tildagi matnlarni tahrir va tahlil qilishda asosiy dasturiy vosita sifatida tayanish ish samaradorligini oshiradi.

Hozirgi kunda, spellcheckerlar ko'p platformalarda va dasturlarda o'rtacha va mustahkam bir qismini tashkil etib, matndagi so'zlarning qo'llanishlari hamda imlo xatolarini aniqlash va to'g'rilash usullari eng yaxshi natijalarni olish uchun yanada yaxshilanmoqda.

Foydalanilgan adabiyotlar:

1. Абжалова М.А. Ўзбек тилидаги матнларни компьютер воситасида таҳрир қилиш дастурининг лингвистик таъмини // Магистр илмий даражасини олиш учун ёзилган диссертация// -Тошкент, 2011.
2. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari /monografiya / Toshkent: Nodirabegim, 2020. – 176 b.
3. Хамроева Шахло Ўзбек тили морфологик анализаторининг лингвистик таъминоти: филол.фан. доктори (DSc)...дис. автореф. – Фаргона, 2021.– 76 б.
4. Rahimov A. Kompyuter lingvistikasi asoslari. -Toshkent, 2011.
5. Po'latov A., Muhamedova S. Kompyuter lingvistikasi.-Toshkent, 2007



6. https://en.wikipedia.org/wiki/Spell_checker
7. https://en.wikipedia.org/wiki/Spell_checker. Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980.
8. <http://web.stanford.edu/~learnest/les/> Earnest, Les. Visible Legacies for Y3K.
9. <http://en.wikipedia.org>. Computational linguistics.
10. <https://financesonline.com/top-20-grammar-checker-software-solutions/>