

O‘ZBEK TILIDA TAHRIRLOVCHI DASTURLAR YARATISHNING UMUMIY TAVSIFI

Nazira G‘anijon qizi Sobirova

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

nazirasobirova9@gmail.com

ANNOTATSIYA

Kompyuter lingvistikasi fanining maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini yaratishdir. Kompyuter lingvistikasining asosiy vazifalari esa tillarda o‘qitish, bilimlarni tekshirish, matnlarni tahrirlash va mashina tarjima dasturlarini ishlab chiqishdan iboratdir.

Bugungi kunda o‘zbek tilshunosligining kompyuter bilan bog‘liq holda hal etilishi lozim bo‘lgan vazifalari, xususan, o‘zbek tilining kompyuter uslubini yaratish, axborot matnlaridagi qoliplilik, qisqalik standartlarini ishlab chiqish, internet saytlarini yaratish me‘yorlarini belgilash, kompyuter izohli va tarjima lug‘atlarini yaratish, o‘zbek tili va adabiyoti darsliklarining elektron versiyasini ishlab chiqish, kompyuterda tarjima dasturlarini ishlab chiqish, kompyuterdagi matnlarni avtomatik tahrirlash dasturlarini yaratish kabi masalalar dolzarblik kasb etadi. Ushbu maqolada avtomatik tahrir va tahlil jarayonlari muhokama qilingan.

Kalit so‘zlar: avtomatik tahrir, orfografiya, lingvistik ta‘minot, statk tahlil, dinamik tahlil, imlo qoidalari, tahlil metodlari, grafematik tahrir, morfologik tahrir, spellchecker.

ABSTRACT

The goal of computational linguistics is to create computer programs for solving linguistic problems. The main tasks of computer linguistics are teaching languages, testing knowledge, editing texts and developing machine translation programs.

Today, tasks of Uzbek linguistics that need to be solved in connection with the computer, in particular, the creation of a computer style of the Uzbek language, the development of standards for formality and brevity in information texts, the establishment of standards for creating Internet sites, computer annotated and Issues such as creating translation dictionaries, developing electronic versions of Uzbek language and literature textbooks, developing computer translation programs, and creating automatic text editing programs on computers are gaining relevance. This article discusses automatic editing and analysis processes.

Keywords: automatic editing, orthography, linguistic support, statistical analysis, dynamic analysis, spelling rules, analysis methods, graphematic editing, morphological editing, spellchecker.



АННОТАЦИЯ

Цель компьютерной лингвистики — создание компьютерных программ для решения лингвистических задач. Основными задачами компьютерной лингвистики являются обучение языкам, проверка знаний, редактирование текстов и разработка программ машинного перевода.

Сегодня задачи узбекского языкознания, которые необходимо решить в связи с компьютером, в частности, создание компьютерного стиля узбекского языка, разработка стандартов формальности и краткости информационных текстов, установление стандартов создания Интернета. сайты, компьютерные аннотации и такие вопросы, как создание переводческих словарей, разработка электронных версий учебников узбекского языка и литературы, разработка программ компьютерного перевода, создание программ автоматического редактирования текста на компьютерах. В этой статье обсуждаются процессы автоматического редактирования и анализа.

Ключевые слова: автоматическое редактирование, орфография, лингвистическое обеспечение, статистический анализ, динамический анализ, правила правописания, методы анализа, графематическое редактирование, морфологическое редактирование, проверка орфографии.

KIRISH

Kompyuter lingvistikasi tilning kompyuter nuqtayi nazaridagi ilmiy talqinidir. Texnologiya asrida taraqqiyot sur'atining tezlashgani, intellektual salohiyat, texnikaning yuqori darajada yuksalishi kishilik jamiyati oldiga qator vazifalarni qo'ymoqda. Mintaqamizdagi globallashuv jarayoni barcha sohalarda tezkorlikni, jadal taraqqiyotni talab etmoqda. Kompyuter tizimi qulayliklar, imkoniyatlar majmuyiga va tarkibiy qismiga aylandi.[Rahimov, 2011]

Avtomatik tahrir - kompyuter lingvistikasining asosiy yo'nalishlaridan biri bo'lib, u kompyuterdagi matn muharrirlari (Microsoft Word, Exsel, Wordpad, Lexicon) rivojlanishi bilan bog'liqdir. Avtomatik tahrir deganda, kompyuterga kiritilayotgan matnning mexanik xatolari avtomatik tarzda to'g'rilanishi va xato ekanligi haqidagi signallarning foydalanuvchiga taklif etilishi nazarda tutiladi.

ADABIYOTLAR TAHLILI

Xorijda kompyuter lingvistikasi sohasida matnlarni avtomatik tahlil va tahrir qilish masalasi ko'plab olimlar tomonidan o'rganilgan.[Peterson, James, 1980] J.Peterson xatolarni tekshiruvchi va ularni tuzatuvchi dasturiy ta'minot (DT)lar borasida so'z yuritgan. Matnning avtomatik



tahlil jarayonini lingvistik aspektida o'rganish D.Jurafskiy nomi bilan bog'liq.

Avtomatik tahrirlash yo'nalishida avtomatik tahrir qiluvchi dasturlar ishlab chiqilgan, ular ruscha, inglizcha, nemischa, fransuzcha, italyanча matnlarni tahrirlashga yo'naltirilgan. Hatto tatar, boshqird, qozoq, qirg'iz, ozarbayjon kabi turkiy tillarga mansub matnlar uchun morfologik va sintaktik analizatorlar yaratilgan, MS WORD sistemasiga avto-tahrir 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.Nojov, 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.

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

XX asrning 50-yillaridan til va matn muammolari bilan bog'liq masalalar kompyuter texnologiyalari yordamida hal qilinmoqda.[Rahimov, 2011, 2020]

TADQIQOT METODOLOGIYASI

Maqolada matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik protsessorini yaratish masalasi muhokama qilinadi. Uning asosiy tarkibiy omillari va qismlari tavsiflanadi. O'zbek tilidagi matnlar tahririga mo'ljallangan kompyuter dasturlarini yaratish mezonlari tahlili qilindi. Jahon tajribasidagi mavzu bo'yicha yondashuvlar o'rganildi.

TAHLILLAR VA NATIJALAR

O'zbek tilshunosligi sohasida XXI asrning boshlarida kompyuterdan foydalanish natijasida yangi, zamonaviy yo'nalish bo'lmish kompyuter lingvistikasi yuzaga keldi. Kompyuter lingvistikasining asosiy o'rganish doirasini aynan ijtimoiy



taraqqiyot asosidagi ehtiyoj - mashina tushuna oladigan va qayta ishlashi mumkin bo'lgan tabiiy til modelini yaratish tashkil etadi.

“Prezidentimiz tomonidan 2020-yil 20-oktabrdagi “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi Farmoni bilan tasdiqlangan “2020-2030-yillarda o‘zbek tilini rivojlantirish va til siyosatini takomillashtirish Konsepsiyasi”da belgilangan davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta‘minlash maqsadida o‘zbek tilini Internet jahon axborot tarmog‘ida ommalashtirish, unda munosib o‘rin egallashini ta‘minlash, dasturiy mahsulotlarning o‘zbekcha ilovalarini yaratish, o‘zbek tilini o‘rgatuvchi kompyuter dasturlarini keng miqyosda amaliyotga tatbiq qilish hamda o‘zbek tilidagi matnlar tahririga mo‘ljallangan kompyuter dasturlarini yaratish kabi vazifalarining ko‘rsatib berilishi bugungi kun tilshunos-dasturchi mutaxassislariga katta mas‘uliyat yukladi. Shuning uchun ham o‘zbek 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 ta‘minotini yaratish singari masalalar asosiy bajarilishi shart bo‘lgan vazifalar qatoridan o‘rin egallagan.

Avtomatik tahrir matndagi texnik xatolarning avtomatik tarzda to‘g‘rilanishi va xato ekanligi haqidagi signallarning userga taklif etilishi bilan bog‘liq. Microsoft Word dasturida so‘zlarning yozilishi bilan bog‘liq xatolar bo‘lsa, ularning ostiga qizil yoki ko‘k rangda chizilishi, katta harf bilan yozilishi kerak bo‘lgan so‘zlarning avtomatik tarzda bosh harfga o‘tkazilishi, hujjatga avtomatik numeratsiya qo‘yilishi kabilar avtomatik tahrir erishgan yutuqlar hisoblanadi.

Matnlarni tahrir qilish ikki usulda amalga oshiriladi:

1. Inson tomonidan amalga oshiriladigan tahrir (bevosita tahrir – muharrirlik ishi).
2. Zamonaviy texnika vositasida amalga oshiriladigan tahrir (bilvosita, ya‘ni avtomatik tahrir).

Avtomatik tahrirlash yo‘nalishida avtomatik tahrir qiluvchi dasturlar ishlab chiqilgan, ular ruscha, inglizcha, nemischa, fransuzcha, italyancha matnlarni tahrirlashga yo‘naltirilgan. Hatto tatar, boshqird, qozoq, qirg‘iz, ozarbayjon kabi turkiy tillarga mansub matnlar uchun morfologik va sintaktik analizatorlar yaratilgan, MS WORD sistemasiga avto-tahrir paketlari joylashtirilgan.[Орехов, Слободян, 2010] Bunday tahrirlash dasturlari, asosan, imloni tekshirishga mo‘ljallangan.

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 soʻzlarning toʻgʻri yozilishini tekshiradi. _ *SpellLink -Enlisʻh*, *SpellLink French*, *SpellLink Italian*. *SpellLink Spanish*. *Ruta 1.0* rus va ukrain tillaridagi soʻzlarning imlosini tekshirishga yordam beradi), *Asutyp*. *Spellex*, *Spell atche* kabi. Mazkur dasturlar Word dasturiga qoʻshimcha tarzda oʻrnatiladi.

Imloni tekshirish tizimi (spell-cheker ingl. spell checker) – kompyuter dasturi boʻlib, kiritilgan matnning orfografik tahririni amalga oshiradi. Aniqlangan imlo xatolari maxsus belgilanadi, yaʼni xato yozilgan leksema tagiga chiziladi. Koʻp hollarda matn teruvchiga imloviy xatolarga ishora qilishdan tashqari dastur maxsus eslatmasi sifatida soʻzning toʻgʻri yozilish variantlarini ham taklif qiladi. Shuningdek, matnga qanday tuzatish kiritish mumkinligiga izohlar ham beriladi.

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

Kompyuter dasturi ikkita taʼminot asosida ishlaydi:

- 1) texnik taʼminot
- 2) lingvistik taʼminot (LT).

Lingvistik taʼminot qanchalik aniq, batafsil yaratilgan boʻlsa, dasturning mukammalligi shunchalik chuqur boʻladi. Lingvistik dasturlarning LTlarini yaratishda adabiy til meʼyorlari asos vazifasini bajaradi.

Oʻzbek tilida ham imlo tekshirish dasturlari yaratish bugun uchun muhimlik kasb etadi. Avtomatik tahrir faqat matnlardagi mexanik xatolarni bartaraf etibgina qolmay, matnning savodli yozilishini ham taʼminlaydi, matn kirituvchiga jumlada xato soʻz borligi haqida xabar berib, xatoni bartaraf etishga undaydi. Avvalo, bunday dasturlarni yaratishda biz LTni mukammal etib yaratishimiz shart.

Taʼminot asosini lingvistik qoidalar va meʼyorlar tashkil etadi. Shuningdek, LTdan muayyan tilning lingvistik va filologik lugʻat boyligi ham oʻrin oladi. Shu bois lingvistik protsessorni yaratishda qayta ishlanayotgan tabiiy til haqida yetarli darajada maʼlumotga ega boʻlish talab qilinadi. Bu oʻz navbatida lingvistik protsessorning mukammal boʻlishini taʼminlaydi. Lingvistik maʼlumotlar tabiiy tilning adabiy meʼyorlaridir.

Lingvistik taʼminot tarkibi toʻldirib borilishi 2 xil:

1. dasturini turgʻun (statik)
2. dinamik (oʻzgaruvchan)



Statik tahlil dasturi L_{Ti} faqat mutaxassis tomonidan yaratiladi, uning neologizmlar, yangi lingvistik modellar, til me'yorlari bilan bog'liq o'zgarishlar dasturning keyingi versiyalariga kiritib boriladi. Ma'lum bo'lganidek, L_Tga foydalanuvchi tomonidan to'g'ridan-to'g'ri so'zshakllarni kiritish imkoniyati bo'lmaydi, L_Tga foydalanuvchi tomonidan to'g'ridan-to'g'ri so'zshakllarni kiritish imkoniyati bo'lmaydi.

Dinamik muharrir dasturining vazifasi doimiy ravishda *leksikografik* ta'minotni yangi so'zlar bilan to'ldirib, boyitib borish hisoblanadi. Ko'pincha ushbu vazifa foydalanuvchi tomonidan amalga oshiriladi, ya'ni matnda qo'llanilayotgan muayyan neologizm, yangi termin maxsus so'rovlarga binoan ta'minotning zarur moduliga kiritilishi natijasida keyingi o'rinlarda uning turli shakllarda qo'llanilganda tahrir va tahlil qilinishiga erishiladi. Bu matn muharrir dasturi rivoji uchun xizmat qiladi, chunki til doimiy harakatda, so'zlar yangi-yangi ma'nolar kasb etishi, neologizm, arxaizm, istorizmlarning paydo bo'lishini kuzatib borish imkonini beradi.

XULOSA

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 strukturasiidagi 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.

O'zbek tilidagi matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik protsessorini yaratishda o'zbek adabiy tilining lingvistik lug'atlari va grammatik qoidalar jamlanmasi zaruriy vositalaridir.

REFERENCES

1. Rahimov A. Kompyuter lingvistikasi asoslari. -Toshkent, 2011.
2. Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980
3. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari /monografiya / Toshkent: Nodirabegim, 2020. – 176 b.
4. Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка [Текст] // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28–31 октября 2010 г.)

