

AVTOMATIK TAHRIR ASOSIY YO‘NALISH SIFATIDA

Nazira G‘anijon qizi Sobirova

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

nazirasobirova9@gmail.com

ANNOTATSIYA

Dunyo tilshunosligida muayyan tillarda yaratilgan og‘zaki va yozma manbalarni raqamlashtirish hamda ulardan tabiiy tillarni qayta ishlashda foydalanish muhim masalalardan biriga aylandi. Kompyuter lingvistikasiga doir qator izlanishlar olib borilmoqda, jumladan, mashina tarjimasini, 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. Zamonaviy texnologiyalarning kundalik turmushda muhim o‘rin egallashi, ilm-fanning rivojlanishi avtomatik tahrir va tahlilning ahamiyatini oshirmoqda.

Kalit so‘zlar: avtomatik tahrir, orfografiya, grafematik tahrir, morfologik tahrir, tahlil metodlari, semantik tahrir, lug‘at

ABSTRACT

Digitization of oral and written resources created in certain languages and their use in processing natural languages has become one of the important issues in world linguistics. A number of researches are being conducted in computer linguistics, including machine translation, corpus technology, computer lexicography, and the development of linguistic foundations of software applications, lexicographic text processing, and linguostatistical analysis. The importance of modern technologies in everyday life, the development of science increases the importance of automatic editing and analysis.

Keywords: automatic editing, orthography, graphematic editing, morphological editing, analysis methods, semantic editing, dictionary

АННОТАЦИЯ

Оцифровка устных и письменных ресурсов, созданных на отдельных языках, и использование их при обработке естественных языков стала одной из важных проблем мировой лингвистики. Ведется ряд исследований в области



компьютерной лингвистики, включая машинный перевод, корпусные технологии, компьютерную лексикографию, разработку лингвистических основ программных приложений, лексикографическую обработку текста и лингвостатистический анализ. Значение современных технологий в повседневной жизни, развитие науки повышают значимость автоматического редактирования и анализа.

Ключевые слова: автоматическое редактирование, орфография, графематическое редактирование, морфологическое редактирование, методы анализа, семантическое редактирование, словарь

Kompyuter lingvistikasida avtomatik tahrir yo'nalishi matnlarni takomillashtirish uchun keng qo'llanilmoqda. Uning asosiy yo'nalishlari matn xatolarini tuzatish, matn uslubini yaxshilash va matn tuzilishini yaxshilashdan iborat. Ushbu yo'nalishda turli xil statistik, lingvistik va mashinali o'rganish usullaridan foydalanilmoqda. Avtomatik tahrir tizimlarining samaradorligi doimiy ravishda oshib bormoqda va ular kundalik hayotimizda keng qo'llanilmoqda. Avtomatik tahrir yo'nalishi XX asr o'rtalariga kelib rivojlandi. Ma'lumki, kompyuter yaratilmasdan avval matnlar yozuv mashinkalarida terilgan, bu jarayonda orfografik, punktuatsion, stilistik, grammatik xatolarga yo'l qo'yilgan bo'lsa, matnni boshdan oxirigacha qayta yozib chiqishga to'g'ri kelgan, bu esa foydalanuvchiga qiyinchilik tug'dirgan. Kompyuterda matnni tahrirlash birmuncha qulay bo'lib, unda foydalanuvchi dialogli rejimda ishlaydi va o'z o'rnida xatolarni to'g'rilash imkoniyatiga ega bo'ladi. Shu o'rinda alohida ta'kidlash zarurki, avtomatik tahrirlash tizimi faqat matn muharrirlari dasturlari bazasiga qo'shilgan ingliz, rus, nemis, fransuz kabi tillarda amal qiladi. Agar bazada mavjud bo'lmagan tilda matn kiritilsa, uni kompyuter avtomatik tahrirlay olmaydi, ayrim internatsional so'zlar bundan mustasno.

Avtomatik tahrirlash jarayonida MS Word lug'atida mavjud bo'lmagan so'z uchrasa, ostiga qizil chiziq bilan ajratib ko'rsatiladi. Mazkur holatda dastur so'z imlosi haqida variant taklif eta olmaydi, chunki Ms Word lug'atida bu so'z mavjud emas. Ehtimol u so'z tog'ri yozilgan bo'lishi mumkin, bunda insonning ishtiroki zarur bo'ladi. Agar so'zning imlosi to'g'ri bo'lsa, matnda tez-tez qayta ishlatilsa, uning avtokorreksiyada to'g'ri deb sanalishi uchun Ms Word lug'atiga «kiritib qo'yish» (kontekst menyudagi «add» yoki «добавить» komandasi yordamida) mumkin. Matn kiritish davomida so'z yoki so'zlar birikmasi ko'k chiziq bilan ajratib qo'yiladi, bunda so'zning imlosida xato mavjud bo'ladi. Bu holda foydalanuvchi xatoni o'zi tog'rilab qo'yishi zarur yoki ostiga chizilgan so'zda sichqonchani o'ng tomoni bosilsa, kontekst menyusida so'zning



to'g'ri varianti taklif etiladi va uni tanlashi mumkin.[Xolmanova, 2019]

Kompyuter dasturi ikkita ta'minot asosida ishlaydi:

- 1) texnik ta'minot;
- 2) lingvistik ta'minot (LT).

Lingvistik ta'minot qanchalik aniq, batafsil yaratilgan bo'lsa, dasturning mukammalligi shunchalik chuqur bo'ladi. Lingvistik dasturlarning LTlarini yaratishda adabiy til me'yorlari asos vazifasini bajaradi.

Mashina tahririda, asosan, imloviy xatolarni tekshirish ko'zda tutiladi va asosiy e'tibor so'zshakllarga qaratiladi. Ushbu maqsadga erishish uchun quyidagi ishlar amalga oshirilishi talab qilinadi: 1) tilning lug'at tarkibini kompyuterga kiritish, ya'ni elektron lug'atni tuzish; 2) lug'atdagi adabiy tilga mansub so'zlarni ajratib olish; 3) ajratib olingan so'zlarni uslubiy (ilmiy, badiiy, publitsistik, rasmiy-idoraviy va so'zlashuv) jihatdan guruhlash; 4) adabiy so'zlarni turkumlarga ajratish; 5) so'zlarni turkum doirasida guruhlash (masalan, ot turkumidagi so'zlarni shaxs otlari, o'simlik nom (fitonim)lari, joy nom (toponim)lari kabi guruhlarga ajratish); 6) guruhlangan so'zlarning affikslar kombinatsiyasini tuzish; bunda affikslar kombinatsiyasining amalda adabiy til doirasida foydalaniladigan hajmi qamrab olinadi. Kombinatsiyalar qatorini tuzishda affikslar ketma-ketligiga e'tibor qaratiladi. M-n, ot turkumida ko'plik+egalik +kelishik kategoriyasi va h.k. [Абжалова, 2019]

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 ravish o'rganilmoqda. Bunda mazkur ko'p aspektli muammoni semantik maydon, freym yonda tezauruslar tizimini boyitish, so'z va so'z birikmalari o'rtasidagi paradigmatic munosaba (sinonimiya, gipo-giperonimik munosabat – tur-jins munosabati)ni modellashtirishga tatil etish asosida hal etish mumkin. Chunki inson bunday assosiativ munosabatlarni o'z intellektiga tayanib o'rnatadi, biroq bu jarayon kompyuter uchun murakkabdir.



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.

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. Bunday tahrirlash dasturlari, asosan, imloni tekshirishga mo'ljallangan.

Matnlarni to'liq avtomatik tahrir va tahlil qilish uchun dastur quyidagi modullar asosidagi lingvistik tizimdan tarkib topishi lozim:

***Grafematik tahrir** bosqichida matndagi xatboshi, raqam, punktuatsion va boshqa ramziy belgilar aniqlanadi, ya'ni so'z va so'zshakllarga qadar tekshiruv amalga oshiriladi. Grafematik tahrir bosqichining maqsadi – matndagi eng kichik birliklarni ham aniqlash va tasniflashdan iborat. Bunday birliklarga quyidagilar kiradi: so'z, xatboshi, tinish belgilari, sanalar, pul birliklarining ramzlari, so'z-raqamli birikmalar, sonlar, IP-manzillar va fayl nomlari, telefon raqamlari. Grafematik tahrir qilish dasturining ilk bosqichi hisoblanib, u keyingi bosqichlarning to'laqonli va samarali bo'lishi uchun zamin yaratadi. Morfologik tahrir bosqichi asos va shu asosdan yuzaga kelgan so'zshakllarni tekshiradi.

***Morfologik tahrirning** maqsadi – faqat orfografik xatolarni tekshirishdan iborat, deb hisoblash uning vazifa ko'lamini toraytiradi. Ushbu bosqichda kiritilayotgan so'zshaklning qanday asosiy shakldan, qaysi lingvistik parametrlardan yuzaga kelgani aniqlanadi. Bu keyinchalik bir so'zning barcha shakllari ustida ishlashga emas, balki asosiy shaklning qanday lingvistik parametrlarni olish imkoniyatiga egaligi ustida ish olib BORISHga turtki beradi. Masalan, ot turkumidagi kitob so'zining qanday shakl hosil qiluvchi va sintaktik shakl yasovchi qo'shimchalarni qabul qilishi hamda ularning kombinatsiyasini biriktirib, so'zshakllarni hosil qilish imkoniyati, ya'ni parametrlari dastur bazasiga kiritiladi. Natijada kitob so'zining barcha shakllari kiritilib, dastur ta'minotidan katta hajm egallanmaydi, balki asosiy shakl va unga birikishi me'yoriy hisoblangan lingvistik parametrlar kiritilib, hajm iqtisod qilinadi.

Avtomatik tahrir afzalliklari

Avtomatik tahrirning eng muhim afzalliklari quyidagilardan iborat:



**Tezkorlik va samaradorlik*: Kompyuter tizimlarining katta hajmdagi matnni qisqa muddat ichida tahrir qila olish qobiliyati, matnni ishlov berisp jarayonini sezilarli darajada tezlashtiradi;

**Xatolarni kamaytirish*: Avtomatik tahrir dasturlari nutqiy va grammatik xatolarni aniqlab, tuzatish imkoniyatiga ega. Bu esa matn sifatini yaxshilashga xizmat qiladi;

**Konsistentlik*: Avtomatik tahrir tizimlarida qo'llaniladigan qoidalar va standartlar doimo bir xil bo'lgani uchun, matnlardagi uslubiy va shakliy yaxlitlikni ta'minlaydi;

**Masshtablanish*: Avtomatik tahrir jarayonini katta hajmdagi matnlarga qo'llash mumkin. Bu esa nashriyot tashkilotlari uchun katta qiymatga ega;

**Xarajatlarni kamaytirish*: Avtomatik tahrir jarayoni insonlar tomonidan bajarilayotgan tahrirchilik xizmatlariga qaraganda arzonroq.

Avtomatik tahrir tizimlarining asosiy cheklovi ular tomonidan yuksak darajadagi ijodiy va mantiqiy tahlil talab qilinuvchi vazifalarni bajara olmasliklaridir. Shuningdek, avtomatik tahrir tizimlari kontekstga bog'liq bo'lgan xatolarni aniqlay olmaydi yoki matnning maqsadiga mos kelmaydigan o'zgarishlarga yo'l qo'yishi mumkin. Bu kabi cheklovlar tufayli avtomatik tahrir asosiy yo'nalish bo'lishiga qaramasdan, insonlar tomonidan amalga oshiriladigan qo'shimcha tahrir jarayonlariga muhtoj bo'ladi.

O'zbek tilshunosligida ham avtomatik tahrir yo'nalishida izlanishlar olib borilgan. Professor H.Arziqulov rahbarligida Samarqand davlat chet tillari institutida injener tilshunosligi markazi tashkil etilgan. Bu markazda davlat tilidagi matnlarga kompyuter yordamida qayta ishlov berish masalalari ustida ilmiy izlanishlar olib borilgan. Bunda matnlarga qayta ishlov beradigan kompyuter dasturlarini yaratish, o'zbek tilining barcha sathlariga oid matnlarni yig'ish, ularni muayyan bir tizimga keltirish, algoritmlarni tuzish, ularga statistik jihatdan ishlov berish, inglizcha-o'zbekcha mashina tarjimasini yaratish kabi sohalarda ilmiy-tadqiqot ishlari amalga oshirilgan. O'zbekcha matnlarga qayta ishlov beruvchi avtomatik tizim "Uzlington", lotin alifbosidagi matnlarni kirill alifbosiga va krill alifbosidan lotin alifbosiga o'tkazuvchi "Spellchecker", kompyuter ko'magida korrektura qiladigan, imloviy xatolarni tuzatadigan kompyuter dasturlarini yaratishga birinchilardan bo'lib kirishilgan.[Rahimov, 2011]

O'zbek tilidagi matnlarni avtomatik ravishda tahrir va tahlil qilish dasturini yaratish bugungi kunning dolzarb masalasiga aylangan,[Abjalova, 2015] Bunday dastur tizimining katta imkoniyat va yuqori sifati lingvistik modullarning mukammal ishlab chiqilgani bilan belgilanadi, ular o'z navbatida dasturning bosqichlarini tashkil etadi.



XULOSA

Texnologik rivojlanish va mashinali o'rganish usullarining takomillashishi bilan avtomatik tahrir doimiy ravishda takomillashib, insonlar tomonidan amalga oshiriladigan tahrir jarayonini to'ldirib, yengillashtirmoqda.

Avtomatik tahrirning pragmatik ahamiyati shundaki, u kelajakda taraqqiy etsa, yetarli ma'lumotlar bazasi yaratilsa va maxsus dasturlar ishlab chiqilsa, muharrir faoliyati yengillashadi,[Xolmanova, 2019] xususan, o'zbek tili asosida ko'p bosqichli avtomatik tahrir dasturining lingvistik ta'minotini mukammal darajada ishlab chiqish bugungi kun tilshunos-dasturchilari oldidagi asosiy masaladir.

REFERENCES

1. Xolmanova Z. kompyuter lingvistikasi. –T., 2019.
2. Абжалова М. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (расмий ва илмий услубдаги матнлар таҳрири дастури учун). Филол. фан...(PhD) дисс.-Тошкент, 2019.-Б.15.
3. Rahimov A. Kompyuter lingvistikasi asoslari. –T., 2011, 65-bet
4. Abjalova M.A. Matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik bosqichlari//Yosh olim va talabalarning respublika ilmiy-amaliy anjumani materiallari. –T., 2015, 7-bet
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.

