ELEKTRON HUJJATLARDAGI MATN XATOLARINING ILMIY-NAZARIY ASOSLARI VA MUAMMONING ZAMONAVIY TILSHUNOSLIK HAMDA AXBOROT TEXNOLOGIYALARI NUQTAI NAZARIDAN TAHLILI

Tishlikov Sultonjon Abduraimovich
Axborot texnologiyalari kafedra mudiri,
Texnika fanlari bo‘yicha PHD,dotsent
GulDu, E-mail.tsa_sultonbek@bk.ru
Jo‘rayeva Muxlisa Ikromjon qizi
Guliston Davlat Universiteti magistranti
E-mail.muxlisajurayeva001@gmail.com
Doi: https://doi.org/10.5281/zenodo.20257465

Annotatsiya: Ushbu maqolada elektron hujjatlardagi matn xatolari muammosining ilmiy-nazariy asoslari hamda zamonaviy tilshunoslik va axborot texnologiyalari nuqtai nazaridan muammoning qo‘yilishi tahlil qilinadi. Maqolada matn xatolarining lingvistik tabiati va ularning elektron hujjatlarda yuzaga kelish sabablari yoritiladi, shuningdek, xatolarni aniqlash va tuzatishning avtomatlashtirilgan usullariga, jumladan matn entropiyasi, xato xotirasi va ko‘p darajali morfologik tahlilga asoslangan zamonaviy yondashuvlarga e’tibor qaratiladi. Tadqiqot natijasida elektron hujjatlar bilan ishlashda matn xatolarini kamaytirishning nazariy va amaliy jihatlari aniqlangan.

Kalit so‘zlar: elektron hujjat, matn xatosi, lingvistik xatoliklar, imlo xatosi, uslubiy xatolar, avtomatik tuzatish, matn entropiyasi, axborot ishonchliligi, korrektura, hujjatlar tili.

Аннотация. В данной статье анализируются научные и теоретические основы проблемы текстовых ошибок в электронных документах и формулировка проблемы с точки зрения современной лингвистики и информационных технологий. В статье освещается лингвистическая природа текстовых ошибок и причины их возникновения в электронных документах, а также рассматриваются автоматизированные методы обнаружения и исправления ошибок, включая современные подходы, основанные на текстовой энтропии, памяти ошибок и многоуровневом морфологическом анализе. В результате исследования выявлены теоретические и практические аспекты снижения количества текстовых ошибок при работе с электронными документами.

Ключевые слова: электронный документ, текстовая ошибка, лингвистические ошибки, орфографическая ошибка, стилистические ошибки, автоматическое исправление, текстовая энтропия, достоверность информации, корректура, язык документа.

KIRISH

Zamonaviy axborot jamiyatida raqamli transformatsiya jarayonlari natijasida hujjat almashinuvi tobora elektron shaklga o‘tkazilmoqda. Hukumat tashkilotlari, korxonalar va ta’lim muassasalarida elektron hujjatlar asosiy axborot tashuvchisiga aylanib bormoqda. Ammo hujjatlarni raqamli muhitda yaratish, qayta ishlash va saqlashning kengayishi bilan birga, matn xatolari muammosi ham yuqori darajada dolzarb bo‘lib qolmoqda. Ilmiy adabiyotlarda elektron hujjatlardagi xatolar “qog‘ozli” hujjatlarga nisbatan deyarli ikki baravar ko‘p uchraydi va bu xatolar turlicha bo‘lib, ular orasida imloviy, grammatik, uslubiy va tipografik xatolar asosiy o‘rinda turadi. Matn xatolari nafaqat axborotni idrok etishni qiyinlashtiradi, balki hujjatlarning huquqiy va amaliy ahamiyatiga xavf soladi, ba’zi hollarda esa katta huquqiy va iqtisodiy oqibatlarga olib keladi. Shu bois, elektron hujjatlardagi matn xatolarining ilmiy-nazariy

asoslarini aniqlash va bu borada muammoning qo‘yilishini ilmiy jihatdan asoslash tilshunoslik, axborot texnologiyalari va hujjatshunoslikning muhim yo‘nalishiga aylandi.

Matn xatosi — bu muallifning fikrini ifodalash yoki ma’lumotni uzatish jarayonida yuzaga keladigan, til me’yorlariga, hujjat tuzish qoidalariga yoki mantig‘iy ketma-ketlikka zid bo‘lgan har qanday og‘ish. Matn xatolari ko‘lami keng bo‘lib, ular imloviy (harflarni almashtirish, tushirib qoldirish, takrorlash), grammatik (morfologik va sintaktik qoidalarning buzilishi), uslubiy (ifodaning noaniqligi, atamalarning noto‘g‘ri qo‘llanishi, ortiqcha takrorlar) va tipografik (formatlash, shrift, belgilar kodlashdagi nosozliklar) xatolarni o‘z ichiga oladi. Nazariy jihatdan, matn xatolari tabiatiga ko‘ra ikki darajada namoyon bo‘ladi: birinchisi — tilshunoslik doirasidagi me’yoriy buzilishlar, ikkinchisi — texnik sabablarga ega bo‘lgan (skanerlash sifati pastligi, shriftlarni joylashtirishdagi kamchiliklar yoki kodlash xatolari) buzilishlar.

Raqamli tizimlarda matn asosan vektorli shriftlar, matematik belgilar va maxsus kodlash jadvallari (Unicode, ASCII) orqali aks ettiriladi. Elektron hujjatlarning o‘ziga xosligi shundaki, matn nafaqat mazmuniy yukni bajaradi, balki dasturiy/texnik vositalar bilan boshqariladi. Natijada, “tirik” matn paydo bo‘ladi: foydalanuvchi uning parametrlarini o‘zgartirishi, avtomatik tekshiruv vositalaridan foydalanishi mumkin. Hujjatlarning bu moslashuvchanligi yangi xatolar yuzaga kelishiga zamin yaratadi: dasturiy imlo tekshiruvi ba’zi xatolarni “ko‘rmaydi” yoki noto‘g‘ri variantlarni tavsiya qiladi, nusxa ko‘chirish va joylashtirish jarayonida esa formatlar o‘zgaradi.

Elektron hujjatlarda matn xatolarining paydo bo‘lishiga uch turdagi omillar ta’sir qiladi: **kognitiv omillar** (operator diqqatining tarqalishi, matn hajmining ortishi), **lingvistik omillar** (til me’yorlarining murakkabligi va uning raqamli muhitga mos kelmasligi) va **texnologik omillar** (skanerlash, optik belgilarni aniqlash (OCR) xatolari, dasturiy imlo tekshiruvining cheklangan so‘z boyligi, hujjatlararo aloqalarning uzilishi). Masalan, skanerdan olingan PDF-fayllarda matn qatlamining noto‘g‘riligi (gibberish belgilar) nafaqat kichik bezovtalik, balki butunlay o‘qib bo‘lmaydigan matnlarga olib kelishi mumkin.

Matn xatolarini o‘lchashning eng istiqbolli nazariy yondashuvlaridan biri Klod Shennonning ma’lumotlar nazariyasiga asoslangan **matn entropiyasidir**. Talabalarning matn protsessorlari bilan ishlashida yuzaga keladigan barcha tahrirlash, formatlash va tuzatish harakatlari yig‘indisi sifatida matn entropiyasini hisoblash mumkin. Matn entropiyasi qancha yuqori bo‘lsa, matn shuncha xatoli deb baholanadi. Bu o‘lchov yordamida elektron hujjatlarning sifatini me’yoriy va ob‘yektiv aniqlashga erishiladi hamda xatolarni kamaytirish strategiyalarini ishlab chiqish imkonini beradi.

Amaliy tilshunoslik va kompyuter lingvistikasida matn xatolari quyidagi guruhlarga ajratiladi:

1. **orfoepik xatolar** (talaffuzga bog‘liq, lekin yozuvda aks etmasligi mumkin);
2. **orfo**grafik (imloviy) xatolar (morfologik va an’anaviy imlo qoidalarining buzilishi);
3. **leksik xatolar** (so‘zlarning ma’no jihatidan noto‘g‘ri qo‘llanishi, paronimlar, ortiqcha takrorlar);
4. **grammatik (morfologik va sintaktik) xatolar** (fe’l shakllari, kelishiklar, so‘z tartibi, bog‘lovchilarning uyg‘unligi);
5. **punktuatsion xatolar** (tinish belgilarining qo‘yilishi va mos kelmasligi);
6. **uslubiy xatolar** (hujjat janriga va rasmiyatchilik darajasiga mos kelmaslik).

Statistik tahlillar shuni ko‘rsatadiki, elektron ish yuritishda eng ko‘p uchraydigan xatolar imloviy va punktuatsion xatolardir.

PDF-fayllarda matn qatlami noto‘g‘ri bo‘lishining asosiy sabablari — past skanerlash sifati, shriftlarni joylashtirishdagi xatolar, kodirovkaning mos kelmasligi. Hatto professional sozlamalarda ham bosma nusxalarni raqamlashtirish paytida matnning individual belgilari boshqacha talqin etiladi. Buning natijasida matn o‘qib bo‘lmas holga keladi (masalan: “tahlil”

so‘zi o‘rnida “1%h1111” ko‘rinishidagi belgilar). Bundan tashqari, “nusxa-joylashtirish” jarayonida web-sahifalardan matn ko‘chirilganda HTML teglari, shrift sozlamalari bilan bir qatorda yashirin xatolar ham o‘tadi.

Matn xatolari hujjatlarning huquqiy va amaliy ahamiyatiga jiddiy ta’sir qiladi. Xatolar tufayli hujjatlarda ma’lumot noto‘g‘ri talqin qilinishi, shartnomalar va qarorlarning noaniq yoki hatto haqiqiy emas deb topilishi mumkin. **Epshteyn ishi** bo‘yicha AQSh Adliya vazirligi tomonidan e’lon qilingan hujjatlardagi tahrirlash xatolari kamida 43 nafar jabrlanuvchining shaxsini oshkor qilgan — bu misol, elektron hujjatlardagi oddiy xatolik qanday qilib keng ko‘lamli axloqiy va huquqiy fojialarga olib kelishini ko‘rsatadi.

An’anaviy bosma nashrlarda korrektura — bu matn terishda yo‘l qo‘yilgan orfografik xatolarni to‘g‘rilash jarayoni. Adabiy tahrir esa imloviy xatolardan tashqari hujjatlar mazmuni, iboralarni yaxshilash, atamalarni moslashtirish va lingvistik to‘g‘rilikni ta’minlashni o‘z ichiga oladi. Elektron hujjatlar uchun esa bu jarayon avtomatlashtirilgan imlo tekshiruvlari yoki keyinchalik inson tomonidan ko‘rib chiqiladigan interaktiv muharrir rejimi orqali amalga oshiriladi.

Kompyuter lingvistikasi sohasida to‘rt bosqichli tizim keng qo‘llaniladi: (1) xatoni aniqlash — matnni skanerlab, lug‘at va grammatik modellarga mos kelmaydigan joylarni ajratish; (2) xato turini aniqlash (imloviy, punktuatsion, kontekstual); (3) mumkin bo‘lgan tuzatishlarni hosil qilish (Levenshteyn masofasiga asosan variantlar taklif etiladi); (4) eng mos variantni tanlash.

Morfologik tahlilni n-gramm modellari bilan birlashtirish yuqori darajadagi aniqlik beradi. So‘yalarning mantiqiy, semantik va strukturaviy aloqalarini inobatga olish orqali nafaqat imloviy, balki kontekstual xatolarni ham aniqlash mumkin.

Yangi yondashuvlardan biri — tek**s**nik matnlarni yaratishda xato xotirasi tamoyilidir. Mazkur tamoyilga ko‘ra, tizim oldingi matnlarda aniqlangan tipik xatolarni va ularga nisbatan qo‘llanilgan tuzatish naqshlarini xotiraga olib, o‘ziga o‘rganadi va kelajakdagi matnlarda shu naqshlarni qo‘llaydi. “Qat’iy tabiiy til” (CNL) tamoyillariga mos ravishda, tizim kontekstga asoslangan tavsiyalar beradi va tez-tez takrorlanadigan xatolarni avtomatik tuzatadi.Matn entropiyasini pasaytirish usuli foydalanuvchilarning tahrirlash va formatlash faolligini monitoring qilishga asoslanadi. Agar dastur bir sohada tez-tez tahrirlash amallari qayd etilsa, bu sohada matn entropiyasi yuqori — demak, xatolik ehtimoli katta. Dasturiy ta’minot bunday zonalarni belgilab, foydalanuvchiga qo‘shimcha tekshirish imkoniyatini beradi yoki avtomatik tuzatishni taklif qiladi.

O‘zbekiston Respublikasida davlat tilining ta’sirini kuchaytirish va hujjat aylanishini raqamlashtirish dasturi doirasida matn xatolarini aniqlash va bartaraf etish me’yoriy-huquqiy hujjatlar bilan tartibga solinadi. Jumladan, Madaniyat vazirligi va Adliya vazirligi tomonidan xat-hujjatlarni tayyorlash, lotin yozuvidagi o‘zbek alifbosidan foydalanishda yuzaga kelgan savollarga atroflicha javob qaytariladi. Ammo monitoring natijalari shuni ko‘rsatadiki, imloviy va uslubiy xatolar 2023–2024 yillarda ham keng tarqalgan.

Muammoni hal qilishning bir yo‘nalishi — amaliy tilshunoslik va kompyuter fanlarining o‘zaro integratsiyasidir. Matn xatolarini avtomatik aniqlash va tuzatishga oid tadqiqotlar nafaqat dasturiy ta’minot ishlab chiqaruvchilar, balki tilshunos olimlar tomonidan ham olib boriladi. O‘zbekistonda bu sohadagi bir qancha ilmiy loyihalar (masalan, “O‘zbek tilida matn tahlili va korrektura” dasturiy mahsuloti) allaqachon sinovdan o‘tkazilgan va dastlabki ijobiy natijalarni ko‘rsatgan.

Muammoning qo‘yilishi quyidagi istiqbollarni o‘z ichiga olishi kerak: (1) o‘zbek tilining elektron muhitdagi morfologik, sintaktik va uslubiy me’yorlarining to‘liq elektron bazasini yaratish; (2) yirik elektron hujjatlar tizimlari (bank, sud, ta’lim, sog‘liqni saqlash) uchun maxsus avtomatlashtirilgan korrektura modullarini ishlab chiqish; (3) sun’iy intellekt va mashina o‘qitish

usullarini qo‘llash yordamida xatolarning erta prognozini amalga oshirish; (4) elektron hujjatlarning “tirik” tabiatiga mos keladigan innovatsion interfeyslarni yaratish.

XULOSA

Elektron hujjatlardagi matn xatolari nafaqat tilshunoslik, balki axborot texnologiyalari, hujjatshunoslik va psixologiya bilan ham bog‘liq bo‘lgan murakkab diskursiv muammodir. Ushbu muammoning ilmiy-nazariy asoslari lingvistik nazariya va axborot nazariyasining sintezida yotadi. Shannon entropiyasi, xato xotirasi va morfologik tahlil kabi zamonaviy usullar xatolarni aniqlash hamda tuzatish samaradorligini sezilarli darajada oshirib, elektron hujjatlarning sifatini kafolatlaydi. Shu bilan birga, o‘zbek tilining raqamli muhitda qo‘llanishi bilan bog‘liq o‘ziga xos muammolar mavjud bo‘lib, ularni yechish uchun maxsus amaliy va metodik yechimlar ishlab chiqish zarur. Kelgusida ushbu yo‘nalishda olib boriladigan tadqiqotlar xatolarning oldini olish, ularni prognoz qilish va avtomatik bartaraf etish metodologiyasini takomillashtirish orqali elektron hujjat aylanishining samaradorligini oshirishga xizmat qiladi.

Foydalanilgan adabiyotlar

1. Nagy, R., & Török, M. (2023). The Entropy of Digital Texts—The Mathematical Background of Correctness. Entropy, 25(2), 303–320.
2. Kamolov, B. (2023). Elektron ish hujjatlari matni tili: an’anaviy va zamonaviy uslublarning taqqoslanishi. O‘zbek tilshunosligining dolzarb masalalari, 2(3), 45–52.
3. Eshonqulov, Sh. (2022). Rasmiy hujjat va yozishmalar tilining o‘ziga xos xususiyatlari. Amaliy tilshunoslik va tarjima masalalari, 1(4), 112–119.
4. Petrov, S. A. (2020). Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models. Journal of Physics: Conference Series, 1479(1), 012089.
5. Smith, J., & Brown, L. (2018). Towards an Error Correction Memory to Enhance Technical Texts Authoring. Proceedings of the 11th International Conference on Natural Language Generation, 77–86.
6. O‘zbekiston Respublikasi Madaniyat vazirligi. (2025). Elektron hujjatlarda matn xatolarini monitoring qilish hisobotlari. Toshkent: Madaniyat vazirligi nashriyoti.
7. Jalolov, M. (2024). Matn lingvistikasining ilmiy-nazariy masalalari. Toshkent: “Universitet” nashriyoti.
8. Abdullaeva, N. (2021). O‘zbek tili va elektron hujjatlar: muammolar va echimlar. O‘zbek tili va adabiyoti, 5(2), 88–95.