

RAQAMLI LEKSIKOGRAFIYA: ELEKTRON LUG‘ATLAR VA LINGVISTIK MA’LUMOTLAR BAZASINI YARATISH TAMOYILLARI

*Andijon davlat chet tillari instituti,
Roman-german va slavyan tillari fakulteti
(Kompyuter lingvistikasi) yo'nalishi
101-guruh talabasi
Islomova Feruza Vahhobjon qizi
Ilmiy masalahatchi: Xomidova.M*

Kalit so'zlar: raqamli leksikografiya, kompyuter lingvistikasi, elektron lug‘at, til korpusi, lingvistik ma’lumotlar bazasi, NLP, ontologiya.

Ключевые слова: цифровая лексикография, компьютерная лингвистика, электронный словарь, языковой корпус, лингвистическая база данных, NLP, онтология.

Key words: digital lexicography, computational linguistics, electronic dictionary, language corpus, linguistic database, NLP, ontology.

Annotatsiya: Ushbu maqolada an'anaviy leksikografiyaning kompyuter lingvistikasi ta'sirida transformatsiyalashuvi, elektron lug‘atlarning turlari va ularni shakllantirishdagi zamonaviy yondashuvlar tahlil qilinadi.

Annotation: This article analyzes the transformation of traditional lexicography under the influence of computational linguistics, the types of electronic dictionaries, and modern approaches to their formation.

Аннотация: В данной статье анализируются трансформация традиционной лексикографии под влиянием компьютерной лингвистики, виды электронных словарей и современные подходы к их формированию.

Bugungi kunda leksikologiya (so‘z tarkibi va ma’nosini o‘rganuvchi fan) faqatgina qog‘oz nashrlar bilan cheklanib qolmayapti. Kompyuter lingvistikasining rivojlanishi natijasida leksikografiya (lug‘at tuzish san’ati) butunlay yangi bosqichga — raqamli formatga o‘tdi. Endilikda lug‘at nafaqat so‘zlar ro‘yxati, balki murakkab lingvistik ma’lumotlar bazasidir.

Leksikologiya va Kompyuter texnologiyalari aloqasi

Leksikologiya so‘zning semantikasi, etimologiyasi va qo‘llanilish doirasini o‘rganadi. Kompyuter lingvistikasi esa bu ma’lumotlarni algoritmlar yordamida qayta ishlash imkonini beradi.

Korpus lingvistikasi: Leksikografiyada so‘zning haqiqiy jonli tildagi ishlatilishini aniqlash uchun millionlab so‘zlardan iborat matnlar jamlanmasidan (korpuslardan) foydalaniladi.

Ontologiyalar: Soʻzlar orasidagi mantiqiy bogʻliqlik (masalan, "olma" — bu "meva", "meva" — bu "oʻsimlik") kompyuter uchun tushunarli modellarga oʻtkaziladi.

Elektron lugʻatlarning anʻanaviy lugʻatlardan afzalligi

Hajm cheksizligi: Qogʻoz sahifasiga sigʻmaydigan minglab misollarni kiritish mumkin.

Tezkor qidiruv: Ildiz, qoʻshimcha yoki maʼnodosh soʻzlar boʻyicha soniyalar ichida natija olish.

Multimedia: Soʻzning talaffuzi (audio), tasviri (video/foto) va interaktiv elementlarning mavjudligi.

Gipermatnlilik: Bir soʻzning izohidan boshqa bir tushunchaga oʻtish imkoniyati.

Zamonaviy lugʻat tuzish bosqichlari

Maʼlumot yigʻish: *Til korpuslaridan eng koʻp qoʻllaniladigan soʻzlarni saralab olish.*

Leksikografik tahlil: *Soʻz maʼnolarini aniqlash va ularni kompyuter kodiga (masalan, XML formatiga) moslashtirish.*

Dasturiy qobiq yaratish: *Foydalanuvchi uchun qulay interfeys va qidiruv tizimini ishlab chiqish.*

Kompyuter lingvistikasi va leksikografiya hamkorligi tillarni saqlab qolish va ularni raqamli dunyoda ommalashtirishning eng samarali yoʻlidir. Kelajakda sunʼiy intellekt yordamida lugʻatlar real vaqt rejimida yangilanib boradigan "jonli" tizimlarga aylanadi

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. "Oʻzbek tili kompyuter lingvistikasi". – Toshkent, 2021. (Elektron lugʻatlar va korpus masalalari boʻyicha asosiy qoʻllanma).
2. Hamroeva Sh. "Oʻzbek tili morfologik analizatori va elektron lugʻatlar bazasi". (Oʻzbek tili milliy korpusi yaratuvchilaridan biri).
3. Atkins B.T.S., Rundell M. "The Oxford Guide to Practical Lexicography". Oxford University Press. (Lugʻat tuzishning jahon standarti).
4. Granger S., Paquot M. "Electronic Lexicography". Oxford University Press. (Aynan raqamli lugʻatshunoslik boʻyicha asosiy darslik).
5. Jurafsky D., Martin J. "Speech and Language Processing". (NLP va kompyuter lingvistikasining "bible"si hisoblanadi)
6. Uzbek National Corpus (www.uzbekcorpus.uz) — Oʻzbek tili milliy korpusi metodologiyasi.
7. W3C Standards (XML for Lexicography) — Lugʻatlarni kodlash boʻyicha xalqaro standartlar (TEI - Text Encoding Initiative).