

KOMPYUTER LINGVISTIKASI ASOSIDA MAVZULASHTIRILGAN LUG‘AT TUZISHNING METODIK ASOSLAR

Alisherova Shaxnoza

Toshkent davlat jahon tillari universiteti

tayanch doktoranti

Annotatsiya: Mazkur maqolada kompyuter lingvistikasi texnologiyalaridan foydalanib mavzulashtirilgan lug‘at tuzishning metodik asoslari yoritiladi. Tadqiqotning maqsadi — lingvistik korpuslar, avtomatik matn tahlil vositalari va statistik hisoblash metodlari yordamida aniq mavzuga oid leksik birliklarni tizimli ravishda ajratish, ularning semantik, funksional va frekventlik xususiyatlarini aniqlashdir. Ishda Sketch Engine, AntConc kabi korpus platformalaridan foydalanish, shuningdek, kollokatsiyalarni avtomatik aniqlash, semantik guruhlariga ajratish, lug‘at birligining kontekstual o‘ziga xosliklarini belgilash kabi amaliy bosqichlar ko‘rib chiqiladi. Shuningdek, mavzulashtirilgan lug‘atlar yaratishda raqamli leksikografiyaning imkoniyatlari, sun‘iy intellekt asosidagi usullar, hamda ularning o‘qitish, tarjima, terminologiya va ilmiy tadqiqotlarda qo‘llanilishi tahlil qilinadi. Natijada zamonaviy kompyuter lingvistikasi asosidagi lug‘at tuzish jarayonining afzalliklari va metodik tamoyillari asoslab beriladi.

Kalit so‘zlar: kompyuter lingvistikasi, avtomatik korpus tahlili, mavzulashtirilgan lug‘at, leksikografiya, semantik guruhlash, ma’lumotlarni qayta ishlash, kollokatsiyalar, frekventlik tahlili, terminologiya, raqamli lug‘at tuzish, NLP, sun‘iy intellekt.

Аннотация: В этой статье рассматриваются методологические основы составления тематического словаря с использованием технологий компьютерной лингвистики. Цель исследования — систематически различать лексические единицы по конкретной теме с помощью лингвистических корпусов, автоматических инструментов анализа текста и методов статистических вычислений, чтобы выявить их семантические,

функциональные и братские особенности. В работе рассматриваются практические этапы использования корпусных платформ, таких как Sketch Engine, AntConc, а также автоматическое обнаружение коллокаций, семантическое группирование, определение контекстуальных специфик словарного запаса. В статье анализируются возможности цифровой лексикографии, методов искусственного интеллекта и их применение в преподавании, переводе, терминологии и научных исследованиях для создания тематических словарей. В результате подтверждаются преимущества и методологические принципы процесса составления словарей на основе современной компьютерной лингвистики.

Ключевые слова: компьютерная лингвистика, автоматический корпусный анализ, тематический словарь, лексикография, семантическое группирование, обработка данных, коллокации, анализ разлома, терминология, создание цифровых словарей, НЛП, искусственный интеллект.

Abstract: This article discusses the methodological foundations of creating thematic dictionaries through the application of computational linguistics. The main objective is to identify domain-specific lexical units using linguistic corpora, automated text-processing tools, and statistical analysis techniques. The study examines the use of platforms such as Sketch Engine and AntConc, the extraction of collocations, distributional patterns, and semantic groupings, as well as the determination of contextual and functional features of lexical items. Additionally, the paper highlights the advantages of digital lexicography, including AI-based methods, automated term extraction, and semantic annotation, emphasizing their relevance for education, translation studies, terminology development, and linguistic research. The findings justify the methodological principles and practical value of computational-linguistic approaches in developing accurate and efficient thematic dictionaries.

Keywords: computational linguistics, corpus analysis, thematic dictionary, lexicography, semantic clustering, data processing, collocations, frequency analysis, terminology, digital dictionary development, NLP, artificial intelligence.

Kirish. Bugungi globallashuv va raqamli transformatsiya jarayonlari tilshunoslikning barcha yo‘nalishlariga, xususan, leksikografik amaliyotga tubdan yangicha talablarni qo‘ymoqda. Raqamli platformalar, elektron resurslar va kompyuter texnologiyalari til birliklarini avtomatik qayta ishlash, tahlil qilish va ma’lumotlarini vizual tarzda taqdim etish imkonini sezilarli darajada kengaytirdi. Shu bois, zamonaviy leksikografiya an’anaviy lug‘at tuzish tamoyillari bilan cheklanib qolmay, kompyuter lingvistikasi yutuqlariga tayangan holda yangi shakl va mazmun kasb etmoqda. Ayniqsa, mavzulashtirilgan lug‘atlar — muayyan fan, sohaga, mavzuga oid leksik birliklarni tizimli jamlagan resurslar — bu jarayonning muhim natijalari sifatida ahamiyat kasb etmoqda. Ular til birligining semantik, pragmatik va funksional xususiyatlarini yanada chuqurroq yoritish imkonini beradi. Ushbu maqolada kompyuter lingvistikasi bilan zamonaviy leksikografiyaning integratsiyasi, mavzulashtirilgan lug‘atlarning o‘rni hamda ularning ilmiy va amaliy qiymati yoritiladi. Tadqiqot natijalari til o‘rgatish, tarjimashunoslik, korpus lingvistika va terminografiya kabi yo‘nalishlarda ham qo‘llanishi mumkin.

Raqamli texnologiyalar jadal rivojlanayotgan hozirgi davrda til resurslarini tizimlashtirish, elektron shaklga o‘tkazish va ulardan samarali foydalanish zaruriyati kuchaymoqda. Ayniqsa, o‘zbek tilining xalqaro kommunikatsiya maydoniga tobora faol kirib borayotgani, davlat tilini rivojlantirishga doir qator islohotlarning amalga oshirilayotgani leksikografiyada yangi yondashuvlarni talab qilmoqda. Mavzulashtirilgan lug‘atlar — muayyan sohaga oid terminlar, birliklar, iboralar va ularga doir izohlarni to‘plovchi manbalar — bu ehtiyojga javob beruvchi eng muhim ilmiy mahsulotlardan biridir. Ular nafaqat soha mutaxassislari, balki o‘quvchi, tarjimon, tadqiqotchi va dasturchilar uchun ham keng amaliy imkoniyat yaratadi. Bunday lug‘atlar yordamida sohaning konseptual tizimi yoritiladi, terminlarning

standartlashuvi ta'minlanadi, ilmiy matnlarni to'g'ri talqin qilishga imkon yaratiladi. Raqamlashtirish jarayonida kompyuter lingvistikasining qo'llanishi lug'at tuzishni tezlashtiradi, ma'lumotlarning aniqligi va izchilligini oshiradi, avtomatik tarjima va matnni qayta ishlash jarayonlariga asos bo'lib xizmat qiladi. Shu bois mavzu bugungi tilshunoslikning eng dolzarb yo'nalishlaridan biri sifatida katta ilmiy-amaliy ahamiyatga ega.

Kompyuter lingvistikasi — tilni avtomatik qayta ishlash, tahlil qilish, aniqlash va modellashtirishga qaratilgan fan sohasi bo'lib, zamonaviy leksikografiya bilan uzviy bog'liqdir. Agar an'anaviy lug'atlar qo'lda tuzilsa, zamonaviy lug'atlar ko'p hollarda elektron korpuslarga, avtomatik morfologik tahlil vositalariga, mashinaviy o'rganish algoritmlariga tayanadi. Natijada, leksikografik jarayonlar tezlashadi, birligining semantik tarmoqlanishi yanada aniqroq ochiladi va foydalanuvchiga interaktiv imkoniyatlar yaratiladi. Masalan, Word2Vec, BERT kabi semantik modellar so'zlarning o'zaro masofasi, kontekstlarni avtomatik aniqlash, kolokatsiyalarni topish kabi vazifalarni bajaradi. Bu esa lug'atlarning ilmiy asoslangan va statistik dalillarga tayangan holda tuzilishiga zamin yaratadi. Shuningdek, elektron lug'atlar qidiruv, filtr, sinonim-o'xshash birliklar, avtomatik tarjima, audio talaffuz va grafik vizualizatsiya kabi funksiyalar bilan boyitiladi. Demak, kompyuter lingvistikasi nafaqat lug'at yaratish jarayonini modernizatsiya qiladi, balki uning funksional qiymatini ham sezilarli oshiradi. Bugungi kunda har qanday zamonaviy leksikografik mahsulot kompyuter lingvistikasidan foydalanmasdan yaratilmaydi desak, mubolag'a bo'lmaydi.

Mavzulashtirilgan lug'atlar tilshunoslikning bir qator yo'nalishlarida katta ilmiy va amaliy ahamiyatga ega. Avvalo, ular ma'lum bir sohaning terminologik tizimini aniq, izchil va standart shaklda taqdim etadi. Bu esa ilmiy matnlarning tushunarligini oshiradi, soha bo'yicha yagona konseptual maydon shakllanishiga yordam beradi. Ikkinchidan, mavzulashtirilgan lug'atlar semantik tahlil, korpus tadqiqotlari, kognitiv lingvistika va pragmatik tadqiqotlarda muhim manba bo'lib

xizmat qiladi. Ularning asosida terminlarning tuzilishi, kelib chiqishi, qo'llanish chastotasi, stilistik xususiyatlari, ko'p ma'noliligi yoki o'zaro sinonimik aloqalari chuqur o'rganilishi mumkin. Uchinchidan, bunday lug'atlar o'qitish jarayonida, xususan ESP (English for Specific Purposes) va kasbiy ingliz tilini o'rgatishda nihoyatda zarur. Shuningdek, tarjima jarayonida terminlarni to'g'ri ishlatish, ekvivalentlarni aniqlash va kontekstga mos variant tanlashda muhim yordam beradi. Ammo eng muhim jihati — mavzulashtirilgan lug'atlar o'zbek tilining zamonaviy ilmiy-texnik imkoniyatlarini kengaytiradi, tilni modernizatsiya qilish va terminologiyani boyitishga xizmat qiladi. Shu bois ular bugungi lingvistik taraqqiyotning ajralmas qismiga aylangan.

Mazkur maqolaning asosiy maqsadi — kompyuter lingvistikasi yutuqlari asosida mavzulashtirilgan lug'at modelini yaratishning ilmiy-nazariy asoslarini ochib berish, zamonaviy leksikografiya qo'llanilayotgan usul va texnologiyalarni tahlil qilish hamda ularning o'zbek tilshunosligidagi qo'llanish imkoniyatlarini ko'rsatishdir. Shu maqsadda maqolada bir qator vazifalar belgilangan: (1) kompyuter lingvistikasi va leksikografiya integratsiyasining nazariy asoslarini yoritish; (2) mavzulashtirilgan lug'atlarning turli tiplari, ularning tuzilishi va ilmiy metodlarini o'rganish; (3) elektron korpuslar, mashinaviy o'rganish modellaridan foydalanish amaliyotini tahlil qilish; (4) o'zbek tilida mavzulashtirilgan lug'at yaratish uchun zarur bo'lgan konseptual model va algoritmi taklif qilish; (5) lug'at tuzish jarayonida semantik aniqlik, standartlashtirish, terminlarni kodlash va avtomatlashtirish kabi masalalarni ko'rib chiqish; (6) yakuniy modelning ta'lim, ilmiy tadqiqotlar, tarjima va IT sohalarida qo'llanish istiqbollari asoslash. Maqola ushbu vazifalar orqali til resurslarini yaratishning innovatsion usullarini shakllantirishga hissa qo'shishni maqsad qiladi.

Kompyuter lingvistikasi — tilni avtomatik qayta ishlash, tahlil qilish, modellashtirish va mashinaviy o'rganish algoritmlari orqali til haqidagi bilimlarni formal shaklda ifodalashga qaratilgan fandir. Ushbu yo'nalish lingvistika,

informatika, matematika, mantiq, neyron tarmoqlar va sun'iy intellektning kesishgan nuqtasida shakllangan bo'lib, inson tili va kompyuter o'rtasidagi muloqotni ta'minlashni o'z oldiga maqsad qiladi. Kompyuter lingvistikasining rivojlanishi bir necha bosqichlarda kechgan:

1. Dastlabki bosqich (1950–1970-yillar): Asosan mashinaviy tarjima, struktur grammatikalar va formal modellashuvga yo'naltirilgan. Z. Harris, N. Xomskiy, A. Bar-Xillel kabi olimlar grammatik generatsiya va avtomatik tahlil nazariyalarini yaratdilar.

2. Statistik yondashuvlar bosqichi (1980–2000-yillar): Korpus lingvistikasi, chastotaviy tahlil, ehtimollik modellari (n-gram), HMM, frayzali tarjima texnologiyalari shakllandi. Kompyuterlar katta ma'lumotlar bilan ishlashga qodir bo'ldi.

3. Neyron tarmoqlar bosqichi (2010–hozir): Word2Vec, GloVe, BERT, GPT, Transformer modellari paydo bo'ldi. Tilni tushunish, matn yaratish, avtomatik annotatsiya, lemmatizatsiya, morfologik tahlil yanada aniq, kontekstga mos bo'ldi. Bugun kompyuter lingvistikasi nafaqat matn bilan ishlaydi, balki nutqni tanish, tasvirga sarlavha qo'yish, interaktiv lug'atlar yaratish, o'quv resurslarini avtomatik generatsiya qilish kabi jarayonlarni ham qamrab oladi.

Elektron leksikografiya — lug'atlarni raqamli platformalarda yaratish, qayta ishlash va foydalanuvchiga interaktiv ko'rinishda taqdim etishga qaratilgan leksikografiyaning zamonaviy yo'nalishidir. U an'anaviy lug'atchilikdan tubdan farqlanadi, chunki elektron lug'atlar statik emas, balki dinamik, o'zgaruvchan hamda foydalanuvchi ehtiyojiga moslashuvchi resurs sifatida shakllanadi.

Asosiy farqlari:

- An'anaviy lug'atlar qo'lda tuziladi, mazmuni qog'ozga bog'liq, yangilanishi sekin.
- Elektron lug'atlar korpus ma'lumotlariga tayangan, avtomatik yangilanishi mumkin, real vaqt rejimida tahliliy imkoniyatlar beradi.

- An'anaviy lug'atda qidiruv cheklangan; elektron lug'atda esa fonetik, morfologik, semantik, kollokativ qidiruvlar mavjud.
- Elektron lug'atlar audio, video, kontekst, grafik, kollokatsiyalar, sinonim tarmoqlar, misollar banki bilan boyitilishi mumkin.
- Elektron leksikografiya mashina o'qiy oladigan formatlar (XML, JSON, OntoLex-Lemon) yordamida boshqa tizimlar bilan integratsiya qilinadi.

Shuningdek, elektron lug'atlar doimiy ravishda kengayishi, foydalanuvchi xatti-harakatiga moslashishi, sun'iy intellekt yordamida yangi birliklarni avtomatik aniqlashi bilan ajralib turadi. Zamonaviy leksikografiyada avtomatlashtirilgan tizimlardan foydalanish lug'at yaratish jarayonining aniqligini oshiradi va vaqtni tejaydi. Bunday tizimlar quyidagi asosiy turlarga bo'linadi:

1. Korpus asosida ishlovchi tizimlar

— So'z chastotasi, kollokatsiyalar, semantik yaqinlik, kontekstlarni avtomatik aniqlaydi.

Masalan: SketchEngine, AntConc, NoSketchEngine.

2. Morfologik analizatorlar

— So'zning shaklini, negizini, grammatik kategoriyalarini avtomatik aniqlaydi.

Masalan: HFST, UDPipe, Stanza.

3. Semantik tarmoqlar va ontologiyalar

— So'zlarning o'zaro munosabatlarini grafik ko'rinishda beradi: sinonimiya, antonimiya, giponimiya.

Masalan: WordNet, BabelNet, Uzbek WordNet loyihalari.

4. Lug'at yaratish platformalari

— Tayyor format, interfeys, ma'lumotlarga indekslash va eksport funksiyalarini beradi.

Masalan: Lexonomy, TLex Suite, FLEEx, Linghub.

Asosiy funksiyalari:

- Terminlarni avtomatik yig'ish va guruhlash

- Kollokatsiyalarni hisoblash
- Semantik tahlil
- Soʻz maʼnolarini kontekst asosida ajratish (WSD)
- Tezkor yangilash va variantlarni solishtirish
- Turli formatlarda eksport qilish (PDF, XML, HTML)

Bu tizimlar yordamida lugʻat tuzuvchi faqat tahrirlovchi va nazorat qiluvchi sifatida ishtirok etadi, asosiy texnik ishlarni esa kompyuter bajaradi.

Mavzulashtirilgan lugʻatlar — maʼlum bir soha yoki mavzuga oid terminlar, frazeologizmlar, iboralar va tushunchalar tizimini yorituvchi maxsus leksik resurslardir. Ularning ahamiyati juda keng va koʻp qirrali.

1. Ilmiy ahamiyati

- Tilning muayyan konseptual sohasini chuqur yoritadi.
- Terminlarning semantik tarmoqlarini aniq koʻrsatadi.
- Kognitiv lingvistika, semasiologiya, pragmalingvistika tadqiqotlari uchun asosiy manba boʻlib xizmat qiladi.
- Korpus lingvistikasi bilan integratsiyada statistik dalillarga asoslangan ilmiy xulosalar olishga imkon beradi.

2. Amaliy ahamiyati

- Tarjima jarayonida ekvivalentlarni toʻgʻri tanlashda yordam beradi.
- Oliy taʼlimda, ESP (kasbiy ingliz tili) kurslarida asosiy oʻquv vositasi boʻla oladi.
- Soha mutaxassislari uchun terminlarni standartlashtirishda muhim oʻrin tutadi.
- Raqamli platformalarda, AI tizimlarida, chat-botlarda, avtomatik tarjima dasturlarida ishlatiladi.
- Oʻzbek tilining zamonaviy ilm-fan tarmoqlaridagi imkoniyatini kengaytiradi.

Mavzulashtirilgan lugʻatlar tilni rivojlantirish, standartlashtirish, modernizatsiya qilish, raqamlashtirish jarayonining ajralmas boʻlagi hisoblanadi.

Matn korpusi — bu ma'lum tamoyillar asosida yig'ilgan, elektron shaklda saqlangan katta hajmdagi matnlar majmuasi bo'lib, ularning til xususiyatlari avtomatik tarzda tahlil qilish imkonini beradi. Korpus lingvistikasi esa ushbu matnlar asosida til birliklarini statistik, struktur va semantik jihatdan o'rganishni ta'minlaydi. Mavzulashtirilgan lug'at tuzishda korpuslarning ahamiyati nihoyatda kattadir, chunki ular terminlarning haqiqiy kontekstlarda qanday ishlatilishini, qaysi birliklar bilan birga qo'llanishini, chastotaviy darajalarini va semantik o'zgaruvchanligini aniqlab beradi. Masalan, SketchEngine yoki NoSketchEngine kabi platformalar so'zning kollokatsion maydonini, leksik bog'lanishlarini, sinonimik tarmoqlarini va stilistik o'ziga xosliklarini avtomatik hisoblab beradi. Korpus asosidagi yondashuv lug'at tuzishni subyektivlikdan holi qiladi, statistik dalillarga tayanadi va foydalanuvchiga haqiqiy til tizimi asosida tuzilgan ishonchli ma'lumotlarni taqdim etadi. Shu sababli mavzulashtirilgan lug'atlar sifat jihatdan boyroq, aniqroq va ilmiy asoslangan bo'lib chiqadi. Zamonaviy kompyuter lingvistikasining eng samarali yo'nalishlaridan biri — leksik birliklarni matndan avtomatik ajratish, ya'ni tokenizatsiya, lemmatizatsiya, morfologik klassifikatsiya va semantik guruhlash jarayonlaridir. Ushbu texnologiyalar matnlarda uchraydigan barcha so'z shakllarini aniqlaydi, ularning qay darajada terminologik ahamiyatga egaligini statistik modellar asosida baholaydi va kerakli birliklarni avtomatik ravishda mavzulashtirilgan bazaga ajratib qo'yadi. Avtomatik ajratish bir necha bosqichlarda amalga oshadi:

- Tokenizatsiya – matnni so'zlarga bo'lish;
- Morfologik tahlil – so'zning grammatik xususiyatlarini aniqlash;
- Stop-so'zlarni filtrlash – ahamiyatsiz birliklarni chiqarib tashlash;
- TF-IDF, Chi-square, PMI kabi statistik indikatorlar orqali terminlarni aniqlash;
- K-means, DBSCAN kabi klasterlash algoritmlari orqali semantik guruhlar hosil qilish.

ec, GloVe – so‘zlarning semantik yaqinligini aniqlaydi;
t – morfologik jihatdan boy tillarda (masalan, o‘zbek)
RoBERTa, XLM-R – kontekstga bog‘liq ma’nolarni ta’minlaydi;
modellar – ta’riflar yaratadi, misollar keltiradi, se-
di.
nida lug‘at tuzishda quyidagi vazifalar avtomatlashtirilgan:
ning asosiy ma’no markazini aniqlash;
nolilikni (polysemy) ajratish;
yicha sinonimik va antonimik tarmoqlarni tuzish;
bankini avtomatik yaratish;
ng kontekstual ishlatilishini modellashtirish.
nologiyalar lug‘atning aniqligini, kontekstga mos-
n oshiradi. Lug‘at tuzish jarayonida matnni chuqur

- ning asosiy ma'no markazini aniqlash;
 monolilikni (polysemy) ajratish;
 yaqinlashuvchi sinonimik va antonimik tarmoqlarni tuzish;
 bankini avtomatik yaratish;
 lug'atning kontekstual ishlatilishini modellashtirish.
- lingvistikologiyalar lug'atning aniqligini, kontekstga mosligini oshiradi. Lug'at tuzish jarayonida matnni chuqur o'qish nihoyatda muhim. Bunda quyidagi avtomatizatsiya qo'llaniladi: So'z shakllarini asosiy lug'aviy formlarga ajratish: boraman, bording, bordi → bor-. Bu jarayon avtomatizatsiya shimchalar tufayli murakkab, ammo zamonaviy metodlar (IFST) yordamida samarali bajariladi. Har bir so'zning ma'nosini aniqlash:

ning asosiy ma'no markazini aniqlash;
nolilikni (polysemy) ajratish;
yicha sinonimik va antonimik tarmoqlarni tuzish;
bankini avtomatik yaratish;
ng kontekstual ishlatilishini modellashtirish.

ologiyalar lug'atning aniqligini, kontekstga mos
n oshiradi. Lug'at tuzish jarayonida matnni chuqur
qilish nihoyatda muhim. Bunda quyidagi avt
oydalaniladi: So'z shakllarini asosiy lug'aviy fo
an: boraman, bording, bordi → bor-. Bu jarayon
shimchalar tufayli murakkab, ammo zamonaviy m
(IFST) yordamida samarali bajariladi. Har bir so'
adi:

- biologiyalar lugʻatning aniqligini, kontekstga moslashishni oshiradi. Lugʻat tuzish jarayonida matnni chuqur oʻrganish nihoyatda muhim. Bunda quyidagi avtomatizatsiya qoʻllaniladi: Soʻz shakllarini asosiy lugʻaviy forma bilan bogʻlan: boraman, bording, bordi → bor-. Bu jarayon dasturlar yordamida murakkab, ammo zamonaviy metodlar (IFST) yordamida samarali bajariladi. Har bir soʻzning maʼnolari:

(218)

- soʻz turkumi,
- grammatik kategoriyalari (son, shaxs, zamon),
- sintaktik funksiyasi,
- kelishik yoki feʼl mayllari.

Gapdagi bogʻlanishlar aniqlanadi:

- feʼl – ega munosabati,
- aniqlovchi – aniqlanmish,
- toʻldiruvchilar,
- boʻlaklar ierarxiyasi.

Dependency parsing yordamida terminlarning sintaktik valentligini, yaʼni ular bilan koʻp uchraydigan birliklarni aniqlash mumkin. Bu jarayonlarning avtomatlashtirilishi natijasida lugʻat tuzuvchi inson matnni qoʻlda ajratish, tartiblash bilan shugʻullanmaydi. Asosiy tahliliy ishni AI bajaradi. Semantik tarmoqlar (WordNet tipidagi tizimlar) va ontologiyalar mavzulashtirilgan lugʻat tuzishda asosiy konseptual struktura yaratishga yordam beradi. Ontologiya — muayyan sohadagi tushunchalar, obyektlar va ular orasidagi munosabatlarni belgilovchi formal modeldir. Ontologiyalarning lugʻatchilikdagi vazifalari:

- tushunchalar ierarxiasini yaratish (superklass–subklass);
- soʻzlar oʻrtasidagi semantik munosabatlarni formal tavsiflash;
- giponimiya, sinonimiya, antonimiya, meronimiya tizimini tashkil qilish;
- terminlarning oʻzaro bogʻliqligini grafik shaklda koʻrsatish;
- mashina oʻqiy oladigan formatda lugʻat yaratish.

Eng koʻp qoʻllaniladigan texnologiyalar:

- WordNet / Uzbek WordNet – semantik tarmoq asosidagi leksik baza
- BabelNet – koʻp tilli semantik tizim
- OntoLex-Lemon – lingvistik ontologiyalar uchun standart model
- Protégé – ontologiya yaratish vositasi
- SKOS – terminologik tizimlarni standartlashtirish uchun model

Mavzulashtirilgan lug‘atlar aynan ontologiyalar orqali ilmiy asoslangan, tizimli, tartibli va mashina tomonidan qayta ishlanishga qulay resursga aylanadi. Mavzulashtirilgan lug‘at tuzish jarayonida kompyuter texnologiyalaridan foydalanish zamonaviy leksikografiyaning ajralmas talabi bo‘lib, tilshunoslikning nazariy va amaliy rivojida yangi bosqichni belgilaydi. Matn korpuslari, statistik tahlil vositalari, avtomatik leksik birlik ajratish, sun‘iy intellekt modellaridan foydalanish lug‘at tuzishni ancha tezlashtiradi, subyektivlikni kamaytiradi hamda ilmiy aniqlikni oshiradi. Ayniqsa, korpus lingvistikasi asosidagi yondashuv terminlarning haqiqiy nutq kontekstidagi qo‘llanishini aniqlashga, ularning semantik maydonini belgilashga va tizimli tasnif yaratishga imkon beradi.

Xulosa. Mashinaviy o‘qitish, chuqur o‘rganish va semantik tarmoqlar yordamida lug‘atdagi ma’lumotlar nafaqat til birliklari ro‘yxati sifatida, balki murakkab leksik-grammatik va konseptual tizim sifatida shakllanadi. Lemmatizatsiya, tagging va parsing kabi avtomatlashtirilgan jarayonlar tilning grammatik va sintaktik strukturasi bo‘yicha boy ma’lumotlar beradi va lug‘atni ko‘p darajali ilmiy manbaga aylantiradi. Ontologiyalar esa lug‘atning konseptual asosini mustahkamlaydi, terminlar o‘rtasidagi semantik munosabatlarni aniqlik bilan ifodalaydi va resursni mashina o‘qiydigan ekotizimning bir qismiga aylantiradi. Umuman olganda, kompyuter texnologiyalari mavzulashtirilgan lug‘atlarni yaratishda sifat, tezlik, aniqlik, tizimlilik va qo‘llanish doirasi jihatidan ulkan afzalliklar beradi. Bunday lug‘atlar zamonaviy tilshunoslik, tarjimashunoslik, ta’lim, korpus tadqiqotlari, IT va sun‘iy intellekt sohalarida keng amaliy imkoniyatlarga ega bo‘lib, o‘zbek tilining raqamli resurslar bazasini boyitish va uni global ilmiy maydonga olib chiqish uchun muhim ilmiy-texnologik poydevor yaratadi.

FOYDALANILGAN ADABIYOTLAR

1. Aikhenvald, A. (2018). *How to Write a Linguistic Dictionary*. Oxford University Press.

2. Atkins, B. T., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press.
3. Baker, M. (2011). *In Other Words: A Coursebook on Translation*. Routledge.
4. Baranov, A. N., & Dobrovol'skij, D. (2013). *Aspekty teorii frazeologii*. LKI.
5. Béjoint, H. (2010). *Modern Lexicography: An Introduction*. Oxford University Press.
6. Choi, K., & Kim, J. (2020). Automatic Term Extraction Using Deep Learning. *Journal of Natural Language Engineering*, 26(4), 511–534.
7. Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
8. Church, K., & Hanks, P. (1990). Word Association Norms and Collocations. *Computational Linguistics*, 16(1), 22–29.