TADQIQOT KORPUSINI SHAKLLANTIRISH VA SEMANTIK EKVIVALENTLIKNI KODLASH METODIKASI

Alisherova Shahnoza Asqarovna
O‘zbekiston davlat jahon tillari universiteti
tayanch doktoranti

Annotatsiya: Mazkur maqolada ikki va ko‘p tilli terminologik tadqiqotlar uchun tadqiqot korpusini shakllantirish, materialni tanlash, metama’lumot bilan ta’minlash, termin birliklarini ajratish hamda tillararo semantik ekvivalentlikni izchil kodlash metodikasi ishlab chiqiladi. Korpusning reprezentativligi faqat matn hajmi bilan emas, balki soha, janr, kommunikativ vazifa, vaqt, manba turi va tarjima maqomi bo‘yicha muvozanat bilan belgilanishi asoslanadi. Termin juftliklarini baholashda konseptual mazmun, denotativ qamrov, funksional moslik, uslubiy-registr xususiyati va kontekstual barqarorlikni alohida ko‘rsatkichlar sifatida qayd etish taklif qilinadi. Ushbu ko‘rsatkichlar asosida to‘liq, qisman, funksional-kontekstual va nol ekvivalentlik kodlari ajratiladi. Kodlash jarayonida annotatorlar uchun aniq yo‘riqnoma, mustaqil ikki bosqichli belgilash, kelishmovchiliklarni ekspert muhokamasi orqali bartaraf etish va annotatorlararo moslikni statistik tekshirish tamoyillari yoritiladi. Taklif etilgan metodika terminografik baza, parallel yoki qiyosiy korpus, tarjima tadqiqotlari hamda elektron lug‘at yaratish ishlarida qayta qo‘llanishi mumkin bo‘lgan nazorat qilinadigan va tekshiriladigan tadqiqot modelini ta’minlaydi.

Kalit so‘zlar: tadqiqot korpusi, korpus lingvistikasi, terminologik birlik, semantik ekvivalentlik, annotatsiya, kodlash, terminografiya, parallel korpus, qiyosiy tahlil, validatsiya.

KIRISH

Korpusga asoslangan terminologik tadqiqotda ilmiy xulosaning ishonchliligi avvalo materialning qanday tanlangani va til birliklari qanday mezonlar asosida kodlanganiga bog‘liq. Ayniqsa, ikki tilli terminologiyada bir terminning ikkinchi tildagi muqobilini shunchaki lug‘aviy moslik sifatida qabul qilish yetarli emas: ekvivalent birliklar bir xil konseptni ifodalashi, o‘xshash sohaviy vazifani bajarishi, muayyan kontekstlarda almashish imkoniga ega bo‘lishi va terminologik tizimdagi mavqei jihatidan qiyoslanishi kerak. Zamonaviy korpusshunoslikda korpusning maqsadga muvofiq dizayni, matnlarni qatlamlar bo‘yicha tanlash, tokenlash, lemmalash, teglash va metama’lumotni saqlash muhim bosqichlar sifatida qaraladi. O‘zbek korpusshunosligida Sh. Hamroyeva, N. Abduraxmonova, G. Toirova, M. Abjalova, V. Zaxarov, B. Mengliyev kabi olimlarning ishlari matn bazasini tashkil etish, lingvistik razmetka va korpusdan foydalanish tamoyillarini ishlab chiqishga xizmat qilgan. Xorijiy tadqiqotlarda esa parallel va qiyosiy korpuslarda terminlarni avtomatik yoki yarim avtomatik aniqlash, qo‘lda annotatsiya qilingan “gold standard” to‘plamlarini yaratish hamda annotatorlararo moslikni tekshirish masalalariga katta e’tibor qaratiladi. Shundan kelib chiqib, mazkur maqolaning maqsadi terminologik tadqiqot uchun korpusni bosqichma-bosqich shakllantirish va semantik ekvivalentlikni qayta tekshirish mumkin bo‘lgan kodlar tizimi orqali belgilash metodikasini taklif etishdan iborat.

Tadqiqot korpusini tuzishning birinchi sharti tadqiqot obyektini aniq chegaralashdir. “Terminologik korpus” umumiy til korpusining tasodifiy qismi emas, balki muayyan ilmiy yoki kasbiy soha diskursini maqsadli aks ettiruvchi, oldindan belgilangan mezonlarga ko‘ra yig‘ilgan matnlar majmuidir. Korpus tuzuvchi tadqiqotchi avval soha chegarasini, tillarni, vaqt oralig‘ini, matn turlarini va tahlil birligini belgilaydi. Masalan, inglizcha-o‘zbekcha terminologik tadqiqotda korpus bir tildagi manbalarni ikkinchi tildagi tasodifiy matnlar bilan tenglashtirmasligi, balki mavzu va kommunikativ vazifasi o‘xshash

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

materiallardan tuzilishi lozim. Parallel korpusda original va tarjima matnlar o‘zaro bog‘langan bo‘ladi; qiyosiy korpusda esa matnlar to‘g‘ridan-to‘g‘ri tarjima bo‘lmasligi mumkin, biroq soha, janr, davr va auditoriya bo‘yicha qiyoslashga yaroqli bo‘lishi kerak. Martin Volk parallel korpuslarda gap va so‘z darajasidagi moslashtirishning sifati keyingi tarjima variantlarini aniqlash va semantik tahlil natijalariga bevosita ta’sir qilishini ko‘rsatadi. O‘zbek tilidagi korpus tadqiqotlarida ham matnni yagona formatga keltirish, tokenlarni alohida birlik sifatida ajratish, lemma va teglarni bir xil tuzilma asosida saqlash qidiruv hamda tahlilning barqarorligini ta’minlaydi. Shu bois korpusning texnik formati tadqiqotning tashqi masalasi emas, balki lingvistik natijani takrorlash imkonini beruvchi metodologik omil sifatida qaralishi zarur.

Reprezentativlikni ta’minlash uchun korpus qatlamli tanlash asosida tuziladi. Har bir til bo‘yicha material kamida besh metama’lumot maydoni bilan tavsiflanishi maqsadga muvofiq: soha yoki mikrosoha, janr, nashr sanasi, manba turi va matnning tarjima maqomi. Zarur hollarda muallif yoki tashkilot, hudud, maqsadli auditoriya, ilmiy-rasmiy ommaboplik darajasi kabi qo‘shimcha belgilar ham kiritiladi. Matnlar soni bilan birga tokenlar hajmi ham qayd etiladi, chunki 50 ta juda qisqa hujjat bilan 10 ta katta monografiya statistik jihatdan bir xil bazani bermaydi. Korpusda bitta muallif, bitta tashkilot yoki bitta sayt materialining haddan tashqari ko‘payishi leksik va terminologik chastotalarni sun’iy ravishda o‘zgartirishi mumkin; shu sababli dublikatlar, qayta nashrlar va mazmunan bir xil hujjatlar chiqarib tashlanadi. Korpus tayyor bo‘lgach, matnlar UTF-8 kodlashiga o‘tkaziladi, ortiqcha xizmat belgilaridan tozalanadi, gap va token chegaralari belgilanadi, lemmalash hamda morfologik teglash amalga oshiriladi. O‘zbek tilining agglutinativ tuzilishi sababli terminning turli grammatik shakllari bir lemma ostida birlashtirilishi muhim: aks holda bir terminning ko‘plik, egalik yoki kelishik shakllari alohida birlik sifatida sanalib, chastota buziladi. M. Abjalova hamda V. Zaxarov, B. Mengliyev va Sh. Hamroyeva korpus materialining tizimli razmetkasi korpusdan lingvistik va statistik maqsadlarda foydalanishning asosiy sharti ekanini ta’kidlaydi. Shuning uchun terminologik korpusda asl matnni saqlash bilan birga normallashtirilgan shakl, lemma, so‘z turkumi, soha tegi va hujjat identifikatori kabi ma’lumotlar alohida ustunlarda saqlanishi tavsiya etiladi.

Korpus tayyor bo‘lgach, termin nomzodlarini aniqlash bosqichi boshlanadi. Terminlar faqat chastotaga ko‘ra tanlansa, umumiste’mol so‘zlari yoki soha matnlarida tez-tez uchraydigan yordamchi birliklar ro‘yxatga kirib qoladi; faqat ekspert intuitiyasiga tayanilsa esa past chastotali, biroq konseptual jihatdan muhim terminlar chetda qolishi mumkin. Shu sababli kombinatsiyalangan usul qo‘llanadi: chastota va kalit so‘zlik ko‘rsatkichlari, terminning tipik grammatik modeli, soha lug‘atlaridagi mavjudligi, kontekstlar bo‘yicha semantik barqarorligi va ekspert bahosi birgalikda hisobga olinadi. Rigouts Terryn, Hoste va Lefever tomonidan ishlab chiqilgan ko‘p tilli termin annotatsiyasi tajribasi batafsil yo‘riqnoma va qo‘lda tekshirilgan korpusning terminlarni baholashda muhimligini ko‘rsatadi. Har bir ajratilgan termin uchun “termin kartasi” yaratilishi metodik nazoratni kuchaytiradi. Kartada terminning asl shakli, lemma, til kodi, soha/mikrosoha, hujjat identifikatori, kontekst oynasi, chastota, grammatik tuzilma, manba turi, ehtimoliy tarjima ekvivalenti, ekvivalentning manbasi va ekspert izohi bo‘lishi kerak. Agar termin ko‘p ma’noli bo‘lsa, umumiy leksema emas, aynan korpusdagi terminologik ma’no kodlanadi. Bu talab semantik ekvivalentlikni “so‘z–so‘z” darajasida emas, “konsept–kontekst–termin” uchligi asosida tekshirishga imkon beradi. ISO 704:2022 standartida ham terminologik ish obyekt, konsept, ta’rif va belgilash o‘rtasidagi munosabatlarga tayanishi ko‘rsatilgan; demak, ekvivalentlikni baholashda shakliy o‘xshashlikdan ko‘ra konseptual moslik ustuvor bo‘lishi kerak.

Tadqiqot ma’lumotlar bazasida har bir termin juftligi alohida yozuv sifatida saqlanadi. Minimal yozuv tarkibi quyidagicha bo‘lishi mumkin: source_term, source_lemma, target_term, target_lemma, domain, context_source, context_target, definition_source, definition_target, equivalence_code,

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

confidence_score, annotator_id va decision_note. Bunday tuzilma TermBase eXchange (TBX) kabi terminologik resurslarni boshqarish tamoyillariga mos keladi va keyinchalik ma’lumotni elektron lug‘at, tarjima xotirasi yoki terminologik baza formatiga ko‘chirishni yengillashtiradi. Muhimi, ma’lumot kartasida ekvivalentlik kodi bilan birga qaror sababini ko‘rsatuvchi qisqa izoh ham bo‘lishi kerak. Masalan, ikki terminning lug‘aviy shakli bir-biriga to‘g‘ri kelgan bo‘lsa-da, ulardan biri kengroq konseptni anglatsa, “to‘liq” kodini qo‘yish mumkin emas; aksincha, so‘zma-so‘z o‘xshash bo‘lmagan, lekin sohada aynan bir funksional konsept uchun qo‘llanadigan birliklar funksional ekvivalent sifatida belgilanishi mumkin. Shu tariqa kodlash tizimi faqat natijani emas, qarorning lingvistik asosini ham qayd etadi.

Semantik ekvivalentlikni kodlashda bir o‘lchovli “bor/yo‘q” modeli ko‘p hollarda yetarli emas. Terminlararo moslikni kamida besh mezon bo‘yicha tekshirish taklif qilinadi: konseptual yadro, denotativ qamrov, funksional vazifa, registr/uslub va kontekstual almashinuvchanlik. Konseptual yadro ikki terminning asosiy ilmiy tushunchasi qanchalik mos kelishini bildiradi; denotativ qamrov termin qamrab oladigan obyekt yoki hodisalar doirasini solishtiradi; funksional vazifa terminning sohaviy kommunikatsiyadagi rolini tekshiradi; registr rasmiylik, standartlashganlik va terminologik maqomni ko‘rsatadi; kontekstual almashinuvchanlik esa tipik kontekstlarda bir termin ikkinchisining tabiiy muqobili bo‘la olish-olmasligini baholaydi. Har bir mezon 0–2 oralig‘ida kodlanishi mumkin: 0 – mos emas, 1 – qisman mos, 2 – yuqori darajada mos. Biroq yakuniy kod oddiy arifmetik yig‘indi asosida avtomatik berilmasligi kerak; konseptual yadro bo‘yicha 0 baho olingan juftlikni boshqa ko‘rsatkichlar yuqori bo‘lsa ham to‘liq ekvivalent deb tasniflash metodik jihatdan noto‘g‘ri. Shu sababli yakuniy qarorda konseptual moslik “majburiy shart”, qolgan o‘lchovlar esa darajani aniqlovchi omillar sifatida olinadi.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

1-jadval. Semantik ekvivalentlikni kodlash shkalasi

Kod	Ekvivalentlik turi	Asosiy diagnostik belgi	Kodlash qarori
E1	To‘liq ekvivalentlik	Konseptual yadro va denotativ qamrov deyarli bir xil; registr hamda vazifa mos.	Ikki birlik tipik sohaviy kontekstlarda o‘zaro barqaror muqobil bo‘la oladi.
E2	Qisman ekvivalentlik	Asosiy konsept mos, ammo qamrov, komponent, registr yoki qo‘llanish chegarasi farqlanadi.	Muqobillik kontekst yoki mikrosoha bilan cheklanadi; izohda farq ko‘rsatiladi.
E3	Funksional-kontekstual ekvivalentlik	Shakl yoki semantik tarkib to‘liq mos emas, biroq amaliy vazifa va maqsadli kontekst bir xil.	Termin so‘zma-so‘z emas, funksional muqobil sifatida tasdiqlanadi.
E4	Nol ekvivalentlik	Maqsad tilida standart yoki barqaror termin mavjud emas yoxud konseptual moslik yetarli emas.	Transliteratsiya, tavsifiy tarjima yoki yangi termin taklifi alohida belgilanadi.

Jadvaldagi E1–E4 kodlari yakuniy natijani ifodalaydi, ammo kodlashning ichki protokolida qo‘shimcha diagnostik belgilarni saqlash maqsadga muvofiq. Masalan, C (concept) – konseptual moslik, D (denotation) – denotativ qamrov, F (function) – funksional moslik, R (register) – registr, X (context) – kontekstual almashinuvchanlik shaklida 0, 1 yoki 2 ball berilishi mumkin. Shunda E2 kodli ikki juftlikning nima sababdan qisman ekvivalent ekani ham ma’lumotlar bazasida ko‘rinadi: bir juftlikda denotativ qamrov farqi, boshqasida esa registr yoki normativ maqom farqi bo‘lishi mumkin. Bu usul keyinchalik statistik tahlil, xatolar tipologiyasi va terminografik izohlarni avtomatik shakllantirish uchun qulaydir. Brezina korpus ma’lumotlarini tahlil qilishda miqdoriy ko‘rsatkichlarni ma’lumot tuzilishi va tanlanma xususiyatlaridan ajratmaslik kerakligini ta’kidlaydi; semantik kodlashda ham raqamli ball lingvistik izohning o‘rnini bosmasligi, balki uni standartlashtirishi kerak.

Semantik ekvivalentlik inson talqinini talab qiladigan vazifa bo‘lgani sababli kodlashning ishonchliligi alohida nazorat qilinadi. Bir annotator barcha juftliklarni mustaqil kodlasa, ichki izchillik saqlanishi mumkin, biroq subyektiv qarorlar sezilmay qoladi. Shu sababli kamida ikki annotator bir xil tanlamaning ma’lum qismini mustaqil baholashi, so‘ng natijalar solishtirilishi tavsiya etiladi. Dastlab 5–10 foizlik sinov to‘plamida yo‘riqnoma tekshiriladi: ko‘p kelishmovchilik tug‘dirgan holatlar aniqlanadi, kod ta’riflari aniqlashtiriladi va misollar qo‘shiladi. Keyingi bosqichda annotatorlar asosiy tanlamani mustaqil kodlaydi. Kelishmovchiliklar uchinchi ekspert yoki konsensus muhokamasi orqali hal qilinadi, biroq dastlabki kodlar o‘chirib yuborilmaydi; ular audit izi sifatida saqlanishi lozim. Rigouts Terryn va hammualliflar termin annotatsiyasida aniq yo‘riqnoma hamda iterativ kelishish jarayoni annotatorlararo moslikni sezilarli yaxshilashini ko‘rsatadi. Moslikni oddiy foiz bilan birga Cohen kappa yoki bir nechta annotator bo‘lsa Fleiss kappa kabi ko‘rsatkichlar orqali tekshirish mumkin. Tartibli E1–E4 shkalasi uchun weighted kappa qo‘llash ham maqsadga muvofiq, chunki E1 va E2 o‘rtasidagi kelishmovchilik E1 va E4 o‘rtasidagi keskin farq bilan teng emas.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Validatsiya ikki yo‘nalishda olib boriladi. Birinchisi ichki validatsiya bo‘lib, korpusdagi takroriy kontekstlar, termin variantlari va annotatsiya izchilligini tekshiradi. Bir xil termin juftligi turli hujjatlarda turlicha kodlangan bo‘lsa, bu haqiqiy polisemiyaga, mikrosoha farqiga yoki annotatsiya xatosiga bog‘liqligi aniqlanadi. Ikkinchisi tashqi validatsiya bo‘lib, korpusdagi qarorlar normativ terminologik standartlar, sohaviy lug‘atlar, rasmiy hujjatlar va ekspert xulosalari bilan solishtiriladi. ISO 30042:2019 terminologik resurslarni tuzishda ma’lumot kategoriyalarini aniq ajratish va terminologik yozuvlarni tizimli boshqarish zarurligini ko‘rsatadi; bu prinsip ekvivalentlik kodlari, manba ma’lumoti, ta’rif va izohlarni alohida maydonlarda saqlashni qo‘llab-quvvatlaydi. Mahalliy korpus tadqiqotlarida ham matn, teg va metama’lumotni standart formatda saqlash keyingi qidiruv va avtomatik qayta ishlash uchun muhim deb qaraladi. Shuning uchun yakuniy korpus “muzlatilgan” bir martalik fayl emas, versiyalanadigan ilmiy resurs sifatida yuritilishi lozim. Har bir yangilanishda sana, o‘zgargan yozuvlar, qo‘shilgan manbalar va kodlash qoidasidagi o‘zgarishlar qayd etilsa, tadqiqot natijasi qayta tekshiriladigan va boshqa tadqiqotchilar tomonidan takrorlanadigan holga keladi.

Metodikaning amaliy algoritmi quyidagi ketma-ketlikda umumlashtiriladi: avvalo tadqiqot sohasi va tillar chegaralanadi; keyin parallel yoki qiyosiy korpus modeli tanlanadi; manbalar soha, janr, davr va kommunikativ vazifa bo‘yicha balanslanadi; matnlar tozalanib, tokenlanadi, lemmalanadi va metama’lumot bilan ta’minlanadi; termin nomzodlari statistik va lingvistik mezonlar asosida ajratiladi; har bir termin uchun kontekst va ta’rif yig‘iladi; maqsad tilidagi ehtimoliy ekvivalentlar korpus, lug‘at va ekspert manbalari orqali aniqlanadi; C, D, F, R va X ko‘rsatkichlari bo‘yicha dastlabki baho beriladi; E1–E4 yakuniy kodi tayinlanadi; annotatorlararo moslik tekshiriladi va bahsli juftliklar ekspert muhokamasidan o‘tkaziladi; so‘ng ma’lumot bazasi versiyalanib, statistik va terminografik tahlil uchun eksport qilinadi. Bunday ketma-ketlik tadqiqotchi qarorini shaffof qiladi, korpusning qaysi qismi xom material, qaysi qismi annotatsiyalangan ma’lumot va qaysi qismi ekspert tomonidan tasdiqlangan “gold standard” ekanini aniq ajratadi. Natijada semantik ekvivalentlik haqidagi xulosa alohida misollarga emas, boshqariladigan va izchil kodlangan empirik bazaga tayanadi.

XULOSA

Tadqiqot korpusini shakllantirish va semantik ekvivalentlikni kodlash bir-biridan ajratib bo‘lmaydigan metodologik jarayonlardir. Korpus noto‘g‘ri yoki bir tomonlama tuzilsa, eng mukammal kodlash tizimi ham til birliklarining real qo‘llanishini to‘liq aks ettira olmaydi; aksincha, reprezentativ material aniq annotatsiya mezonlarisiz yig‘ilsa, olingan ma’lumotlar qiyosiy tahlil uchun yetarlicha izchil bo‘lmaydi. Shuning uchun terminologik korpus soha, janr, davr, manba turi va tarjima maqomi bo‘yicha balanslangan, texnik jihatdan normallashtirilgan, lemmalangan va metama’lumot bilan ta’minlangan bo‘lishi zarur. Semantik ekvivalentlikni baholashda esa faqat so‘zma-so‘z moslik emas, konseptual yadro, denotativ qamrov, funksional vazifa, registr va kontekstual almashinuvchanlik birgalikda hisobga olinishi lozim. Taklif etilgan E1–E4 shkalasi to‘liq, qisman, funksional-kontekstual va nol ekvivalentlikni ajratadi, C–D–F–R–X diagnostik ko‘rsatkichlari esa yakuniy kodning sababini ma’lumot bazasida saqlash imkonini beradi. Annotatorlar uchun yo‘riqnoma, ikki tomonlama mustaqil kodlash, kelishmovchiliklarni ekspert orqali hal qilish va annotatorlararo moslikni statistik nazorat qilish metodikaning ishonchliligini oshiradi. Shu tarzda tashkil qilingan korpus terminografik lug‘at tuzish, tarjima ekvivalentlarini normallashtirish, parallel va qiyosiy tadqiqotlar, mashina tarjimasi uchun terminologik resurs yaratish hamda keyingi avtomatik tahlil vazifalarida qayta foydalanish mumkin bo‘lgan ilmiy baza vazifasini bajaradi. Metodikaning asosiy afzalligi — har bir ilmiy qarorning manba, kontekst, mezon va kod orqali tekshirilishi mumkinligidadir; bu esa terminologik tadqiqot natijalarining shaffofligi, takrorlanuvchanligi va amaliy qiymatini oshiradi.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Abjalova M.A. Korpus lingvistikasi: uslubiy qo‘llanma. – Toshkent: Nodirabegim, 2022. – 110 b.
2. Zaxarov V.P., Mengliyev B.R., Hamroyeva Sh.M. Korpus lingvistikasi: o‘quv qo‘llanma. – Toshkent: Fan, 2021. – 185 b.
3. Hamroyeva Sh.M. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari: filol. fan. bo‘yicha falsafa doktori (PhD) diss. – Buxoro, 2018. – 265 b.
4. Abduraxmonova N.Z. O‘zbek tili elektron korpusining kompyuter modellari: filol. fan. doktori (DSc) diss. – Toshkent, 2021. – 219 b.
5. Karshiev A.B., Karimov S.A., Tursunov M.S. O‘zbek tili milliy korpusi uchun matnlarni formatlash // Raqamli texnologiyalarning nazariy va amaliy masalalari. – 2022. – № 1(1).
6. Kadirova Z.Z. O‘zbek tili korpusida terminlarning qidiruv va tahlil usullari // Modern Science and Research. – 2024. – Vol. 3, Issue 9. – B. 317–320.
7. Brezina V. Statistics in Corpus Linguistics: A Practical Guide. – Cambridge: Cambridge University Press, 2018. – 296 p. – DOI: 10.1017/9781316410899.
8. Doval I., Sánchez-Nieto M.T. (eds.). Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications. – Amsterdam; Philadelphia: John Benjamins, 2019. – Studies in Corpus Linguistics, Vol. 90.
9. Volk M. Innovations in parallel corpus alignment and retrieval // Parallel Corpora for Contrastive and Translation Studies / eds. I. Doval, M.T. Sánchez-Nieto. – Amsterdam; Philadelphia: John Benjamins, 2019. – P. 79–90. – DOI: 10.1075/scl.90.05vol.
10. Rigouts Terryn A., Hoste V., Lefever E. In no uncertain terms: a dataset for monolingual and multilingual automatic term extraction from comparable corpora // Language Resources and Evaluation. – 2020. – Vol. 54. – P. 385–418. – DOI: 10.1007/s10579-019-09453-9.
11. ISO 704:2022. Terminology work — Principles and methods. – Geneva: International Organization for Standardization, 2022.
12. ISO 30042:2019. Management of terminology resources — TermBase eXchange (TBX). – Geneva: International Organization for Standardization, 2019.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.