

## **O‘ZBEK TILSHUNOSLIGIGA OID DISSERTATSIYALAR BIBLIOGRAFIK KORPUSINI YARATISH METODOLOGIYASI**

Gulyamova Shaxnoza Kaxramonovna  
Toshkent Amaliy fanlar universiteti

**Annotatsiya:** Maqolada 1940-2025-yillarda o‘zbek tilshunosligi va unga aloqador ixtisosliklarda yaratilgan 2 218 ta dissertatsiya hamda avtoreferat haqidagi bibliografik ma’lumotlarni maxsus korpusga aylantirish metodologiyasi ishlab chiqildi. Tarixiy davrlar davomida o‘zgargan ilmiy daraja nomlari, ixtisoslik shifrlari, tashkilot nomlari, yozuv tizimlari va bibliografik tavsif shakllarini yagona tizimga keltirish muammolari tahlil qilindi. Manbalarni inventarizatsiya qilish, yozuvlarni strukturaviy maydonlarga ajratish, dublikatlarni aniqlash, shaxs va tashkilot nomlarini normallashtirish, obyektlarni disambiguatsiya qilish, barqaror identifikatorlar berish, analitik annotatsiyani amalga oshirish, qidiruv indeksini yaratish va sifat nazoratini o‘tkazishdan iborat sakkiz bosqichli algoritm taklif qilindi. Birlamchi yozuvni daxlsiz saqlash, kiritilgan o‘zgarishlarni versiyalash va har bir ma’lumotning manbasini ko‘rsatish korpus ishonchliligining asosiy shartlari sifatida belgilandi. Metodologiya bibliografik ma’lumotlarning tarixiy xususiyatlarini saqlagan holda ularni zamonaviy qidiruv va ilmiy tahlil talablariga moslashtirish imkonini beradi.

**Kalit so‘zlar:** bibliografik korpus, metodologiya, dissertatsiya, normallashtirish, disambiguatsiya, identifikator, dublikat, metama’lumot, sifat nazorati, raqamli bibliografiya

## **METHODOLOGY FOR CREATING A BIBLIOGRAPHIC CORPUS OF DISSERTATIONS ON UZBEK LINGUISTICS**

Gulyamova Shakhnoza Kakhramonovna  
Tashkent University of Applied Sciences

**Abstract:** The article presents a methodology for creating a bibliographic corpus containing records of 2,218 dissertations and dissertation abstracts in Uzbek linguistics and related fields from 1940 to 2025. It examines the challenges of harmonizing historically changing academic degree titles, speciality codes, institutional names, writing systems and bibliographic description practices. An eight-stage workflow is proposed, covering source inventory, record segmentation, duplicate detection, normalization of personal and institutional names, entity disambiguation, analytical annotation, search indexing and quality control.

Preservation of original records, versioning of changes and documentation of data provenance are defined as essential conditions for corpus reliability. The methodology makes it possible to preserve the historical features of bibliographic data while adapting them to contemporary retrieval and research requirements.

**Keywords:** bibliographic corpus, methodology, dissertation, normalization, disambiguation, identifier, duplicate, metadata, quality control, digital bibliography

Kirish. Ilmiy merosni raqamlashtirish mavjud manbalarni elektron shaklga ko‘chirish bilangina yakunlanmaydi. Raqamli muhitda saqlanayotgan ma’lumot muayyan model asosida tuzilmasa, uning qismlari mashina tomonidan farqlanmasa va variant shakllari o‘zaro bog‘lanmasa, undan keng qamrovli ilmiy tahlilda foydalanish qiyin bo‘ladi. Shu sababli yirik bibliografik majmualarni korpusga aylantirish alohida metodologik yondashuvni talab qiladi.

1940-2025-yillardagi o‘zbek tilshunosligiga oid dissertatsiya va avtoreferatlar bibliografiyasi turli davrlar hamda ilmiy tizimlarga mansub ma’lumotlarni birlashtiradi. Undagi yozuvlar yagona tarixiy sharoitda yaratilmagan. Ilmiy darajalar, ixtisoslik shifrlari, muassasalar, bibliografik tavsif talablari va yozuv tizimlarining o‘zgarishi ma’lumotlarning strukturaviy jihatdan notekis bo‘lishiga olib kelgan.

Bibliografiyadagi 2 218 ta yozuvning 1 200 tasi o‘zbek tili, 184 tasi til nazariyasi, amaliy va kompyuter lingvistikasi, 749 tasi qiyosiy va chog‘ishtirma tilshunoslik hamda tarjimashunoslik, 50 tasi Yevropa, Amerika va Avstraliya xalqlari tili, 35 tasi Osiyo va Afrika xalqlari tili va adabiyoti bo‘yicha bajarilgan tadqiqotlarni qamrab oladi [2]. Bunday keng qamrovli materialni korpuslashtirish manbalarni yig‘ish, tekshirish, tozalash, normallashtirish va annotatsiyalash jarayonlarini yagona tizim asosida amalga oshirishni taqozo qiladi.

Bibliometriya yozma ilmiy kommunikatsiyaning miqdoriy jihatlarini matematik va statistik metodlar yordamida o‘rganish sohasi sifatida shakllangan [3: 348-349]. Biroq ishonchli bibliometrik natijaga erishish uchun dastlab tahlil qilinadigan bibliografik ma’lumotlarning to‘liqligi, bir xilligi va aniqligi ta’minlanishi kerak. Muallif yoki tashkilot nomining turli shakllarda berilishi, bir tadqiqotning takroriy qayd etilishi va ixtisosliklar tasnifidagi tarixiy o‘zgarishlar hisobga olinmasa, statistik natijalar ilmiy voqelikni to‘g‘ri aks ettirmaydi.

Tadqiqotning maqsadi o‘zbek tilshunosligiga oid dissertatsiyalar bibliografik korpusini yaratishning manbalarni inventarizatsiya qilishdan sifat nazoratigacha bo‘lgan izchil metodologiyasini ishlab chiqishdan iborat.

Tadqiqot materiali va metodlari. Tadqiqot materialini “XX-XXI asr o‘zbek tilshunosligiga oid dissertatsiya va avtoreferatlar bibliografiyasi (1940-2025-yillar)”ga kiritilgan 2 218 ta bibliografik yozuv tashkil etdi. Yozuvlarda muallif, mavzu, ixtisoslik, ilmiy daraja, rahbar yoki maslahatchi, tashkilot, himoya joyi va

sanasi, sahifa miqdori, UDK yoki UO‘K, inventar raqami hamda manbani izlash manzili kabi ma’lumotlar aks etgan.

Tadqiqotda bibliografik tavsif, qiyosiy tahlil, strukturaviy ajratish, ma’lumotlarni normallashtirish, obyektlarni identifikatsiyalash, konseptual modellashtirish va sifat nazorati metodlari qo‘llandi. Bibliografik yozuvlarning tarixiy shakllarini qiyoslash orqali o‘zgaruvchan va barqaror elementlar aniqlandi.

Metama’lumotlarni shakllantirishda Dublin Core elementlar tizimi, bibliografik obyektlar o‘rtasidagi munosabatlarni belgilashda IFLA Library Reference Model, matniy ma’lumotlarni formal ifodalashda TEI tavsiyalari va raqamli ilmiy ma’lumotlarni boshqarishda FAIR tamoyillarining umumiy qoidalari e’tiborga olindi [1; 4; 5; 6].

Mazkur standartlar tayyor sxema sifatida aynan ko‘chirilmadi. Chunki o‘zbek dissertatsion amaliyotida ilmiy rahbar, ilmiy maslahatchi, ixtisoslik shifri, dissertatsiya kengashi, attestatsiya holati va manbani izlash manzili singari maxsus maydonlarni ifodalash zarur.

Bibliografik ma’lumotlarning metodologik muammolari mavjud. Jumladan:

1. Manbalarning tarqoqligi. Dissertatsiyalar haqidagi ma’lumotlar yagona fond yoki platformada saqlanmaydi. Bir qismi ilmiy kutubxonalarda, yana bir qismi universitetlar fondlarida, elektron saytlarda, dissertatsiya kengashlari e’lonlarida yoki avtoreferatlarda uchraydi. Ayrim manbalarda dissertatsiyaning to‘liq bibliografik tavsifi berilgan bo‘lsa, boshqalarida faqat muallif, mavzu va yil ko‘rsatilishi mumkin. Shuning uchun korpusga kiritilgan har bir ma’lumotning kelib chiqish manbasi qayd etilishi kerak. “Manba” maydoni faqat adabiyot nomini emas, fond, inventar raqami, elektron manzil, kirish sanasi va tekshirilganlik holatini ham qamrab olishi lozim.

2. Bibliografik tavsiflarning turlicha shakllanishi. 1940-2025-yillarda bibliografik tavsif me’yorlari bir necha bor o‘zgargan. Dastlabki davr manbalarida ism va familiyalar rus tilidagi shaklda, keyingi manbalarda o‘zbek kirill yoki lotin yozuvida berilgan. Ayrim yozuvlarda ilmiy rahbar to‘liq ko‘rsatilgan, boshqalarida esa bu ma’lumot mavjud emas.

Ma’lumotning mavjud emasligi bilan uni aniqlash imkoniyati bo‘lmagan holat o‘zaro farqlanishi kerak. Bo‘sh katak ma’lumotning kiritilmaganini anglatishi mumkin. Shu sababli “ma’lumot mavjud emas”, “manbada ko‘rsatilmagan”, “aniqlanmagan” va “tatbiq etilmaydi” kabi holatlar alohida kodlanishi zarur.

3. Takroriy yozuvlar. Bir dissertatsiya to‘liq matn, avtoreferat, elektron nusxa va katalog kartochkasi shaklida bir necha marta qayd etilishi mumkin. Muallif ismining yoki mavzuning yozilishidagi kichik farq bunday qaydlarning avtomatik ravishda alohida tadqiqot sifatida qabul qilinishiga sabab bo‘ladi.

Takroriy yozuvlarni aniqlashda faqat sarlavhaning aynan mos kelishiga tayanish mumkin emas. Muallif nomi, mavzu, himoya yili, ilmiy daraja, ixtisoslik, tashkilot va

ilmiy rahbar belgilarining umumiy mosligi tekshirilishi kerak. Shubhali yozuvlar avtomatik o'chirilmaydi, balki ekspert tekshiruviga yuboriladi.

4. Tarixiy o'zgaruvchanlik. Ilmiy daraja, ixtisoslik va tashkilot nomlari tarixiy kategoriyalardir. Ularni zamonaviy nomlarga mexanik almashtirish manbaning tarixiy xususiyatini yo'qotadi. Aksincha, barcha tarixiy nomlarni alohida va bog'lanmagan birlik sifatida saqlash qiyosiy tahlilni murakkablashtiradi. Shu sababli asl qiymat, normallashtirilgan qiymat va analitik kategoriya alohida maydonlarda saqlanishi kerak. Bunda manbaning tarixiy shakli daxlsiz qoladi, lekin turli davrlarga oid ma'lumotlarni umumiy mezon asosida qidirish imkoniyati yaratiladi.

Bibliografik korpus yaratishning asosiy tamoyillari:

1. Birlamchi ma'lumot daxlsizligi. Korpusga kiritilgan manba yozuvi tahrirsiz shaklda alohida qatlamda saqlanishi lozim. Imlo xatosi, tarixiy yozuv shakli yoki bibliografik nomuvofiqlik aniqlansa ham, asl yozuv o'zgartirilmaydi. Tuzatilgan yoki normallashtirilgan shakl alohida maydonda beriladi. Bu tamoyil tadqiqotchiga korpusdagi qayta ishlangan ma'lumotni uning dastlabki manbasi bilan qiyoslash imkonini beradi. Birlamchi qatlam saqlanmasa, keyinchalik kiritilgan o'zgarishning to'g'riligini tekshirish qiyinlashadi.

2. Ma'lumot manbasini ko'rsatish. Har bir yozuv va, zarur hollarda, alohida maydon uchun ma'lumotning qaysi manbadan olingani ko'rsatiladi. Masalan, himoya sanasi avtoreferatdan, inventar raqami kutubxona katalogidan, elektron manzil esa universitet saytidan olingan bo'lishi mumkin. Bir-biriga zid ma'lumotlar uchraganda ishonchlilik iyerarxiyasi qo'llanadi. Dissertatsiyaning titul varag'i va rasmiy avtoreferat birlamchi, kutubxona katalogi ikkilamchi, boshqa bibliografik ko'rsatkich yoki norasmiy elektron qayd yordamchi manba sifatida baholanishi mumkin.

3. Barqaror identifikatsiya. Har bir dissertatsion tadqiqot, shaxs, tashkilot va hujjatga barqaror identifikator berilishi kerak. Identifikator obyekt nomi o'zgarsa ham o'zgarmaydi. Masalan, universitetning tarixiy nomi yangilanishi mumkin, lekin uning korpusdagi asosiy identifikatori saqlanib qoladi. Identifikatorlardan foydalanish ayni obyektga tegishli turli yozuvlarni birlashtirish, aloqalarni formal ko'rsatish va ma'lumotlarni boshqa raqamli tizimlarga eksport qilish imkonini beradi.

4. Versiyalash. Bibliografik korpus muntazam to'ldiriladigan va tuzatiladigan ochiq tizim bo'lishi kerak. Har bir o'zgarishning sanasi, muallifi, sababi va oldingi holati qayd etiladi. Yangi ma'lumot topilganda oldingi yozuv izsiz o'chirib yuborilmaydi. Versiyalash ilmiy tekshiriluvchanlikni ta'minlaydi. Muayyan tahlil korpusning qaysi versiyasi asosida bajarilgani ko'rsatilsa, uning natijasini qayta tekshirish mumkin bo'ladi.

5. Ma'lumotlarni normallashtirish. Normallashtirishning vazifasi tarixiy yoki grafik variantlarni yo'q qilish emas, ularni qidiruv jarayonida yagona obyekt bilan

bog'lashdir. Korpusda "asl qiymat - normallashtirilgan qiymat - variant shakllar" modeli qo'llanishi maqsadga muvofiq.

6. Shaxs nomlarini normallashtirish. Muallif va ilmiy rahbarlarning ism-familiyalari to'liq ism, bosh harflar, ruscha, o'zbek kirill va o'zbek lotin shakllarida berilishi mumkin. Bir xil bosh harf va familiyaga ega ikki shaxsni avtomatik ravishda birlashtirish ham, bir shaxsning turli yozilgan nomlarini alohida shaxs sifatida qabul qilish ham xato hisoblanadi.

Shaxsni aniqlashda ism-familiyadan tashqari uning muassasasi, ilmiy rahbari, tadqiqot mavzusi, himoya yili va keyingi ilmiy faoliyati haqidagi belgilar hisobga olinadi. To'liq ishonch hosil qilinmagan holat "ehtimoliy moslik" sifatida ekspert ko'rigiga chiqariladi.

7. Tashkilot nomlarini normallashtirish. Universitet va ilmiy muassasalarning tarixiy nomlari alohida saqlanadi. Har bir nom variantiga amal qilgan vaqt oralig'i biriktiriladi. Qayta nomlangan tashkilotning tarixiy shakli uning zamonaviy nomi bilan maxsus munosabat orqali bog'lanadi. Bunday model, bir tomondan, tarixiy bibliografik ma'lumotni o'zgartirmaydi; ikkinchi tomondan, tashkilot faoliyatining uzoq muddatli ko'rsatkichlarini yagona obyekt doirasida hisoblash imkonini beradi.

8. Ixtisosliklarni normallashtirish. Ixtisoslik shifrlari va nomlari vaqt o'tishi bilan o'zgargani sababli korpusda kamida ikki tasnif qatlamidan foydalanish kerak. Birinchi qatlamda dissertatsiya himoya qilingan davrdagi rasmiy shifr va nom saqlanadi. Ikkinchi qatlamda mazkur ixtisoslik umumiy analitik yo'nalish bilan bog'lanadi. Natijada foydalanuvchi muayyan tarixiy davrning rasmiy tasnifi bo'yicha ham, barcha davrlar uchun umumiy ilmiy yo'nalish bo'yicha ham qidiruv o'tkaza oladi.

9. Sarlavhalarni normallashtirish. Dissertatsiya sarlavhasi asl imlo va yozuv tizimida saqlanadi. Qidiruv uchun uning standartlashtirilgan grafik varianti, zarur hollarda transliteratsiyasi va lemmalashtirilgan shakli yaratiladi. Tinish belgilari, qo'shtirnoq, bosh harf va ayrim imlo variantlari qidiruv natijasini sun'iy ravishda ajratmasligi kerak. Sarlavhaning normallashtirilgan shakli dissertatsiya mavzularining leksik o'xshashligini aniqlashda qo'llanadi. Biroq u manbadagi asl sarlavhaning o'rnini egallamaydi.

Korpus yaratishning sakkiz bosqichli algoritmi esa quyidagilarni qamrab oladi:

1-bosqich: manbalarni inventarizatsiya qilish. Dastlab bibliografik ma'lumotlarning barcha manbalari aniqlanadi. Kutubxona fondlari, arxivlar, universitet kataloglari, avtoreferatlar, dissertatsiya kengashi e'lonlari va elektron platformalar alohida ro'yxatga olinadi. Har bir manbaning turi, joylashuvi, qamrovi va ishonchlilik darajasi belgilanadi.

2-bosqich: yozuvlarni strukturaviy ajratish. Yig'ilgan bibliografik matn muallif, sarlavha, ilmiy daraja, ixtisoslik, rahbar, tashkilot, yil, sahifa, indeks va saqlanish

manzili kabi maydonlarga ajratiladi. Avtomatik ajratish natijasi inson tomonidan tekshiriladi, chunki tarixiy yozuvlarda rekvizitlarning joylashish tartibi doimo bir xil emas.

3-bosqich: dublikatlarni aniqlash. Muallif, mavzu, yil, daraja va tashkilot maydonlari asosida ehtimoliy dublikatlar aniqlanadi. To'liq mos kelgan qaydlar bitta dissertatsion tadqiqotga bog'lanadi. Qisman mos keladigan holatlar ekspert tekshiruviga yuboriladi. Takroriy qayd aniqlanganda uning manbasi o'chirib tashlanmaydi. U ayni tadqiqotga oid alohida katalog yoki hujjat yozuvi sifatida saqlanadi.

4-bosqich: normallashtirish va disambiguatsiya. Shaxslar, tashkilotlar, shaharlar, ilmiy darajalar va ixtisosliklar uchun normallashtirilgan shakllar yaratiladi. Bir xil nomli turli shaxslarni ajratish va turli nomlarda qayd etilgan bitta obyektni birlashtirish jarayoni amalga oshiriladi.

5-bosqich: bibliografik verifikatsiya. Majburiy maydonlarning to'liqligi tekshiriladi. Yetishmayotgan ma'lumotlar ishonchli birlamchi yoki ikkilamchi manbalardan to'ldiriladi. Bir-biriga zid ma'lumotlar mavjud bo'lsa, ularning har ikkisi manbasi bilan qayd etilib, asosiy deb qabul qilingan variantning tanlanish sababi ko'rsatiladi.

6-bosqich: lingvistik va mazmuniy annotatsiya. Dissertatsiyalarga tadqiqot obyekti, til sathi, paradigma, metod, til, material, davr, hudud va amaliy natija bo'yicha teglar beriladi. Teglar sarlavha, annotatsiya, avtoreferat yoki to'liq matndan olingan dalilga asoslanadi. Tegning manbasi va ishonchlilik darajasi qayd etiladi.

7-bosqich: indekslash va qidiruv tizimini yaratish. Normallashtirilgan maydonlar va annotatsiya teglari qidiruv indeksiga kiritiladi. Oddiy qidiruv muallif va sarlavha bo'yicha, kengaytirilgan qidiruv esa yil, daraja, ixtisoslik, tashkilot, rahbar, obyekt, til sathi va metod kabi belgilar kombinatsiyasi bo'yicha ishlaydi. Ma'lumotlarning ochiq almashinuvi uchun ularni JSON-LD, XML yoki boshqa mashina o'qiydigan formatlarda eksport qilish imkoniyati nazarda tutiladi.

8-bosqich: sinov va sifat nazorati. Korpus dastlab turli yillar va ixtisosliklarni qamrab olgan kichik namuna asosida sinovdan o'tkaziladi. Sinov orqali maydonlarning yetarliligi, teglar chegarasining aniqligi, qidiruv so'rovlarining ishlashi va foydalanuvchi duch keladigan muammolar aniqlanadi. Model aniqlashtirilgandan keyin u to'liq bibliografik majmuaga tatbiq etiladi. Bunday bosqichlilik katta hajmdagi ma'lumotni noto'g'ri model asosida qayta ishlash xavfini kamaytiradi.

Bibliografik korpusning barcha qismini bitta texnik formatda saqlash shart emas. Bibliografik maydonlar relatsion yoki hujjatga yo'naltirilgan ma'lumotlar bazasida, matniy fragmentlar TEI/XML formatida, obyektlar o'rtasidagi aloqalar esa JSON-LD yoki RDF shaklida ifodalanishi mumkin.

TEI tavsiyalari bibliografik yozuv, shaxs, tashkilot, joy va sana birliklarini formal belgilash imkonini beradi [5]. JSON-LD yordamida "shaxs - dissertatsiya muallifi", "olim - ilmiy rahbar", "dissertatsiya - muassasada bajarilgan" va "fayl - tadqiqotning avtoreferati" kabi munosabatlar bog'langan ma'lumot shaklida ifodalanadi.

Korpus ma'lumotlari quyidagi to'rt qatlamda saqlanishi maqsadga muvofiq: 1) manbadan olingan o'zgarmas birlamchi yozuv; 2) strukturaviy qismlarga ajratilgan va normallashtirilgan bibliografik ma'lumot; 3) lingvistik va mazmuniy annotatsiya; 4) qidiruv hamda tahlil uchun indekslangan ma'lumot. Ushbu arxitektura birlamchi manba bilan tahrirlangan ma'lumotni aralashtirib yubormaydi. Shuningdek, keyingi davrda annotatsiya tizimi kengaytirilsa, dastlabki bibliografik yozuvni qayta yaratishga ehtiyoj qolmaydi.

Bibliografik korpusning sifati yozuvlar soni bilangina emas, ularning to'liqligi, aniqligi, izchilligi va tekshiriluvchanligi bilan baholanadi. Sifat nazorati korpus tayyor bo'lgandan keyin bajariladigan bir martalik ish emas, balki barcha bosqichlarni qamrab oluvchi uzluksiz jarayon bo'lishi kerak.

To'liqlik majburiy maydonlarning qay darajada to'ldirilganini anglatadi. Har bir yozuvda muallif, mavzu, daraja, ixtisoslik va yil kabi asosiy ma'lumotlar mavjudligi tekshiriladi. Aniqlik ma'lumotning birlamchi manbaga mosligini bildiradi. Tasdiqlanmagan ma'lumot qat'iy fakt sifatida berilmaydi. Izchillik bir xil turdagi ma'lumotlarning yagona qoida asosida shakllantirilishini anglatadi. Sana, ilmiy daraja, ixtisoslik shifri va til kodlari bir xil formatda saqlanishi kerak. Bir ma'nolilik bitta identifikatorning faqat bitta obyektga tegishli bo'lishini talab qiladi. Bir shaxs yoki tashkilot uchun takroriy identifikatorlarning yaratilishi statistik tahlilni buzadi. Tekshiriluvchanlik har bir yozuvni uning manbasi va tahrir tarixiga bog'lash imkoniyatini anglatadi. Foydalanuvchi ma'lumot qayerdan olingani va qachon yangilanganini ko'ra olishi kerak.

Formal maydonlarning to'liqligi, sanalar formati, shifrlar tuzilishi va majburiy identifikatorlar avtomatik tekshirilishi mumkin. Murakkab mazmuniy teglar esa ekspert tomonidan tanlama yoki to'liq tekshiruvdan o'tkaziladi.

Raqamli ilmiy ma'lumot topiladigan, foydalanish shartlari aniq, boshqa tizimlar bilan muvofiq ishlaydigan va qayta foydalanishga yaroqli bo'lishi kerak. Ushbu talablar FAIR tamoyillarida umumlashtirilgan [6: 1-9].

Bibliografik korpusda ochiq metama'lumot, barqaror identifikator, versiya raqami, eksport formati, foydalanish sharti va manba ko'rsatilishi lozim. Dissertatsiyalarning to'liq elektron nusxalarini ochiq taqdim etish mualliflik huquqi yoki fond qoidalari bilan cheklanishi mumkin. Biroq bibliografik tavsif, hujjatning saqlanish joyi va undan foydalanish tartibi qonuniy doirada ko'rsatilishi mumkin. Korpusning muntazam yangilanib borishi ham metodologiyaning bir qismi

hisoblanadi. Yangi himoya qilingan dissertatsiyalarni kiritish, aniqlangan xatolarni tuzatish va yangi manbalarni bog'lash uchun mas'uliyat, davriylik va tahrir qoidalari oldindan belgilanishi kerak.

Xulosa. 1940-2025-yillarda o'zbek tilshunosligi va unga aloqador ixtisosliklarda yaratilgan 2 218 ta dissertatsiya hamda avtoreferat haqidagi bibliografik ma'lumotlar maxsus ilmiy korpus yaratish uchun yetarli empirik asosni tashkil etadi. Biroq ushbu majmuani korpusga aylantirish ma'lumotlarni elektron shaklga ko'chirish bilan cheklanmaydi. Bibliografik ma'lumotlarning tarixiy va strukturaviy notekisligi birlamchi yozuvni saqlash, variantlarni normallashtirish, obyektlarni identifikatsiya qilish, dublikatlarni aniqlash va har bir ma'lumotning manbasini ko'rsatishga asoslangan metodologiyani talab qiladi.

Tadqiqotda manbalarni inventarizatsiya qilish, yozuvlarni strukturaviy maydonlarga ajratish, dublikatlarni aniqlash, normallashtirish va disambiguatsiya, bibliografik verifikatsiya, lingvistik annotatsiya, qidiruv indeksini yaratish hamda sifat nazoratini o'tkazishni qamrab oluvchi sakkiz bosqichli algoritm taklif etildi.

Mazkur metodologiya bibliografik ma'lumotning tarixiy shaklini saqlash bilan birga, uni zamonaviy qidiruv, statistik tahlil va raqamli ilmiy kommunikatsiya talablariga moslashtiradi. Natijada korpus o'zbek tilshunosligi taraqqiyotini o'rganish, ilmiy maktablar va mavzuviy yo'nalishlarni aniqlash, tadqiqot takrorlarini kuzatish hamda istiqbolli ilmiy muammolarni belgilashga xizmat qiladigan ishonchli raqamli resursga aylanishi mumkin.

### **Foydalanilgan adabiyotlar**

1. ISO 15836-1:2017. Information and Documentation - The Dublin Core Metadata Element Set - Part 1: Core Elements. - Geneva: International Organization for Standardization, 2017.
2. Mengliyev B., Gulyamova Sh. XX-XXI asr o'zbek tilshunosligiga oid dissertatsiya va avtoreferatlar bibliografiyasi (1940-2025-yillar). - Toshkent, 2026. - 504 b.
3. Pritchard A. Statistical Bibliography or Bibliometrics? // Journal of Documentation. - 1969. - Vol. 25, No. 4. - P. 348-349.
4. Riva P., Le Boeuf P., Žumer M. IFLA Library Reference Model: A Conceptual Model for Bibliographic Information. - Den Haag: IFLA, 2017. - 101 p.
5. TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.10.0. - 2025.
6. Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship // Scientific Data. - 2016. - Vol. 3. - Article 160018. DOI: 10.1038/sdata.2016.18.

7. Donthu N., Kumar S., Mukherjee D., Pandey N., Lim W.M. How to Conduct a Bibliometric Analysis: An Overview and Guidelines // Journal of Business Research. - 2021. - Vol. 133. - P. 285-296. DOI: 10.1016/j.jbusres.2021.04.070.

8. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. - Cambridge: Cambridge University Press, 2012. - 294 p.