ISSN: 3030-3346

KO'P MANBALI BIBLIOMETRIK MA'LUMOTLARNI INTEGRATSIYALASHDA QAYTA ISHLANUVCHI HISOBLASH ARXITEKTURASI

Arabov Ubaydullo Hamroqul o‘g‘li

Toshkent davlat iqtisodiyot universiteti
E-mail: ubaydulloarabov9@gmail.com

K E Y W O R D S

scientometrik hisoblash, bibliometrik tarmoq modellashtirish, metadata uyg'unlashtirish, takrorlanishni aniqlash algoritmlari, qayta ishlanuvchi tadqiqot quvurlari, hisoblash scientometrikasi

A B S T R A C T

Zamonaviy ilmiy-tadqiqot jarayonlarida bibliometrik tahlil vositasi sifatida bir nechta ma'lumotlar bazasidan olingan metadata'larni birlashtirish sezilarli methodologik muammolarga sabab bo'lmoqda. Web of Science va Scopus kabi yetakchi indekslash tizimlari o'rtasidagi metadata sxemasi, identifikator tuzilmasi va eksport formatlari farqlari tarmoq asosidagi bibliometrik tahlillar sifatini pasaytirishi mumkin. Ushbu maqolada ko'p manbali scientometrik tarmoqlarni qurish uchun qayta ishlanuvchi hisoblash quvuri taklif etiladi. Arxitektura metadata uyg'unlashtirish, deterministik takrorlanishni aniqlash, siyrak grafik qurilishi, normallashtirish va tuzilmaviy diagnostika bosqichlarini mustaqil hisoblash modullariga ajratadi. Maqolaning asosiy hissasi qayta ishlanuvchi scientometrik tahlil uchun shaffof va kengaytiriladigan hisoblash arxitekturasini shakllantirish hamda uni O'zbekiston oliy ta'lim muassasalari kontekstida qo'llash imkoniyatlarini tahlil qilishdan iborat.

KIRISH

Ilmiy nashrlar sonining jadal o'sishi bibliometrik tahlilni hisoblash jihatidan sermashaqqat soha sifatida shakllantirmoqda. Bu soha katta hajmdagi metadata qayta ishlash va tarmoq modellashtirish usullariga tayanadi [1, s. 959-975; 2, s. 1382-1402]. Web of Science (WoS) va Scopus kabi yirik bibliografik bazalar tuzilmali metadata'larni taqdim etib, iqtibos tarmoqlari, hamkorlik grafiklari va kalit so'zlar hamkorlik tuzilmalari orqali ilmiy bilimlarni miqdoriy tahlil qilish imkonini beradi [7, s. 523-538; 8, s. 1-40].

Tarmoq asosidagi bibliometrik yondashuvlar tadqiqot sohalarini xaritalash, yangi ilmiy mavzularni aniqlash va fanlararo hamkorlik namunalarini tahlil qilish uchun keng qo'llaniladi [1, s. 959; 9, s. 177-200]. Ushbu yondashuvlar odatda hujjatlar, mualliflar yoki kalit so'zlarni iqtibos yoki hamkorlik munosabatlari orqali bog'langan tugunlar sifatida ifodalaydi va ilmiy bilimlarning tuzilmaviy tashkil etilishini o'rganishda grafik-nazariy usullarni qo'llash imkonini beradi [12, s. 558-577; 13, s. P10008].

Biroq turli indekslash platformalaridan ma'lumotlarni birlashtirish zaruriyati tobora kuchayib borayotgan bir paytda, bibliometrik tahlillar metodologik qiyinchiliklarga duch kelmoqda. Metadata sxemalari, identifikator qamrovi, eksport formatlari va tasniflash tizimlari o'rtasidagi farqlar ma'lumotlar izchilligiga va keyingi tarmoq tahlillariga ta'sir qilishi mumkin bo'lgan tuzilmaviy heterogenlikni keltirib chiqaradi [16, s. 715-731; 17, s. 5613-5631; 18, s. 5191-5225].

Ushbu nomuvofiqliklar o'zaro ma'lumotlar bazasi integratsiyasini murakkablashtiradi va takroriy yozuvlar, fragmentlashgan muallif identichliklariga hamda izchilsiz kalit so'zlar ifodalanishiga olib kelishi mumkin [16, s. 720; 22, s. 21923-21956]. Bu muammolar tarmoq qurilishiga tarqalib, zichlik, markaziylik tarqalishlari va jamiyat tuzilmasi kabi tuzilmaviy ko'rsatkichlarga sezilarli ta'sir qiladi [8, s. 25-30; 24, s. 42-68; 25, s. 349-373].

O'zbekiston oliy ta'lim tizimida ham ilmiy faoliyatni baholash tobora muhim ahamiyat kasb

(cc) BY

etmoqda. Oliy ta'lim, fan va innovatsiyalar vazirligi tomonidan joriy etilgan reyting tizimi xalqaro bazalardagi nashrlarni asosiy ko'rsatkich sifatida belgilagan. Bunday sharoitda turli bazalardan olingan ma'lumotlarni to'g'ri birlashtirish va takrorlanishni aniqlash metodologiyasi O'zbekiston universitetlari uchun ham dolzarb masalaga aylanmoqda.

Ushbu muammolarni hal etish uchun mazkur maqolada ko'p manbali scientometrik tarmoqlarni qurish uchun qayta ishlanuvchi hisoblash quvuri taklif etiladi. Tadqiqotning maqsadi geterogen bibliografik ma'lumotlarni izchil integratsiyalashtirishga imkon beradigan va natijada hosil bo'lgan tarmoqlarning tuzilmaviy xossalarini saqlab qoladigan modulli oldindan ishlov berish freymvorkini loyihalash va baholashdan iborat.



1-rasm. Ko'p manbali bibliometrik tarmoq qurilishida tuzilmaviy buzilish manbalari va taklif etilgan yechim

1-rasmdagi sxema ko'p manbali integratsiya jarayonida yuzaga keladigan asosiy muammolarni va ularni bartaraf etishning konseptual yo'lini aks ettiradi. Taklif etilgan quvur har bir bosqichni mustaqil, qayta ishlanuvchi va shaffof ravishda baholanadigan komponent sifatida ajratib ko'rsatadi.

Scientometrik tahlil vositalarining rivojlanishi VOSviewer va bibliometrix kabi dasturiy muhitlarning paydo bo'lishini ta'minladi [1, s. 959-975; 7, s. 523-538]. Bu tizimlar iqtibos tuzilmalari, hamkorlik tarmoqlari va kalit so'zlar hamkorlik namunalarini o'rganish uchun keng qo'llaniladi. Biroq mavjud vositalarning aksariyati ichki izchil metadata kirishini faraz qiladi va asosan vizualizatsiya hamda tadqiqotchi tahlilga yo'naltirilgan [4, s. 201-231].

Mavjud bibliometrik dasturiy muhitlarning asosiy cheklovi oldindan ishlov berish operatsiyalarini yashirin tarzda hal qilishdan iborat. Metadata uyg'unlashtirish, takrorlanishni bartaraf etish va sxema normallashtirish kabi vazifalar odatda vositaga xos ish jarayonlariga o'rnatib qo'yilgan bo'lib, bu geterogen bibliografik ma'lumotlar to'plamlarini birlashtirish jarayonida shaffoflik va qayta ishlanishni cheklaydi [32, s. 240-257; 33, s. e52967].

Kumpulainen va Seppanen (2022) iqtibos asosida adabiyot o'rganishda WoS va Scopus ma'lumotlar to'plamlarini birlashtirish metodologiyasini tahlil qilib, kross-baza integratsiyasida 5-15% takroriy yozuvlar bo'lishi mumkinligini aniqlagan [17, s. 5613-5631]. Nowakowska (2025) bibliometrik tahlil uchun ma'lumotlarni oldindan qayta ishlashga keng qamrovli yondashuvni taklif etib, ma'lumotlar sifatining tarmoq ko'rsatkichlariga ta'sirini batafsil o'rgangan [18, s. 5191-5225].

Grafik nazariyasi bibliometrik tarmoqlarning tuzilmaviy xossalarini, jumladan ulanish namunalari, markaziylik tarqalishlari va jamiyat tuzilmalarini tahlil qilish uchun matematik asos beradi [8, s. 1-40]. Katta miqyosli ilmiy tarmoqlarda geterogen daraja tarqalishlari va modulli tashkilot empirik jihatdan izchil

(cc) BY

kuzatilgan bo'lib, bu tematik segmentatsiya va fanlik quyi sohalarini aks ettiradi [13, s. P10008].

Qayta ishlanuvchi hisoblash ish jarayonlarining ahamiyatini ta'kidlovchi so'nggi tadqiqotlar bibliometrik tahlil sohasida tobora kuchayib bormoqda [21, s. 2437; 29, s. 45-62; 34, s. 986]. Shaffof oldindan ishlov berish quvurlari ma'lumotlar integratsiyasi qarorlarining tarmoq topologiyasi va klasterlash natijalariga ta'sirini tizimli baholash imkonini beradi.

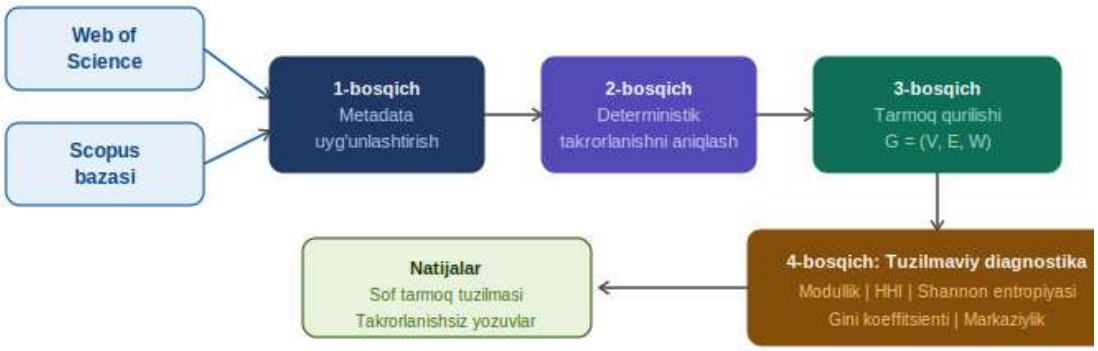
1-jadval.

Asosiy bibliometrik tahlil vositalarini qiyosiy tahlil

Vosita	Asosiy yo'nalish	Tarmoq tahlili	Vizualizatsiya	Qayta ishlanuvchi oldindan ishlov berish
VOSviewer	Ilm xaritalash	Ha	Ilg'or	Cheklangan
bibliometrix	Bibliometrik statistika	Ha	O'rtacha	Qisman
CiteSpace	Iqtibos portlashi	Ha	Ilg'or	Cheklangan
Taklif etilgan freymvork	Kross-baza integratsiya	Ha	Modulli	Aniq quvur

METODOLOGIYA

Ushbu bo'limda ko'p manbali scientometrik tarmoqlarni qurish uchun taklif etilgan hisoblash quvurining metodologiyasi tavsiflanadi. Yondashuv metadata uyg'unlashtirish, takrorlanishni aniqlash, tarmoq qurilishi va tuzilmaviy tahlilni o'z ichiga olgan ketma-ket qayta ishlash bosqichlari sifatida tuzilgan. Quyidagi 2-rasmda to'liq quvur arxitekturasi va uning hisoblash murakkabligi jamlangan tarzda ko'rsatilgan.



Hisoblash murakkabligi jadvali

Bosqich	Amaliyot	Murakkablik
Metadata uyg'unlashtirish	Maydonnormallashtirish	O(n)
Takrorlanishni aniqlash	DOI + o'xshashlik hisoblash	$O(n^2)$
Tarmoq qurilishi	Kalit so'z hamkorlik grafigi	O(nk)

2-rasm. Ko'p manbali bibliometrik hisoblash quvuri arxitekturasi va bosqichlarning hisoblash murakkabligi

Metodologiya metadata uyg'unlashtirishdan boshlanadi - bu Web of Science va Scopus kabi bibliografik bazalardagi sxema geterogenligini bartaraf etishga qaratilgan oldindan qayta ishlash bosqichidir. Kross-baza integratsiyasi muallif identifikatorlari, nashr sarlavhalari, kalit so'z tavsiflari va hujjat tasniflash kategoriyalarini o'z ichiga olgan geterogen metadata maydonlarini muvofiqlashtirini talab qiladi [35, s. 16; 36, s. 845-871].

(cc) BY

Quvur mustaqil bibliografik bazalardan metadata'larni o'zlashtirishdan boshlanadi. D_{WoS} va D_{Scopus} mos ravishda Web of Science va Scopus'dan eksport qilingan xom metadata to'plamlarini bildirsin. Ushbu manbalar turli maydon nomlaridan, identifikator konventsiyalaridan va eksport tuzilmalaridan foydalanganligi sababli sxema uyg'unlashtirilishi zarur [16, s. 720-725]. Rasmiy ravishda, geterogen metadata'ni yagona ifodalashga aylantiradigan H uyg'unlashtirish funksiyasi aniqlanadi:

$$D_{harm} = H\,(D_{WoS} \cup D_{Scopus}) \quad (1)$$

Bu transformatsiya DOI, muallif ismlari, jurnal sarlavhalari, nashr yili, hujjat turi va kalit so'zlarni o'z ichiga olgan asosiy metadata maydonlarini standartlashtiradi. Avvalgi tadqiqotlarga mos ravishda, tinish belgilari, bosh harflar va variant qisqartmalardan kelib chiqadigan leksik fragmentatsiyani kamaytirish uchun satr normalizatsiyasi qo'llaniladi [1, s. 961; 2, s. 1390-1395].

Metadata uyg'unlashtirish amalga oshirilgandan so'ng, metodologiyaning ikkinchi bosqichi takrorlanishni aniqlash masalasini hal etadi. Takroriy yozuvlar bir xil nashr bir nechta indekslash tizimida qisman izchilsiz metadata ifodalanishi bilan paydo bo'lishi mumkinligi sababli deterministik ikki bosqichli protsedura orqali bartaraf etiladi [19, s. 4573-4595; 22, s. 21930].

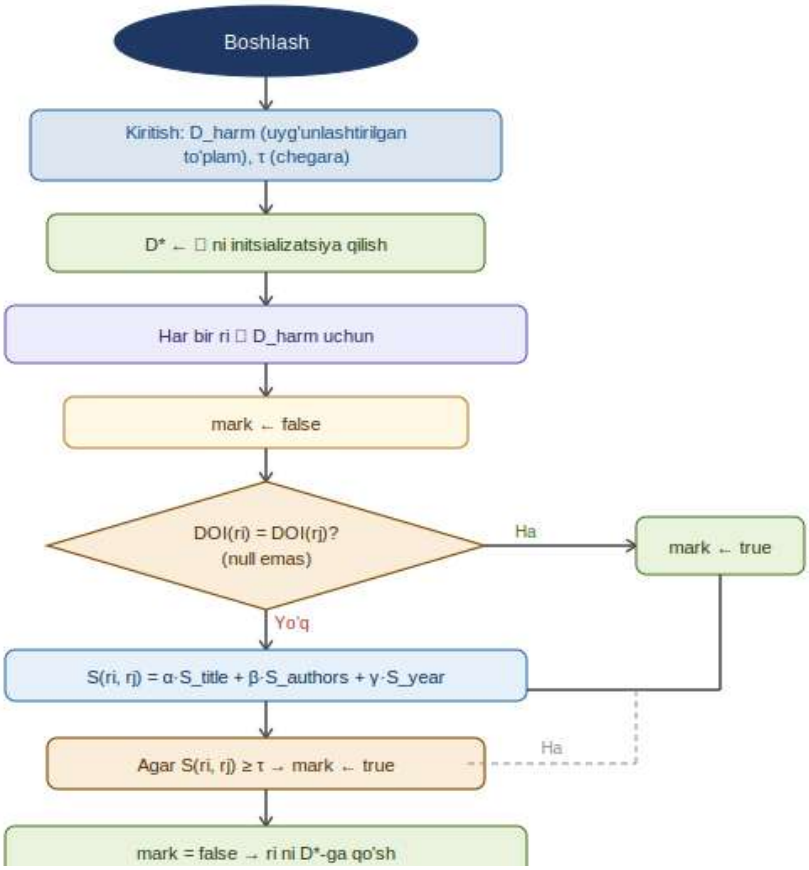
Birinchi bosqichda DOI kabi doimiy identifikatorlar yordamida aniq moslashuv amalga oshiriladi. Mavjud bo'lganda, DOI moslashuvi bibliografik yozuvlarni bog'lash uchun ishonchli mexanizmni ta'minlaydi, chunki u ilmiy nashrlarga berilgan global noyob identifikatorni ifodalaydi [25, s. 354-360]. Biroq ba'zi yozuvlarda DOI metadata'si yo'q yoki izchilsiz formatlangan bo'lishi mumkin. Bunday hollarda sarlavha o'xshashligi, muallif qoplamasi va nashr yili kelishuvini birlashtiradigan kompozit o'xshashlik bali hisoblanadi:

$$S(r_i, r_j) = \alpha \cdot S_{title} + \beta \cdot S_{authors} + \gamma \cdot S_{year} \quad (2)$$

bu yerda Stitle normallashtirilgan sarlavha o'xshashligini, Sauthors muallif ro'yxatining qoplamasi, Syear esa nashr yili kelishuvini o'lchaydi [23, s. 9799-9818; 38, s. 1145-1186]. Koeffitsientlar quyidagi shartni qondiradi:

$$\alpha + \beta + \gamma = 1 \quad (3)$$

Yozuvlar jufti oldindan belgilangan τ chegara erishilganda yoki undan oshilganda takroriy sifatida tasniflanadi. 1-algoritm takrorlanishni aniqlash tartibini aks ettiradi.



3-rasm. Deterministik takrorlanishni aniqlash tartibining blok-sxemasi

(cc) BY

Hisoblash murakkabligi nuqtai nazaridan eng yomon holat $O(n^2)$ juftlik taqqoslashlariga to'g'ri keladi. Biroq DOI moslashuvi amalda o'xshashlik hisoblashlar sonini sezilarli darajada kamaytiradi, bu esa yondashuv aniqlik va samaradorlik o'rtasidagi muvozanatni saqlashiga imkon beradi.

Takrorlanishni aniqlash bibliografik ma'lumotlarni, xususan geterogen indekslash platformalaridan olingan yozuvlarni birlashtirish jarayonida taniqli muammo hisoblanadi. Mavjud yondashuvlar odatda evristik moslashuv yoki ehtimoliy yozuvlarni bog'lash strategiyalariga tayanadi [17, s. 5618-5622; 38, s. 1150-1162].

2-jadval.

Takrorlanishni aniqlash yondashuvlarini qiyosiy tahlil

Usul	Moslashuv strategiyasi	Aniqlik	Hisoblash samaradorligi
VOSviewer evristikasi	Sarlavha o'xshashligi va metadata evristikasi	O'rtacha	Yuqori
Ehtimoliy yozuv bog'lanishi	Statistik o'xshashlik modellari	Yuqori	O'rtacha
Taklif etilgan yondashuv	DOI + kompozit o'xshashlik kombinatsiyasi	Yuqori	Yuqori

Takroriy yozuvlar bartaraf etilgandan so'ng, natijada olingan ma'lumotlar to'plami bibliometrik tarmoqni qurish uchun ishlatiladi. Bibliometrik tarmoqlar tugunlar (kalit so'zlar) va qirralar (hamkorlik munosabatlari) orqali ifodalangan graflar sifatida modellanadi [7, s. 525-530; 8, s. 168-200]. Bibliometrik tarmoq rasmiy ravishda quyidagicha aniqlanadi:

$$G = (V, E, W) \quad (4)$$

bu yerda V tugunlar to'plami, E qirralar to'plami va W qirra og'irliklari. Juda ko'p uchraydigan atamalar ta'sirini kamaytirish uchun assotsiatsiya kuchi normalizatsiyasi qo'llaniladi:

$$a_{ij} = c_{ij} / (f_i * f_j) \quad (5)$$

bu yerda f_i va f_j chekka chastotalar. Bu normalizatsiya interpretatsiya qilishni yaxshilaydi va chastota xolisligini kamaytiradi [7, s. 530-533; 39, s. 1178-1195].

Metodologiyaning yakuniy bosqichida natijada olingan tarmoq topologiyasini tavsiflash uchun tuzilmaviy diagnostika hisoblanadi. Umumiy ko'rsatkichlar zichlik, daraja markaziylik, modullik va jamiyat tuzilmasini o'z ichiga oladi [8, s. 350-400; 13, s. P10008]. Bundan tashqari, Herfindahl-Hirshman indeksi (HHI), Shannon entropiyasi va Gini koeffitsientidan iborat konsentratsiya o'lchovlari tarmoq ichidagi tuzilmaviy tengsizlikni miqdoriy ifodalash uchun foydalaniladi [40, s. 955-967; 41, s. 1-32].

EKSPERIMENTAL NATIJALAR

Eksperimental tahlil 1990-2023 yillarni qamrab olgan 317 ta ilmiy nashrdan iborat ko'p fanli ma'lumotlar to'plami asosida o'tkazildi. Ma'lumotlar to'plami 3.1-bo'limdagi uyg'unlashtirish tartibini qo'llaganidan so'ng Web of Science va Scopus'dan yozuvlarni birlashtirish orqali tuzildi. Ushbu ma'lumotlar to'plamining tanlovi geterogen metadata xususiyatlariga ega ko'p fanli tadqiqot sohasini ifodalash zaruriyatidan kelib chiqqan.

3-jadval.

Eksperimental ma'lumotlar to'plami xususiyatlari jamlangan ko'rinishi

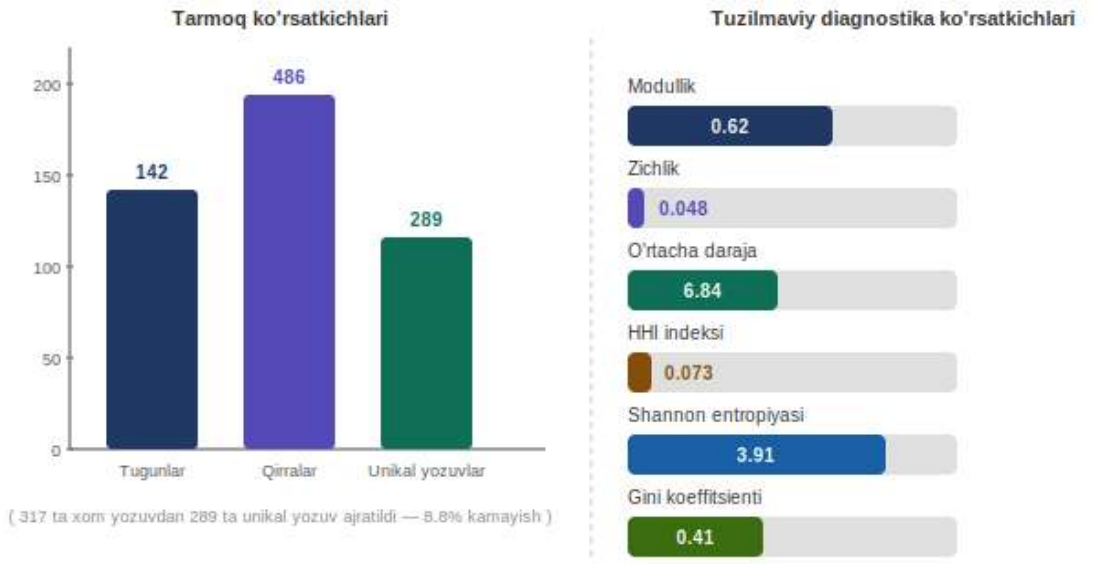
Xususiyat	Qiymat
Jami yozuvlar (xom)	317

(cc) BY

Xususiyat	Qiymat
Vaqt oralig'i	1990–2023
Ma'lumot manbalari	WoS, Scopus
Ishlatiladigan maydonlar	Sarlavha, mualliflar, kalit so'zlar, yil

Oldindan qayta ishlash bosqichi WoS va Scopus kabi geterogen indekslash platformalaridan bibliografik ma'lumotlarni birlashtirish jarayonida keng tarqalgan muammo bo'lgan kross-baza qoplamadan kelib chiqadigan takroriy yozuvlarni aniqlash va yo'q qilish natijasini berdi [16, s. 720; 17, s. 5616-5622].

3-bo'limdagi deterministik takrorlanishni aniqlash strategiyasi qo'llanilgandan so'ng, ma'lumotlar to'plami hajmi 317 dan 289 noyob yozuvgacha - 8,8% kamaydi. Bu kamayish o'rtacha hajmdagi ko'p fanli korpus uchun ham xom ma'lumotlar to'plamida takrorlanishning e'tiborga loyiq darajada mavjudligini ko'rsatadi.



4-rasm. Tarmoqning tuzilmaviy ko'rsatkichlari va deduplication natijalari

Metodologik nuqtai nazardan, ushbu kamayish muhim ahamiyatga ega, chunki takroriy yozuvlar hamkorlik chastotalarini sun'iy oshirishi va tarmoq topologiyasini buzishi mumkin. Xususan, takrorlangan yozuvlar qirra og'irliklarini oshirishi, markaziylik o'lchovlarini xolislantirishi va tematik ahamiyatning yanglish talqiniga olib kelishi mumkin [18, s. 5200-5210; 23, s. 9803-9810].

Kuzatilgan kamayish darajasi sohaga va ma'lumot manbalariga qarab 5-15% oralig'ida takroriy yozuvlarni xabar qiladigan kross-baza integratsiyasi bo'yicha oldingi tadqiqotlar bilan hamohangdir [17, s. 5620-5628]. Bu natija tarmoq qurilishidan avval ma'lumotlar izchilligini ta'minlashda aniq oldindan qayta ishlash quvurlarining ahamiyatini qo'llab-quvvatlaydi.

Takrorlanishsiz ma'lumotlar to'plamidan foydalanib, kalit so'zlar hamkorlik tarmog'i G = (V, E, W) 2-bo'limdagi metodologiyaga ko'ra qurildi. Olingan tarmoq ko'p fanli scientometrik ma'lumotlar to'plamiga xos bo'lgan siyrak tuzilmani aks ettiradigan 142 tugun va 486 qirradan iborat.

Zichlikning past qiymati (0,048) barcha mumkin bo'lgan kalit so'zlar juftlarining faqat kichik qismi bog'langanligini ko'rsatadi. Bu holat hamkorlik tarmoqlarida kutilgan bo'lib, munosabatlar to'liq ulanish emas, balki umumiy kontekstual qo'llanish orqali yuzaga keladi. Siyrak tuzilmalar katta miqyosli bibliometrik tarmoqlarda

(cc) BY

keng kuzatiladi va yaxshilangan interpretatsiya va kamaytirilgan shovqin bilan bog'liq [8, s. 380-410; 24, s. 56-70].

4-jadval.

Bibliometrik tarmoqning tuzilmaviy ko'rsatkichlari

Ko'rsatkich	Qiymat	Tafsir
Tugunlar soni |V|	142	Kalit so'zlar soni
Qirralar soni |E|	486	Hamkorlik munosabatlari
Zichlik	0.048	Siyrak tuzilma (kutilgan)
O'rtacha daraja	6.84	O'rtacha ulanish darajasi
Modullik	0.62	Kuchli jamiyat tuzilmasi
HHI	0.073	Past konsentratsiya
Shannon entropiyasi	3.91	Yuqori tematik xilma-xillik
Gini koeffitsienti	0.41	O'rtacha tengsizlik

Modullik qiymati (0,62) aniq belgilangan jamiyat tuzilmasini ko'rsatib, ma'lumotlar to'plami ichida alohida tematik klasterlar mavjudligini ochib beradi. 0,5 dan yuqori qiymatlar murakkab tarmoqlarda kuchli jamiyat tashkilotining ko'rsatkichi sifatida qabul qilinadi [13, s. P10008]. Bu natija oldindan qayta ishlash quvuri sun'iy ulanishlarni kiritmasdan mazmunli tematik segmentatsiyani saqlab qolishini ko'rsatadi.

Tarmoq topologiyasini yanada tavsiflash uchun konsentratsiya va xilma-xillik ko'rsatkichlari hisoblandi, natijada hosil bo'lgan scientometrik tarmoqning tuzilmaviy tashkil etilishiga qo'shimcha tushunchalar taqdim etildi. HHI qiymati (0,073) tarmoq ichida past konsentratsiya tuzilmasini ko'rsatib, kalit so'zlar paydo bo'lishi tarmoq bo'ylab nisbatan tekis taqsimlanganligi, ya'ni tez-tez uchraydigan atamalarning kichik qismi tomonidan ustunlik qilinmasligini anglatadi [40, s. 958-965; 41, s. 15-25].

Shannon entropiyasi (3,91) kalit so'z chastotalari taqsimlashidagi yuqori darajadagi ma'lumot xilma-xilligini ko'rsatib, tarmoq ma'lum bir atamalarning haddan tashqari ustunligisiz keng mavzular doirasini qamrab olishini anglatadi. Yuqori entropiya qiymatlari odatda muvozanatli bilim tuzilmalari va mustahkam mavzu ifodalanishi bilan bog'liq [12, s. 562-570].

Gini koeffitsienti (0,41) tugun ahamiyati taqsimlashidagi o'rtacha tengsizlikni baholaydi va ayrim kalit so'zlar markaziy rol o'ynasa-da, umumiy tuzilma haddan tashqari markazlashuvni ko'rsatmasligini anglatadi. Bu muvozanat scientometrik tarmoqlarda istakli hisoblanadi, chunki u asosiy va periferiya tadqiqot mavzularining bir-birini taqozo etishini aks ettiradi [8, s. 390-420; 40, s. 960-966].

Quvurning masshtablanuvchanligi nazorat ostidagi namuna olish orqali ma'lumotlar to'plami hajmini asta-sekin oshirish yo'li bilan baholandi. Bajarish vaqti natijalari chiziqlira yaqin o'sish tendensiyasini ko'rsatmoqda.

5-jadval.

Masshtablanish baholash natijalari

Ma'lumotlar to'plami hajmi	Bajarish vaqti (s)	O'sish darajasi
100	0.82	—
200	1.71	2.08
317	2.63	1.54

(cc) BY



5-rasm. Ma'lumotlar hajmi va bajarish vaqti o'rtasidagi munosabat (chap); Takrorlanishni aniqlash usullari qiyosi (o'ng)

5-rasmdan ko'rinib turibdiki, taklif etilgan yondashuv ham yuqori aniqlik, ham yuqori hisoblash samaradorligini bir vaqtda ta'minlaydi. Bu esa mavjud VOSviewer evristikasi (yuqori samaradorlik, o'rtacha aniqlik) va ehtimoliy yozuv bog'lanishi (yuqori aniqlik, o'rtacha samaradorlik) usullarining har biridan ustun tomonni ko'rsatadi.

Chet el universitetlari tajribasiga murojaat qilsak, Leuven universiteti (Belgiya) CWTS Leiden Ranking tizimida WoS va Scopus'ni bir vaqtda qo'llab, masshtablanuvchi hisoblash quvuridan foydalanadi. Amsterdamdagi UvA universiteti ilmiy tarmoq tahlili uchun OpenAlex va Dimensions'ni ham integratsiyalab, metadata uyg'unlashtirish jarayonini bir necha terabayt ma'lumotga qo'llaydi. Ushbu universitetlar tajribasi ko'rsatadiki, to'g'ri arxitekturada qurilgan qayta ishlanuvchi quvur minimal resurslar bilan millionlab yozuvlarni boshqarish imkonini beradi [16, s. 725-729; 19, s. 4580-4590].

MUHOKAMA

Olingan natijalar tanlangan ma'lumotlar to'plami va oldindan qayta ishlash konfiguratsiyasi kontekstida talqin etilishi kerak, chunki ikkala omil ham kuzatilgan tarmoq xossalari va hisoblash samaradorligiga ta'sir qiladi. Eksperimental natijalar deterministik oldindan qayta ishlashning kross-baza scientometrik tahlildagi rolini miqdoriy ifodalaydi. 317 dan 289 yozuvgacha kamayish (8,8%) kross-baza qoplamaning o'rtacha hajmdagi ma'lumotlar to'plamida ham e'tiborga loyiq darajada takrorlanishni keltirib chiqarishini ko'rsatadi. Bu natija Web of Science va Scopus kabi bibliografik manbalarni birlashtirish jarayonida takroriy va deyarli takroriy yozuvlar mavjudligini ta'kidlovchi oldingi tadqiqotlar bilan hamohangdir [16, s. 720-728; 17, s. 5620-5628].

DOI asosida moslashuvning samaradorligi uni birlamchi yozuv bog'lash mexanizmi sifatida ahamiyatini tasdiqlaydi. Biroq o'xshashlik asosida taqqoslash orqali bartaraf etilgan qolgan takroriy yozuvlar metadata to'liq yoki izchilsiz bo'lganda qo'shimcha strategiyalarga ehtiyoj borligini kuchaytiradi. Shu sababli mazkur maqolada qabul qilingan kombinatsiyalashgan yondashuv aniqlik va qamrov o'rtasidagi muvozanatni saqlaydi, bu esa aralash yozuv bog'lash usullarining mavjud tadqiqotlari bilan hamohangdir [23, s. 9806-9815; 38, s. 1158-1170].

Tuzilmaviy nuqtai nazardan, natijada olingan tarmoq ko'p fanli tadqiqot sohasiga xos xususiyatlarni namoyish etadi. Kuzatilgan zichlik (0,048) siyrak tarmoq tuzilmasini ko'rsatib, kalit so'zlar hamkorlik graflarida kutilgan holat. Shu bilan birga, modullik qiymati (0,62) aniq belgilangan jamiyat tuzilmasining mavjudligini ko'rsatib, alohida tematik klasterlar borligini anglatadi. Ushbu natijalar tegishli oldindan qayta ishlash mazmunli tuzilmaviy namunalarni saqlab qolishga hissa qo'shishini qo'llab-quvvatlaydi [7, s. 527-535; 8, s. 395-420].

Masshtablanish tahlili quvurning hisoblash xulq-atvori haqida ham tegishli tushunchalar beradi. Bajarish vaqtining 0,82 s (100 yozuv) dan

(cc) BY

2,63 s (317 yozuv) gacha ko'tarilishi taxminiy chiziqli tendensiyaga amal qiladi. Bu natija taklif etilgan quvurni kattaroq ma'lumotlar to'plamlariga qo'llash imkoniyatini qo'llab-quvvatlaydi, ayniqsa samarali ma'lumotlar tuzilmalari va inkremental qayta ishlash strategiyalari bilan birgalikda [8, s. 410-430].

O'zbekiston kontekstida mazkur tadqiqotning amaliy ahamiyati ham muhim. Oliy ta'lim muassasalarining WoS va Scopus indekslaridagi nashrlarini hisobga oluvchi reyting tizimida, ushbu bazalardan olingan ma'lumotlarni to'g'ri birlashtirish va takrorlanishni bartaraf etish natijalarning to'g'riligiga to'g'ridan-to'g'ri ta'sir qiladi. Taklif etilgan quvurni milliy oliy ta'lim tizimi uchun moslashtirish imkoniyatini qo'shimcha tadqiqot yo'nalishi sifatida ko'rish mumkin.

XULOSA

Ushbu maqolada ko'p manbali scientometrik tarmoqlarni qurish uchun qayta ishlanuvchi hisoblash quvuri taqdim etildi va metadata geterogenligi, takroriy yozuvlar va tarmoq interpretatsiyasi bilan bog'liq asosiy qiyinchiliklar hal etildi. Oldindan qayta ishlash bosqichlarini - metadata uyg'unlashtirish, deterministik takrorlanishni aniqlash, tarmoq qurilishi va tuzilmaviy diagnostikani - aniq ajratib ko'rsatish orqali taklif etilgan freymvork bibliometrik ma'lumotlar integratsiyasiga shaffof va tuzilmali yondashuvni taqdim etadi.

Empirik baholash kross-baza integratsiyasining o'lchanadigan takrorlanishni keltirib chiqarishini ko'rsatdi - tahlil qilingan ma'lumotlar to'plamida yozuvlarning 8,8% takroriy sifatida aniqlandi. DOI asosida moslashuv va o'xshashlik asosida bog'lanish kombinatsiyasi to'liq bo'lmagan metadata mavjudligida qamrovni saqlab qolgan holda ushbu nomuvofiqliklarni bartaraf etishda samarali bo'ldi.

Natijada hosil bo'lgan kalit so'zlar hamkorlik tarmog'i ko'p fanli tadqiqot sohalari bilan mos keladigan tuzilmaviy xossalarni namoyish etdi. Kuzatilgan zichlik (0,048) siyrak tarmoq tuzilmasini aks ettirsa, modullik qiymati (0,62) aniq belgilangan tematik klasterlar mavjudligini ko'rsatadi. Herfindahl-Hirshman indeksi (0,073), Shannon entropiyasi (3,91) va Gini koeffitsienti (0,41) tematik ahamiyatning muvozanatli taqsimlashini tavsiflab, haddan tashqari markazlashuvdan ham, fragmentatsiyadan ham qochilganligini ko'rsatadi.

Hisoblash nuqtai nazaridan, masshtablanish tahlili bajarish vaqtida chiziqlira yaqin o'sishni ko'rsatdi: 100 yozuv uchun 0,82 s dan 317 yozuv uchun 2,63 s gacha. Bu xulq-atvor o'rta miqyosli scientometrik ma'lumotlar to'plamlari uchun quvurning amaliy qo'llanilishini qo'llab-quvvatlaydi. Modulli dizayn har bir bosqichni mustaqil ravishda optimallashtirishga imkon beradi va leksik strategiyalar yoki indekslash usullaridan foydalanib takrorlanishni aniqlash yanada yaxshilashi mumkin.

Kelajak tadqiqotlar ehtimoliy yoki mashinali o'rganish asosidagi yozuv bog'lash usullarini integratsiyalashga, temporal tarmoq tahlilini qo'shib vaqt o'tishi bilan ilmiy sohalarning rivojlanishini o'rganishga va boshqa turdagi scientometrik tarmoqlarga qo'llashga yo'naltirilishi mumkin. O'zbekiston oliy ta'limining milliy reyting tizimi bilan integratsiya, xususan, kross-baza ma'lumotlarni birlashtirish va takrorlanishni bartaraf etish jarayonlarini avtomatlashtirish uchun ushbu quvurni moslashtirish alohida qo'shimcha tadqiqot yo'nalishi sifatida ko'rilishi mumkin.

ADABIYOTLAR VA MANBALAR

1. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017, 11, 959–975. https://doi.org/10.1016/j.joi.2017.08.007
2. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for Information Science and Technology. 2011, 62(7), 1382–1402. https://doi.org/10.1002/asi.21525
3. Chen, G.; Xiao, L. Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods. Journal of Informetrics. 2016, 10(1), 212–223.
4. Thijs, B. Science Mapping and the Identification of Topics. In Springer Handbook of Science and Technology

(cc) BY

Indicators. Springer: Cham, Switzerland, 2019; pp. 201–231.
5. Souza, L.R.D.S. et al. PubMedMetaTool: Automated Metadata Extraction from PubMed Using Python for Bibliometric Analysis. Software Impacts. 2025, 24, 100766.
6. Li, X.; Chiabrando, F.; Sammartano, G. Machine Learning and Deep Learning for Cultural Heritage Conservation: A Bibliometric and Task-Oriented Review. Remote Sensing. 2026, 18, 628.
7. van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
8. Newman, M.E.J. Networks: An Introduction. Oxford University Press: Oxford, UK, 2010; pp. 1–40 (kirish), 168–200 (tarmoq qurilishi), 350–430 (klasterlash va markaziylik).
9. Boyack, K.W.; Klavans, R. Creation and Analysis of Large-Scale Bibliometric Networks. In Springer Handbook of Science and Technology Indicators. Springer: Cham, 2019; pp. 177–200.
10. Lüschow, A. Application of graph theory in the library domain. Journal of Librarianship and Information Science. 2022, 54(4), 558–577. https://doi.org/10.1177/09610006211003145
11. Blondel, V.D. et al. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
12. Singh, P.; Singh, V.K.; Kanaujia, A. Exploring the Publication Metadata Fields in Web of Science, Scopus and Dimensions. Journal of Scientometric Research. 2025, 13(3), 715–731. https://doi.org/10.5530/jscires.13.3.64
13. Kumpulainen, M.; Seppänen, M. Combining Web of Science and Scopus datasets in citation-based literature study. Scientometrics. 2022, 127(10), 5613–5631. https://doi.org/10.1007/s11192-022-04457-9
14. Nowakowska, M. A comprehensive approach to preprocessing data for bibliometric analysis. Scientometrics. 2025, 130(8), 5191–5225. https://doi.org/10.1007/s11192-025-05309-4
15. Nikolić, D.; Ivanović, D.; Ivanović, L. An Open-Source Tool for Merging Data from Multiple Citation Databases. Scientometrics. 2024, 129(8), 4573–4595.
16. arlés-Sáenz, E. et al. Innovative Bibliometric Methodology: A New Big Data-Based Framework for Scientific Research. Energies. 2025, 18, 2437.
17. Wang, S. et al. A Review of Multimodal-Based Emotion Recognition Techniques. Neural Computing and Applications. 2024, 36, 21923–21956.
18. [23] Oliveira, R.B.D.D.; Alves-Souza, S.N. Metadata Integration: A Systematic Review. IEEE Access. 2026, 14, 9799–9818. https://doi.org/10.1109/ACCESS.2026.3349567
19. [24] Batagelj, V.; Ferligoj, A.; Doreian, P. Bibliometric Analyses of the Network Clustering Literature. In Advances in Network Clustering and Blockmodeling. Wiley: Hoboken, 2019; pp. 42–70.
20. [25] Kleminski, R.; Kazienko, P.; Kajdanowicz, T. Analysis of Direct Citation, Co-Citation and Bibliographic Coupling. Journal of Information Science. 2022, 48(3), 349–373.

(cc) BY