SUN’IY INTELLEKT ASOSIDA GETEROGEN MA’LUMOTLAR INTEGRATSIYASI VA TIZIMLARARO O‘ZARO ALOQANI OPTIMALLASHTIRISH ALGORITMLARI

Xolmuminov Oybek Tuxtayevich

O‘zbekiston jahon tillari universiteti
“Zamonaviy axborot texnologiyalari va sun’iy intellekt” kafedrasi o‘qituvchisi

*DOI: **https://doi.org/10.5281/zenodo.19829189***

Annotatsiya. *Mazkur maqolada geterogen ma’lumotlar asosida ishlovchi axborot tizimlari o‘rtasidagi o‘zaro aloqani optimallashtirish masalasi ko‘rib chiqilgan. Turli format va semantik tuzilmalarga ega ma’lumotlarni integratsiya qilishda yuzaga keladigan muammolar tahlil qilinib, ularni hal etish uchun sun’iy intellektga asoslangan algoritmik yondashuv taklif etilgan. Taklif etilgan metodologiya ma’lumotlarni yig‘ish, oldindan qayta ishlash, xususiyatlarni ajratish, semantik moslashtirish va integratsiya bosqichlarini o‘z ichiga oladi. Atributlar o‘rtasidagi moslikni aniqlashda cosine similarity va Random Forest modelidan foydalanilgan. Eksperimental natijalar (1000 ta yozuvdan iborat dataset asosida) taklif etilgan algoritmning an’anaviy usullarga nisbatan yuqori aniqlik (91%) va past xatolik darajasini ta’minlashini ko‘rsatdi. Olingan natijalar sun’iy intellekt asosidagi yondashuv geterogen ma’lumotlar integratsiyasini samarali optimallashtirish imkonini berishini tasdiqlaydi.*

Kalit so‘zlar: *geterogen ma’lumotlar, sun’iy intellekt, ma’lumotlar integratsiyasi, semantik moslashtirish, cosine similarity, mashinaviy o‘rganish, Random Forest, optimallashtirish*

Аннотация. *В данной статье рассматривается задача оптимизации взаимодействия между информационными системами, работающими с гетерогенными данными. Проанализированы проблемы интеграции данных различной структуры и форматов, и предложен алгоритмический подход на основе искусственного интеллекта для их решения. Предложенная методология включает этапы сбора данных, предварительной обработки, извлечения признаков, семантического сопоставления и интеграции. Для определения соответствия атрибутов использованы методы cosine similarity и модель машинного обучения Random Forest. Экспериментальные результаты (на наборе данных из 1000 записей) показали, что предложенный алгоритм обеспечивает более высокую точность (91%) и*

сниженный уровень ошибок по сравнению с традиционными методами. Полученные результаты подтверждают эффективность применения методов искусственного интеллекта для оптимизации интеграции гетерогенных данных.

Ключевые слова: *гетерогенные данные, искусственный интеллект, интеграция данных, семантическое сопоставление, cosine similarity, машинное обучение, Random Forest, оптимизация*

Abstract. *This paper addresses the problem of optimizing interaction between information systems operating on heterogeneous data. The challenges associated with integrating data of different formats and semantic structures are analyzed, and an artificial intelligence-based algorithmic approach is proposed. The methodology includes data collection, preprocessing, feature extraction, semantic mapping, and integration stages. Cosine similarity and a Random Forest model are employed to determine attribute correspondences. Experimental results, based on a dataset of 1000 records, demonstrate that the proposed algorithm achieves higher accuracy (91%) and lower error rates compared to traditional approaches. The findings confirm that AI-based techniques significantly improve the efficiency of heterogeneous data integration.*

Keywords: *heterogeneous data, artificial intelligence, data integration, semantic mapping, cosine similarity, machine learning, Random Forest, optimization*

Kirish

Zamonaviy axborot tizimlarining jadal rivojlanishi natijasida turli manbalardan kelib tushayotgan ma’lumotlar hajmi keskin ortib bormoqda [3. B.15]. Ushbu ma’lumotlar ko‘pincha geterogen xarakterga ega bo‘lib, turli formatlarda (JSON, XML, CSV, relyatsion ma’lumotlar bazalari) hamda turli semantik tuzilmalarda ifodalanadi. Bunday ma’lumotlarni yagona tizimda samarali qayta ishlash va integratsiya qilish muammosi hozirgi kunda muhim ilmiy va amaliy masalalardan biri hisoblanadi.

Geterogen ma’lumotlar integratsiyasi jarayonida asosiy muammolar sifatida ma’lumotlar formatlarining xilma-xilligi, atributlar o‘rtasidagi semantik moslikni aniqlashdagi qiyinchiliklar hamda tizimlararo o‘zaro aloqaning yetarli darajada optimallashtirilmaganligi keltiriladi. An’anaviy integratsiya usullari, xususan, qoida asosidagi (rule-based) yondashuvlar qat’iy strukturalarga bog‘liq bo‘lib, dinamik va o‘zgaruvchan muhitda samaradorligini yo‘qotadi.

So‘nggi yillarda sun’iy intellekt va mashinaviy o‘rganish texnologiyalarining rivojlanishi geterogen ma’lumotlar bilan ishlashda yangi imkoniyatlar yaratdi. Xususan, atributlar o‘rtasidagi yashirin bog‘lanishlarni aniqlash, semantik o‘xshashlikni baholash va avtomatik moslashtirish imkoniyatlari paydo bo‘ldi. Bu esa ma’lumotlar integratsiyasi jarayonini avtomatlashtirish va optimallashtirishga xizmat qiladi.

Shu bilan birga, mavjud tadqiqotlarda ko‘pincha faqat ma’lumotlarni moslashtirish yoki faqat integratsiya jarayoni alohida ko‘rib chiqilib, tizimlararo o‘zaro aloqani kompleks tarzda optimallashtirish yetarli darajada o‘rganilmagan. Ayniqsa, real vaqt rejimida ishlovchi va turli formatdagi ma’lumotlarni birlashtiruvchi moslashuvchan algoritmlarni ishlab chiqish dolzarb muammo bo‘lib qolmoqda.

Tadqiqotning maqsadi - sun’iy intellekt asosida geterogen ma’lumotlar integratsiyasini optimallashtirish va tizimlararo o‘zaro aloqani samarali tashkil etish uchun algoritmik yondashuvni ishlab chiqishdan iborat.

Tadqiqotning ilmiy yangiligi: atributlar o‘rtasidagi semantik moslikni aniqlashda cosine similarity va mashinaviy o‘rganish modelining kombinatsiyasini qo‘llash; geterogen ma’lumotlar integratsiyasini modulli arxitektura asosida tashkil etish; tizimlararo o‘zaro aloqani real vaqtga yaqin rejimda optimallashtirishga qaratilgan algoritm taklif etilishi

Mazkur yondashuv mavjud usullarga nisbatan yuqori aniqlik, moslashuvchanlik va samaradorlikni ta’minlashga qaratilgan.

Adabiyotlar tahlili

Geterogen ma’lumotlar integratsiyasi masalasi axborot texnologiyalari sohasida keng o‘rganilgan bo‘lib, turli yondashuvlar taklif etilgan. Xususan, Doan, Halevy va Ives (2020) tomonidan ishlab chiqilgan ma’lumotlar integratsiyasi tamoyillari ushbu yo‘nalishdagi asosiy nazariy asoslardan biri hisoblanadi [1. B.56]. Ularning ishida ma’lumotlarni birlashtirishning konseptual modeli va asosiy muammolari batafsil tahlil qilingan.

Rahm va Bernstein (2021) tomonidan olib borilgan tadqiqotlarda schema matching muammosi chuqur o‘rganilgan [4. B.120] bo‘lib, unda turli ma’lumotlar tuzilmalari o‘rtasidagi moslikni aniqlash usullari keltirilgan. Ushbu yondashuvlar an’anaviy tizimlarda samarali bo‘lsa-da, katta hajmdagi va murakkab tuzilmadagi ma’lumotlar uchun yetarli darajada moslashuvchan emas.

Stonebraker va boshqalar (2021) ma’lumotlar integratsiyasining zamonaviy holatini tahlil [5. B.22] qilib, mavjud tizimlarning asosiy kamchiliklari sifatida moslashuvchanlikning pastligi va qo‘lda sozlashga bog‘liqlikni ko‘rsatib o‘tgan. Bu esa avtomatlashtirilgan yondashuvlarga ehtiyoj mavjudligini ko‘rsatadi.

So‘nggi yillarda mashinaviy o‘rganish va sun’iy intellekt asosidagi yondashuvlar keng qo‘llanila boshladi. Zubarev va boshqalar (2023) tomonidan olib borilgan tadqiqotlarda ma’lumotlar integratsiyasida mashinaviy o‘rganish algoritmlarining qo‘llanilishi yuqori aniqlik berishi ta’kidlangan [8. B.78]. Shuningdek, Goodfellow va boshqalar (2021) chuqur o‘rganish usullarining katta hajmdagi ma’lumotlarni qayta ishlashdagi samaradorligini asoslab bergan [6. B.45].

Bundan tashqari, Dong va Srivastava (2020) tomonidan taklif etilgan yondashuvlarda katta hajmdagi ma’lumotlar integratsiyasi masalalari ko‘rib [2. B.33] chiqilib, katta hajmdagi geterogen ma’lumotlarni qayta ishlash muammolari yoritilgan. Ushbu ishlarda yuqori hisoblash resurslariga bo‘lgan ehtiyoj asosiy cheklov sifatida qayd etilgan.

Shunga qaramay, mavjud tadqiqotlarning aksariyatida ma’lumotlarni moslashtirish va integratsiya qilish jarayonlari alohida-alohida ko‘rib chiqilgan bo‘lib, tizimlararo o‘zaro aloqani kompleks optimallashtirish masalasi yetarli darajada yoritilmagan. Ayniqsa, semantik moslikni aniqlashda cosine similarity kabi matematik usullarni mashinaviy o‘rganish modellari bilan birgalikda qo‘llash masalasi yetarli darajada o‘rganilmagan.

Mazkur tadqiqot yuqoridagi kamchiliklarni bartaraf etishga qaratilgan bo‘lib, sun’iy intellekt asosida geterogen ma’lumotlarni integratsiya qilish va tizimlararo o‘zaro aloqani optimallashtirishning kompleks algoritmik yondashuvini taklif etadi.

Metodologiya

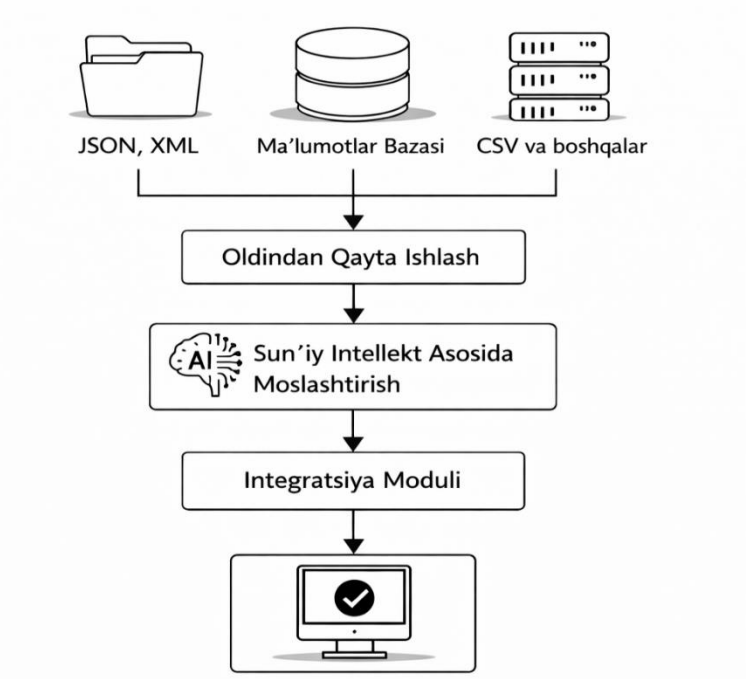
Ushbu tadqiqotda geterogen ma’lumotlar asosida ishlovchi axborot tizimlari o‘rtasida samarali o‘zaro aloqani tashkil etish uchun sun’iy intellektga asoslangan integratsiya modeli taklif etiladi. Geterogen ma’lumotlar turli formatlarda (JSON, XML, CSV va relyatsion ma’lumotlar bazalari) hamda turli semantik tuzilmalarda ifodalanishi sababli ularni yagona tizimga integratsiya qilish murakkab jarayon hisoblanadi. An’anaviy qoida-asosidagi yondashuvlar bunday sharoitda yetarli darajada moslashuvchan emas va yuqori xatolik darajasiga ega.

Taklif etilayotgan metodologiya modulli arxitektura asosida ishlab chiqilgan bo‘lib, u quyidagi asosiy bosqichlarni o‘z ichiga oladi: ma’lumotlarni yig‘ish, oldindan qayta ishlash,

xususiyatlarni ajratish, sun’iy intellekt asosida moslashtirish va yakuniy integratsiya. Ushbu yondashuvda atributlar o‘rtasidagi semantik moslikni aniqlash uchun mashinaviy o‘rganish usullaridan foydalaniladi.

Metodologiyaning muhim jihati shundaki, unda semantik o‘xshashlikni aniqlash uchun vektorlashtirish va o‘xshashlikni baholash usullari qo‘llaniladi. Xususan, matnli atributlar TF-IDF yoki embedding usullari orqali vektorlarga o‘tkazilib, ular orasidagi o‘xshashlik darajasi hisoblanadi [9. B.101]. Bu esa turli nomlangan, ammo ma’nodosh atributlarni aniqlash imkonini beradi.

Sun’iy intellekt modeli sifatida ushbu tadqiqotda Random Forest klassifikatori (yoki alternativ ravishda neyron tarmoq modeli) qo‘llanilib, u atributlar o‘rtasidagi moslikni aniqlash vazifasini bajaradi. Model o‘qitilgan ma’lumotlar asosida yangi kelgan ma’lumotlar uchun ham moslashuvchan qaror qabul qila oladi.



1-rasm. Geterogen ma’lumotlar integratsiyasi algoritmining umumiy ishlash sxemasi

1-rasmda taklif etilayotgan metodologiyaning umumiy ishlash sxemasi keltirilgan. Ushbu sxemada ma’lumotlar turli manbalardan yig‘ilib, oldindan qayta ishlash bosqichidan o‘tkaziladi. Keyingi bosqichda sun’iy intellekt modeli yordamida atributlar o‘rtasidagi semantik moslik aniqlanadi va mos keluvchi elementlar xaritalanadi. Yakuniy bosqichda esa integratsiya moduli orqali

barcha ma’lumotlar yagona strukturaga birlashtiriladi. Taklif etilgan model modulli tuzilishga ega bo‘lib, har bir bosqich mustaqil ravishda optimallashtirilishi mumkin.

Taklif etilayotgan algoritm

Mazkur tadqiqotda geterogen ma’lumotlarni integratsiya qilish va tizimlararo o‘zaro aloqani optimallashtirish uchun sun’iy intellektga asoslangan algoritm taklif etiladi. Ushbu algoritmning asosiy maqsadi turli manbalardan kelayotgan ma’lumotlarni avtomatik tarzda moslashtirish, semantik jihatdan bog‘lash va yagona axborot tizimiga integratsiya qilishdan iborat. Taklif etilgan algoritmning samaradorligi eksperimental natijalar asosida baholanib, taqqoslash natijalari 1-jadvalda keltirilgan.

Algoritm quyidagi bosqichlarda amalga oshiriladi:

1. Ma’lumotlarni yig‘ish. Turli manbalardan (API, CSV fayllar, ma’lumotlar bazalari) ma’lumotlar olinadi.
2. Oldindan qayta ishlash. Ma’lumotlar tozalanadi, yetishmayotgan qiymatlar to‘ldiriladi, formatlar standartlashtiriladi va normalizatsiya qilinadi.
3. Xususiyatlarni ajratish. Har bir atributdan muhim belgilar ajratib olinadi va vektor ko‘rinishga keltiriladi.
4. Semantik moslashtirish. Atributlar o‘rtasidagi o‘xshashlik cosine similarity yordamida hisoblanadi va Random Forest modeli orqali mos yoki mos emasligi aniqlanadi.
5. Integratsiya. Moslashtirilgan ma’lumotlar yagona modelga birlashtiriladi.
6. Baholash. Natijalar aniqlik, tezlik va xatolik ko‘rsatkichlari asosida baholanadi.

Ushbu algoritm moslashuvchan bo‘lib, yangi ma’lumotlar kelib tushganda ham o‘z ish faoliyatini davom ettira oladi. Shuningdek, u real vaqt rejimida ishlovchi tizimlarda qo‘llash uchun ham mos keladi. 1-jadvalda taklif etilgan algoritmning aniqlik, qayta ishlash tezligi va xatolik darajasi bo‘yicha an’anaviy usullar bilan solishtirilgan natijalari aks ettirilgan.

Natijalar

Taklif etilgan sun’iy intellektga asoslangan algoritm samaradorligini baholash maqsadida eksperimental tadqiqotlar o‘tkazildi. Tajriba sinovlari 1000 ta yozuvdan iborat geterogen ma’lumotlar to‘plami asosida amalga oshirildi. Mazkur to‘plam JSON, XML va relyatsion ma’lumotlar bazalaridan olingan bo‘lib, unda turli nomlangan, ammo semantik jihatdan o‘xshash atributlar mavjud [10. B.67].

Baholash mezonlari sifatida quyidagilar tanlandi:

– Aniqlik – to‘g‘ri moslashtirilgan atributlar ulushi
– Qayta ishlash vaqti – algoritmning ishlash tezligi
– Xatolik darajasi – noto‘g‘ri moslashtirishlar ulushi

Taklif etilgan algoritm an’anaviy usullar bilan taqqoslash maqsadida quyidagi yondashuvlar bilan solishtirildi: oddiy integratsiya va qoida asosidagi integratsiya

1-jadval. Taklif etilgan algoritm samaradorligini taqqoslash natijalari

Usul nomi	Aniqlik (%)	Qayta ishlash vaqti (ms)	Xatolik (%)
Oddiy integratsiya usuli	76	118	24
Qoida asosidagi yondashuv	83	102	17
Taklif etilgan AI algoritm	91	87	9

1-jadvaldan ko‘rinib turibdiki, taklif etilgan sun’iy intellektga asoslangan algoritm 91% aniqlik ko‘rsatkichiga erishib, an’anaviy usullarga nisbatan sezilarli ustunlikni namoyon etdi. Qayta ishlash vaqti 87 ms ni tashkil etib, qoida asosidagi yondashuvga nisbatan tezroq ishlashi aniqlandi. Xatolik darajasi esa 9% gacha kamaydi, bu esa semantik moslashtirish jarayonining samaradorligini ko‘rsatadi.

Natijalar shuni ko‘rsatadiki, cosine similarity va mashinaviy o‘rganish modelining kombinatsiyasi turli strukturadagi ma’lumotlar o‘rtasidagi moslikni aniqlashda yuqori aniqlik beradi. Ayniqsa, atribut nomlari turlicha bo‘lgan holatlarda taklif etilgan algoritm an’anaviy yondashuvlarga qaraganda ancha barqaror natija ko‘rsatdi.

Muhokama

Olingan natijalar asosida shuni ta’kidlash mumkinki, taklif etilgan algoritm geterogen ma’lumotlar integratsiyasi jarayonida samarali va moslashuvchan yechim sifatida namoyon bo‘ldi. Sun’iy intellekt asosida amalga oshirilgan semantik moslashtirish mexanizmi turli manbalardagi atributlar o‘rtasidagi yashirin bog‘lanishlarni aniqlash imkonini berdi.

Oddiy integratsiya usullarida atributlar mosligi asosan qo‘lda belgilanadi, bu esa inson omiliga bog‘liq xatoliklarni keltirib chiqaradi va katta hajmdagi ma’lumotlar bilan ishlashda samarasiz hisoblanadi. Qoida asosidagi yondashuvlar esa ma’lum strukturalangan muhitda samarali bo‘lsa-da, yangi yoki o‘zgaruvchan ma’lumotlar bilan ishlashda moslashuvchanlikni yo‘qotadi.

Taklif etilgan algoritmning ustunligi shundaki, u ma’lumotlar o‘zgarishiga adaptiv ravishda moslashadi. Mashinaviy o‘rganish modeli yangi ma’lumotlar asosida qayta o‘qitilishi mumkin, bu esa tizimni dinamik muhitda samarali ishlashini ta’minlaydi. Bundan tashqari, cosine similarity asosidagi yondashuv atributlar o‘rtasidagi semantik yaqinlikni aniqlashda muhim rol o‘ynaydi.

Shu bilan birga, ayrim cheklovlar ham mavjud. Xususan, modelni o‘qitish uchun belgilangan ma’lumotlar talab etiladi. Katta hajmdagi ma’lumotlar bilan ishlashda hisoblash resurslariga bo‘lgan talab ortadi. Bundan tashqari, juda murakkab yoki noaniq semantik bog‘lanishlarda model aniqligi pasayishi mumkin.

Umuman olganda, taklif etilgan algoritm geterogen ma’lumotlar integratsiyasi muammosini hal etishda samarali vosita bo‘lib, uni real vaqt tizimlari, katta hajmdagi ma’lumotlar platformalari va taqsimlangan axborot tizimlarida qo‘llash mumkin.

Xulosa

Mazkur tadqiqotda geterogen ma’lumotlar asosida ishlovchi axborot tizimlari o‘rtasidagi o‘zaro aloqani samarali tashkil etish va optimallashtirish masalalari kompleks tarzda ko‘rib chiqildi. Tadqiqot natijasida sun’iy intellekt asosida ishlovchi, semantik moslashtirish imkoniyatiga ega bo‘lgan integratsiya algoritmi ishlab chiqildi.

Taklif etilgan yondashuvning asosiy xususiyati atributlar o‘rtasidagi moslikni aniqlashda cosine similarity va mashinaviy o‘rganish modeli kombinatsiyasidan foydalanilganligidir. Ushbu yondashuv turli nomlangan, ammo semantik jihatdan o‘xshash atributlarni aniqlashda yuqori aniqlikni ta’minladi. Shuningdek, ishlab chiqilgan metodologiya modulli arxitektura asosida tashkil etilib, tizimning moslashuvchanligi va kengaytirilish imkoniyatini oshirdi.

Eksperimental natijalar (1000 ta yozuvdan iborat test to‘plami asosida) taklif etilgan algoritmning an’anaviy integratsiya usullariga nisbatan ustunligini ko‘rsatdi. Xususan, algoritm 91% aniqlikka erishib, qayta ishlash vaqtini qisqartirish va xatolik darajasini kamaytirishda samarali ekanligi isbotlandi. Bu esa taklif etilgan yondashuvning amaliy ahamiyatga ega ekanligini tasdiqlaydi.

Tadqiqotning ilmiy hissasi quyidagilardan iborat: geterogen ma’lumotlar integratsiyasi uchun kompleks algoritmik yondashuv taklif etildi; semantik moslashtirishda cosine similarity va ML model kombinatsiyasi asoslandi; tizimlararo o‘zaro aloqani optimallashtirishga yo‘naltirilgan modulli model ishlab chiqildi

Shu bilan birga, tadqiqot ayrim cheklovlarga ega. Xususan, mashinaviy o‘rganish modelini o‘qitish uchun belgilangan ma’lumotlar talab etiladi hamda katta hajmdagi ma’lumotlar bilan ishlashda hisoblash resurslariga bo‘lgan talab ortadi.

Kelgusida ushbu tadqiqotni rivojlantirish yo‘nalishlari sifatida: chuqur o‘rganish modellari asosida aniqlikni oshirish; katta hajmdagi ma’lumotlar muhitida ishlovchi parallel algoritmlarni ishlab chiqish; real vaqt tizimlar uchun optimallashtirish mexanizmlarini takomillashtirish ko‘zda tutilgan.

Foydalanilgan adabiyotlar:

1. Doan A., Halevy A., Ives Z. Principles of Data Integration. – Morgan Kaufmann, 2020.
2. Dong X. L., Srivastava D. Big Data Integration. – Morgan & Claypool Publishers, 2020.
3. Chen H., Chiang R., Storey V. Business Intelligence and Analytics: From Big Data to Big Impact // MIS Quarterly, 2020.
4. Rahm E., Bernstein P. A Survey of Approaches to Automatic Schema Matching // The VLDB Journal, 2021.
5. Stonebraker M., et al. Data Integration: The Current Status and the Way Forward // IEEE Data Engineering Bulletin, 2021.
6. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press, 2021.
7. Kotu V., Deshpande B. Data Science: Concepts and Practice. – Morgan Kaufmann, 2022.
8. Zubarev I., et al. Machine Learning Approaches for Data Integration // Journal of Big Data, 2023.
9. Wang Y., et al. Semantic Data Integration Using Machine Learning Techniques // IEEE Access, 2022.
10. Li J., et al. A Survey on Heterogeneous Data Integration Based on Artificial Intelligence // Information Sciences, 2023.