

MA'LUMOTLARNI TO'PLASH VA ULARNING ZAMONAVIY
YONDASHUVLARI VA AMALIY JIHATLARI

Raximova Umida Ziyadullayevna

Samarqand iqtisodiyot va servis instituti

“Oliy matematika” kafedrası o'qituvchisi

Musayev Azimjon Akramovich

Samarqand iqtisodiyot va servis instituti

"Iqtisodiyot" fakulteti 2-kurs talabasi

Annotatsiya: Ushbu maqolada zamonaviy ilmiy tadqiqotlar, biznes-analitika va sun'iy intellekt tizimlari uchun ma'lumotlarni to'lash hamda ularning sifatini tekshirish metodlari tizimli tahlil qilinadi. Tadqiqotda ma'lumot to'lashning birinchil va ikkilamchi manbalari, avtomatlashtirilgan va qo'lda boshqariladigan yondashuvlar, shuningdek, ma'lumot sifatini baholashning xalqaro standartlari (ISO 8000, DAMA-DMBOK, FAIR tamoyillari) ko'rib chiqiladi. Sifat nazorati jarayonida qo'llaniladigan statistik, qoidaga asoslangan va mashinaviy o'qitishga asoslangan anomaliya aniqlash usullari solishtirma tahlil qilingan. Maqolada ma'lumotlarni tozalash, dublikatlarni yo'qotish, yetishmayotgan qiymatlarni to'ldirish va validatsiya quvurlarini (data pipelines) avtomatlashtirish bo'yicha amaliy tavsiyalar keltirilgan. Natijalar shuni ko'rsatadiki, sifatli ma'lumot bazasini shakllantirish nafaqat texnik jarayon, balki metodologik intizom, etik me'yorlar va uzluksiz monitoringni talab qiladigan kompleks tizimdir.

Kalit so'zlar: ma'lumotlarni to'plash, ma'lumotlar sifati, validatsiya, anomaliya aniqlash, ma'lumotlarni tozalash, FAIR tamoyillari, DAMA-DMBOK, sun'iy intellekt, reproduktivlik.

KIRISH

Raqamli iqtisodiyot, sun'iy intellekt va katta ma'lumotlar (Big Data) davrida ilmiy va amaliy qarorlarning asosi sifatida ma'lumotlar hal qiluvchi ahamiyat kasb etmoqda. Biroq, "chiqindi kirdi, chiqindi chiqadi" (Garbage In, Garbage Out) printsipli zamonaviy analitikada ham o'z kuchini yo'qotmagan. Sifatsiz, noto'g'ri yoki noqulay shakldagi ma'lumotlar nafaqat model natijalarini buzadi, balki ilmiy ishonchlilikni, biznes strategiyalarini va davlat boshqaruvi samaradorligini ham pasaytiradi.

Shu sababli, ma'lumotlarni to'plash metodikasini ilmiy asoslash va ularning sifatini tizimli tekshirish mexanizmlarini ishlab chiqish dolzarb ilmiy-amaliy masala hisoblanadi. Ushbu maqolaning maqsadi – ma'lumot to'lash va sifat nazorati metodlarini zamonaviy standartlar, texnologik imkoniyatlar va amaliy cheklovlar nuqtai nazaridan tahlil qilish, tadqiqotchilar va ma'lumot muhandislari uchun takomillashtirilgan yondashuvni taklif etishdan iborat.

Ma'lumotlarni to'plash metodikasining azifalar:

- Ma'lumot to'lash usullarining tasnifi va afzallik/cheklovlarini yoritish;
- Ma'lumot sifatini baholashning xalqaro mezonlari va metrikalarini tahlil qilish;
- Sifatni tekshirish, tozalash va validatsiya qilish metodlarini solishtirma ko'rib chiqish;
- AI/ML va avtomatlashtirilgan quvurlarning sifat nazoratidagi o'rnini baholash;
- Amaliyot uchun metodik tavsiyalar ishlab chiqish.

ADABIYOTLAR TAHLILI

Ma'lumot to'lash va sifat nazorati masalalari klassik ilmiy metodologiya (Creswell, 2018; Babbie, 2020) va zamonaviy ma'lumot muhandisligi (DAMA International, 2017; ISO 8000, 2023) doirasida keng yoritilgan.

Klassik yondashuvlar: Tadqiqot maqsadiga qarab birinchil (so'rovnoma, kuzatuv, eksperiment) va ikkilamchi (ma'muriy yozuvlar, ochiq ma'lumotlar portallari, ilmiy bazalar) manbalardan foydalanish tamoyillari shakllangan.

Sifat standartlari: DAMA-DMBOK ma'lumot boshqaruvi bo'yicha 11 ta funksiya, shu jumladan "Ma'lumot sifati"ni alohida yo'nalish sifatida belgilaydi. ISO 8000:2023 esa ma'lumot sifatini o'lchash, sertifikatlash va almashtirish talablarini standartlashtiradi.

FAIR tamoyillari: Wilkinson va hamkorlar (2016) tomonidan taklif etilgan Findable, Accessible, Interoperable, Reusable printsiplari ilmiy ma'lumotlarni boshqarishning global standarti sifatida qabul qilingan.

Zamonaviy tendensiyalar: So'nggi yillarda (2023–2025) ma'lumot sifatini tekshirishda "Data Observability", "Self-Healing Pipelines" va LLM-asosli semantik validatsiya tez rivojlanmoqda (Barr & Moore, 2024; OECD AI Data Guidelines, 2025).

Shunday bo'lsada, metodologik adabiyotlar va amaliy data engineering yondashuvlari o'rtasidagi uzilish, shuningdek, milliy kontekstda standartlashtirilgan sifat nazorati protokollarining yetishmasligi tadqiqotlar reproduktivligini cheklab kelmoqda.

METODOLOGIYA

Tizimli adabiyotlar tahlili: Scopus, Web of Science, IEEE Xplore va milliy ilmiy bazalardagi 2018–2026 yillardagi 80+ manba tanqidiy ko'rib chiqildi.

Qiyosiy tahlil: An'anaviy statistik validatsiya, qoidaga asoslangan tekshiruv va ML-anomaliya aniqlash usullari samaradorlik, narx va moslashuvchanlik mezonlari bo'yicha solishtirildi.

Framework sintezi: DAMA-DMBOK, ISO 8000, FAIR va Data Quality Assessment Framework (DQAF) integratsiyasi asosida amaliy metodik model ishlab chiqildi.

Amaliy tekshiruv: Open-source vositalar (Great Expectations, Pandera, Deequ, Scikit-learn) yordamida sintetik va real ochiq ma'lumot to'plamlarida sifat nazorati

quvuri testdan o'tkazildi.

Ma'lumotlarni to'lash usullari: tasnif va xususiyatlari.

Usul turi	Misollar	Afzalliklari	Cheklovlari
Birlamchi (Primary)	So'rovnoma, intervyu, kuzatuv, eksperiment, sensor/IoT	Maqsadga mos, yangi, nazorat qilinadigan	Vaqt/xarajat yuqori, namuna xatosi, etik talablar
Ikkilamchi (Secondary)	Ochiq ma'lumotlar (data.gov, World Bank), ma'muriy yozuvlar, ilmiy repozitoriylar, web-scraping	Tez, arzon, keng qamrovli, tarixiy qiyoslash imkoni	Sifat noma'lum, format nomuvofiqlik, litsenziya cheklovlari
Avtomatlashtirilgan	API integratsiya, log yig'ish, IoT telemetriya, LLM-assisted labeling	Real-vaqt, masshtablanuvchan, inson xatosi kam	Infrastruktura talab qiladi, shovqin filtratsiya zarur

Tadqiqot dizayniga qarab gibrid yondashuv (masalan, ochiq ma'lumotlar + maqsadli so'rovnoma validatsiyasi) eng yuqori ishonchlikni beradi.

Sifatni tekshirish va tozalash metodlari.

- Qoidaga asoslangan validatsiya (Rule-based)
- Schema tekshiruvi (turlar, majburiy maydonlar)
- Diapazon va mantiqiy cheklovlar ($age > 0$, $sum \geq 0$)
- Referensial yaxlitlik (foreign keys, lookups)
- Vositalar: Great Expectations, Pandera, AWS Deequ, custom SQL checks

Statistik va anomaliya aniqlash.

- Z-score, IQR method, Mahalanobis distance
- Isolation Forest, Local Outlier Factor (LOF), Autoencoder-based anomaly detection

• Qo'llanilishi: Sensor ma'lumotlari, tranzaksiya loglari, vaqt qatorlari

Yetishmayotgan qiymatlarni boshqarish. (Missing Data Handling)

• O'chirish: MCAR shartida (bias xavfi yuqori)

• To'ldirish (Imputation): Mean/Median, KNN, MICE, MissForest, model-based prediction

• Belgilash: Missing indicator flag, probabilistic assignment

• Zamonaviy yondashuv: Embedding-based semantic matching + human-in-the-loop review

Sun'iy intellekt va avtomatlashtirilgan sifat nazorati.

2024–2026 yillarda ma'lumot sifatini boshqarishda quyidagi innovatsiyalar joriy etilmoqda:

LLM-assisted validation: Tabiiy til qoidalarini kodga aylantirish, noaniq yozuvlarni semantik tekshirish, metadata avto-generatsiya

Self-monitoring pipelines: Data drift, concept drift, schema drift ni real vaqtda aniqlash va alert berish

Active learning loop: Model xatolari orqali ma'lumot sifati past joylarni avtomatik belgilash va qayta nazoratga olish

Ethical & Bias checks: Fairness metrics (demographic parity, equalized odds) orqali training data'dagi tizimli og'ishlarni aniqlash

XULOSA

Ma'lumotlarni to'lash va sifatini tekshirish jarayoni ilmiy tadqiqotlar, biznes-analitika va AI tizimlarining poydevori hisoblanadi. Tadqiqot natijalari quyidagi xulosalarni tasdiqladi:

Ma'lumot to'lash metodini tanlash tadqiqot maqsadi, resurslar, etik me'yorlar va ma'lumot hayot sikliga bog'liq bo'lib, gibril yondashuv eng yuqori sifatni ta'minlaydi.

Ma'lumot sifati ko'p o'lchovli tushuncha bo'lib, uni faqat "tozalash" orqali emas, balki profilash, qoidalar, monitoring va audit sikli orqali boshqarish kerak.

Zamonaviy texnologiyalar (ML-anomaliya aniqlash, LLM-validatsiya, data observability) sifat nazoratini tezlashtiradi, lekin shaffoflik, reproducibility va inson nazorati talablarini buzmaslik shart.

Milliy va tashkilot darajasida FAIR tamoyillari, ISO 8000 va DAMA standartlariga asoslangan yagona ma'lumot sifati protokolini joriy etish ilmiy ishonchlilik va texnologik samaradorlikni oshiradi.

Ma'lumot – XXI asrning eng qimmatli resursi, ammo uning qiymati faqat sifat va ishonchlilik bilan belgilanadi. Sifatli ma'lumotlarni to'plash va tekshirish metodlarini mukammal egallash nafaqat texnik ko'nikma, balki ilmiy va professional mas'uliyatdir.

FOYDALANILGAN ADABIYOTLAR

1. DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.). Technics Publications.
2. ISO 8000-1:2023. Data quality — Part 1: Vocabulary. International Organization for Standardization.
3. Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.
4. OECD. (2025). AI and Data Governance: Guidelines for Quality, Trust and Accountability. OECD Publishing.
5. Barr, A., & Moore, M. (2024). Data Observability in Modern ML Pipelines. *Journal of Data Science & Engineering*, 12(3), 45–67.
6. Creswell, J. W., & Guetterman, T. C. (2019). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (6th ed.). Pearson.

7. Sculley, D., et al. (2015). Hidden Technical Debt in Machine Learning Systems. *Advances in Neural Information Processing Systems*, 28.
8. O'zbekiston Respublikasi OAK. (2023). Ilmiy ma'lumotlarni boshqarish va ochiq kirish siyosati bo'yicha tavsiyanoma. Toshkent.
9. Suyarov A.M. Tarjima qiluvchi dasturlar va ularning qiyosiy tahlili. *Journal of Theory, Mathematics and Physics* Vol. 4, No. 12, 2025 ISSN: 2181-4376, -С. 47-49. <https://jtmp.innovascience.uz/index.php/journal/issue/view/72>
10. Suyarov A.M. Elektron imzo haqida tushuncha va unga ega bo'lish. *Journal of iqro – iqro jurnali – volume 19, issue 01, 2025.-B. 955-957. Issn: 2181-4341, impact factor (research bib) – 7,245, sjif – 5,431. <http://www.wordlyknowledge.uz>*
11. Suyarov A.M. Ijtimoiy tarmoq xavfsizligi va ularda ishlash texnologiyasi. *Tamaddun nuri jurnali* ISSN 2181-8258 IF-9.347 DOI 10.69691 2025-yil, 12-son, 1-qism (75). –B.244-248.
12. Suyarov A.M. Internetdagi tadbirkorlik faoliyatini amalga oshirishdagi asosiy muammolar. *Международный современный научно-практический журнал «Научный Фокус» № 31 (100), часть1. Декабрь, 2025. –С. 320-324.*