XORIJIY TALABALAR MA'LUMOTLARINING HAYOTIY SIKLI VA SHAKLLANISH JARAYONLARINING KO'P QATLAMLI KONSEPTUAL MODELI

Abdullayev Javoxir Jaxongirovich
Buxoro davlat texnika universiteti o'qituvchisi, i.f.f.d. (PhD)
https://doi.org/10.5281/zenodo.22930575

Annotatsiya. Maqolada oliy ta'lim muassasalarida xorijiy talabalar to'g'risidagi ma'lumotlarning paydo bo'lishidan boshlab talaba bitirib ketganidan keyingi davrgacha bo'lgan hayotiy sikli tahlil qilingan. Ma'lumot olish nuqtalari uch qatlamga ajratilib, har bir qatlamdagi manbalar, atributlar va ularning ishonchlilik darajasi tavsiflangan. Shaxsni identifikatsiya qilish va ma'lumot to'g'riligi tushunchalari formal jihatdan farqlangan hamda atribut to'g'riligini ko'p manbali tasdiqlash sharti taklif etilgan. Taklif etilgan model xorijiy talabalar ma'lumotlarini boshqarish axborot tizimini loyihalash uchun konseptual asos bo'lib xizmat qiladi.
Kalit so'zlar: axborot tizimi, ma'lumotlar hayotiy sikli, konseptual model, ma'lumot manbalari, identifikatsiya, ma'lumot to'g'riligi, xorijiy talabalar.

Аннотация. В статье анализируется жизненный цикл данных об иностранных студентах в высших учебных заведениях — от момента их возникновения до периода после окончания обучения. Точки получения данных разделены на три уровня, для каждого описаны источники, атрибуты и степень их достоверности. Формально разграничены понятия идентификации личности и корректности данных, предложено условие многоисточникового подтверждения корректности атрибута. Модель служит концептуальной основой для проектирования информационной системы управления данными иностранных студентов.
Ключевые слова: информационная система, жизненный цикл данных, концептуальная модель, источники данных, идентификация, корректность данных, иностранные студенты.

Abstract. The paper analyses the lifecycle of data on international students in higher education institutions, from the moment the data first appear to the period after graduation. Data acquisition points are grouped into three layers, and the sources, attributes and reliability of each layer are described. The notions of identity resolution and data correctness are formally separated, and a multi-source confirmation condition for attribute correctness is proposed. The model provides a conceptual basis for designing an information system for managing international student data.
Keywords: information system, data lifecycle, conceptual model, data sources, identity resolution, data correctness, international students.

Kirish

So'nggi yillarda O'zbekiston oliy ta'lim muassasalarida tahsil olayotgan xorijiy talabalar soni sezilarli darajada oshdi. Ular haqidagi ma'lumotlar kontingent hisobi, o'quv jarayonini rejalashtirish, migratsiya organlari bilan hamkorlik va turli hisobotlarni tayyorlashda foydalaniladi. Biroq amaliyotda bu ma'lumotlar bir nechta mustaqil manbada, turli vaqtda va turli shaklda shakllanadi. Natijada bitta talaba haqida bir-biriga mos kelmaydigan yozuvlar paydo bo'ladi.

Mavjud axborot tizimlari asosan talaba muassasaga qabul qilinganidan keyingi davrda, ya'ni ma'lumot tizimga kiritib bo'linganidan so'ng ishlaydi. Talaba kelishidan oldingi

bosqich (agentlik, konsullik, hujjatlarni tan olish) va bitirganidan keyingi bosqich (bandlik, migratsiya, keyingi ta'lim) odatda e'tibordan chetda qoladi. Shu bois ma'lumotning qayerda va qachon buzilgani aniqlanmay qoladi. Holbuki "Raqamli O'zbekiston – 2030" strategiyasida ta'lim sohasidagi axborot tizimlarini integratsiyalash va ma'lumotlar sifatini oshirish raqamli transformatsiyaning muhim yo'nalishi sifatida belgilangan [1].

Ushbu maqolaning maqsadi xorijiy talabalarga oid ma'lumotlarning to'liq hayotiy siklini qamrab oluvchi konseptual modelni ishlab chiqish va unda identifikatsiya hamda ma'lumot to'g'riligi tushunchalarini formal farqlashdan iborat.

Adabiyotlar tahlili

Ma'lumotlar sifati masalasi axborot tizimlari nazariyasida R. Wang va D. Strong tomonidan ma'lumotdan foydalanuvchi nuqtai nazaridan tizimli ko'rib chiqilgan [2]. Y. Wand va R. Wang sifat o'lchamlarini ontologik asosda talqin qilib, ma'lumot real olamdagi holatni to'g'ri aks ettirishi kerakligini ko'rsatgan [3]. C. Batini va M. Scannapieco sifat o'lchamlari, ularni baholash va yaxshilash usullarini yagona tizimga keltirgan [4]. Ma'lumotlarni boshqarishning amaliy tamoyillari DAMA-DMBOK bilimlar to'plamida jamlangan bo'lib, unda ma'lumotlar hayotiy sikli boshqaruvning asosiy obyekti sifatida qaraladi [5].

Xorijiy talabalarni kuzatish bo'yicha eng rivojlangan amaliy tizimlardan biri AQShdagi SEVIS bo'lib, u talabani viza bosqichidan bitiruvdan keyingi amaliyotgacha kuzatadi [6]. Ammo bu tizim asosan holatni qayd etishga yo'naltirilgan, qayd etilgan ma'lumotning haqiqatga mosligini alohida tekshirmaydi.

Bir nechta manbadan kelgan qarama-qarshi qiymatlar ichidan haqiqiysini aniqlash masalasi xorijiy adabiyotda "haqiqatni aniqlash" (truth discovery) yo'nalishi doirasida o'rganiladi. Bu yondashuvda har bir manbaning ishonchliligi baholanadi va qiymatlar shu ishonchlilikka qarab tortiladi [7]. X. Dong va hammualliflar esa manbalar bir-biridan ma'lumot ko'chirishi mumkinligini, ya'ni ularning o'zaro bog'liqligini ham hisobga olish zarurligini ko'rsatgan [8].

Mahalliy tadqiqotlar orasida oliy ta'lim tizimida axborot oqimlarini boshqarishning integratsiyalashgan korporativ axborot tizimi [9] va korporativ tizimlararo axborot almashish samaradorligini oshirish [10] masalalari o'rganilgan. Ular tizim ichidagi ma'lumot harakatini yaxshilashga qaratilgan bo'lib, ma'lumot manbasining o'zidagi xatolar va tizimdan tashqaridagi bosqichlar ko'rib chiqilmagan.

Tadqiqot metodologiyasi

Tadqiqotda tizimli tahlil, ma'lumotlar oqimini tavsiflash va konseptual modellashtirish usullaridan foydalanildi. Dastlab OTMning xorijiy talabalar bilan ishlovchi bo'linmalari faoliyati, ular yuritadigan hujjatlar va o'zaro almashiladigan ma'lumotlar o'rganildi. So'ngra har bir manba bo'yicha qaysi atributlar shakllanishi, ular qachon yangilanishi va kim tomonidan yuritilishi aniqlandi. Olingan tavsif asosida manbalar va atributlar grafi qurildi.

Tahlil va natijalar

Hayotiy sikl qatlamlari. Xorijiy talaba haqidagi ma'lumotlar uch qatlamda shakllanadi (1-jadval).

1-jadval. Ma'lumot olish nuqtalarining qatlamlari

Qatlam	Asosiy manbalar	Tipik atributlar
L1. Kelishdan oldingi	ta'lim agentligi, qabul komissiyasi, konsullik (viza), hujjatlarni tan olish organi	F.I.Sh., fuqarolik, pasport, oldingi ta'lim hujjati, viza turi va muddati
L2. O'qish davri	dekanat, registrator ofisi, xalqaro bo'lim, yotoqxona, migratsiya ro'yxati	yo'nalish, kurs, o'zlashtirish, yashash joyi, ro'yxatga olish muddati, holat

Qatlam	Asosiy manbalar	Tipik atributlar
L3. Bitiruvdan keyingi	bitiruvchilar bo‘limi, ish beruvchilar, magistratura qabul komissiyalari	diplom, bandlik holati, mamlakatdan chiqish, keyingi ta’lim

Jadvaldan ko‘rinib turibdiki, bir xil atribut (masalan, fuqarolik yoki pasport) bir necha qatlamda qayta paydo bo‘ladi. Aynan shu takrorlanish ma’lumotni tekshirish imkonini beradi: agar atribut qiymati turli manbalarda mos kelsa, uning to‘g‘riligiga ishonch ortadi.

Formal model. Modelni quyidagicha ifodalaymiz. $S = \{s1, s2, ..., sn\}$ — manbalar to‘plami, $A = \{a1, a2, ..., am\}$ — atributlar to‘plami. Har bir manba uchun $r(s) \in [0; 1]$ ishonchlilik koeffitsiyenti belgilanadi. $G = (S \cup A, E)$ graf manba va atributlar orasidagi bog‘lanishni ko‘rsatadi: $(s, a) \in E$, agar s manbada a atribut qayd etilsa. $v(s, a)$ — s manbadagi a atributning qiymati.

Identifikatsiya va to‘g‘rilik. Amaliyotda bu ikki tushuncha ko‘pincha aralashtiriladi. Identifikatsiya — yozuv qaysi shaxsga tegishli ekanini aniqlash. U yagona identifikator yoki yozuvlarni moslashtirish orqali hal qilinadi:

$$ID(x) = ID(y) \Leftrightarrow x \text{ va } y \text{ bitta shaxsga tegishli} \quad (1)$$

To‘g‘rilik esa yozuv ichidagi har bir atribut qiymati haqiqatga mosligini bildiradi. Biz uni manbalarning vaznli kelishuvi orqali baholashni taklif qilamiz:

$$C(a, v) = \Sigma\, r(s)\cdot\delta(v(s,a) = v) / \Sigma\, r(s), \quad s \in S(a) \quad (2)$$

Bu yerda $S(a)$ — a atributni qayd etuvchi manbalar, δ — qiymatlar mos kelsa 1, aks holda 0 ga teng funksiya. (2) ifoda haqiqatni aniqlash usullaridagi vaznli ovoz berish sxemasining soddalashtirilgan ko‘rinishidir [7]. Atribut qiymati tasdiqlangan hisoblanadi, agar kamida k ta mustaqil manba mavjud bo‘lsa va $C(a, v) \geq \theta$ shart bajarilsa. Dastlabki bosqichda $k = 2$ va $\theta = 0{,}7$ qiymatlari tavsiya etiladi; ular keyingi empirik tadqiqotlarda aniqlashtiriladi.

Muhim xulosa shundaki, yozuv to‘g‘ri identifikatsiya qilingan bo‘lsa ham, uning ayrim atributlari noto‘g‘ri bo‘lishi mumkin. Masalan, pasport raqami bo‘yicha shaxs aniq topiladi, lekin fuqarolik maydoniga xato qiymat kiritilgan bo‘ladi. Bunday holat (1) shart bajarilib, (2) shart bajarilmagan holatga mos keladi va an’anaviy tizimlarda aniqlanmaydi.

Muhokama

Taklif etilgan modelning mavjud yondashuvlardan farqi ikki jihatda. Birinchidan, u ma’lumotni muassasa ichidagi holati bilan cheklamay, kelishdan oldingi va bitiruvdan keyingi qatlamlarni ham qamrab oladi. Ikkinchidan, u to‘g‘rilikni yagona manba ma’lumotiga emas, manbalarning vaznli kelishuviga asoslaydi. Bu yondashuv ma’lumot sifatini bir martalik tekshiruvdan doimiy nazoratga o‘tkazish imkonini beradi.

Modelning cheklovlari ham bor: ishonchlilik koeffitsiyentlarini aniqlash uchun real ma’lumotlar talab etiladi, L3 qatlamidagi manbalar esa hozircha to‘liq shakllanmagan. Bundan tashqari, real ma’lumotlar bilan ishlash shaxsga doir ma’lumotlarni himoya qilish talablariga qat’iy rioya qilishni taqozo etadi [11]. Bu masalalar keyingi tadqiqotlarda ko‘rib chiqiladi.

Xulosa

Xorijiy talabalar ma’lumotlarining hayotiy sikli uch qatlamdan iborat ekani asoslandi va har bir qatlamdagi manbalar hamda atributlar tavsiflandi. Identifikatsiya va ma’lumot to‘g‘riligi formal jihatdan farqlandi, atribut to‘g‘riligini ko‘p manbali vaznli kelishuv asosida baholash sharti taklif etildi. Model xorijiy talabalar ma’lumotlarini boshqarish axborot tizimining ma’lumotlar bazasi va tekshiruv modullarini loyihalash uchun asos bo‘ladi.

Foydalanilgan adabiyotlar

1. O‘zbekiston Respublikasi Prezidentining 2020-yil 5-oktabrdagi PF-6079-son "«Raqamli O‘zbekiston – 2030» strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida"gi Farmoni.
2. Wang R.Y., Strong D.M. Beyond accuracy: What data quality means to data consumers // Journal of Management Information Systems. – 1996. – Vol. 12, No. 4. – P. 5–33. – DOI: 10.1080/07421222.1996.11518099.
3. Wand Y., Wang R.Y. Anchoring data quality dimensions in ontological foundations // Communications of the ACM. – 1996. – Vol. 39, No. 11. – P. 86–95. – DOI: 10.1145/240455.240479.
4. Batini C., Scannapieco M. Data and Information Quality: Dimensions, Principles and Techniques. – Cham: Springer, 2016. – 500 p.
5. DAMA International. DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. – Basking Ridge: Technics Publications, 2017. – 628 p.
6. U.S. Immigration and Customs Enforcement. Student and Exchange Visitor Information System (SEVIS). – URL: https://www.ice.gov/sevis
7. Li Y., Gao J., Meng C., Li Q., Su L., Zhao B., Fan W., Han J. A survey on truth discovery // ACM SIGKDD Explorations Newsletter. – 2016. – Vol. 17, No. 2. – P. 1–16. – DOI: 10.1145/2897350.2897352.
8. Dong X.L., Berti-Equille L., Srivastava D. Integrating conflicting data: The role of source dependence // Proceedings of the VLDB Endowment. – 2009. – Vol. 2, No. 1. – P. 550–561. – DOI: 10.14778/1687627.1687690.
9. Abdullaev A.I. Oliy ta’lim tizimida axborot oqimlarini boshqarishning integratsiyalashgan korporativ axborot tizimi va modellarini yaratish: texnika fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati. – Toshkent: TATU, 2024.
10. Asqaraliev O.U. Ko‘p pog‘onali korporativ tizimlararo axborot almashish jarayonlari samaradorligini oshirish: texnika fanlari doktori (DSc) dissertatsiyasi avtoreferati. – Toshkent: TATU, 2025.
11. O‘zbekiston Respublikasining 2019-yil 2-iyuldagi O‘RQ-547-son "Shaxsga doir ma’lumotlar to‘g‘risida"gi Qonuni.