

ELEKTRON HUJJATLARDA JADVALLAR TUZILISHINI TANIB OLIISH

<https://doi.org/10.5281/zenodo.8195204>

Kenjayev Xamdam Bazarbayevich

Muhammad al-Xorazmiy nomidagi TATU Nukus filiali "Kompyuter tizimlari"

kafedra assistenti. E-mail: tatunukus2019@gmail.com

Annotatsiya

Elektron jadvallar katta hajmdagi biznes ma'lumotlarini to'plash va taqdim etishning eng mashhur vositalaridan biridir. Afsuski, aksariyat hollarda elektron jadvallarning tuzilishi aniqlanmagan. Natijada, ishlov berish vositalari ma'lumotlarni avtomatik ravishda, inson ishtirokisiz qabul qila olmaydi. Elektron jadvallar yaratilganda, odamlar sarlavhalar, ma'lumotlar va yig'ish katakchalarini to'g'ri ko'rsatish uchun vizual formatlashni amalga oshiradilar. Taklif etilayotgan usul insonlar kabi katakcha formatidagi vizual elementlarning farqiga asoslangan elektron jadval tuzilishini taniydi. Ushbu usul asosida ishlab chiqilgan dasturiy ta'minot elektron jadvalni kirish sifatida qabul qiladi va oddiy ob'ekt yozuvi shakliga asoslangan chiqishni ishlab chiqaradi. Usul bir necha bosqichlarni o'z ichiga oladi. Birinchidan, ish maydoni bo'sh bo'lmagan katakchalarning kvadrat maydoni sifatida aniqlanadi. Shundan so'ng, butun ish maydoni bitmaplar to'plamiga o'zgartiriladi, bu yerda 1 qiymati katakchada ma'lum formatlash mavjud bo'lganda ifodalanadi. Ikkinchi bosqich biznes ma'lumotlaridagi ma'lumotlar ketma-ketligini aniqlashni o'z ichiga oladi - yuqoridan pastga yoki chapdan o'ngga. Ushbu bosqichda biz Xaf usulini belgilangan bitmaplarning kichik to'plamiga qo'llaymiz. Keyinchalik, oddiy statistika usullaridan foydalangan holda ma'lumotlar va sarlavha belgilarini topamiz. Biz eng ko'p ishlatiladigan belgilarni ma'lumotlar katakchalari sifatida belgilaymiz. Subtitrlar va kiritilgan sarlavhalar uchun maxsus ishlov beriladi. Bunday subtitrlar olingan ob'ektlarga qo'shimcha xususiyat yaratadi. Usul turli biznes ma'lumotlarini, masalan, narxlar ro'yxatini o'z ichiga olgan elektron jadvallarning katta to'plamida sinovdan o'tkazildi.

Kalit so'zlar

katakchalarni avtomatik qayta ishlash, tuzilmani aniqlash, elektron jadvallar.

Elektron hujjat aylanish tizimlari keng qo'llaniladi va unda elektron jadvallar eng ommalashgan hujjat formati hisoblanadi. Ushbu turdagi hujjatlarga avtomatik ishlov berish odatda ma'lumot olish, bir nechta hujjatlardan ma'lumotlarni yig'ish va taqqoslash, elektron jadvallar asosida ma'lumotlar bazasini yaratish va to'ldirish kabilar bilan bog'liq ko'p vaqt talab qiladigan vazifalarni hal qilishga yordam

beradi. Bunda asosiy muammo - jadval tuzilishini tanish, ya'ni uning sarlavhalari, quyi sarlavhalari va ma'lumotlarning o'zi kabi qismlarini aniqlashdir. Keltirilgan muammoni hal qilishga bir qator ishlar tadqiq qilingan.

Jadvaldagi formulalarni tahlil qilishga asoslangan yechim [1] da taklif qilingan. Mualliflar elektron jadvallarni tavsiflash tilini kiritishgan va Prolog mantiqiy dasturlash tili yordamida strukturani aniqlash muammosini hal qilishgan. Elektron jadvallarning yashirin strukturaviy xususiyatlarini izlash va bu xususiyatlardan elektron jadvallarni tahrirlashda foydalanuvchiga avtomatik taklif qilish [2] ishda muhokama qilinadi. Bunda mualliflar hujjatning relyatsion modeli kabi ma'lumotlar to'plamlari orasidagi funktsional bog'liqliklarni qarashgan. Elektron jadvallar va relyatsion ma'lumotlar bazasi aro oson o'tish usuli [3] maqolada tavsiflangan. Ushbu usulning asosi elektron jadvallar va ma'lumotlar bazalari o'rtasidagi munosabatlarni aniqlash bilan bog'liq degan gipotezaga kelgan. Bunday munosabatlarni izlash uchun mualliflar ma'lumotlar orasidagi funktsional bog'liqlikni izlaydilar. Funktsional bog'liqliklarni hisoblash formulalari bizga aniqlik kiritish bosqichlarini shakllantirish imkonini beradi. Ma'lumotlarni aniqlash nazariyasi yondashuvidan [4] maqolada foydalanilgan.

Yuqorida keltirilgan usullar jadval tuzilishini tanish muammosiga turli tomonlardan yondashadi. Ba'zilar qo'shimcha tuzilma ma'lumotlarini talab qiladi, boshqalari formulalar orqali ma'lumotlarni bog'lashga tayanadi. Ushbu ishda taklif qilingan usul jadval tuzilishini tahlil qilishda vizual yondashuvga asoslangan. U jadvalning qat'iy tuzilishiga emas, balki dastlabki elektron jadval faylidan olingan jadvalning grafik tasviriga tayanadi. Dastlabki ma'lumotlarga XLS va XLSX formatidagi elektron jadval hujjatlar kiradi. Usul skanerlangan hujjatlar yoki boshqa grafik tasvirlardan jadvallarni tanib olishga mo'ljallanmagan.

Jadval tuzilishi haqida tushuncha

Jadvalning tuzilishi Gost2.105-95 xalqaro standartiga ko'ra unda bosh, yon tomon va matritsa yadrosi ajralib turadi. Jadval boshi ustun sarlavhalarini o'z ichiga oladi, ular ko'p darajali bo'lishi mumkin. Yon tarafda qator sarlavhalari mavjud. Shuningdek o'rnatilgan sarlavhalardan foydalanish mumkin. Jadvalning tepasida uning nomi va raqami, uning ostida boshqa qo'shimcha ma'lumotlar joylashtirilishi mumkin, masalan, yaratilgan sana, muallifning ismi yoki o'lchov birligi. Standart jadvallarni loyihalash qoidalarini tartibga soladi, lekin jadvalda saqlangan ma'lumotlarga ta'sir qilmaydi.

Yana bir yondashuv jadvallarni rasmiy ravishda argumentlar va bog'liq o'zgaruvchilar to'plami bo'lgan axborot ob'ektlari nuqtai nazaridan tavsiflash [5]. Ushbu yondashuv relyatsion modelga yaqin bo'lib, unga ko'ra jadval qatori

katakchalar orasidagi semantik munosabatlar to'plami bilan ifodalanadi [6]. Bunday modelda jadval ma'lum bir nomlangan xususiyatlarga (atributlarga) ega bo'lgan ob'ektlar to'plami haqidagi ma'lumotlarni saqlaydi. Bunda har bir yozuvni atribut-qiymat juftligiga aylantirish mumkin. Atribut nomlari ustun sarlavhalarida, qiymatlar esa katakchalarda ko'rsatiladi. Ob'ekt jadvalning bitta qatorini egallaydi. Hujjatlarda atributlar ierarxiyaga birlashtirilib, bir xil munosabatlarning turli xil vizual tasvirlarini shakllantirishi mumkin.

Jadval konstruksiyasini ajratib olish usuli

Qoida bo'yicha bir xil toifali jadvalning yo'nalishiga qarab ob'ektlar qatorlar yoki ustunlar bo'yicha joylashtiriladi. Odatda ob'ektlarning xususiyatlarini qatorlar bo'yicha ko'rsatish ko'proq tarqalgan. Qator orientatsiya holatida bitta qator bitta ob'ektni tavsiflaydi va sarlavhalar barcha ob'ektlar uchun umumiy bo'lgan xususiyatlarning nomlarini o'z ichiga oladi. Sarlavhalar va ma'lumotlarning joylashishini, ya'ni jadval tuzilishini aniqlash masalasini hal qilgandan so'ng, ob'ektlarning xususiyatlarini va ularning qiymatlar to'plamini ajratib olish mumkin. Jadval tuzilishini aniqlash quyidagi bosqichlardan iborat:

- 1-bosqich - jadval maydonining chegaralarini aniqlash;
- 2-bosqich - ma'lumotlar yo'nalishini aniqlash;
- 3-bosqich - ma'lumotlar va sarlavhalar andozasini qidirish;
- 4-bosqich - sarlavhalarni qayta ishlash.

1-bosqich. Jadval ish maydonining chegaralarini aniqlash. Elektron jadval hujjatlarida "varaq" atamasi hujjatning bir sahifasiga murojaat qilish uchun ishlatiladi. Umumiy holda, bir varaqda bir nechta jadvallar joylashishi mumkin, ammo ko'rib chiqilayotgan usul har bir varaqda faqat bitta jadval borligini nazarda tutadi.

Yangi elektron jadval hujjat yaratilganda, ilova (ms: excel) to'liq katakchalar bilan to'ldirilgan standart o'lchamdagi bo'sh varaqni ko'rsatadi. Biroq, katakchalar aslida foydalanuvchi tomonidan tahrirlanganda yaratiladi. Shunday qilib, hujjatda o'zgartirilmagan katakchalar mavjud emas, bu esa saqlangan ma'lumotlarning hajmini kamaytiradi. Biroq, ba'zi hollarda foydali ma'lumotlarga ega bo'lmagan katakchalar yaratiladi. Misol uchun, agar siz qator boshida bo'lmagan katakchani tahrir qilsangiz, satr boshidan o'sha katakchagacha bo'lgan barcha katakchalar yaratiladi. Bundan tashqari, butun ustunni tanlashva unda formatlash bajarilganda, ustunning barcha katakchalari standart varaq balandligiga mos ravishda yaratiladi.

Jadvalning ish maydoni - bu foydali ma'lumotlarni o'z ichiga olgan va qayta ishlanishi kerak bo'lgan katakchalar diapazonidir. Yuqorida ko'rsatilgan xususiyatlar bilan bog'liq holda, birinchi bosqichda ko'rib chiqilayotgan

katakchalar to'plamidan bo'shlarini olib tashlash orqali kamayadi. Ish maydoni ishchi varaqning barcha bo'sh bo'lmagan katakchalarni o'z ichiga olgan R_w dan R_h gacha bo'lgan minimal to'rtburchaklar diapazoni sifatida aniqlanadi.

2-bosqich. Ma'lumotlarning yo'nalishini aniqlash. Agar sarlavhalari ustunlarda joylashgan jadval ustunli jadval deyilsa, sarlavhalar qatorlarda bo'lgan jadval qatorli jadval deyiladi. Ma'lumotlar yo'nalishi jadvaldagi ob'ektlarning joylashishini anglatadi. Ular qatorlar yoki ustunlar bo'yicha bo'lishi mumkin.

Usul jadval strukturasi vizual xususiyatlaridan (masalan, struktura elementlarining nisbiy joylashuvi, formati va katakcha ma'lumot turlari) foydalanadi. Har bir katakcha katakchani o'ziga va uning ichidagi ma'lumotlarga tegishli atributlar to'plamiga ega. Ma'lumotlar yozuv formati va turiga ega, masalan, ma'lumotlar turi "raqam", yozuv formati esa "foiz" bo'lishi mumkin. Katakcha va ma'lumotlar matn va fon rangi, chegaralar, tekislash kabilar dan iborat formatga ega bo'lishi mumkin. Har bir atributda mumkin bo'lgan qiymatlar to'plami mavjud, masalan, shrift atributi Arial, Verdana, Georgia va h.k.

Ushbu bosqichda har bir atributning har bir qiymati uchun bitmap tuziladi. Bitmapning o'lchamlari ish maydonining o'lchamlari bilan bir xil bo'lib, berilgan katakchani berilgan qiymatga va atributga ega ekanligini anglatadi. 1-rasmda kitob do'konida mavjud bo'lgan kitoblar haqidagi ma'lumotlarga ega jadval namunasi ko'rsatilgan.

Jadvalning o'ng tomonida "ma'lumotlar turi" atributi uchun bitmaplarning namunasi mavjud. Birinchi kartada u "matn", ikkinchisida - "raqamli", uchinchisida - "sana" qiymatini oladi. Barcha atributlarning barcha qiymatlari uchun tuzilgan bitmaplar to'plami jadvalning o'ziga xos grafik tasvirini ifodalaydi. Jadvaldagi ob'ektlarning joylashuvi bitmaplarda olingan tasvirga ta'sir qilishini ko'rish mumkin.

Ob'ektlar bo'lsa (bizning holda, kitoblar) satrlarda joylashgan bo'lib, ularning har xil turdagi va formatlarga ega bo'lgan, ustunlar bo'yicha tashkil etilgan xususiyatlari bitmaplarda vertikal chiziqlar sifatida paydo bo'ladi.

"Asosiy" kichik sarlavhalar		Kichik sarlavhalar			1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
		02-00-85-00			0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
ID	Kichik sarlavha	Mavzu	Yili	Naxsi	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Chet el adabiyoti																			
134289	ABE	Barbu Kralin	2020	49000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
362565	Kirchikina shaxsida	Anton de Sent Ekzopieri	2019	89000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
432903	Uchi qalbi	Jekil London	2022	35000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
496756	1985	Blizani Buzanin	2021	49000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
673060	Graf Mirante Kristin	Aldegarck Drama	2019	69000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
325367	Chinotuz	Fransuz Dantsevalov	2019	35000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
O'zbek adabiyoti																			
56149	Otkan kumuk	Abdulla Qodiriy	2019	39000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
135824	Ichki ishlar xizmati	O'zbekiston	2018	35000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
325266	Shaxs	Tolib Mollit	2019	139000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
268658	Uchi	Said Ahmad	2019	59000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
786410	Uchi boshqaruvi	O'zbekiston	2021	29000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
154988	Savon davlati o'limi	Xudobekov To'xtabekov	2019	89000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
Poziyalar																			
123471	Roman va Injil	Utkan Shaxs	2022	25000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
388729	Fant	Yuliam Volfigar Gant	2022	49000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
254869	Yozuvchi Otkan	A.S. Roshkin	2021	89000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
208653	O'zbekiston	Blizni Yuliam	2019	35000	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0

1-rasm. Jadval uchun bitmap hosil qilish namunasi

Shuning bilan birga, asosiy va o'rnatilgan sarlavhalar gorizontall chiziqlar ko'rinishida bo'ladi.

Biz ikkita muhim xususiyatni ta'kidlaymiz. Birinchidan, jadvalning yo'nalishidan qat'i nazar, ichki sarlavhalar gorizontall ravishda joylashtiriladi. Ikkinchidan, sarlavhalar odatda boshqa ma'lumotlardan formatlash orqali farqlanadi va deyarli har doim qatorli ma'lumotlar turiga ega bo'ladi, obyekt xossalari esa turli xil ma'lumotlar turlari va yozuv formatlarini o'z ichiga oladi, lekin kamdan-kam hollarda formatlash bilan farqlanadi. Ya'ni, ma'lumotlar turlari va yozuv formatlari xaritalarida joylashgan chiziqlar deyarli bir ma'noda jadvalning yo'nalishini ko'rsatadi va formatlash xaritalarida hech narsani anglatmaydigan gorizontall chiziqlar topilishi ehtimolini oshiradi.

Ta'riflangan funktsiyalarga asoslanib, jadvalning yo'nalishini aniqlash uchun faqat ma'lumotlar turlari va yozuv formatlarining xaritalari qo'llaniladi. Xaritalar to'plamida vertikal chiziqlar ustunli jadvalni, gorizontall chiziqlar - kichik harflarni ko'rsatadi.

Bitmapdagi chiziqlarni aniqlash uchun Xaf usuli [7, 8] qo'llaniladi, bu umumiy holatda parametrik tenglama bilan berilgan tasvirdagi ixtiyoriy egri chiziqlarni topish masalasini hal qiladi. To'g'ri chiziqlarni qidirishda funktsiya qutb koordinatalarida to'g'ri chiziq tenglamasi bo'ladi. Qutb koordinatalarni tanlash vertikal chiziqlarni ifodalash zarurati bilan bog'liq. Xaf usuliga ko'ra, to'g'ri chiziq tenglamasining ikkita parametri uchun matritsa tuziladi -bunda boshlang'ich chiziqgacha bo'lgan masofa (R) va moyillik burchagi (th). Bu holda kerakli to'g'ri chiziqning moyillik burchagi 0 yoki 90 daraja qiymatini olishi mumkin va R masofasi bir qadam bilan $R_{max} = \max(R_w, R_h)$ ish maydonining o'lchovlaridan 0 dan maksimalgacha o'zgaradi (2-rasm). Jadvalning ahamiyatga ega joylarining

tanlanishi ko'rsatilgan, ularning har biri uchun ushbu nuqtadan o'tadigan barcha chiziqlar qidiriladi va Xaf matritsasi to'ldiriladi (2-rasm). Muayyan bitmap uchun ahamiyatli nuqtalari sifatida 1 qiymatiga ega bo'lgan katakchalar va birgalikda 0 qiymatiga ega bo'lgan kamida bitta katakcha yoki bitmap uchun chegara katakchalari hisoblab chiqiladi. 3-rasmدا jadvalning ahamiyatga ega joylari quyuuq kulrang rangga bo'yalgan.

R	0	1	2	3	...	R _{max}
0					...	
90					...	

2-rasm. To'g'ri chiziqlarni izlash uchun Xaf usulining qo'llanilishi

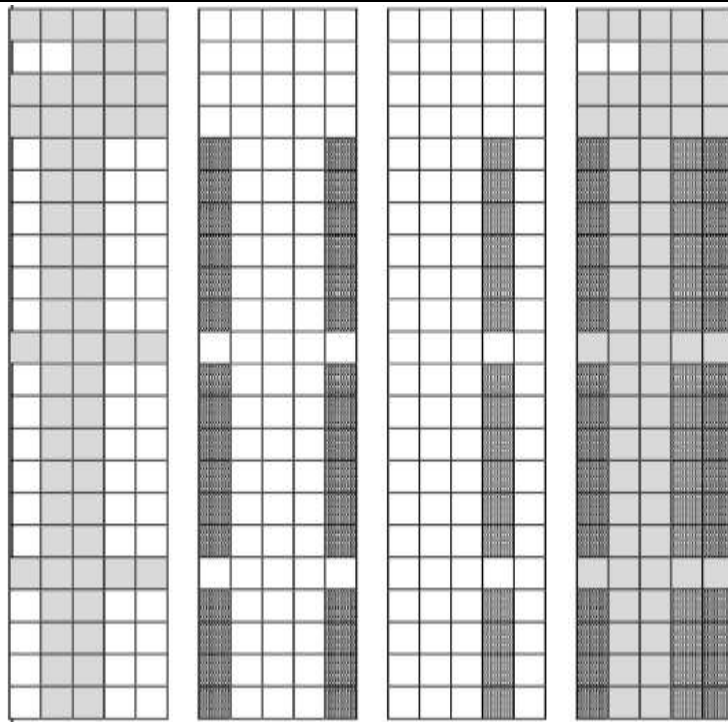
1	1	1	1	1	0	1
0	1	1	1	1	0	1
0	1	1	1	1	0	1
0	1	1	1	1	0	1
0	1	1	1	1	0	0
0	1	1	1	1	0	1
0	1	1	1	1	0	1
0	1	1	1	1	0	1

3-rasm. Jadvalning ahamiyatga ega nuqtalarni tanlash namunasi

To'ldirilgan Xaf matritsasi qaysi chiziqlar eng ko'p nuqtalarni o'z ichiga olganligini va shunga mos ravishda ular tasvirda eng ko'p bo'lishini ko'rsatadi. Olingan nuqtalar matritsasining har bir satri ma'lum o'lcham uchun mumkin bo'lgan maksimal qiymatga (ya'ni jadvalning balandligi yoki kengligi) normallashtiriladi. Keyin qator qiymatlari birlashtiriladi, tartiblanadi va o'rtacha arifmetik qiymat hisoblab chiqiladi.

3-bosqich. Ma'lumotlar andozasi va sarlavhalarini topish. Qoida tariqasida, jadvaldagi bitta ob'ekt bitta qatorga mos keladi. Biroq, jadvaldagi har bir qatorda ob'ekt mavjud emas. Qator asosiy sarlavhani, o'rnatilgan sarlavhani, jamlangan ma'lumotlarni o'z ichiga olishi mumkin yoki jadvalni vizual ravishda bloklarga bo'lish uchun bo'sh bo'lishi mumkin. Bundan tashqari, ish maydoniga nafaqat jadval, balki uning atrofidagi ma'lumotlar ham, jadvalning yuqorisida ham, pastida ham bo'lishi mumkin va bu ob'ekt emas.

Bu ob'ektlarni o'z ichiga olgan chiziqlarni boshqa barcha qatorlardan ajratish muammosini keltirib chiqaradi. Ob'ektlarning bir xilligi va ularning ko'pligidan kelib chiqqan holda, bu muammoni algoritmning oldingi bosqichida yaratilgan bitmaplar yordamida statistik jihatdan hal qilish mumkin. Barcha ob'ektlar uchun bir xil ustunlar bir xil ma'lumotlar turiga tegishli va formatlashda o'xshash bo'lgan mos qiymatlarni o'z ichiga oladi. Bitmaplarni bir butun sifatida ko'rib chiqsak, biz ma'lumotlar va o'rnatilgan sarlavhalarga mos keladigan qatorlarning takrorlanuvchi belgilarini topishimiz mumkin (4-rasm).



4-rasm. Bitmaplar to'plami bo'ylab ma'lumotlar namunasini qidirish (bo'yalgan hujayralar qiymati 1 ga teng)

Andozalarni qidirishda barcha mavjud bitmaplarni hisobga olish tavsiya etilmaydi, ko'p hollarda ma'lumotlar turlari va formatlari xaritalari bilan cheklanishi mumkin. Bitmaplarning birlashuvini tahlil qilib, biz barcha noyob chiziqlarni tanlaymiz va ularning takrorlanish chastotasini hisoblaymiz. Keyin biz D dan ko'p bo'lmagan turli xil katakchalar bilan noyob qatorlarni belgilarga guruhlaymiz. Andozalar qatorlardan farq qiladi, chunki ular qiymatlari aniqlanmagan katakchalarni o'z ichiga olishi mumkin, ya'ni u 0 yoki 1 bo'lishi mumkin. D parametri qatordagi katakchalar soniga (n) qarab quyidagi qiymatlarni oladi:

$$D(n) = \begin{cases} 0, n < 3, \\ 1, 3 < n < 10, \\ \lceil n/5 \rceil, n > 10, \end{cases} \quad (1)$$

Manba jadvalidagi eng ko'p qatorlar soniga mos keladigan belgi ma'lumotlar namunasi sifatida qabul qilinadi.

Ikkinchi eng keng tarqalgani - sarlavha qo'yish belgisidir. Sarlavhalar bir necha darajali bo'lsa, bunday andozalar bir nechta bo'lishi mumkin. O'rnatilgan sarlavha belgisiga mos keladigan barcha qator ma'lumotlari orasida paydo bo'lishi kerak. Aks holda, ularni o'rnatilgan sarlavhalar deb hisoblash mumkin emas. Eng kichik guruh jadval sarlavhasining asosiy sarlavhalarini o'z ichiga oladi, ular odatda bitta nusxada uchraydi, shuningdek, yuqoridagi toifalarga kirmaydigan

satrlar ham bo'lishi mumkin. Barcha aniqlangan belgilar elektron jadval hujjatlaridan olingan ma'lumotlar asosida tuzilgan ob'ektlarni qurishda qo'llaniladi.

4-bosqich. Sarlavhalar va o'rnatilgan sarlavhalar bilan ishlash. Jadvalning sarlavhalarini tanib olish bir nechta taxminlarga asoslanadi. Birinchidan, jadval sarlavhasi to'g'ridan-to'g'ri ma'lumotlarning birinchi qatoridan, ya'ni birinchi ob'ektdan yuqorida bo'lishi kerak. Sarlavha va birinchi ob'ekt o'rtasida faqat o'rnatilgan sarlavha qo'yilishi mumkin. Ikkinchidan, sarlavha ierarxik tuzilishga ega. Agar sarlavha bir nechta satrdan iborat bo'lsa va ierarxiyaning yuqori darajalari hujjiralarni birlashtirish orqali shakllantirilsa, quyi darajadagi har bir katakcha yuqori darajadagi to'liq bitta ajdodga ega bo'lishi kerak. Uchinchidan, ob'ektning har bir maydoni tepasida aynan bitta sarlavha bo'lishi kerak. Bu shuni anglatadiki, sarlavhada bo'sh katakchalar bo'lishi mumkin emas, bir nechta sarlavhalar bitta maydonga mos kelmaydi va bitta sarlavha bir nechta maydonlarga mos kelmaydi.

Ushbu taxminlarni hisobga olgan holda, biz sarlavhani qidirishda harakatlar ketma-ketligini quyidagicha shakllantiramiz.

- manba jadvalidagi ma'lumotlar andozalaridan biriga mos keladigan birinchi qatorni qidiramiz;
- birinchi ma'lumotlar qatori ustida joylashgan qatorni tanlaymiz va uni yuqorida tavsiflangan qoidalarga muvofiqligini tekshiramiz;
- agar chiziq qoidalaridan biriga zid bo'lsa, biz mos keladiganini topgunimizcha, uning ustidagi keyingi qatorga o'tamiz;
- mos keladigan qator topilsa, u pastki darajadagi sarlavhalar sifatida qabul qilinadi.

Sarlavha ko'p darajali yoki bunday emasligini aniqlash uchun uning ustidagi qatorni tekshiramiz. Agar asosiy satr yuqorida tavsiflangan qoidalarga javob beradigan quyi darajadagi sarlavhalar bilan izchil ierarxiyani tashkil qilsa, u ko'p darajali sarlavhaning keyingi darajasi sifatida qabul qilinadi. Bu bosqich ierarxiyaga mos kelmaydigan qator topilmaguncha takrorlanadi. Oxirida sarlavhani qidirish to'xtaydi.

Yuqorida tavsiflangan qadamlar natijasida biz belgilangan jadvalni olamiz, uning har bir satri jadval strukturasi bilan bog'liq. Endi jadvaldan foydali ma'lumotlarni, ya'ni undagi ob'ektlarni to'g'ridan-to'g'ri ajratib olish mumkin. Yuqorida aytib o'tilganidek, ob'ekt xususiyatlar to'plami va ularning qiymatlari bilan ifodalanadi, shuning uchun bir nechta sarlavha sathlaridan ob'ekt xususiyati nomi sifatida xizmat qiladigan bitta qatorni yaratishingiz kerak. Buni,

masalan, shunday qilish mumkin: "Sar.1: Sar.2 ... Sar.N". Ya'ni, ajratuvchi sifatida ikki nuqtadan foydalanib, pastki darajadagi sarlavhaga to'liq yo'lni yozishimiz mumkin. Ob'ekt xususiyati qiymatlari ma'lumotlar andozasiga mos keladigan qatorlardan tanlanadi. O'rnatilgan sarlavhalar, shuningdek, ob'ektning toifaga tegishli ekanligini ko'rsatadigan haqiqiy yoki noto'g'ri xususiyatlar sifatida ham ko'rib chiqilishi mumkin. Qabul qilingan ob'ektlarni saqlash uchun JSON (JavaScript Object Notation) formatiga o'xshash formatdan foydalanishimiz mumkin.

Ishlab chiqilgan algoritmi sinovdan o'tkazish

Ishlab chiqilgan algoritmi turli fanlar bo'yicha jadvallar to'plamida sinovdan o'tkazildi. Tanib olishning barcha mumkin bo'lgan nozik jihatlarini hisobga olish uchun sinovlarda yangi hujjatlar yaratilmadi, lekin turli odamlar tomonidan tuzilgan va bepul mavjud bo'lgan haqiqiy ma'lumotlarga ega jadvallardan foydalanildi. Barcha foydalanilgan jadvallarni quyidagi guruhlarga bo'lish mumkin:

- oddiy (bir darajali sarlavha va ma'lumotlar);
- murakkab sarlavhali (ko'p darajali sarlavhalar);
- o'rnatilgan sarlavhali;
- ko'p darajali o'rnatilgan sarlavhali.

Jadval tuzilmasini aniqlash algoritmining samaradorligi jadvallarning har bir guruhi va element turlari uchun alohida baholandi. Har bir satr jadvalning ma'lum elementi sifatida tan olinishi mumkin. Ular ma'lumotlar, i-daraja sarlavha, i-darajadagi sarlavha osti sarlavha va boshqalar ko'rinishida bo'ladi. To'g'ri tanib olingan jadval elementlarining ushbu turdagi elementlarning umumiy soniga nisbati muayyan jadval ishlash mezonini ko'rsatadi:

$$k = \frac{E_{rec}}{E_{all}} * 100\%, \quad (2)$$

bu yerda E_{rec} - to'g'ri tanib olingan elementlar soni; E_{all} - ushbu turdagi elementlarning umumiy soni; k - taniy olish koeffitsienti.

Butun jadval uchun umumiy tanib olish darajasi quyidagi formula bo'yicha aniqlandi:

$$K = \frac{\sum k_i}{N}, \quad (3)$$

Bu yerda N - turli elementlarning soni, $i \in [0, N]$.

Usulni sinovdan o'tkazish natijalari 1-jadvalda ko'rsatilgan. Jadvallarning har bir guruhi uchun birinchi qatorda ma'lum turdagi to'g'ri tanib olingan elementlar sonining test jadvallaridagi ushbu turdagi elementlarning umumiy soniga nisbati ko'rsatilgan. Ikkinchi satr noto'g'ri berilgan turdagi elementlar deb tanib olingan boshqa turdagi elementlarning soni ko'rsatilgan.

1-jadval. Ishlab chiqilgan algoritmnin sinovdan o'tkazish natijalari

Jadvallarguruxi	Asosiy sarlavhalar darajalari			O'rnatilgan sarlavhalar darajalari			Ma'lumotlar obyekt	Boshqalar	Tanish ko'rsatkichi, %
	1-й	2-й	3-й	1-й	2-й	3-й			
Oddiy jadvallar (32 dona)	30/32	0/0	0/0	0/0	0/0	0/0	12362/12878	41/43	95
	0	2	0	493	0	0	1	24	
Murakkab sarlavhali jadvallar (10 dona)	10/10	9/10	4/4	0/0	0/0	0/0	3597/3910	8/9	94
	0	0	1	281	0	0	0	33	
O'rnatilgan sarlavhali (23 dona)	22/23	0/0	0/0	183/197	0/0	0/0	4202/4511	15/17	92
	0	0	0	17	306	2	0	1	
Ko'p darajali o'rnatilgan sarlavhali(15 dona)	15/15	0/0	0/0	51/51	85/104	5/12	1637/1742	7/7	86
	0	0	0	19	7	97	0	8	
Turi bo'yicha tanilgan elementlar ko'rsatkichi, %	96	90	100	94	81	42	95	94	

Jadvalning asosiy sarlavhasi to'g'ri tuzilgan va bo'sh katakchalarni o'z ichiga olmasa, usul bilan to'g'ri taniy oladi. Ko'p darajali asosiy sarlavhalarni tanib olishdagi xatolar, uning ustidagi jadval haqidagi qo'shimcha ma'lumotlar, ba'zi hollarda, sarlavhalar ierarxiasining keyingi darajasi sifatida talqin qilinishi mumkin bo'lgan tarzda formatlanganligi bilan bog'liq.

O'rnatilgan sarlavhalar bilan ishlashda noaniqliklar ierarxiyaning bir darajasini boshqasidan farqlash uchun formatlash mavjud bo'lmaganda, masalan, ierarxiya katakcha matnida ko'p darajali raqamlash orqali yaratilganda sodir bo'ldi. Qisman to'ldirilgan jadvallar, barcha ma'lumotlar katakchalari qiymatlarini o'z ichiga olmaydi, shuningdek, ma'lumotlar katakchalari va o'rnatilgan sarlavhalarni tanib olishga salbiy ta'sir ko'rsatdi.

Shunga qaramay, ishlab chiqilgan usul jadvalning strukturaviy elementlarini tanib olishning yuqori aniqligini ko'rsatdi. Har xil turdagi jadvallar uchun to'g'ri tan olingan qatorlar ulushi o'rtacha 92% ni tashkil etdi.

Shunday qilib, maqolada jadval tuzilishini aniqlashning zamonaviy usullari muhokama qilindi. Jadvalning yo'nalishini, ma'lumotlar belgilari va sarlavhalarini aniqlaydigan ko'rinishning vizual xususiyatlariga asoslangan jadval tuzilishini tanib olish usuli taklif etiladi. Ishlab chiqilgan usul XLS va XLSX formatidagi hujjatlar bilan ishlash uchun POI kutubxonasidan foydalangan holda Java-da amalga oshiriladi. Usul ko'p sonli ob'ektlarni o'z ichiga olgan jadvallar tuzilishini tanib olishda yuqori aniqlikni ko'rsatdi. Usulni yanada takomillashtirish ma'lumotlarning semantik o'xshashligi asosida yaratilgan bitmaplardan foydalanish, shuningdek, katakchalar tarkibini semantik tahlil qilish orqali amalga oshirilishi mumkin.

FOYDALANGAN ADABIYOTLAR:

1. Paine J. Spreadsheet structure discovery with logic programming. *Proc. of EuSpRIG 2004 Conf. Risk Reduction in End User Computing: Best practice for spreadsheet users in the new Europe*. 2004, 11 p.
2. Cunha J., Saraiva J., Visser J. From spreadsheets to relational databases and back. *Proc. of the 2009 ACM SIGPLAN workshop on Partial evaluation and program manipulation*. 2009, Savannah, GA, USA.
3. Cunha J., Saraiva J., Visser J. Discovery-based Edit Assistance for Spreadsheets. *Proc. of 2009 IEEE Symp. on Visual Languages and HumanCentric Computing (VL/HCC 2009)*. 2009, pp. 233–237.
4. Oliveira J.N. Transforming data by calculation. *Generative and Transformational Techniques in Software Engineering II*. R. Lammel, J. Saraiva, J. Visser (Eds). International Summer School, vol. 5235, pp. 134–195, Springer Publ., Heidelberg, 2008.
5. Ivanov Yu.N., Emelyanov N.E., Sotnikova R.A. *Dokumenty: tipy, opisaniya* [Documents: Types, Descriptions]. Moscow, VNIISI Publ., 1987, 62 p.
6. Codd E.F. *The relational model for database management: ver. 2*. Addison-Wesley Longman Publ. Co., 1990.
7. Degtyareva A., Vezhnevets V. Preobrazovanie Khafa (Hough Transform). *Kompyuternayagrafikai multimedia* [Computer Graphics and Multimedia]. 2003, iss. 1 (2). Available at: <http://cgm.computergraphics.ru/content/view/36> (accessed June 9, 2016).

8. Duda R.O., Hart P.E. Use of the Hough transformation to detect lines and curves in pictures. *Communicationsofthe ACM*. 1972, vol. 15.1, pp. 11-15.