



10-ma'ruza: XML va XML-ma'lumotlar bazasi

Reja:

1. XML
2. XMLni ma'lumotlar bazasini boshqarish tizimida saqlash

XML

XMLning paydo bo'lgan yilini 1996 yil deb hisoblasa ham bo'ladi. Hammasi 1986 yilda paydo bo'lgan SGML tilidan boshlanadi.

SGML (Standard Generalized Markup Language ♦ umulashgan standart belgilash tili) belgilash tillarini xosil qiluvchi meta-til sifatida paydo bo'lgan.

SGML boshqa belgilash tillarini yaratishda keng qo'llaniladi.

Aynan shu til yordamida gipertekst hujjatlarni belgilash tili – HTML ni yaratishda ishlatilgan.

XML (eXtensible Markup Language ♦ kengaytirilgan belgilash tili) ♦ Butun dunyo o'rgimchak to'ri Konsorsiumi tomonidan tavsiya etilgan, strukturalangan ma'lumotlarni saqlashda, matn hujjatlarni shakllantiruvchi til. Shuningdek dasturlar aro ma'lumot almashishda ham ishlatilishi mumkin.

XML yaratishdan maqsad strukturalangan ma'lumotlarni har-xil tizimlar orasida internet orqali o'zaro ma'lumot almashishda mutanosiblikni ta'minlash bo'lgan.

XML tili hamHTML tiliga o'xshash til xisoblanadi. HTML dan farqli tomoni shundaki, XML da dasturchi o'zining shaxsiy teglarini yaratadi va ular orasiga ma'lumotlar joylashtiradi. XML-teglar harflar katta kichikligini farqlaydi. HTML teglari hujjatni ekranda ko'rinishini ifodalaydi. XML teglari hujjatdagi ma'lumotlarni tavsiflash uchun ishlatiladi. Undan tashqari XML yordamida yangi teglarni yaratish mumkin. XMLda ma'lumotlar tuzilmaviy holda saqlanadi. XML asosan ma'lumotlar almashinuvida ko'p ishlatiladi, chunki XML platformadan mustaqil bo'lib, HTTP orqali ishlashi juda qulay.

Quyida XMLda tuzilgan dastur bilan tanishamiz:

Bu dastur bloknotga yoziladi va .xml ko'rinishida saqlanadi. Bu dasturni ishga tushirish natijasida quyidagi oynaga ega bo'lamiz:

Buni quyidagicha qisqartiriladi:

XML hujjatlar turi

Standart bo'yicha to'g'ri tuzilgan ikki xil darajadagi XML hujjatlar aniqlangan:

To'g'ri tuzilgan (Well-formed). To'g'ri tuzilgan hujjat XML sintaksisining barcha umumiy qoidalariga mos keladi. Masalan, XML-hujjatning ochilgan tegi o'ziga mos yopilgan tegiga ega bo'lmasa, noto'g'ri tuzilgan hujjat hisoblanadi. Noto'g'ri tuzilgan hujjat XML-hujjat hisoblanmaydi. Bu holatda XML-protessor fatal xatolikni keltirib chiqarishi kerak.

Ishonchli (Valid). Ishonchli hujjat qo'shimcha yana semantik qoidalarga ham mos kelishi kerak. Bu foydalanuvchi tomonidan oldindan aniqlangan tashqi qoidalarga (masalan, struktura va matni) mos kelishi yanada jiddiyroq tekshiruvdan o'tkazilgan hujjat. Odatda bunday qoidalar hujjatning strukturasi, elementlari, atributlari va b. aniq yozib chiqilgan maxsus sxema – fayllarda saqlanadi.

XML sintaksisi

Bu bo'limda faqat to'g'ri tuzilgan XML-hujjatlar sintaksisi haqida so'z yuritiladi.

XML – ixtiyoriy ma'lumotni o'zida saqlashga mo'ljallangan matn shaklidagi ierarxik struktura.

Bunday struktura tashqi ko'rinishidan elementlar daraxti tarzida tasavvur etilish mumkin. XML elementlari teglar tarzida aniqlanadi. Misol tarzida XML yordamida belgilangan oddiy kulinariya retseptini ko'rishimiz mumkin:

```
Muka
Drojni
Tyoplaya voda
Sol
```

Smeshat vse ingredient? i t?atelno zamesit.

Zakr?t tkanyu i ostavit na odin chas v tyoplom pome?enii.

Zamesit ye?yo raz, polojit na protiven i postavit v duxovku.

XML e'lonlar

XML-hujjatning birinchi satri XML e'lon deyiladi (XML declaration). Bu zarur bo'lmagan satrda XML standart versiyasi va simvollar kodirovkasi e'lon qilinadi.

XML-protessor albatta UTF-8, UTF-16 yunikod-kodirovkalarni tanishi talab qilinadi. Boshqa keng ishlatiladigan ISO/IEC 8859 standartga asoslangan Windows-1251, KOI-8 ruscha kodirovkalarni ham tanishi mumkin.

Asosiy element

Eng muhim sintaksik talab – hujjat faqat bitta asosiy elementga (root element) ega bo'lishi kerakligidan iborat.

Bu hujjatning matni va barcha boshqa ma'lumotlari ochilgan va yopilgan asosiy teg orasida joylashishini anglatadi.

Eto kniga: "Knijechka"

Quyidagi kelgirilgan misoldagi hujjat to'g'ri tuzilgan XML-hujjat hisoblanmaydi:

nimadir №1

nimadir №2

Kommentariy

Daraxning ixtiyoriy joyida kommentariy elementini joylashtirish mumkin.

XML-kommentariy simvollar bilan tugagan oraliqda joylashtirilishi mumkin. Ikkita defis belgisi (--) kommentariy ichiga joylashtirilishi mumkin emas.

Kommentariy ichiga joylashtirilgan birorta ham teg ishlamaydi.

Teglar

XML-hujjatning qolgan qismi ichma-ich joylashgan elementlardan iborat bo'ladi. Ularning ba'zilarining ichida atributlari bo'lishi mumkin.

Element odatda ochilgan va yopilgan teglar oralig'idagi matn yoki boshqa elementlardan iborat bo'ladi. Ochilgan element burchak qavslarga o'ralgan element nomidan iborat bo'ladi. Yopilgan teg burchak qavslarga o'ralgan xuddi shu element nomi va slesh belgidan iborat bo'ladi.

Element nomi va atributlari nomida probel ishlatilishi mumkin emas. Lekin element nomi XML-hujjat qo'llab quvvatlaydigan ixtiyoriy kodirovkadagi tilda yozilishi mumkin.

Element matni (content) ochilgan va yopilgan teg oralig'idagi ixtiyoriy matn va jolashtirilgan boshqa elementlardan iborat bo'lishi mumkin.

bu yerda ixtiyoriy matn bo'lishi mumkin.

Elementning matnidan tashqari ochilgan tegida, element nomidan keyin, atribut nomi=qiymati juftligidan iborat bo'lgan atributi ham bo'lishi mumkin. Atribut qiymati doimo qo'sh tirnoqqa olinib yoziladi.

Muka

Yuqorida keltirilgan misoldagi «ingredient» elementining «3» va «stakan» qiymatli ikkita amount va unit atributlari bor ekanligi ma'lum bo'ladi.

Odiy tekstli matndan tashqari boshqa elemntlar ham bo'lishi mumkin:

Smeshat vse ingredient? i t?atelno zamesit.

Zakr?t tkanyu i ostavit na odin chas v tyoplom pome?enii.

Zamesit ye?yo raz, polojit na protiven i postavit v duxovku.

Bu yerda «instructions» elementi o'zida uchta «step» elementlarini saqlaydi.

XML o'zida bir birini yopib yuboradigan elementlarni ishlatishga yo'l qo'ymaydi. Quyidagi misolda noto'g'ri XML-hujjat ko'rsatilgan. Bu yerda «em» va «strong» elementlar bir birini o'zaro yopib yuborayapdi.

Ob?chn?y aksentirovann?y **v?deleenn?y i aksentirovann?y v?deleenn?y**

Quyida bo'sh elementning uch xil ko'rinishi ishlatilgan:

Maxsus simvollar

XML-hujjatda maxsus ishlatiladigan simvollar aniqlangan.

Maxsus simvollar boshqa simvollardan oldidan ampersend belgisi (&) va ortidan keyin nuqta vergul (;) qo'yib ajratiladi.

AT&T

Oldindan aniqlangan maxsus simvollar & (&), < (<), > (>), ' (') i " (") dan iborat. – uzun probelni anglatadi.

Ixtiyoriy simvolni uning yunikod kodini (&#yunikod; tarzida) yozib ham ko'rsatish mumkin:

AT&T

Afzalliklari

XML yunikodda ishlaydi;

XML formatda yozuvlar, ro'yxatlar va daraxtlar kabi ma'lumotlar strukturalarni ishlatish mumkin;

XML qat'iy aniqlangan strukturaga ega. Shu bilan birga ishlatishda oddiy, samarali va ziddiyatsiz bo'lib qoladi. Shu bilan birga har xil uslublarni ishlatishdan chegaralanmagan (masalan, qiymatlarni teg parametrlariga yoki teg tanasiga joylashtirilib, ma'lumotlarni modellash tirish mumkin, teglarni nomlashda turli xil tillardan foydalanish mumkin);

XML  xalqaro standartlarga asoslangan format;

XML ierarxik strukturasi amalda ixtiyoriy tipdagi hujjatlarni yozish imkonini beradi, faqat audio, video va multimedia potoklarni, rastrli suratlarni, tarmoqli ma'lumotlar strukturasi va ikkilik ma'lumotlardan tashqari;

XML litsenziyalash va har-xil cheklovlardan xoli bo'lgan oddiy tekst;

XML hech qanday platformalarga bog'liq emas;

Binar formatlardan farqli o'laroq, XML ob'ektni ta'riflaydigan nomlar, tiplar va sinflar haqida meta ma'lumotlarga ega, unga ko'ra ixtiyoriy ilova noma'lum strukturali hujjat bilan ishlayverishi mumkin (maslan, dinamik interfeyslarni qurishda);

XML barcha zamonaviy dasturlash tillari uchun sintaksik analizatorga (parserga) ega;

XMLda brauzerlarda, operatsion tizimlarda, veb-serverlarda amalga oshirilgan XSLT o'zgarish standart mexanizmiga ega;

XML zamonaviy qurilmalarda apparat, mikrodastur va dastur quyi darajasida qo'llab quvvatlanadi.

Kamchiliklari

XML sinteksisida keragidan ortiqchalik bor.

XML-hujjat o'lchami xuddi shu ma'lumotlarni binar shaklidan sezilarli darajada (10 baravar) katta.

XML-hujjat o'lchami xuddi shu ma'lumotlarni axborot uzatishning alternativ matn formatlari shaklidan (masalan JSON[2], YAML) katta.

XMLning keragidan ortiqcha sintaksisi ilovani ishlatishning samaradorligiga ta'sir qiladi. Ma'lumotlarni saqlash, qayta ishlash va uzatish narxi oshib ketadi.

XML maydonlar, sinflar nomi, strukturalarning ichma ich joylashuvi kabi metama'lumotlarni o'zida jamlagan. Shu bilan birga XML – ochiq tizimlarning o'zaro muloqot tili hamdir. Tizimlar orasida ko'psonli bir xil tipdagi (strukturali) ma'lumotlarni uzatilganda, har safar metama'lumotlarni takroran uzatishdan ma'no yo'q. Chunki bunday ma'lumot XMLning har bir nus'hasi ichida bor.

Ko'p sonli masalalar uchun XMLning butun boshli sintaksisining barcha kuchini ishlatishning keragi yo'q. Ko'pincha oddiyroq biror qismidan foydalanilsa yetarli.

Modellashtirishi bitta yechimga ega emas.

XML da modellashtirishning umumiy kelishilgan yagona uslubi yo'q. Aksincha relyatsion va ob'ektga yo'naltirilgan modellarda bunday vositalar ishlab chiqilgan va tizimli yondoshuv va tizimli tahlilda relyatsion algebraga tayanadi.

Tilning o'ta moslashuvchanligi va qat'iy cheklovlarning yo'qligi natijasida, bitta struktura ko'p usulda taqdim etilishi mumkin. Masalan, bir xil qiymat teg atributi yoki teg tanasi tarzida yozilishi mumkin.

yoki yoki 11 yoki yoki va h.k.

XML tiliga ma'lumotlar tipini qo'llab quvvatlashni kirmaydi. Unda ma'lumotlarni butun son, satr, sana, mantiqiy kabi tiplarga qat'iy ajratish yo'q.

XML tavsiya qiladigan ma'lumotlarning ierarxik modeli, relyatsion model va ob'ektga-yo'naltirilgan graflar va tarmoq modellariga nisbatan chegaralangan.

Ierarxik ma'lumotlar ifodasi (masalan graflar) qo'shimcha harakatlarni talab qiladi.

XML ismlar makonidan foydalanish va ularni XML-parserlarda qo'llash murakkab vazifalardan.

Inson tomonidan o'qilishi ancha qulay bo'lgan, XMLga o'xshash imkoniyatlarga ega bo'lgan, ma'lumotlarning boshqa tekst formatlari mavjud (YAML, JSON, SweetXM, XF).

XMLni WWW internet tarmog'ida aks ettirish

XML-hujjatni internet brauzerlarda foydalanuvchi o'qiy olishi uchun, uni o'zgartirishning uch xil usuli keng tarqalgan:

CSS stillarni qo'llash;

XSLT o'zgarishlarni tatbiq qilish;

Ixtiyoriy dasturlash tilida XML-hujjatga ishlov berish dasturini tuzib chiqish.

CSS yoki XSL ishlatilmasa, XML–hujjat ko'pchilik veb-brauzerlarda odiy tekst bo'lib ko'rinadi. Internet Explorer va Mozilla Firefox kabi ba'zi brauzerlar XML-hujjat strukturasi daraxt ko'rinishida ko'rsatadi.

CSS stillarni tatbiq qilish

Jarayon xuddi HTML-hujjatni aks ettirish uchun CSS stillarni qo'llagan bilan bir xil. XML-hujjatni brauzerda aks ettirishda CSS ni qo'llash uchun uchun, hujjat o'zida stillar jadvaliga ko'rsatkich havolasini saqlashi kerak bo'ladi. Masalan:

XSLT o'zgarishlarni tatbiq etish

XSL – XML-hujjat ma'lumotlarini qanday shakllantirish yoki o'zgartirishni aniqlab beradigan texnologiya hisoblanadi.

Hujjat brauzerda aks ettirishga yaroqli formatga o'giriladi.

XSL texnologiya eng ko'p brauzerlarda qo'llaniladi, lekin XSL yordamida ixtiyoriy boshqa formatlarga, masalan VRML, PDF tekst formatlariga ham o'girish mumkinligini unutmaslik kerak.

XSL transformatsiya (XSLT) vazifasini bermoq uchun, tarmoqning mijoz tomonida XMLning quyidagi ko'rinishdagi instruksiyasi bo'lishi talab qilinadi;

XML lug'ati

XML o'ziga xos abstrakt til bo'lgani uchun, XML lug'ati ishlab chiqilgan.

Mazkur lug'at dasturchilarga teglar nomi va ularning atributlarining cheklangan to'plamini ishlatishga kelishish imkoniyatini beradi.

XHTML – ko'pchilik brauzerlar tushunadigan, ana shunday ilk lug'atlardan biri hisoblanadi.

XHTML dan ko'pincha CMS kontentlarini saqlash va taxrirlashda foydalaniladi.

Ma'lumotlarni uzatishda ishlatiladigan CommerceML, xCBL va cXML kabi tijorat lug'atlariyam mavjud.

Bundan tashqari, insonlarga mo'ljallanmagan va ancha murakkab SOAP ma'lumotlarni uzatish protokoli kabi maxsus lug'atlar ham ishlab chiqilgan.

XMLni ma'lumotlar bazasini boshqarish tizimida saqlash

Axborot olamida o'zaro XML ko'rinishidagi ma'lumotlar almashinishi soni ortib borishi bilan, saqlanadigan XML-hujjatlar hajmi ham tez sur'atlarda o'sib bormoqda.

Ko'pgina veb-saytlar ma'lumotlarni saqlanadigan joyda XML dan foydalanadilar. saqlanayotgan ma'lumot foydalanuvchilarga ko'rsatilishidan oldin HTML yoki XHTML o'zgartiriladi. Microsoft InfoPath va W3C XForms kabi dasturiy mahsulot va texnologiyalarning yangi avlodlari Microsoft Access yoki SQL Server larni o'zining ma'lumotlarni saqlash joyi sifatida tanlab, XML-ma'lumotlar bilan ishlashni boshlayapdilar.

Shuningdek, ruxsatsiz foydalanishdan muxofazalangan konfidensial ma'lumotlarni saqlashda ko'pincha XML dan foydalaniladi.

Ma'lumotlarni XML ko'rinishida saqlashning o'zi kifoya qilmaydi, shuningdek ma'lumotlar xavfsizligi va kengayishini ham tadbirini qilib borish lozim.

XML ma'lumotlarni saqlash joyi

XML ko'rinishidagi ma'lumotlarni saqlash haqida gap ketganda, ish samarasi muhim omillardan biri hisoblanadi.

Agar XML hujjat oddiy matnli hujjat sifatida saqlanadigan bo'lsa, qanday qilib unga samarali ishlov berishni ta'minlash mumkin?

Saqlanayotgan ma'lumotlar hajmi gigabaytlar bilan o'lchanganda, xotirada butunlay joylashadigan hujjatning ob'ekt modelini (DOM) hosil qilish samarasiz hisoblanib, o'zgacha yondashuvlarni izlash kerak bo'ladi.

Saqlanayotgan XML ma'lumotlar hajmini o'sib borishi bilan, katta hajli ma'lumotlar orasidan kerakligini izlab topish samarasini oshirish yanada muhimroq ahamiyat kasb etayapdi.

XML ma'lumotlarni saqlashda qanday usul qo'llanishidan qat'iy nazar, ma'lumotni izlab topish tezligini oshirish uchun, ularni indekslash zarurligi yaqqol ma'lum.

Saqlanayotgan ma'lumotlar orasidan biror-bir to'plamni ajratib olish amaliyoti ko'p qo'llaniladi.

Internet tarmog'ida ishlayotganda, ma'lumot almashish tezligi muhim ahamiyatga ega. XML ko'rinishdagi ma'lumotlar bazasidan foydalanilganda, bunday ma'lumotlar bazasi kerakli tezlikni ta'minlab bera oladimi?

Bundan tashqari XML ma'lumotlar bazasi yetarlicha ishonchli ishlashi talab qilinadi.

XML va relyatsion ma'lumotlar

XML dan zamonaviy ma'lumotlar bazasida foydalanish masalasiga o'tishdan avval, relyatsion va XML ma'lumotlar strukturasini qisqacha solishtirib chiqamiz.

Relyatsion ma'lumotlar bazasida axborot ustun va satrlardan iborat jadvallarda saqlanadi. Ustunlarda jadvalning barcha yozuvlari bo'yicha bir xil ko'rinishdagi ma'lumotlar saqlanadi. Jadvalning har-bir yozuvi satr ko'rinishida taqdim etiladi.

Jadvaldagi satrlar tartibi ma'lumotlarning hech qanaqa tartibi bilan bog'liq emas. Aksincha, XMLda hujjatning joriy ichki tartibi XPath ga o'xshash funksiyalar qaytargan position() ma'lumotlarga o'z ta'sirini o'tkazadi.

Faqat eng sodd relyatsion ma'lumotlarga bitta jadvalda saqlanishi mumkin. Odatda relyatsion ma'lumotlar bazasi murakkab mantiqiy bog'lanishlar bilan o'zaro bog'langan jadvallardan iborat bo'ladi.

Har xil jadvaldagi ma'lumotlar o'zaro kalit maydonlar bilan bog'lanadi.

Odatda bunday kalit maydon sifatda yozuvning ID nomerli maydoni qatnashishi mumkin.

Ma'lumotlar orasidagi bog'lanish "birga-bir" (masalan, bitta farzandning bitta otasi bo'ladi), "birga-ko'p" (masalan, bitta otaning bir nechta farzandi bo'lishi mumkin), "ko'pga-ko'p" (masalan, bitta tovar bir nechta buyurtmalarda bo'lishi va bitta buyurtmada har xil tovar bo'lishi mumkin) tarzida bo'lishi mumkin.

Har qanaqa bunday bog'lanishlar ma'lumotni bir-nechta o'zaro bog'liq jadvallarda bo'lib saqlanishidan kelib chiqadi.

XPath, DOM yoki XML Infoset ma'lumotlar modelidan ko'rinib turganidak, ichki ierarxiyaga ega bo'lgan XML-hujjatlardan farqli o'laroq, relyatsion ma'lumotlar bazasi ierarxiyaga ega emas.

XML ma'lumotlar quyidagi misoldan ko'rinib turganidek ichki tartibga ega:

```
Value="1234.56" Currency="US Dollars" />
```

```
Value="4300.12" Currency="US Dollars" />
```

```
Value="10000.00" Currency="US Dollars" />
```

XMLning ichki ierarxiyasi XML-hujjatni to'g'ri tuzilganligini aniqlaydigan sharti hisoblanadi.

Shu kabi oddiy ma'lumotlarni relyatsion ma'lumotlar bazasida saqlash ularning tartibini yo'qolishiga olib keladi.

Indekslar

Hujjatlarni saqlaydigan joyi uchun o'rinli bo'lgan yana bir masala – bu so'rovlardir. Saqlanayotgan barcha ma'lumotlar ichidan faqat ma'lum bir qismini ajratib olish so'rovlar shaklida amalga oshiriladi. So'rovni amalga oshirish uchun, ma'lumot so'ralayotgan atributi bo'yicha indekslangan bo'lishi kerak.

Masalan, agar hujjatlarning katta kolleksiyasi ichidan, ma'lum avtor tomonidan yozilgan barcha hujjatlarni tezda ajratib olish zarur bo'lsa, buning uchun ma'lumotlarning avtori atributi bo'yicha indekslash kerak bo'ladi.

XML va ma'lumotlar bazasi

Relyatsion ma'lumotlar bazasi – ma'lumotlarni saqlashning mashxur usullaridan biri. Bunday usul uzoq vaqtdan beri strukturalangan ma'lumotlarni saqlashda izchil ishlatilib, bu vaqt mobaynida ancha mukammallashtirilgan. Ko'p sonli dasturchilar ulkan hajmdagi ma'lumotlarni saqlashda aynan shu ma'lumotlar bazasidan foydalanishi ham bejiz emas. Shu sabablarga ko'ra XML va relyatsion ma'lumotlar bazasini boshqarish tizimini birgalikda ishlatish mumkin bo'ladi.

10-mavzuga doir savollar:

1. XMLni ta'riflab bering.
2. SGML nima?
3. Ma'lumotlarni saqlashda XMLning roli qanday?
4. XML hujjat ma'lumotlari qanday strukturada bo'ladi?
5. XML va relyatsion MB birgalikda qay tarzda ishlatiladi?