

## 21 - mavzu. Expert tizimlari qurish qiyinchiliklari.

### 1. Strukturali-dog'li(strukturno-pyatenniy) etalon.

### 2. Cognitive Technologies dan mashinali o'qish darslar.

### 3. Qo'lyozma matnlarni tanish.

#### 1. Strukturali - dog'li etalon.

«Fontanli almashtirish» shablonli va strukturali tizimlarning afzalliklarini o'zida birlashtiradi va bizning fikrimizcha, ularning har biriga alohida xos bo'lgan kamchiliklardan qutulishga imkon beradi. Bu texnologiyaning asosida strukturali - dog'li etalonni qo'llash yotadi. U tasvirni belgining strukturasini beradigan bir - biri bilan n-ar munosabatlar orqali bog'langan dog'lar to'plami ko'rinishida tasvirlashga imkon beradi. Bu munosabatlar(ya'ni dog'larning bir - biriga nisbatan joylashishi) belgilarni tashkil etadigan strukturali elementlarni yuzaga keltiradi. Masalan kesma dog'lar orasidagi n-ar munosabatlarning bir turi. Ellips - boshqasi, yoy - uchinchisi. Boshqa munosabatlar belgini tashkil etuvchi elementlarning fazoviy joylashishini beradi.

*Etalonda* quyidagilar beriladi:

1 Nom;

2 majburiy, taqiqlangan va majburiy bo'lmagan strukturali elementlar;

3 strukturali elementlar orasidagi munosabatlar;

4 strukturali elementlarni belgini tavsiflaydigan to'rtburchak bilan bog'laydigan munosabat;

5 strukturali elementlarni ajratishda ishlatiladigan xususiyatlar;

6 elementlar orasidagi munosabatlarni tekshirishda ishlatiladigan atributlar;

7 elementlar va munosabatlarning sifatini baholashda ishlatiladigan atributlar;

8 elementni ajratish boshlanadigan vaziyat;

Tasvirlar sinfi uchun ajratiladigan strukturali elementlar dastlabki va qo'shma bo'lishi mumkin. *Dastlabki strukturali elementlar* - bu dog'lar, *qo'shmalari* - kesma, yoy, xalqa, nuqta. *Qo'shma strukturali elementlar* sifatida etalonda tavsiflangan ixtiyoriy ob'ektlar olinishi mumkin. Bundan tashqari ular dastlabki strukturali elementlar orqali ham boshqa qo'shma strukturali elementlar orqali xam tavsiflanishi mumkin.

Masalan koreyscha ierogliflarni tanishda (bo'g'inli xat) bo'g'inni tavsiflash uchun qo'shma element alohida harflarning tavsiflari hisoblanadi. Qo'shma strukturali elementlarni qo'llash taniladigan ob'ektlar sinflarining ierarxik tavsiflarini qurishga imkon beradi.

*Munosabatlar* sifatida strukturali elementlar orasidagi bog'lanishlar ishlatiladi. Bu bog'lanishlar yoki elementlarning metrik xarakteristikalarini(masalan, <uzunligi katta>) orqali aniqlanadi yoki tavsirda ularning o'zaro joylashishiga(masalan, <chaproqda>, <kesishadi>) qarab aniqlanadi.

Strukturali elementlar va munosabatlarni berishda muayyan sinf etalonida bu elementni ishlatganda strukturali element yoki munosabatni aniqlashga imkon

beradigan aniqlashtiruvchi parametrlar qo'llaniladi. Strukturali elementlar uchun aniqlashtiruvchilar sifatida masalan kesmaning mumkin bo'lgan yo'nalishini beradigan diapazon parametrlari bo'lishi mumkin. Munosabatlar uchun esa xarakterli nuqtalar orasidagi mumkin bo'lgan chegaraviy masofani beradigan parametr bo'lishi mumkin.

*Muayyanlashtiruvchi* parametrlar tasvirdagi muayyan strukturali elementning <sifatini> va berilgan munosabatning bajarilish <sifatini> hisoblashda ham ishlatiladi.

Taniladigan ob'ektlar sinflari uchun strukturali-dog'li etalonlarni qo'rish va sinash murakkab va og'ir jarayon. Tavsiflarni sozlash uchun ishlatiladigan tasvirlar bazasi har bir grafema uchun yaxshi va yomon namunalarga ega bo'lishi kerak. Bazadagi tasvirlar o'rgatuvchi va nazorat qiluvchi to'plarga ajratiladi. Tavsifni yaratuvchi strukturali elementlar va ular orasidagi munosabatlarni oldindan beradi. Tasvirlar bazasi asosida o'rgatuvchi tizim avtomatik ravishda elementlar va munosabatlarning parametrlarini hisoblaydi. Hosil qilingan etalon berilgan grafemalarning tasvirlarini nazoratli tanlovi asosida tekshiriladi va to'g'irlanadi. Nazoratli tanlov asosida tanish natijalari tekshiriladi, ya'ni gopotezani tasdiqlash sifati baholanadi.

*Strukturali - dog'li* etalonni qo'llash orqali tanish quyidagicha sodir bo'ladi. Etalon tasvir ustiga qo'yiladi va tasvirda ajratilgan dog'lar orasidagi munosabatlar etalondagi dog'larning munosabatlari bilan taqqoslanadi. Agar tasvirda ajratilgan dog'lar va ular orasidagi munosabatlar qandaydar belgining etalonini qanoatlantirsa, u holda bu belgi kiruvchi tasvirni tanish haqidagi gipotezalar ro'yxatiga qo'shiladi.

## **2. Cognitive Technologies dan mashinali o'qish darslari.**

Tizim <bitta tugma> tamoyili asosida ishlaydi. Bu shuni anglatadiki, <(Skanerla va tani)Skaniruy i Raspoznavay> tugmasini bosganda xujjatni qayta ishlash jarayoni ishga tushadi: skanerlash, sahifani matnli va grafik bloklarga ajratish, matnni tanish, orfografiyani tekshirish va chiquvchi faylni shakllantirish. Buning barchasida nima turadi? Ongli algoritm hujjatning foniga bog'liq ravishda skanerning optimal yorug'ligini avtomatik tanlashga(adaptiv skanerlash), illyustratsiyalarni saqlashga(yoki yechiladigan masalaga bog'liq ravishda keraksiz grafik elementlarini o'chirish) imkon beradi. Cunie Form da bunga o'xshash mosliklarning bir qancha usullari ishlatiladi. Birinchidan, har bir belgining shakli alohida elementar hodisalarga yoyiladi. Masalan kesishishning bir chizig'idan boshqasigacha bo'lgan qism hodisa hisoblanadi. Hodisalar majmui belgining ixcham tavsifini o'zida ifodalaydi.

Boshqa usullar alohida belgilar elementlari <og'irlik> larining o'zaro nisbatlariga va ularning xarakterli alomatlarini tavsiflashga asoslanadi. Bu tavsiflarning har biriga mos etalonlar topiladigan ma'lumotlar bazasi mavjud. Tasvirning qayta ishlashga beriladigan elementi etalon bilan taqqoslanadi. SHundan so'ng bu taqqoslashga asoslanib hal qiluvchi funktsiya tasvirning aniq belgiga mosligi haqida hukm chiqaradi. Bundan tashqari past sifatli matnlar bilan ishlashga imkon beradigan algoritmlar ham mavjud. Masalan «yopishib qolgan» belgilarni

ajratish uchun optimal ajratishni baholash usuli mavjud. Aksincha, «sochilgan» elementlarni birlashtirish uchun ularni birlashtirish mexanizmlari ishlab chiqilgan.

Cunie Form 96 da biz birinchi marta o'z-o'zini o'qitish (yoki adaptiv tanish) algoritmini qo'lladik. Ularning ishlash printsipti quyidagidan iborat. Har bir matnda aniq va noaniq bosilgan belgilar mavjud. Agar tizim matnni tanigandan keyin aniqlik chegaradan pastda ekanligi aniqlansa, yaxshi bosilgan belgilarning tizim generatsiya qilgan shriftiga asoslangan holda matnni qayta tanish amalga oshiriladi. Bu yerda ishlab chiqaruvchilar ikki turdagi tanish tizimining afzalliklarini birlashtirishgan: birinchisi ixtiyoriy shriftni qo'shimcha o'qitmasdan tanish imkonini beradi, ikkinchisi past sifatli matnlarni tanishda chidamli. Cunei Form 96 qo'llash natijalari shuni ko'rsatdiki, o'z - o'zini o'qitadigan algoritmlarning qo'llanilishi past sifatli matnlarni tanish aniqligini 4-5 marta oshirishga imkon beradi. Asosiysi o'z - o'zini o'qitadigan tizimlar tanish aniqligini oshirishda katta potentsialga ega.

*Sintaktik va lug'atli* tanish usullari muhim rol o'ynayda va mohiyatiga ko'ra geometrik tanishni taminlashda kuchli vosita bo'lib xizmat qiladi. Lekin ularni samarali qo'llash uchun ikkita muhim masalani yechish kerak bo'ladi. Birinchidan katta lug'atga (100000 so'z) tez murojatni amalga oshirish. Natijada so'zlarni saqlash tizimini qurishga erishildi va bunda har bir so'zni saqlash uchun bir baytdan oshmaydi, murojat esa minimal vaqtda amalga oshirildi. Boshqa tomondan hodisalarning alternativligiga yo'naltirilgan tanish natijalarini to'g'irlyadigan tizimni qurishga talab qilindi. O'z o'zidan tanish natijalarining alternativligi aniq va harflar kolleksiyasining <moslik bahosi> bilan birga saqlanishiga bog'liq. Lug'atli nazorat esa lug'at bazasini qo'llab, bu baholarni o'zgartirishga imkon berdi. Natijada lug'atni qo'llash belgilarni qayta tanish sxemasini amalga oshirishga imkon berdi.

Hozirgi kunda tanish aniqligini oshirish masalalari bilan birga tanish texnologiyalari bilan arxivli tizimlarni birlashtirish orqali OCR-texnologilarni qo'llanilish sohalarini kengaytirish masalalari oldingi o'ringa chiqmoqda. Boshqacha aytganda, biz matnni kiritishni amalga oshiradigan monoprogrammadan mijozning hujjatni qayta ishlash sohasidagi masalasini yechadigan avtomatlashtirilgan kompleksga o'tyapmiz. Mana yarim yildirki Cunei Form tashkilotlarda ma'lumotlarni birgalikda kiritishga mo'ljallangan Cunei Form OCR Server tanish serveri bilan chiqarilmoqda, tanish modulini o'z ichiga olgan <Evfrat> elektron arxivi esa qisqa vaqt ichida katta shuhrat qozondi.

Shunday mo'ljal bilan umuman tanish tizimlari haqidagi tasavvurlarni tubdan o'zgartirgan Cunei Form96 Professional komplekti yaratildi.

### **3. Qo'lyozma matnlarni tanish**

Qo'lyozma matnlarni tanish masalasi bosma matnlarni tanishga qaraganda ancha qiyin. Bosma matnlarni tanishda biz shrift tasvirlarining chegaralangan miqdori bilan ishlasak, qo'lyozma matnlarda esa shablonlar soni o'lchab bo'lmas darajada ko'p. Tasvir elementlari chiziqli o'lchamlarining boshqa munosabatlari qo'shimcha qiyinchiliklarni keltirib chiqaradi.

Lekin baribir qo'lyozma matnlarni tanish texnologiyasini ishlab chiqarishning asosiy bosqichlaridan o'tilganligini tan olishimiz mumkin. Cognitive Technologies

zaxirasida barcha asosiy turdagi matnlar: stilizatsiyalangan raqamlar, bosma matnlar va qo'lyozma belgilarni tanish texnologiyalari mavjud. Ammo qo'lyozma matnlarni kiritish texnologiyalarini moslashish(adaptatsiya) bosqichidan o'tishi talab qilinadi. SHundan so'ng xujjatlarni uzluksiz arxivga kiritish uchun vositalar to'liq amalga oshirilganligini e'lon qilish mumkin bo'ladi. Yangi kompyuter texnologiyalarning dinamik rivojlanishi elektron xujjat ayirboshlash sektorining holatiga ham o'z aksini topdi. Oldinlari klaviaturasiz kiritish texnologiyalarini rivojlantirish shaxsiy foydalanishning afzalliklariga qaratilgan bo'lsa, hozirgi kunda xujjatlarni kiritish va qayta ishlash texnologiyalaridan birgalikda va ratsional foydalanish afzalliklari oldingi o'ringa chiqmoqda. Hozirgi kunda birgina tinish tizimiga ega bo'lish yetarli emas. Tanilgan matnli fayllar bilan nimadir qilish kerak: ma'lumotlar bazasiga saqlash, ularni qidirash, lokal tarmoq orqali uzatish va h.k. Bir so'z bilan aytganda xujjat bilan ishlashning arxivli yoki boshqa tizimi bilan birgalikda ishlash talab qilinadi. SHunga ko'ra tanish tizimi xujjatlar bilan ishlashning arxivli yoki boshqa tizimi uchun utilitga aylanadi. Xujjatlarni skanerlash va tanish tizimlari tarmoq versiyalarining paydo bo'lishi bilan bizning kompaniyamizda turli xil tashkilotlarda bu texnologiyalardan birgalikda foydalanishning ba'zi bir afzalliklarini amalga oshirishga erishildi.