

## **20 - mavzu. Instrumental vositalarni tanlash.**

### **1. Kirish.**

### **2. Asosiy tamoyillar va idrok etishning yaxlitligi.**

### **3. Belgilarni tanish.**

Tayanch iboralar: to'liqlik tamoyili, yaxlit tavsiflash, shablonli tavsiflash, shablonli tizimlar, strukturali tizimlar, belgili tizimlar, strukturali-dog'li etalon, strukturali elementlar, hodisalarning muqobilligi, monoprogrammalar.

### **1. Kirish**

«Kompyuterga matnni avtomatik kiritadigan» maxsus tizimlarning mavjudligini hatto boshlang'ich foydalanuvchilar ham bilishadi. Ko'rinishidan bu oddiy va mantiqiy. Skaner qilingan tasvirdan tizim xarflarni «taniydigan» qismlarni topadi va bu tasvirlarni haqiqiy harflar bilan, boshqacha aytganda ularni mashina kodi bilan almashtiradi. Matn tasviridan «haqiqiy» matnga o'tish shunday amalga oshiriladi. Bunga qanday erishiladi?

«Bit» kompaniyasi tomonidan «Fontanli almashtirish» deb nomlangan belgilarni tanishning maxsus texnologiyasi ishlab chiqarilgan. Bu texnologiya asosida yuqori bahoga ega bo'lgan savdo maxsuloti ishlab chiqarilgan. Bu Fine Reader optik tanish tizimidir. Hozirgi paytda uning uning uchinchi versiyasi ishlab chiqarilgan bo'lib, nafaqat matnlar bilan, balki shakllar va jadvallar bilan ham ishlash imkoniga ega. Ishlab chiqaruvchilar esa uning to'rtinchi versiyasi nafaqat bosma matnlarni, hatto qo'lyozma matnlarni ham taniy oladigan bo'lishini bashorat qilishmoqda.

### **2. Asosiy tamoyillar va idrok etishning yaxlitligi.**

«Fontanli almashtirish» asosida yaxlitlik tamoyili yotadi. Unga muvofiq ixtiyoriy idrok etiladigan ob'ekt yaxlit, ma'lum munosabatlarda o'zaro bog'langan qismlardan iborat holda qaraladi. Masalan bosma sahifa bo'limlardan tashkil topgan, bo'limlar sarlavhalar va kolonkalar, kolonkalar abzatlardan tashkil topgan. Abzats satrlardan, satrlar so'zlardan, so'zlar haflardan tashkil topgan. Bunda matnning barcha sanab o'tilgan elementlari bir - biri bilan ma'lum fazoviy va til munosabatlar orqali bog'langan. Yaxlitlikni ajratish uchun uning qismlarini aniqlash kerak. Qismlarni o'z navbatida faqat yaxlitlikning tarkibida qarash mumkin. SHuning uchun idrok etishning yaxlit jarayoni faqat idrok etiladigan ob'ekt haqidagi gipotezalar doirasida yaxlit sodir bo'lishi mumkin. Idrok etiladigan ob'ekt haqidagi faraz o'rtaga tashlangandan keyin uning qismlari ajratiladi va interpretatsiya qilinadi. SHundan so'ng boshlang'ich gipotezaning to'g'riligini tekshirish uchun ulardan yaxlitlikni «yig'ish» ga harakat qilinadi. SHubhasiz, idrok etiladigan ob'ekt kattaroq yaxlitlik doirisida interpretatsiya qilinishi mumkin.

Biror gapni o'qib inson harflarni taniydi, so'zlarni idrok etadi, ularni sintaktik konstruktsiyalarga bog'laydi va uning ma'nosini tushunadi.

Texnik tizimlarda matnni tanishdagi ixtiyoriy yechimga birdaniga kirishilmaydi, balki gipotezalarni ketma - ket tekshirish va oldinga surish hamda tadqiq qilinadigan ob'ekt haqidagi bilimlarni ham, umumiy kontekstdagi bilimlarni ham jalb qilish orqali kirishiladi. Idrok etiladigan ob'ekt sinflarining yaxlit tavsifi ikkita shartga javob beradi: birinchidan, berilgan sinfdagi barcha ob'ektlar bu tavsifni qanoatlantiradi, ikkinchidan boshqa sinfdagi hech qanday ob'ekt uni qanoatlantirmaydi. Masalan, «K» harfining tasvirlari sinfi shunday tavsiflanishi kerakki «K» harfining ixtiyoriy tasviri unga tushsin, boshqa barcha xarflarning tasvirlari esa unga tushmasin. Bunday tavsif aks etish xususiyatiga ega bo'ladi, ya'ni tavsiflanadigan ob'ektlarni qayta ishlab chiqishni ta'minlaydi: OCR tizimlari uchun xarfning etaloni xarfni vizual qayta ishlab chiqishga imkon beradi, nutqni tanish uchun so'zlar etaloni so'zlarni talaffuz etish imkonini beradi, sintaktik analizatoridagi gapning strukturali tavsifi to'g'ri gapni sintez qilish imkonini beradi. Amaliy nuqtai nazardan aks etish katta rol o'ynaydi, modomiki tavsiflarning sifatini effektiv nazorat qilishga imkon beradi.

Yaxlit tavsiflashning ikki ko'rinishi mavjud: shablonli va strukturali.

*Birinchi holda* tavsiflash vektorli yoki rastri ko'rinishdagi tasvirni o'zida aks ettiradi va almashtirishlar sinfi beriladi(masalan, takrorlash, masshtablashtirish va x.k.)

*Ikkinchi holda* tavsiflash graflar ko'rinishida aks ettiriladi. Grafning tugunlari kiruvchi ob'ektning tashkil etuvchi elementlaridan iborat, yoylari esa ular o'rtasidagi fazoviy munosabatlardan iborat. O'z navbatida elementlar murakkab bo'lishi mumkin(ya'ni o'zining tavsifiga ega bo'lishi mumkin).

Albatta, shablonli tavsiflashni strukturali tavsiflashga qaraganda amalga oshirish ancha oson. Lekin uni yuqori o'zgarish darajasiga ega bo'lgan ob'ektlarni tavsiflash uchun qo'llab bo'lmaydi. SHablonli tavsiflashni masalan, faqat bosma belgilarni tanish uchun, strukturali tavsiflashni esa qo'lyozma matnlarni tanishda ham qo'llash mumkin.

*Idrok etishning to'liqligi* ikkita muhim arxitekturali yechimlarni taklif qiladi. Birinchidan, barcha bilimlar manbai imkon qadar bir vaqtda ishlashi kerak. Masalan, avval sahifani tanib, so'ngra uni lug'at va kontekst qayta ishlashga berish mumkin emas, modomiki bu holda konteks qayta ishlashdan tanishga qayta aloqani amalga oshirish mumkin bo'lmaydi. Ikkinchidan, tadqiq qilinadigan ob'ekt imkon qadar yaxlit holda aks etishi va qayta ishlanishi kerak.

Idrok etishning *birinchi qadami* - bu idrok etiladigan ob'ekt haqidagi gipotezani shakllantirishdan iborat. Gipoteza ob'ektning aprior modeli, konteksti va oldingi gipotezalarning natijalarini tekshirish asosida ham(«yuqoridan-quyiga» jarayoni), ob'ektni oldindan analiz qilish asosida ham(«quyidan - yuqoriga») shakllanishi mumkin. Ikkinchi qadam - idrok etishni chuqurlashtirish(gipotezani tekshirish). Bu holda ob'ektni ilgari surilgan gipoteza dorasida qo'shimcha analizi amalga oshiriladi va to'liq kuchni kontekst jalb qiladi.

Idrok etish qulay bo'lishi uchun ob'ektni oldindan qayta ishlashni amalga oshirish zarur. Lekin bu holda ob'ekt haqidagi ma'lumot yo'qolmasligi kerak. Odatda ob'ektni boshlang'ich qayta ishlash kiruvchi ob'ektni keyingi ishlar uchun qulay bo'lgan tasavvurga almashtirishga olib kelinadi(masalan, tasvirni

vektorlashtirish) yoki kiruvchi ob'ektning barcha mumkin bo'lgan segmentlash variantlarini olishga olib kelinadi va ularning ichidan gipotezalarni ilgari surish va tekshirish orqali to'g'risi tanlanadi. Gipotezalarni o'rtaga tashlash va tekshirish jarayoni dastur arxitekturasida yaqqol aks etishi lozim. Har bir gipoteza uni baholash yoki boshqasi bilan taqqoslash mumkin bo'lishi uchun ob'ekt bo'lishi kerak. SHuning uchun odatda gipotezalar ketma - ket ravishda o'rtaga tashlanadi, shundan so'ng ro'yxatga birlashtiriladi va oldindan baholash orqali saralanadi. Gipotezani oxirgi tanlashda kontekst va boshqa qo'shimcha bilimlar manbai faol ishtirok etadi.

Hozirgi kunda genetik dasturlash sohasidagi peshqadamlardan biri Stenford universitetida professor Djon Koza rahbarligadi ishlaydigan tadqiqotchilar guruhi hisoblanadi. Genetik dasturlash Djon Makkarti guruhi tomonidan ro'yxatlarni qayta ishlash va funktsional dasturlash uchun mo'ljallangan, allaqachon unutilgan LISP(List Processing) tiliga yangi hayot bag'ishladi. Aynan shu til AQSHda sun'iy ong masalalari uchun keng tarqalgan dasturlash tillaridan bo'lgan va bo'lib qolmoqda.

### **3. Belgilarni tanish.**

Hozirgi kunda belgilarni tanishda uchta yondashuv ma'lum - *shablonli, strukturali va belgili*. Lekin yaxlitlik tamoyiliga faqat birinchi ikkitasi javob beradi.

*SHablonli tavsiflashni* amalga oshirish uchun oson, ammo, strukturaliga qaraganda u, shakllarning turli - tumanligiga ega bo'lgan murakkab ob'ektlarni tavsiflash imkonini bermaydi. Aynan shuning uchun shablonli tavsiflash faqat bosma belgilarni tanish uchun, ayni vaqtda strukturali tavsiflash ko'proq shakl variantlariga ega qo'lyozma belgilarni tanishda qo'laniladi.

*3.1. SHablonli tizimlar.* Bunday tizimlar aloxida belgining tasvirini rastrliga almashtiradi, uni bazada mavjud bo'lgan barcha shablonlar bilan taqqoslaydi va kiruvchi tasvirdan eng kam nuqtalar bilan farq qiluvchi shablonni tanlaydi. SHablonli tizimlar tasvir kamchiliklariga yetarlicha bardoshli va kirituvchi ma'lumotlarni qayta ishlashda yuqori tezlikka ega, ammo shabloni unga ma'lum bo'lgan shriftlarnigina yaxshi taniy oladi. Agar taniladigan shrift etalondan ozgina farq qilsa, shablonli tizimlar hatto yuqori sifatli tasvirlarni qayta ishlashda ham xato qilishi mumkin.

*3.2. Strukturali tizimlar.* Bunday tizimlarda ob'ekt graf ko'rinishida tavsiflanadi. Grafning tugunlarini kiruvchi ob'ektning elementlari, yoylarini esa ular o'rtasidagi fazoviy munosabatlar tashkil qiladi. Bunday yondashuvni amalga oshiradigan tizim, odatda vektorli tasvirlar bilan ishlaydi. Belgini tashkil etuvchi chiziqalar strukturali elementlar hisoblanadi. Masalan «r» harfi uchun bu vertikal kesma va yoy.

Strukturali tizimlarning kamchiligiga ularning tasvir kamchiliklariga sezuvchanligining yuqoriligini kiritish mumkin. Bundan tashqari bu tizimlar uchun shablonli va belgili tizimlardan farqli ravishda xozirgacha samarali avtomatlashtirilgan o'qitish protseduralari yaratilmagan. SHuning uchun Fine Reader uchun strukturali tavsiflarni qo'lda yaratishga to'g'ri keldi.

3.3. *Belgili tizimlar.* Bu tizimlarda har bir belgining o'rtacha tasviri n-o'lchovli belgilar fazosidagi ob'ekt sifatida aks ettiriladi. Bu yerda kiruvchi tasvirni tanishda qiymati hisoblanadigan belgilar alifbosi tanlanadi. Hosil qilingan n-o'lchovli vektor etalon bilan taqqoslanadi va tasvir ularning ichidan ko'proq mos keladiganiga tegishli bo'ladi. Belgili tizimlar yaxlitlik tamoyiliga javob bermaydi. Ob'ektlar sinfini tavsiflashni yaxlitligining zarur, ammo yetarli bo'lmagan sharti shundan iboratki, berilgan sinfdagi barcha ob'ektlar tavsifni qanoatlantirishi kerak. Modomiki, belgilarni hisoblashda axborotning ma'lum qismi yo'qolar ekan, faqat berilgan sinfga qarashli ob'ektlarni kiritishga kafolat berish qiyin.