

15-dars. PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

Python dasturlash tilida inson nutqini tanish va tushunish texnologiyasini yaratish boshqa dasturlash tillariga nisbatan oson va samarali kechadi. Bunga asosiy sabab – dasturlash tilida kodlarining soddaligi hamda ko‘plab qo‘shimcha kutubxonalarning mavjudligidir. Tabiiy tilni qayta ishlash (Natural language processing – inson nutqini tanish, tushunish, tahlil qilish va javob berish) sun‘iy intellektning bir qismi sifatida mashinalar va insonlar o‘rtasida ovozli yoki yozma muloqotni amalga oshirmoqda.

TAYANCH TUSHUNCHALAR

- **Tokenizatsiya** – matn ko‘rinishidagi ma‘lumotni qismlarga ajratib, ma‘lum bir belgilar ko‘rinishida ifodalash.
- **Nutqni qismlarga bo‘lish va belgilash** – ovozli ma‘lumotni qo‘llanilayotgan algoritim asosida qayta ishlash va tahlil qilish uchun ma‘lum bir qismlariga ajratish jarayoni.
- **Nutqni tanib olish** – foydalanuvchi ovozini tahlil qilish orqali uning shaxsini aniqlash jarayoni.
- **Foydalanuvchi his-tuyg‘ularini tahlil qilish** – dasturning foydalanuvchi his-tuyg‘ularini aniqlash orqali uning ovozli ma‘lumotlariga ma‘lum bir ma‘no berish jarayoni.
- **So‘z vektorlari** – ma‘lumotda qo‘llanilayotgan so‘zlarni ma‘no jihatidan vektorli guruhlariga bo‘lish orqali undan foydalanish jarayoni.
- **Wordnet integratsiyasi** – ingliz tilidagi so‘zlar birikmalaridan tashkil topgan ma‘lumotlar bazasi. U foydalanuvchiga so‘z birikmalarining o‘zgarmas ma‘nosini taqdim etadi.

Sun‘iy intellektning hisoblash lingvistikasi va kompyuter fanlari imkoniyatlarini birlashtirish natijasida mashina tabiiy tilni tushuna boshladi. Bu degani, foydalanuvchi o‘z kompyuterini og‘zaki muloqot orqali boshqarishi natijasida ish samaradorligini bir necha barobar oshirish imkoniyatiga ega bo‘ldi. Ma‘lumki, so‘zlashuv insonlar o‘rtasida asosiy muloqot manbayi hisoblanadi. Tabiiy tilni qayta ishlash (TTQI) orqali so‘zlashuv jarayonini inson va mashina o‘rtasidagi asosiy muloqot manbayiga aylantirish, kelgusida bu orqali sun‘iy intellekt texnologiyalarining taraqqiyotini yangi bosqichga olib chiqish mumkin.

PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

Dastlabki TTQI dasturi 1950-yillarda yaratilgan bo'lib, dastur 60 ga yaqin so'zlardan iborat jummalarni rus tilidan ingliz tiliga avtomatik tarzda tarjima qilish imkoniyatiga ega bo'lgan. Shundan so'ng ko'plab nazariyalar yaratildi va sun'iy intellekt bilan birgalikda TTQI texnologiyasi rivojlanib bordi. Ammo TTQI dasturlarini yaratish talay qiyinchiliklarni keltirib chiqardi. Masalan, bir so'zning turli holatlarda turlicha talqin qilinishi (ot–jonivor, ot–fe'l) yoki gap ma'nosining vaziyatga qarab o'zgarishi (gulxanning yonishi, lampaning yonishi) yaratilayotgan dastur natijalarida xatoliklar kelib chiqishiga sababchi bo'lgan. Bugungi kunda TTQI texnologiyasi Google tarjimon, Yandex tarjimon, Amazon Assistant, Internet qidiruv tizimlari (Google, Yandex), humanoid robotlari, harbiy texnika va boshqalarda samarali qo'llanilmoqda.

TTQIga ikki xil yondashuv mavjud:

1) **qoidalarga asoslangan tabiiy tilni qayta ishlash.** Bu turdagi dastur ma'lumotlarni qayta ishlashda aql-idrokdan foydalanadi, ya'ni gaplarning mantiqiy tuzilishini inson kabi tahlil qilgan holda yaratadi. Masalan, bunday yondashuv asosida ishlovchi dastur muzlashi inson hayotiga xavf solishini inobatga olgan holda gap mazmunini shakllantirishi mumkin. Qoidalarga asoslangan yondashuv bo'yicha ishlovchi dasturlar sekin ishlaydi va ko'p vaqt talab etadi. Shu sababli bu yondashuv dasturchilar orasida ommalashmagan;

2) **statistik tabiiy tilni qayta ishlash.** Bunday yondashuv asosida katta hajmdagi ma'lumotlar mashinani o'qitish texnologiyasiga asoslangan algoritmlar yordamida qayta ishlanadi. Bu esa ma'lumotlarning tezkor qayta ishlanishiga hamda foydalanuvchi bilan to'xtovsiz aloqani ta'minlashga olib keladi. Bunda dastur sifati dasturchi tomonidan qo'llanilayotgan algoritmlar imkoniyatiga chambarchas bog'liqdir. Ya'ni dasturchining optimal algoritmlardan foydalanishi sifatli TTQI dasturining yaratilishiga olib keladi.

Tabiiy tilni qayta ishlash komponentlari (tarkibiy qismlari)

Leksik tahlil. Leksik tahlil matnning butun qismini paragraf, jumla va so'zlarga ajratadi. O'rganilayotgan matnning kichik qismlarga ajratilishi uning tuzilishini aniqlash va tahlil qilish imkoniyatini beradi. Bu orqali dastur so'zlarni shakl va ma'no jihatidan o'rganadi. Tahlil qilingan matn qayta ishlanishi natijasida inson tushunishi mumkin bo'lgan holda foydalanuvchiga taqdim etiladi. Masalan, TTQI dasturi o'rnatilgan robotga biror buyruq berilsa, u buyruqni matn holatiga o'tkazadi, leksik jihatdan tahlil qiladi va unga ma'lum bir javobni beradi. Dastlabki tahlilning to'g'riligi esa robot tomonidan qaytariladigan javobning tushunarli bo'lishini ta'minlaydi.

Sintaktik tahlil. Tahlilning bu turi berilgan matndagi so'z va gaplarni yasash jihatidan tahlil qiladi. So'zlarda so'z birikmalarining to'g'ri qo'llanilishi va gap yasashidagi grammatik

PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

qoidalar sintaktik tahlil asosida o'rganiladi. Grammatik jihatdan noto'g'ri gaplarni qabul qilmaydi. Masalan, "Daraxt kemaga bordi" jumlasini sintaktik jihatdan xato hisoblanadi. Shuning uchun tuzilayotgan TTQI dasturida sintaktik tahlil muhim ahamiyat kasb etadi.

Semantik tahlil. Semantik tahlil so'zlar ma'nosiga asoslangan holda jumla yoki gaplar mazmundorligini tahlil qiladi. Semantik tahlil sintaktik tahlil bilan birgalikda matnlarning to'g'ri qo'llanilishiga yordam beradi. Masalan, "qaynoq muz bo'lagi" kabi jumlar tahlildan o'tmaydi, chunki muz bo'lagi qaynoq bo'lishi mumkin emas.

TTQI texnologiyasi inson kabi mukammal so'zlashuv imkoniyatini taqdim etishi uchun unda yuqoridagi tarkibiy qismlar to'laonli aks etishi lozim. Bugungi kungacha yaratilgan dasturlarda ayrim komponentlar qo'llanilmagan. Shu sababli TTQI texnologiyasiga ega bo'lgan dastur hozircha kutilgan natijani bermayapti.

Python dasturlash tilida TTQI texnologiyasidan foydalanishga yordam beruvchi bir qancha kutubxonalar mavjud. Bularga NLTK, spaCy, Gensim, Pattern, TextBlob kabilarni misol keltirishimiz mumkin.

NLTK (Natural Language Toolkit) Python dasturlash tili TTQI texnologiyasi uchun maxsus yaratilgan modul bo'lib, jarayonlarni modellashtirish uchun bir qancha qo'shimcha imkoniyatlarga ega. Xususan, tokenizatsiya, nutqni qismlarga bo'lish va belgilash, nutqni tanib olish, foydalanuvchi his-tuyg'ularini tahlil qilish va chat-bot to'plamlaridan iborat. Shuningdek, NLTK kutubxonasi nutqni aniqlash bilan bog'liq loyihalarda faol qo'llanilmoqda.

spaCy TTQI uchun ommabop Python kutubxonasi bo'lib, ochiq manbali vosita hisoblanadi. spaCy ko'proq ishlab chiqarish jarayonida foydalanishga mo'ljallangan. Kutubxona quyidagi imkoniyatlarni taqdim etadi: tokenizatsiya, nutqni tanib olish, uni qismlarga bo'lish va belgilash, foydalanuvchi his-tuyg'ularini tahlillash, qarama-qarshilikni tahlil qilish va so'z vektorlari bilan ishlash. Bundan tashqari, mazkur kutubxona kiritilgan ma'lumotlarni avtomatik ravishda to'ldirish va tuzatish, sharhlarni tahlil qilish va tahlil qilinayotgan ma'lumotni umumiy xulosalash kabi qo'shimcha imkoniyatlarni ham taklif etadi. Bunday imkoniyatlar dasturchilar TTQI dasturlarini yaratishlarida uchrashi mumkin bo'lgan qiyinchiliklarning kamayishiga olib kelmoqda.

Gensim kutubxonasi Pythonda berilgan ma'lumotni modellashtirish hamda o'xshashlikni aniqlash vazifasini bajaradi. Ya'ni u kiritilgan ma'lumotlarni kuchli tahlillash imkoniyatini taqdim etadi. Gensim umumiy maqsadli TTQI kutubxonasi bo'lmasa-da, keng qamrovli tahlillash imkoniyati orqali yaratilayotgan dasturlarning sifatiga ijobiy ta'sir ko'rsatishi bilan ahamiyatlidir.

Pattern TTQI jarayonida ma'lumotlarni avtomatik ravishda sintaktik tahlil qilish imkoniyatini taqdim etishi bilan boshqa shu kabi vazifalarni bajaruvchi Python kutubxonalaridan ajralib

PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

turadi. Mazkur Python kutubxonasi asosiy e'tiborni ilmiy ma'lumotlarni qayta ishlash jarayoniga qaratgan. Kutubxona foydalanuvchilarga spaCy kabi ko'plab qo'shimcha imkoniyatlarni taqdim etgan.

TextBlob Pythonning matnli ma'lumotlarni qayta ishlashga mo'ljallangan kutubxonasi hisoblanadi. Kutubxona nutqni qismlarga bo'lish va belgilash, nomlarni tahlil qilish, foydalanuvchi his-tuyg'ularini tahlillash, tarjima qilish va tilni aniqlash, wordnet integratsiyasi, imlolarni tuzatish kabi imkoniyatlarga ega.

Sun'iy intellekt texnologiyasi tabiiy tilni qayta ishlash imkoniyati bilan birgalikda yangi kasblarning rivojlanishiga yordam beradi. Kelajakda insonlarning kasbiy faoliyati bevosita qaysidir robotlashgan texnologiyaga asoslanadi. Robotlar esa insonlar bilan qaysidir tilda muloqot qilishi lozim. Ishonch bilan aytish mumkinki, kelgusida TTQI texnologiyasi insonlar faoliyatida eng ommalashgan hamda barcha bilishi lozim bo'lgan bilim sohasi sifatida rivojlanadi.

WordNet – 200 dan ziyod tilda so'zlar o'rtasidagi semantik va leksik munosabatlarni to'g'ri shakllantirgan ma'lumotlar bazasi. WordNet so'zlarni semantik munosabatlarga, jumladan, sinonim, giponim va meronimlarga bog'laydi. Sinonimlar qisqa ta'riflar va ulardan foydalanish misollari bilan sinsetlarga birlashtirilgan.

TAKRORLASH UCHUN SAVOLLAR

1. TTQI texnologiyasi qaysi sohalarning rivojlanishiga sabab bo'lishi mumkin?
2. TTQI texnologiyasi qaysi dasturlash tili bilan birga rivojlanmoqda?
3. TTQI jarayoni nima uchun ma'lum bir yondashuvlar asosida o'rganiladi?
4. Pythonda TTQI texnologiyasi uchun qanday kutubxonalar mavjud?
5. TTQI texnologiyasi qanday tarkibiy qismlar (komponentlar)dan iborat bo'lishi lozim?
6. Kompyuter o'zbek tilida gapirishi va uni o'quvchilarning tushunishi dars jarayoniga qanday ta'sir qilishi mumkin?

AMALIY FAOLIYAT

Sun'iy intellekt asosida chet tillarini o'rganishga yordam beruvchi Duolingo dasturidan

PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

foydalanish jarayoni bilan tanishish.

Duolingo sun'iy intellektga asoslangan bo'lib, unda TTQI texnologiyasi qo'llanilgan. Dastur ingliz tili kabi bir qancha chet tillarini tez va samarali o'rgatish maqsadida 2011-yilda AQSHda yaratilgan. Bugungi kunda foydalanuvchilar dastur orqali 37 tadan ziyod chet tilini o'rganishi mumkin. Ma'lumotlarga ko'ra, turli tillarni o'rganish maqsadida dasturdan 300 mln dan ortiq foydalanuvchi ro'yxatdan o'tgan. Dasturda chet tilini o'rganish jarayoni qiziqarli o'yin tariqasida tashkil etilgan bo'lib, chat-bot uning asosiy muloqot oynasi hisoblanadi. Unda o'rganuvchilar uchun qiziqarli ma'lumot beriladi, muloqot oynasida ularning bilimlari tekshiriladi va mustahkamlanadi.

Duolingo dasturidan 3 xil usulda foydalanish mumkin:

- 1) kompyuter yoki mobil telefondagi Internet brauzeri yordamida duolingo.com saytiga kirish orqali;
 - 2) Android tizimi uchun ishlab chiqarilgan *.app formatdagi dualingo nomli mobil dasturni smartfonga o'rnatish orqali (yuklab olish Google Play Store yordamida);
 - 3) IOS tizimi uchun ishlab chiqarilgan *.ipa fayl formatidagi mobil dasturni smartfonga o'rnatish orqali (yuklab olish manzili App Store).
- Android tizimidagi mobil telefonlar yordamida dasturdan foydalanish jarayoni quyidagicha amalga oshiriladi:

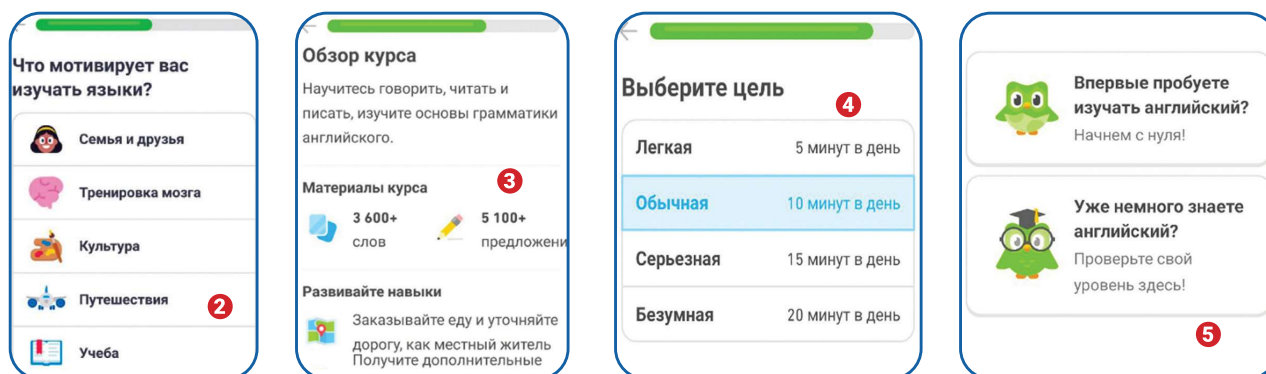
1-qadam. Google Play Store dan Duolingo dasturi yuklanib, mobil telefonga o'rnatiladi;

2-qadam. Dastur ishga tushiriladi. So'ngra dasturdan kerakli til tanlanadi. Til tanlanganidan keyin savollarga quyidagi ketma-ketlikda javob beriladi:

- Qaysi tilni o'rganishni xohlaysiz? **(1)**
- Dastur haqida qayerdan xabar topdingiz?
- Nima uchun til o'rganmoqchisiz yoki qanday motiv sababli bu tilni o'rganyapsiz? **(2)**
- Kurs haqida ma'lumot beriladi **(3).**
- Bir kunda necha daqiqa vaqtingizni chet tilini o'rganishga ajrata olasiz?**(4)**
- Boshlang'ich darajadan (0 dan) o'rganasizmi yoki o'rta darajadan boshlab o'rganasizmi? **(5)**



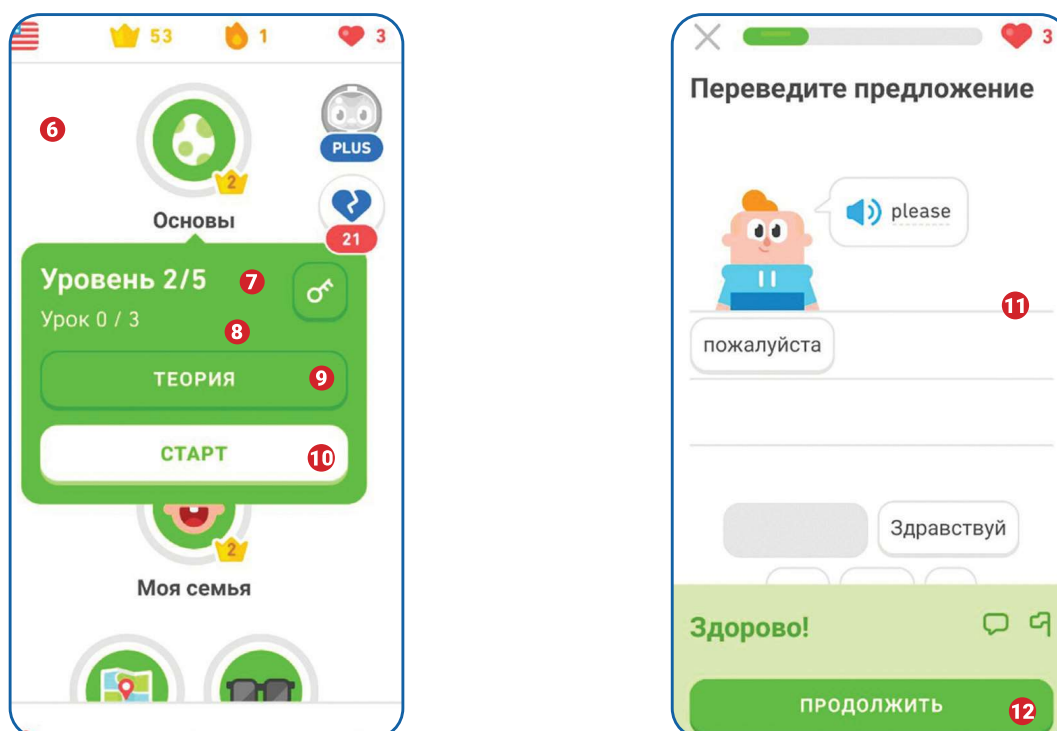
PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI



3-qadam. Dasturning asosiy oynasida tanlangan chet tilini o'rganish jarayoni boshlanadi. Turli mavzularda berilgan ma'lumotlar doira shaklidagi tugmachalarga birlashtirilgan (6).

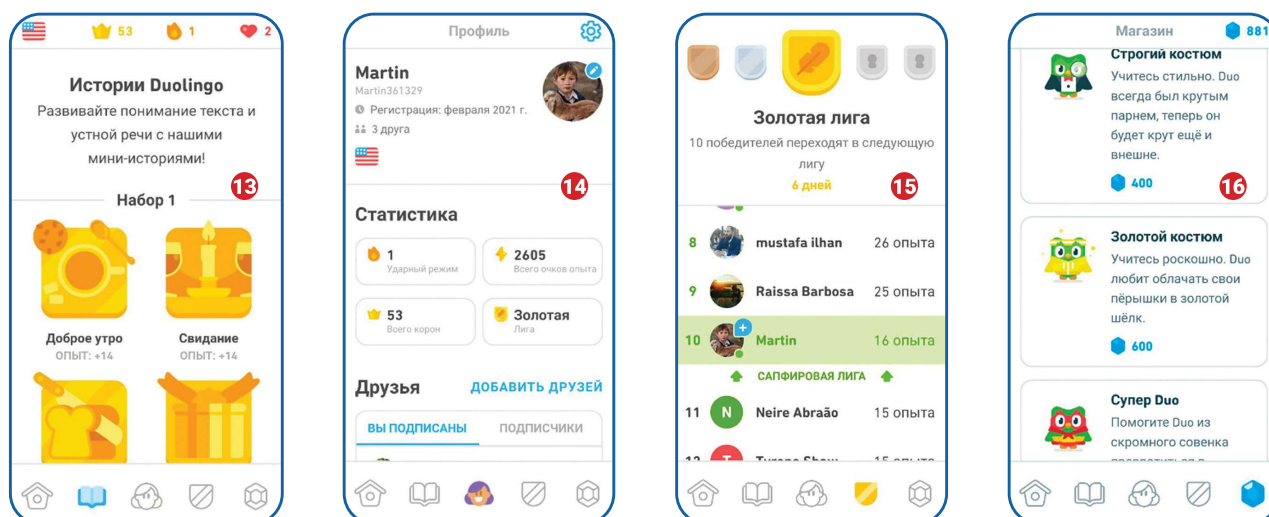
4-qadam. Har bir doirada 5 ta daraja mavjud (7). Bitta daraja uchun esa 3 ta dars birlashtirilgan (8). Har bir dars o'yin shaklidagi qiziqarli savol-javoblardan iborat (10). Shuningdek, savol-javob uchun nazariy ma'lumotlar ham berilgan (9).

5-qadam. СТАРТ (10) tugmachasini bosish orqali chat-bot oynada qiziqarli savol-javob jarayonini boshlash mumkin. Savol-javob jarayoni (11) so'zlarni yodlash, eshitish orqali tushunish, to'g'ri talaffuz qilish, gaplarni tarjima qilish jarayonlarini o'rgatadi. Dasturda savollarning turiga qarab yozma yoki og'zaki javob berish mumkin. Javob berilganidan keyin "Проверить" va "Продолжить" tugmachalari (12) bosiladi. Shundan so'ng keyingi savolga o'tish mumkin.



PYTHONDA TABIIY TILNI QAYTA ISHLASH TEXNOLOGIYASI

6-qadam. Duolingo dasturida 5 ta asosiy oyna mavjud. Bular: asosiy o'rganish oynasi **(13)**, qiziqarli hikoyalar oynasi **(14)**, o'rganuvchi profil oynasi **(15)**, o'yin reytingi oynasi **(16)** va o'yin uchun qo'shimcha jihozlar oynasi (do'kon) **(17)**. Qiziqarli hikoyalar oynasida sodda, kundalik turmushda ko'p duch kelinadigan holatlar bo'yicha chet tilidagi dialoglar mavjud. Ular o'rganuvchining so'z boyligini oshirishga hamda eshitib tushunish ko'nikmasining rivojlanishiga yordam beradi. Profil oynasida esa qo'lga kiritilgan natijalar haqida ma'lumot bilan tanishish mumkin. "Reyting" yoki "liga" oynasida sizning, o'rganuvchi darajasi bilan teng bo'lgan boshqa foydalanuvchilarning reyting jadvali mavjud bo'lib, ko'proq o'rganish orqali reytingda birinchi o'ringa chiqish mumkin. Do'kon oynasidan esa to'plangan ballga turli qo'shimchalar olish mumkin.



UYGA VAZIFA

1. Mobil smartfonga Duolingo dasturini o'rnatish.
2. O'rganmoqchi bo'lgan chet tilini tanlang va qiziqarli o'yinni boshlang.
3. Dasturning talaffuzingizni TTQI texnologiyasi asosida tekshirishiga e'tibor bering.
4. Ingliz yoki rus tilini tanlasangiz, shu fanlar darslarida o'qituvchingiz bilan dasturni muhokama qiling.