

TABIY TILNI QAYTA ISHLASHNING HOZIRGI HOLATI, RIJOJLANISHTENDENSIYASI VA TEXNOLOGIYALARI

Oliy ta'lim tizimida va insonlarni ilm olish jarayonida til asosiy vosita hisoblanadi. Tabiiy tilni qayta ishlash yoki NLP - bu kompyuterlar va insonlarning o'zaro aloqasi uchun juda muhim ahamiyatga ega bo'lgan sun'iy aqlning bir bo'lagi. Minglab va minglab yillar mobaynida ishlab chiqilgan inson tili, ko'pincha so'zlarni ko'pincha ifodalaydigan juda ko'p ma'lumotlarga ega bo'lgan noyob aloqa shakliga aylandi. NLP insoniy aloqa va raqamli ma'lumotlar o'rtasidagi tafovutni bartaraf etishda muhim texnologiyalarga aylanadi. Tabiiy tilni qayta ishlash (NLP – Natural language Processing) sohasi sun'iy intellekt rivojlanishi bilan muvofaqiyat qozonmoqda (1-rasm). Inson tilini hisoblash turli sohalarda o'z ifodasini topgan. Mashina tarjimasi, elektron pochta spamlarini aniqlash, ma'lumotlarni olish, umumlashtirish, tibbiy, savol-javob berish va boshqalar.

Tabiiy tilni qayta ishlash foydalanuvchi ishini yengillashtirish va uni yechish uchunpaydo bo'lgan kompyuter bilan tabiiy tilda muloqot qilish jarayonidir. Chunki barcha foydalanuvchilarda bunday bo'lmashligi mumkin, mashinaning o'ziga xos tilini yaxshi biladigan NLP, yetarli bo'lmagan foydalanuvchilarga yordam beradi.



1-rasm. Tabiiy tilni qayta ishlash jarayoni

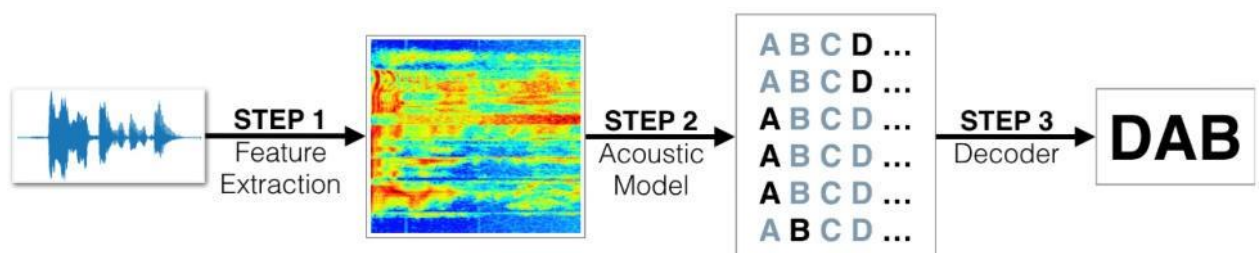
Sintaktik nazariyalarni yaratgan tilshunoslardan biri Noa Xomskiy hisoblanadi. Nazariy tilshunoslik sohasida o'ziga xos mavqeyga ega bo'lgan [1]. Umuman olganda, ikkita yuqori darajaga bo'linishi mumkin. NLPning o'rganilgan vazifalaridan biri avtomatik umumlashtirish, diskurs tahlili, mashina tarjimasi, morfologik segmentatsiya, nomlangan obyektни tanib olish, belgilarni optik aniqlash, nutqni teglash va boshqalar. Qo'shma havola rezolyutsiyasi u jumla yoki undan kattaroq matn to'plamiga ishora qiladi, qaysi so'z xil obyektga tegishli ekanligini

aniqlaydi. Optik belgilarni aniqlash (OCR) bosilgan matnni ifodalovchi rasm, tegishli yoki

bog'liqligini aniqlashga yordam beradi. Nutqning bir qismini teglash u gapni tavsiflaydi, har biri uchun nutq qismini belgilaydi. NLP vazifalari aniq yaqindan o'zaro bog'langan, lekin ular avtomatik umumlashtirish, qo'shma ma'lumotnoma kabi vazifalarning ba'zilarini tahlil qilish va boshqa kichik vazifani bajarish sifatida ishlaydi.

NLP bir nechta mutaxassisliklarga moslashish algoritmi yoki tizimi. Algoritmik tizimda NLP baholash ko'rsatkichi quyidagilarga imkon beradi, integratsiyaning til tushunishi va tillarning avlodini qarash. Tizim eng ko'p tilli tabiiy modulli to'plamini o'z ichiga oladi. Til liniyasi asosiy NLP ishlov berish uchun modullarni birlashtiradi, shuningdek, tillararo nomli obyektlarni bog'lash, semantik kabi yanada rivojlangan vazifalarni normallashtirish vazifalarini bajaradi. Shunday qilib, tillararo ramka hodisalar, ishtirokchilar, joy va vaqtni talqin qilish, shuningdek, o'rtasidagi munosabatlarga imkon beradi [2]. Ushbu alohida tillarning chiqishi tizim uchun kirish sifatida foydalanish uchun mo'ljallangan hodisalarga asoslangan bilim grafiklarini oladi. Ularning barchasi standart kiritish, izohlar qilish va o'z navbatida kirish bo'lgan standart chiqishni ishlab chiqarish vazifasini bajaradi. Bundan tashqari, modulli arxitektura turli xil konfiguratsiyalarga imkon beradi va dinamik ishlaydi.

Tabiiy tilni qayta ishlash bo'yicha ishlarning aksariyati kompyuter olimlari tomonidan amalga oshiriladi, tilshunoslik, psixolog va boshqa turli mutaxassislar ham qiziqish bildirgan. Noaniqlik tabiiy tilning asosiy muammolaridan biri odatda leksik va morfologiya kabi kichik vazifaga ega bo'lgan sintaktik darajada so'z yasalişini o'rganish bilan bog'liq. Ushbu darajalarning har biri ishlab chiqarishi mumkin. Noaniqlikni minimallashtirish, noaniqlikni saqlash kabi turli usullar bilan hal qilinishi mumkin (2-rasm).



2.rasm. Tabiiy tilni qayta ishlash jarayoni

Dunyo axborotlari onlayn bo'lgani uchun, bu ma'lumotlarning mavjudligini ta'minlash vazifasi tobora muhim ahamiyat kasb etadi. Dunyo ma'lumotlarini har bir kishiga, til to'siqlari orqali kirishga majbur qilish, inson tarjimasini o'tkazish qobiliyatini kengaytirdi. Duolingo kabi innovatsion kompaniyalar yangi tilni o'rganish bilan tarjima ishlarini bir-biriga mos keltirish orqali katta miqdordagi odamlarni jalb qilishga intilmoqda. Biroq, mashinaning tarjimasini dunyo axborotlarini uyg'unlashtirish uchun yanada kengroq alternativ taklif qiladi. Google Google tarjima xizmati uchun mulkiy statistik dvigatelni qo'llagan holda, mashinani tarjima qilish bo'yicha birinchi o'rinni egallagan kompaniya. Mashinani tarjima qilish texnologiyalari bilan kurashish so'zlarni tarjima qilishda emas, balki jumlar ma'nosini saqlab qolishda NLP qalbida joylashgan murakkab texnologik muammo.

Spam filtrlar kiruvchi elektron pochta xabarlarini muammosiga qarshi birinchi mudofaa chizig'i sifatida muhim ahamiyat kasb etdi. Ammo e-pochtani keng ishlatadigan deyarli har bir

kishi, hali qabul qilinmagan kiruvchi elektron pochta manzillari yoki filtrda tasodifan tutilgan muhim elektron xabarlar ustidan qayg'uga duch keldi. Spam-filtrlarning noto'g'ri-ijobiy va yolg'on-salbiy masalalari NLP texnologiyasining qalbida bo'lib, yana matnli matnlardan ma'nolarni olish muammosiga aylanadi [3]. Bu elektron usulda spam va spam bo'lmagan e- pochta xatlaridagi odatiy holatlarga nisbatan elektron pochta yozilgan so'zlar soni bo'yicha statistik usul hisoblanadi.

Savollarga avtomatik javob berish qidiruv mexanizmlari dunyodagi boy ma'lumotni barmoq uchida joylashtirmoqda, ammo odamlar tomonidan yaratilgan muayyan savollarga aslida javob berishda odatda juda oddiy. Google bu foydalanuvchilarning sabablarini keltirib chiqargan umidsizlikni ko'rdi, ular tez-tez izlayotgan javobni topish uchun turli xil qidiruv natijalarini sinab ko'rishlari kerak. NLPdagi Google harakatlarining asosiy maqsadi tabiiy tillardagi savollarni tanib olish, ma'nosini olish va javobni berishdir va Google natijalari sahifasining evolyutsiyasi bu yo'nalishni ko'rsatdi. Mutlaqo takomillashib borayotgan bo'lsa-da, bu qidiruv mexanizmlari uchun muhim muammo bo'lib qolmoqda va tabiiy tillarni qayta ishlash tadqiqotlarining asosiy dasturlaridan biri hisoblanadi.

Muhim axborotlarni olish moliya bozorlarida ko‘plab muhim qarorlar inson nazoratini va nazoratidan tobora ortib bormoqda. Algoritmik savdo tobora ommalashib bormoqda, bu texnologiya butunlay nazorat qilinadigan moliyaviy sarmoyalar shaklidir. Biroq, ushbu moliyaviy qarorlarning aksariyat qismi yangiliklarga, ingliz tilida esa asosan taqdim etilayotgan jurnalistikaga ta’sir qiladi. Keyinchalik, NLP ning asosiy vazifasi bu oddiy matnli xabarlarini qabul qilish va tegishli ma’lumotni algoritmik savdo qarorlariga aylantirilishi mumkin bo‘lgan formatda olish bo‘ldi. Misol uchun, kompaniyalar o‘rtasida birlashish haqidagi xabarlar savdo qarorlariga katta ta’sir ko‘rsatishi mumkin, va birlashish, o‘yinchilarning narxlarini, kimni sotib olganini savdo algoritmgiga kiritish mumkin bo‘lgan tezligi, savdo algoritmgiga ta’sir qilishi mumkin. million dollar.

Xulosa o‘rnida shuni aytish kerakki, axborotni haddan tashqari yuklab olish bizning raqamli davrda haqiqiy hodisa bo‘lib, bizning bilim va ma’lumotlarga bo‘lgan kirishimiz buni anglash qobiliyatimizdan ancha ustundir. Bu sekinlashuv belgisi ko‘rsatilmagan tendentsiyadir, shuning uchun hujjatlar va axborotning mazmunini umumlashtira olish qobiliyati tobora muhim ahamiyatga ega. Bu bizga katta hajmdagi ma’lumotlardan tegishli ma’lumotni tanib olish va ularni o‘zlashtirish imkoniyatini berishda muhim ahamiyatga ega. Boshqa istalgan natija, chuqurroq hissiy ma’nolarni tushunishdir