

KOMPYUTER LINGVISTIKASI IMKONIYATLARI VA DASTURIY TEKNOLOGIYALAR

Djurayev Murotali,
Termiz davlat universiteti katta o'qituvchisi
<https://doi.org/10.53885/edinres.2024.02.1.007>

Annotatsiya: Maqolada kompyuter lingvistikasi oldida turgan masalalar axborotni yig'ish, to'plash, qayta ishlash va olish jarayonlarini lingvistik ta'minlash vazifalari o'rganilib chiqilgan. Bu vazifalar, nutqni aniqlash va sintez qilish, matnlarni avtomatik tarjima qilish, elektron matn korpusini yaratish va ulardan foydalanish, matndan ma'lumot imchya olish, o'z-o'ziga havola qilish usullari hisoblanadi.

Kalit so'zlar: Matematik modell, texnologiya, tabiiy tillar, Natural Language Processing, NLP, Kompyuter lingvistikasi, tabiiy tilni yaratish (NLG), tabiiy tilni tushunish (NLU).

ВОЗМОЖНОСТИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ И ПРОГРАММНЫЕ ТЕХНОЛОГИИ

Джураева Муратали,
старший преподаватель Термизского Государственного университета

Аннотация: В статье рассматриваются вопросы, стоящие перед компьютерной лингвистикой, задачи лингвистического обеспечения процессов сбора, сбора, обработки и получения информации. К этим задачам относятся распознавание и синтез речи, автоматический перевод текстов, создание и использование электронных текстовых корпусов, извлечение информации из текста, методы самореференции.

Ключевые слова: математическая модель, технология, естественные языки, обработка естественного языка, NLP, компьютерная лингвистика, генерация естественного языка (NLG), понимание естественного языка (NLU). математическая модель, технология, естественные языки, обработка естественного языка, NLP, компьютерная лингвистика, генерация естественного языка (NLG), понимание естественного языка (NLU).

COMPUTER LINGUISTICS CAPABILITIES AND SOFTWARE TECHNOLOGIES

Djurayev Muratali,
senior teacher of Termiz State University

Abstract: The article discusses the issues facing computer linguistics, the tasks of linguistic support for the processes of collecting, collecting, processing and obtaining information. These tasks include speech recognition and synthesis, automatic translation of texts, creation and use of electronic text corpora, extraction of information from text, and self-reference methods.

Keywords: mathematical model, technology, natural languages, natural language processing, NLP, computational linguistics, natural language generation (NLG), natural language understanding (NLU). mathematical model, technology, natural languages, natural language processing, NLP, computational linguistics, natural language generation (NLG), natural language understanding (NLU).

Kirish. Ko'pchilligimiz har kuni qidiruv tizimlari, ovozli yordamchilar va matn tarjimonlaridan foydalanamiz. Ushbu texnologiyalarning paydo bo'lishi matematik modellar yordamida tabiiy tillarni tavsiflash bilan shug'ullanadigan sun'iy intellekt sohasi hisoblangan lingvistika tufayli paydo bo'lgan. Maqolaning yana bir jihati kompyuter lingvistikasi va tabiiy tilni qayta ishlash texnologiyalari asosida sog'ligida nuqsoni bor shaxslarning imkoniyatlarini kengaytirishga yordam berish yo'llari ham ko'rsatib o'tilgan.

Ajdodlarimiz rivojlanayotgan jamiyat ehtiyojlariga xizmat qilish uchun ming yillar avval tabiiy tilni ixtiro qilgan. Zamonaviy tabiiy tillar o'z qonunlariga ko'ra rivojlanadi, har bir davrda ular insoniy muloqot, his-tuyg'ular, fikrlar va harakatlarni ifodalashning munosib vositasi bo'lib xizmat qilgan.

Tabiiy tilning tuzilishi va qo'llanilishi suhbat ishtirokchilarining tajribalari va bilimlari, shuningdek, dunyoqarashi va xatti-harakatlari o'xshash bo'lgan degan taxminga asoslanadi. Mutlaqo boshqa tabiatdagi mavjudot - kompyuter bilan axborot almashish uchun cheksiz tildan foydalanishi katta muammo edi.

So'nggi 200 yil ichida insoniyat kundalik hayotda jamiyatga xizmat qiladigan mexanik va elektron qurilmalari yordamida ko'plab vazifalarni muvaffaqiyatli avtomatlashtirmoqda. 20-asrning ikkinchi yarmida odamlarning e'tibori tabiiy tillarni qayta ishlashni avtomatlashtirishga qaratildi (Natural Language Processing, NLP). O'shandan beri qurilmalar nafaqat mexanik funksiyalarni, balki intellektual funksiyalarni ham bajarmoqda.

Shulardan kelib chiqqan holda kompyuterlarga yangi vazifa yuklandi: matnni o'qish, uning to'g'riligini tekshirish, belgilangan ko'rsatmalarga rioya qilish yoki ushbu matnning ma'nosiga asoslanib oqilona javob berish bunda odamlar faqat yakuniy qarorlarni o'zlari uchun saqlab qolishgan.

Kompyuter lingvistikasi tabiiy tilni qayta ishlash bilan bog'liq muammolarni hal qiladi. Kompyuter lingvistikasi - bu tabiiy tilni bilish darajasini kompyuterda modellashtirish, matn va nutqni avtomatik qayta ishlashning amaliy muammolarini hal qilish bilan shug'ullanadigan bilim sohasi hisoblanadi.

Mavzuga oid adabiyotlarning tahlili. Kompyuter lingvistikasi tarixi 1950-yillarda mashhur amerikalik tilshunos, publitsist va faylasuf Noam Xomskiyning tabiiy til strukturasi rasmiylashtirishga qaratilgan tadqiqotlari, shuningdek, kompyuter tarjimai bo'yicha sinov tajribalari va tabiiy tilni tushunish uchun birinchi avtomatlashgan axborot dasturiy tizimini ishlab chiqishidir.

Aytish mumkinki, kompyuter lingvistikasi 1954-yil yanvar oyida Jorjtaun universitetida (AQSh) kompyuter tarjimai bo'yicha dunyodagi birinchi ommaviy eksperiment o'tkazilgan paytda paydo bo'lgan. Muhandislar 60 dan ortiq jummalarni rus tilidan ingliz tiliga to'liq avtomatik rejimda tarjima qilishga muvaffaq bo'lishadi.

1980-yillarning oxirida Internetning rivojlanishi bilan elektron shaklda mavjud bo'lgan matnlar hajmi keskin oshdi, bu esa axborotni qidirish texnologiyalarida kvant sakrashiga olib keldi. Tabiiy tildagi matnni qayta ishlash uchun mutlaqo yangi vazifalar paydo bo'ldi. Shu bilan birga, birinchi kompyuter o'rganish algoritmlari va statistik kompyuter tarjimai tizimlari yaratildi.

Tilni qayta ishlash sohasidagi yutuq 2010-yillarda, chuqur o'rganish algoritmlari rivojlana boshlaganda yuz berdi. O'shandan beri hisoblash lingvistikasidagi muammolarni hal qilish uchun ko'plab ishlanmalar paydo bo'ldi va paydo bo'lishda davom etmoqda.[1]

Bugungi kunda kompyuter lingvistikasi tabiiy tilni qayta ishlash, matn va nutqni aniqlash vositalari, tarjima tizimlari, matn muharrirlari, tilni o'rgatuvchi materiallar, ovozli yordamchilar, aqlli chatbotlar va boshqa ko'plab dasturlarni o'z tarkibiga qamrab oladi.

Kompyuter lingvistikasi katta hajmdagi matndan kerakli ma'lumotlarni olish vositasi sifatida ilm fanga kirib keldi.

Tadqiqot metodologiyasi. Matnni aqlli avtomatik qayta ishlash zarurati asosan ikkita sababga ko'ra yuzaga keladi, ularning ikkalasi ham ishlab chiqarilgan matnlar soni bilan bog'liq.

Butun dunyo bo'ylab matnlar bilan ishlaydigan millionlab odamlar hujjatlarni qayta ishlashning zamonaviy standartlariga javob berish uchun yetarli bilim va ma'lumotga ega emaslar. Masalan, ofisdagi kotiba hujjatlar bilan ishlashda har safar yuzlab turli qoidalarni hisobga olmaydi ya'ni xatni o'z tilidan boshqa tilli kompaniyalarga yuborishda ko'plab xatoliklar va matnning mazmunini o'zgartirib yuboradi. Shuning uchun, kompyuter foydalanuvchilarining har bir yangi avlodini o'zlari bajarishga qayta-qayta o'rgatishdan ko'ra, kompyuterni bu ishni bir marta bajarishga

o'rgatish arzonroqdir.[2]

Tahlil va natijalar. Ko'p hollarda, ongli qaror qabul qilish yoki ma'lumotni topish uchun foydalanuvchi katta hajmdagi matnni o'qib chiqishi, tushunishi va uni hisobga olishi zarur bo'ladi. Kompyuter lingvistikasining qayta ishlash ob'ekti tabiiy tildagi matnlar bilan bo'lganligi sababli, uning rivojlanishini umumiy tilshunoslik sohasidagi asosiy bilimlarsiz tasavvur qilib bo'lmaydi.

Tilshunoslik insonning tabiiy tili, uning tuzilishi va faoliyati haqidagi fandır.

Tilshunoslik bir nechta bo'limlarni o'z ichiga oladi:

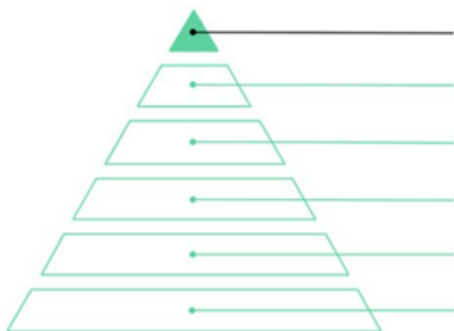
Fonologiya va grafika til belgilarining eshitish yoki vizual idrok etuvchi tomonini, semantikasini esa, aksincha, ularning semantik, tushunarli va tarjima qilinadigan qismini o'rganadi.

Morfologiya nutqiy so'zlarning ichki tuzilishi va tashqi shakli, jumladan, gap bo'laklari va ularning kategoriyalari bilan shug'ullanadi.

Sintaksis gaplar tuzilishini, gapdagi so'zlarning o'zaro kelishik va tartibini hamda til birligi sifatidagi umumiy xususiyatlarini o'rganadi.

Pragmatika belgilarning ularni ishlab chiqaruvchi va sharhlovchi sub'ektlarga munosabatini o'rganadi.

Bu bo'limlar berilgan tabiiy til matnidan grammatik va semantik ma'lumotlarni ajratib olishdan iborat bo'lgan avtomatik matn tahlilining bosqichlariga mos keladi.



Pragmatika odamlarning muayyan vazivatda matn yaratish motivatsivasini o'rganadi

Semantika alohida so'zlarning ma'nosini o'rganadi

Sintaksis gap tuzilmalari va so'z birikmalarini tekshiradi

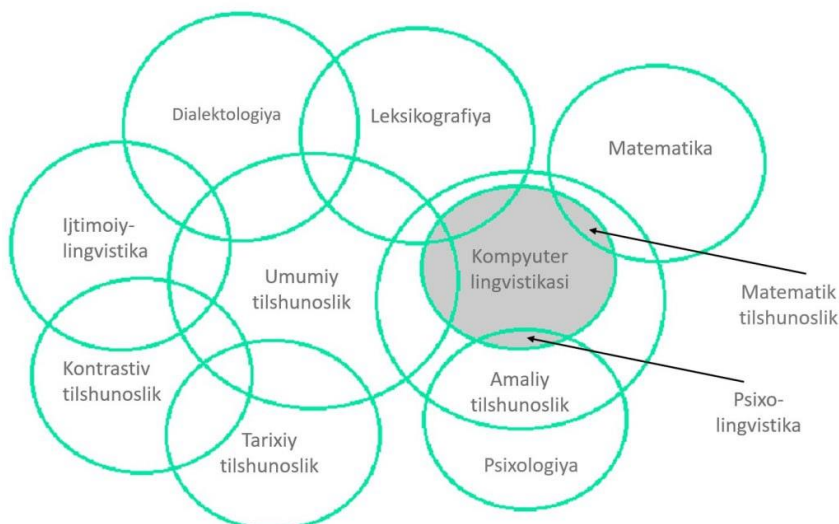
Morfologiya alohida so'zlarning ichki tuzilishini va yangi so'zlarning hosil bo'lish qonunivatlari o'rganadi

Grafika vozma nutq-grafemalari, harflar, ierogliflarning elementar birliklarining xususiyatlarini o'rganadi.

Fonologiya nutqni tashkil etuvchi tovushlarni o'rganadigan fan

Olti yoshga kelib, bola dastlabki uch bosqichni to'liq qamrab oladigan va to'rtinchi va beshinchi bosqichlarni qisman qamrab oladigan til bilimining rivojlanish darajasiga yetadi. Kompyuterni o'rganish hozircha to'rtinchi, sintaktik darajada eng samarali, foydalanuvchi kutganlari esa pragmatik darajada.[4]

Kompyuter lingvistikasi, tilshunoslik, matematika, informatika va sun'iy intellekt tarkibida shakllanadi.



Bu yerda turli fanlar chorrahasida qanday yangi yo'nalishlar shakllanayotganini ko'rishimiz mumkin. Masalan, psixolingvistika psixologiya va tilshunoslik.

Biroq, fan eng ko'p avtomatlashgan axborot sohasi bilan chambarchas bog'liq bo'lib, uning doirasida individual intellektual funksiyalarning dasturiy modellari ishlab chiqiladi. Bunda nutq sintezi va matn yaratish hisoblash lingvistikasining asosiy vazifalari hisoblanadi.

Kompyuter lingvistikasi oldida axborotni yig'ish, to'plash, qayta ishlash va olish jarayonlarini lingvistik ta'minlash vazifasi turibdi. Ulardan eng muhimlari:

Nutqni aniqlash va sintez qilish.

Matnlarni avtomatik tarjima qilish.

Elektron matn korpusini yaratish va ulardan foydalanish.

Matndan ma'lumot olish.

O'z-o'ziga havola.

Kompyuterga matn kiritishda xatolarni avtomatik aniqlaydi va tuzatadi, bunda

savol-javob tizimini yaratish asosiy parametrlardan hisoblanadi, masalan,

elektron lug'atlar, tezauriyalar, ontologiyalarni yaratish masalasi.

NLU va NLG: hisoblash tilshunosligi muammolarini hal qilish uchun yondashuvlar.

NLP ikkita asosiy muammoga duch keladi: tabiiy tilni yaratish (NLG) va tabiiy tilni tushunish (NLU).

Til yaratish - bu, masalan, chatbotlar uchun qisqa matn parchalarini yoki ilovalar uchun maxsus tarkibni yaratishga yordam beradigan xususiyat. NLGni odamlar g'oyalarni yozish yoki nutqqa aylantirganda foydalanadigan jarayon bilan solishtirish mumkin.

Tabiiy tilni tushunish sizga odamdan buyruqlar va so'rovlarni unga qulay shaklda qabul qilish, shuningdek, ma'lumotlarni saralash, asosiy narsani ajratib ko'rsatish, matn mazmunini va uning hissiy ranglarini tahlil qilish imkonini beradi.

Moslashuvchan texnologiyalar yordamida imkoniyati cheklangallarga xizmat qilish tizimlarini kengaytirish.

Keling, kompyuter lingvistikasi nima ekanligini, u qanday muammolarni hal qilishini va sog'ligida nuqsoni borlarga qanday yordam berishini ko'rib chiqaylik.

Kompyuter lingvistikasi imkoniyati cheklanganlarga yordam berish uchun yaratilgan adaptiv texnologiyalarning asosidir. [5]

Jahon sog'liqni saqlash tashkiloti ma'lumotlariga ko'ra, imkoniyati cheklanganlik 1 milliarddan ortiq odamga ta'sir qiladi - dunyo aholisining taxminan 15 foizi. Surunkali kasalliklar va aholining demografik qarishi tufayli sayyoramizda nogironlar soni muttasil ortib bormoqda.

Birlashgan Millatlar Tashkiloti Konventsiyasiga ko'ra, imkoniyati cheklanganlar qatoriga turli to'siqlar bilan munosabatda bo'lganda, insonning jamiyatda boshqalar bilan teng ravishda to'liq va samarali ishtirok etishiga to'sqinlik qiladigan doimiy jismoniy, aqliy, intellektual yoki hissiy nuqsonlari bo'lgan shaxslar kiradi.

Kompyuter tahlili va tabiiy tilni qayta ishlash sohasidagi ishlanmalar aholining ushbu toifasi uchun hayotni osonlashtirishga mo'ljallangan. Bu, ayniqsa, jismoniy, kognitiv va hissiy nuqsonlari bo'lganlar uchun to'g'ri keladi. Yangi adaptiv texnologiyalar nogironlarning imkoniyatlarini sezilarli darajada kengaytiradi va ularga ishda, ko'chada va uyda ko'proq avtonomiya beradi.

Kompyuter lingvistikasi eshitish, ko'rish va nutqida nuqsonlari bo'lgan odamlar uchun muloqotni osonlashtiradi.

Kompyuter lingvistikasi imkoniyati cheklanganlarga yordam berish uchun qanday muammolarni hal qiladi?

1. Jismoniy yoki kognitiv buzilishlarda nutq sintezi tan olinishi.
2. Tushunuvchanlikni yaxshilash uchun nutqni o'zgartirish.
3. Kundalik hayot uchun nutq va til texnologiyalari orqali g'amxo'rlik va cheklangan qaram

yashash muhitini yaxshilash.

4. Yordamchi muqobil aloqa uchun modellashtirish va kompyuterni o'rganishga yangi yondashuvlar.

5. Cheklangan ma'lumotlarga asoslangan yordamchi muloqot va nutq sintezi uchun moslashtirilgan so'zlashuvlar.

6. Yordamchi texnologiyalarga moslashtirilgan multimodal foydalanuvchi interfeyslari va dialog tizimlari.

7. Kognitiv yordam dasturlari uchun NLP.

8. Ko'rish qobiliyati zaif odamlar uchun grafik ma'lumotlar taqdimoti.

9. Nutq va NLP front-end ilovalar uchun qo'llaniladi.

10. Tilni qayta ishlash ilovalari uchun miya-kompyuter interfeyslari.

11. Yordamchi texnologiyalar uchun nutq, tabiiy va multimodal interfeyslar.

12. Webga kirish imkoniyati, matnni soddalashtirish, nutq yoki maxsus belgilar kabi moslashtirilgan taqdimot rejimlari.

13. Klinikalarda nutq va NLP vositalarini qo'llash.

14. Lingvistik manbalar, korpus diagrammalari va izohlari.

NLP bozorining o'sishi bir necha omillar bilan cheklanadi, jumladan:

1. Inson va kompyuter o'rtasida matn ma'lumotlarini idrok etish, tushunish va tan olish nuqtai nazaridan bo'shliq mavjudligi;

2. Tabiiy tillarni qayta ishlash sohasidagi tadqiqotchilarni tayyorlash dasturlari va kadrlar yetishmasligi;

3. Kompyuterda ishlov berish va matnlarning konteksti va ma'nosini tushunish qiyinligi.

NLP segmentidagi muammolardan biri ham bitta tizim yordamida matn bilan ishlashning turli muammolarini hal qiladigan universal til modellari va arxitekturasini yaratishni amalga oshiradi.

Foydalanilgan adabiyotlar

1. Boguraev B. (1994) Mashinada o'qiladigan lug'atlar va hisoblash tilshunosligi tadqiqotlari. In: Hisoblash tilshunosligining dolzarb muammolari: Don Uoker sharafiga. Dordrecht: Springer Niderlandiya. https://doi.org/10.1007/978-0-585-35958-8_7

2. Coscoy Y. (1997) Rasmiy dalillar uchun tabiiy tilda tushuntirish. In: Hisoblash tilshunosligining mantiqiy jihatlari. Berlin, Geydelberg: Springer Berlin Heidelberg.

3. Yo'ldoshev B. Matematik va kompyuter lingvistikasi (uslubiy qo'llanma). – Samarqand: SamDU nashri. 2007. – B. 57-58.

4. Po'latov A., Muhamedova S. Kompyuter tilshunosligida matnni avtomatik tahrir qilish uchun yaratilgan dasturning ba'zi lisoniy asoslari // O'zbek tilshunosligi masalari (ilmiy maqolalar to'plami). – Toshkent: TDPU nashri. 2003. – B. 32-36.

5. To'rayev H. Matematik mantiq va diskret matematika. – Toshkent: O'qituvchi, 2003. – 416 b.