

NLP VAZIFALARI VA YO'NALISHLARI

Adilova Munojot

Mustaqil tadqiqotchi

adilovamunojot@gmail.com

Annotatsiya. Kompyuter lingvistikasi tilga oid muammolarni yechish bo'yicha dasturlarni shakllantirish bilan shug'ullanadi. Tabiiy tilni qayta ishlash (NLP) ancha murakkab jarayon hisoblanib, unda ijtimoiy tabiatga ega tilning barcha hodisalari, umumiy va xususiy jihatlari, istisnoli holatlari, fonetik, morfologik, leksik, grammatik, semantik va hatto orfoepik xususiyatlarini e'tiborga olish muhim sanaladi. Mazkur maqolada matnlarni kompyuter tahlili jarayonidagi ahamiyati, NLPning vazifalari va muammolari haqida so'z yuritiladi, shuningdek, NLPga to'siq bo'luvchi omillar va bajarilishi lozim bo'lgan vazifalar haqida to'xtalib o'tilgan. Tabiiy til matni yoki ovozli ma'lumotlarga mo'ljallangan, ma'nosini aniq belgilaydigan dasturiy ta'minotni yaratishdagi tildagi qiyinchiliklarni bartaraf etish vazifalari ham sanab o'tilgan.

Kalit so'zlar: *tabiiy tilni qayta ishlash, NLP, NER, PoS-tagger, teglash, modellashtirish, tabiiy til generatsiyasi.*

Annotation. Computational linguistics deals with the development of programs for solving language problems. Natural language processing (NLP) is a very complex process in which it is essential to consider all phenomena, general and particular aspects, exceptional cases, phonetic, morphological, lexical, grammatical, semantic, and even orthographic features of the language's social nature. This article talks about the importance of computer analysis of texts, the tasks, and problems of NLP, the factors hindering NLP, and the tasks that need to be performed. It also lists the challenges of overcoming language difficulties in creating meaningful software for natural language text or voice data.

Keywords: *Natural language processing, NLP, NER, PoS-tagger, tagging, modeling, natural language generation.*

Аннотация. Компьютерная лингвистика занимается разработкой программ для решения языковых задач. Обработка естественного языка (NLP) — очень сложный процесс, в котором важно учитывать все явления, общие и частные стороны, исключительные случаи, фонетические, морфологические, лексические, грамматические, семантические и даже орфографические особенности социальной природы языка. В данной статье говорится о важности компьютерного анализа текстов, сложностях и проблемах NLP, а также о факторах, препятствующих работе NLP и задачах, которые необходимо выполнить. Также перечислены задачи по преодолению языковых барьеров при создании значимого программного обеспечения для текстовых или голосовых данных на естественном языке.

Ключевые слова: *обработка естественного языка, NLP, NER, PoS-tagger, тегирование, моделирование, генерация естественного языка.*

Hozirgi raqamli texnologiyalar muhiti rivojlanayotgan damda sun'iy intellekt tizimini yaratishda matn birliklarini raqamli texnologiyalar orqali qayta ishlanishiga erishish muhim vazifa hisoblanadi. Bu jarayonda eng muhim vazifalar sirasiga so'zlar turkumini teglash kiradi. Jahon kompyuter lingvistikasidan ushbu lingvo-texnik tahlilning so'zlar turkumini aniqlash – PoS-tagger, ya'ni so'z turkumlarini teglash, shuningdek, uning matnlarni avtomatik qayta ishlash jarayoni bosqichi ekani ma'lum. Til korpuslarini yaratish uchun boshlangan dastlabki teglash harakatlari bugungi kunga kelib, matn bilan bog'liq dolzarb masalalar yechimini kutmoqda. NLP, ya'ni tabiiy tilni qayta ishlash jarayonida ham so'z turkumlarini teglash birlamchi vazifa hisoblanadi.[1.13-14-b., 3.7-9-b.] Buning natijasida, avvalambor, omonimlikni, omofonlar, istehzolar (sarkazm), idiomalar, metaforalar, grammatika va foydalanishdagi istisnolar, shuningdek, ko'p ma'noli so'zlar semantikasini aniqlash kabi turli lingvistik noaniqliklar matn tarkibida tahlil qilinishiga erishiladi.

Tabiiy tilni qayta ishlash (NLP – *Natural language processing*) kompyuter fanlari sohasiga, aniqlash sun'iy intellekt (SI) yoki SI sohasiga tegishli bo'lib, kompyuterlarga matn va og'zaki (so'zlashuvdagi) so'zlarni xuddi insonlar kabi tushunish imkonini beruvchi tizim hisoblanadi [6.3-6-b., 14.40-48-b.]. NLP kompyuter lingvistikasi – sun'iy intellekt tizimlarini yaratishda inson va hayvonlarning intellektual jarayonlarini matematik va kompyuter modelashtirish sohasidagi ilmiy yo'nalish bo'lib, u tabiiy tillarni tavsiflashda matematik modellardan foydalanish tushuniladi. Inson tilini qoidalarga asoslangan holda modelashtirish – statistik tahlil, mashina tarjima va ma'lumotlarni chuqur o'rganish modellari bilan birlashtiradi. Bu texnologiyalar birgalikda kompyuterlarga inson tilini matn yoki ovozli ma'lumot sifatida qayta ishlash va uning to'liq ma'nosini, jumladan, ma'ruzachi yoki yozuvchining maqsadi va his-tuyg'ularini “tushunish” imkonini beradi.

NLP matnni bir tildan ikkinchi tilga tarjima qiladigan, og'zaki buyruqlarga javob beradigan va katta hajmdagi matnni tez, hatto real vaqtda umumlashtiradigan kompyuter dasturlarini boshqaradi. Ijtimoiy hayotda bu holatlarga juda ko'p murojaat qilinadi: ovozli GPS tizimlari, raqamli yordamchilar, nutqiy sintezator, mijozlarga xizmat ko'rsatish bo'yicha chatbotlar va boshqa iste'molchilar uchun qulayliklar ko'rinishida NLP bilan o'zaro aloqada bo'lishingiz uchun yaxshi imkoniyatlar bor. Ammo NLP, shuningdek, biznes operatsiyalarini tartibga solishga, bank, shifoxona, davlat xizmatlari kabi muassasalarda xodimlarning mahsuldorligini oshirishga va muhim biznes jarayonlarini soddalashtirishga yordam beradigan, korporativ yechimlarda muhim rol o'ynaydi. Bu holat insonga aniqlik, tezlik va navbatlardagi muammolarning yuzaga kelishining oldini oladi. Ba'zi hollarda inson omili qatnashmagan jarayonlar aniq bajariladigan jarayonlar ham uchraydi. Masalan, *bankomatdan pul yechib olasiz yoki aksincha valyutani o'z plastik kartangizga joylaysiz, shuningdek, to'lovlar jadvalida ham shu jarayon aks etadi. Siz kommunal*

yoki davlat xizmatlariga mobil ilova yoki bankomatlar orqali to'lashingiz va sanoqli daqiqalarda elektr yoki gaz avtomatik tarzda ishga tushishi mumkin. Ushbu jarayonlarning amalga oshishi uchun ham texnologiya va inson muloqoti zarur hisoblanadi.

Quyida NLP vazifalari va unda uchraydigan muammolar xususida qisqacha tushuntirib o'tildi.[19.15-25-b., 20.10-18-b.,]

NLPga to'siq bo'luvchi omillar va bajarilishi lozim bo'lgan vazifalari. Tabiiy tili matni yoki ovozli ma'lumotlarning mo'ljallangan ma'nosini aniq belgilaydigan dasturiy ta'minotni yozishni nihoyatda qiyinlashtiradigan noaniqliklarga to'la. Birinchi bo'lib uchraydigan muammolar: omonimlar, omofonlar, istehzolar (sarkazm), idiomalar, metaforalar, grammatika va foydalanishdagi istisnalar, jumlar tuzilishidagi o'zgarishlar inson tilidagi “qonunbuzarlik”larning faqat bir qismi bo'lib, odamlarga yillar davomida o'rganilishi kerak bo'ladi.[29.10-16-b.,] Ammo dasturchilar Tabiiy tildagi ilovalarni tanib olish va tanib olishni o'rgatishlari kerak. Ular ushbu ilovalar foydali bo'ladimi yoki yo'qligini aniq bilishi lozim, va buni bir nechta tajribalar isbotlab beradi. Tabiiy tilni qayta ishlash jarayoni kompyuter lingvistikasidagi muhim vazifalardan biri hisoblanib, til mezonlari kompyuter tushunadigan, raqamli ko'rinishga keltiriladi. Tildagi ijtimoiy hodisalar kompyuter tushunadigan tilda beriladi. Ba'zi NLP vazifalari inson matni va ovozli ma'lumotlarini kompyuterga nimani o'zlashtirayotganini tushunishga yordam bermaydi. Tilimizda shunday omillar borki ular NLP'ga xalaqit beradi. Ushbu vazifalarning ba'zilar quyidagilarni o'z ichiga oladi:

1. **Nutqni tanish** (ingl. *Speech to text*; rus. *Распознавание речи*;) – nutqni matnga aylantirish ham deyiladi, ovozli ma'lumotlarni matn ma'lumotlariga leksik, grammatik xatolarsiz aylantirish vazifasidir. [5.30-40-b., 14.10-16-b.,] Ovozli buyruqlarni bajaradigan yoki ovozli savollarga javob beradigan har qanday ilova uchun nutqni aniqlash talab qilinadi. Nutqni tanib olishni ayniqsa qiyinlashtiradigan narsa – bu odamlarning gapirish usuli: shevalardan foydalanishlari, tez gapirishlari, noaniq talaffuz, so'zlardagi urg'uning noto'g'ri qo'yib aytilishi va intonatsiyalar, orfoepik xatolarga yo'l qo'yilishi kabilar nutqni tanish jarayonida ko'p muammolarni yuzaga keltiradi.

2. **So'z turkumlarini teglash** (ing. *Part of speech tagging*; rus. *Тегирование частей речи*;) – shuningdek, grammatik teglash ham deb yuritiladigan jarayondir. Nutq belgilarining bir qismi ma'lum bir so'z yoki matnning nutq qismini uning qo'llanilishi va kontekstiga qarab aniqlash jarayonidir, ya'ni so'z yoki tekst fragmentining qaysi so'z turkimiga tegishli ekani qo'llanadigan manba yoki kontekstga qarab aniqlanadigan jarayon hisoblanadi.[20.14-15-b., 21.19-29-b.,] So'z turkumlarining identifikatsiyalashishi bo'yicha quyidagi misollari keltirish mumkin:

Kez1 – “ravish, payt holi” – M: Kezi kelganda aytish kerak deyishadi (kezi-palla, payt);

Kez2 – “fe'l” – M: Bog'ni kezayotib uni ko'rib qoldi (kezmoq – sayr qilmoq, aylanib yurmoq) shaklida kechadi. Ushbu so'z qaysi so'z turkimiga mansubligini

faqat kontekstda bilish mumkin. Yoki leksik polisemiyaning asosiy xususiyati shuki, soʻz qanchalik koʻp maʼnoli boʻlmasin, leksik maʼnolar orasida maʼlum munosabat, bogʻlanish boʻladi. Buni “joy” soʻzining leksik maʼnolari misolida koʻrishimiz mumkin.

Joy1. Aravalar joyidan quzgʻalishdi (S.Ahmad). – “Shaxs yoki narsa-predmet ishgʻol qilgan, egallagan, band qilgan yuza, sath, makon” maʼnosida.

Joy2. Nizomjon qurilish boʻlayotgan joyda mashina gurillaganini eshitgan edi. (S.Ahmad) – “Voqea yoki hodisa sodir boʻladigan, ish-harakat bajariladigan yuza, sath, punkt” maʼnosida.

Joy3. Yunusobod degan joydan kelayapman. (Oybek) – “Qandaydir ism bilan ataladigan geografik maydon, yuza, sath, territoriya” maʼnosida.

Joy4. Boylar, boyvachchalar anoyi emas, unga oʻz mehmonxonalaridan joy berib. (Oybek) – “Yashash uchun moʻljallangan yuza, makon” maʼnosida.

Joy5. Choʻlda yigirma sotix joyga oʻzim uchun sholi ekdim. (S.Ahmad) – “Ekin ekiladigan sath, yuza, makon” maʼnosida.

Joy6. Kifoyatxon joy soldi va boshini roʻyjaga burkab yotdi. (A.Qahhor) – “Yotish uchun moʻljallab biror sathga, yuzaga toʻshalgan koʻrpa, yostiqlik va shu kabilarning yigʻindisi” maʼnosida. Bu misollar asosida “joy” soʻzining 6 ta leksik maʼnosi oʻzaro yuza, sath, makon maʼno komponentlarda bir-biri bilan bogʻliq ekanini sezish qiyin emas. Lekin ularning barchasini teglash zarur chunki mashina ularni bir-biridan ajrata olmaydi.

3. Soʻz maʼnosini ajratish (ing. Word Sense Disambiguation (WSD); rus. Устранение неоднозначности смысла слова;) — maʼlum kontekstda koʻp maʼnoli soʻzni aniqlaydigan semantik tahlil jarayoni orqali koʻp maʼnoli soʻzning maʼnosini tanlash.[10.10-12-b.] Masalan, soʻzning maʼnosini ajratib koʻrsatish “surma” (koʻz boʻyogʻi) va “surma” (buyruq feʼl-biror narsani surma)dagi, “surma stol” (sifat-qavatlar surib ochiladigan buyum) soʻzning maʼnosini farqlashga yordam beradi. Bu jihat teglashda eng muhim jarayon hisoblanadi.

4. Nomga ega birliklarni aniqlash (ing. Named Entity Recognition (NER); rus. Распознавание именованных сущностей;) — Bu inson tilini qayta ishlash texnologiyasining yoʻnalishi boʻlib, dasturiy taʼminotni amalga oshirish nutq va matndagi soʻzlar va iboralarning obyektiv toifalarini topishga imkon beradi. [15.2-3-b.] Namuna tariqasida, Farhod bozorini – savdo rastalari joylashgan makon, Farhod esa atoqli ot, ism sifatida soʻzlar bazasiga kiritilishi kerak. Bunday soʻzlar jamlanmasi minglab topiladi va shunday soʻzlar NERni belgilash zarurati eng muhim vazifalardan biri ekanini eslatib turadi.

5. Anoforik tahlil – (ing. Anaphoric analysis; rus. Разрешение совместной ссылки;) — Oʻzidan oldingi soʻz, qaysi birlikka ishora koʻrsatilayotganini aniqlaydi. Bu tahlil ikki soʻz bir xil obyektga tegishli yoki yoʻqligini va qachon ikki xil maʼnoda kelishini aniqlash vazifasi hisoblanadi.[16.2-3-b.] Eng keng tarqalgan misol, maʼlum bir olmosh ishora qiladigan shaxs yoki obyektning aniqlashidir (masalan, “u” = “Abdugʻani”, “u”=“bank yoki bino”), baʼzan bitta soʻz u matndagi metafora yoki

idioma ta'rifini ham o'z ichiga olishi mumkin (masalan, “tulki” bu misol hayvon emas, balki ayyor odam tasviri bo'lishi mumkin).

6. His-hayajon so'zlarni aniqlashtirish va ularning tasnifi (ing. *Sentiment Analysis*; rus. *Анализ настроений, Анализ тональности*;). Tuyg'u va his-hayajon so'zlar va brikmalarni shuningdek, tuyg'uni anglatuvchi so'zlarni tahlil qilish, matndan subyektiv sifatlarni – munosabat, hissiyot, istehzo, chalkashlik, shubhalarni ajratib olishga harakat qilinadi.[13.2-4-b] Tabiiy til makonida his-hayajonni anglatuvchi so'zlarni qo'llamay turib ham his-hayajonni izhor qiluvchi, anglatuvchi so'zlar qo'shilmasi, gaplar mavjud. Buni faqat kontekstni o'qib anglash mumkin, lekin buni mashina tarjima qilayotganda ularni ajratib olishi bu so'zlarni yoki so'zlar qo'shilmasini tasniflash teglash lozim. Masalan, *Bu yilning eng ajoyib va qiziqarli filmlaridan biri deb hisoblayman*; – ing. *One of the most amazing and enjoyable films of the year*; rus. *Один из самых удивительных и приятных фильмов года*; (Muallif bu gapi bilan unga bu film yoqqanini bildirmoqda va bu gap ijobiy his-tuyg'uni anglatyapti). *Men ushbu suv o'tkazmaydigan kamera korpusini sotib oldim, chunki u odatiy kameradan ko'ra bardoshli bo'lishi kerak*; – ing. *I bought this waterproof camera case because it needs to be more durable than the stock one*; rus. *Я купил этот водонепроницаемый чехол для камеры, потому что он должен быть более надежным, чем стандартный*; Kontekstdan ko'rinib turibdiki, bu odam kutgan narsa emas. Gap salbiy ohangga ega, lekin u yashirin. Kayfiyat subyektivlik yoki obyektivlikka bog'liq emas, bu tahlilni murakkablashtirishi mumkin. Ammo fikr-mulohazalar ma'lumotlaridan qimmatli tushunchaga ega bo'lish uchun biz hali ham ifodalangan his-tuyg'ular, baholar yoki munosabatlarga ega bo'lmagan jumalarni ajrata olishimiz kerak.

7. Tabiiy til generatsiyasi – (ing. *Natural Language Generation*, rus. *Генерация естественного языка*;) – ba'zan nutqni tanib olish yoki nutqni matnga aylantirishga qarama-qarshi sifatida tavsiflanadi; tuzilgan ma'lumotni inson tiliga tarjima qilish vazifasidir. Matn va nutqni yaratish. Qarama-qarshi aniqlab olish vazifasi – generatsiya yoki sintez. Algoritm foydalanuvchi matnini yoki nutqiga javob berishi kerak. Bu savolga javob, foydali ma'lumot yoki kulgili ibora bo'lishi mumkin, ammo izoh berilgan mavzuda taaluqli bo'lishi kerak. Nutqni aniqlash tizimlari gaplarni qismlarga ajratadi. Bundan tashqari, ma'lum bir iborani talaffuz qilish uchun kompyuter ularni saqlaydi, o'zgartiradi va ko'paytiradi. Albatta, so'zlarning birikish chegaralarida buzilish bo'lishi mumkin, shuning uchun ovoz ko'pincha g'ayritabiiy eshitiladi. Matn yaratish algoritmgaga kiritilgan shablon javoblari bilan cheklanmaydi. U mashinani o'rganish algoritmlaridan foydalanadi. “So'zlashuvchi” dasturlar haqiqiy ma'lumotlarga tayangan holda o'rganishi mumkin. Mantiqiy tuzilishga ega she'rlar yoki hikoyalar yozish algoritmini olish mumkin, lekin ular odatda juda mazmunga boy bo'lmaydi.[17.45-48-b., 18.17-19-b.] Algoritm ma'lumotlar bazasidagi mavjud faktga tayangan holda savollarga javob beradi. NLG ma'lum kalit so'zlar yoki mavzular asosida odamlar kabi og'zaki matn yaratishga qaratilgan. Hozirda butun dunyoda mashina tarjimasiga alohida urg'u berilmoqda. Agarda

algoritmlar ishlashi uchun ma'lumotlar bazasini yaratsak (meshok slova), so'zlarning barchasi teglansa, tarjima samaradorligi oshadi, shuningdek, tarjima sifatiga ham ijobiy tasir qiladi. Hozirgi kunda ijtimoiy soha obyektlarida sun'iy intellektdan faol qo'llash ommalashdi. Bu chalkashlik va tartibsizliklarning oldini olmoqda. Misol qilib hozirgi kundagi faol jarayonlarni aytishimiz mumkin. NLP imkoniyatlariga ega aqlli chatbot mijozlar bilan mijozlarga xizmat ko'rsatish xodimlari kabi muloqot qilishi mumkin. Face ID tizimi orqali biron binoga kirish yoki biron programmani ishga tushurish mumkin. Barmoq izlari orqali bir nechta operatsiyalarni parol termasdan ochish, yopish, ishga tushurish va aksincha faoliyatini to'xtatish imkoni mavjud. Va bu, albatta, inson omili bilan birga bajariladigan jarayonlar hisoblanib, ijtimoiy hayotning yengillashishiga, qog'ozbozlikning kamayishiga va aniqlikka yordam beradi.

Xulosa

Ushbu tadqiqotdan ko'zlangan maqsad imkoniyati cheklangan insonlar uchun, shuningdek, ta'lim jarayonida, kundalik va rasmiy faoliyatda kompyuterdan foydalanish imkoniyatini oshirish, anjumanlardagi ilmiy munozaralar nutqining sifatli ko'chirilishi, yetkazib berilishi va to'g'ridan to'g'ri yozib olinishi, tarjima samaradorligini oshirish, o'zbek tilidan boshqa tillarga va aksincha, boshqa tillardan o'zbek tiliga matnlarni tarjima qilishda aniqlik va sifatning muhimligini oshirishga qaratilgan dasturiy ta'minotlar uchun tabiiy tilni qayta ishlash zaruratini aniqlash hisoblanadi. Tarjima dasturlarini takomillashtirish, shuningdek, so'z turkumlarini teglash algoritmlarni yaratish avtomatik tahlil, morfoanaliz va tarjimon dasturlari uchun birlamchi shart amalga oshirilsa, matnlarning grammatik jihatdan sifatli tahlil qilinishiga erishiladi. Lingvistik bilimlar kompyuter dasturlari va til korpuslari bazasi uchun eng zarur ma'lumot manbayi hisoblanadi. Lingvistik protsessorda lisoniy ma'lumotlar bazasini shakllantirishda, kundalik hayotda til bilimlaridan foydalanish ko'nikmalariga asoslanilsa, bunday vaziyatlar adabiy til me'yorlariga aylantirilsa, kompyuter dasturlarining tahlil imkoniyati mutaxassis darajasida mukammallikka yetadi. Aynan so'z (asosan, polifunksional, ko'p ma'noli, omonim so'zlar)ning turkumini belgilashda ham maqola avvalida “joy” so'zi misolidagi kabi so'zning pragmatik va kontekstual ma'nolari hamda qo'shimcha semalariga asoslanilsa, raqamli texnologiya dasturlari va tizimlarining amaliy ahamiyati yanada oshadi. Umuman olganda, bugungi kunda lingvistik bilimlar tildan amalda foydalanish imkoniyatiga asoslanib mukammallashtirilsa, nazariy manbalarga tayanilib, kompyuter dasturlari va tizimlari uchun formal o'zbek tili yaratilsa, kelajakda barcha turdagi (o'zbek tarjimon dasturi, nutqni tanish, o'zbek tili morfoanalizatori, kompyuter muloqoti kabi) dastur hamda elektron tizimlarining yuzaga kelishiga zamin yaratiladi.

Foydalanilgan adabiyotlar:

1. Abjalova, M. 2020. Tahrir va tahlil dasturlarining lingvistik modullari. Toshkent: Nodirabegim.

2. Abjalova, M., Yuldashev, A. 2021. "Methods for Determining Homonyms in Linguistic Systems". ACADEMICIA: An International Multidisciplinary 17
3. Abjalova, M., Tabiiy tilni qayta ishlash (NLP)da so'z turkumlarini teglash masalasi. Research Journal 11 (2): 700-715. DOI: 10.5958/2249-7137.2021.00522.
4. Abjalova, M. 2021. "O'zbek tili Milliy korpusida so'zshakllarni leksikografik baza asosida qidiruv imkoniyatlari". Kompyuter lingvistikasi: muammo, yechim, istiqbollar, 12-17. Toshkent: ToshDO'TAU.
5. Abjalova, M. 2021. "O'zbek tili milliy korpusida sinonimayzer yoxud sinonimizatorni yaratish masalasi". O'zbek Milliy va ta'limiy korpuslarining yaratishning nazariy hamda amaliy masalalari, 38-40. Toshkent: ToshDO'TAU. O'zbek tilining izohli lug'ati. 2006. 80 000 dan ortiq so'z va so'z birikmasi. A. Madvaliyev tahriri ostida. 5 jildli. Toshkent.
6. B.Brill, E. 1995. "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging". Computational Linguistics 21: 543-565. <http://acl.ldc.upenn.edu/J/J95/J95-4004.pdf>. Brill, E. 1992. "A simple rule-based part of speech tagger". Proceedings of ANLC 154.
7. Baum, L. E.; Sell, G. R. 1968. "Growth transformations for functions on manifolds". Pacific Journal of Mathematics 27 (2): 211-227.
8. Elov B., Tabiiy tilni qayta ishlash (NLP)da SpaCY modulidan foydalanish, 42-43. Toshkent: ToshDO'TAU.
9. Ernst, Jason; Kellis, Manolis. 2012. "ChromHMM: automating chromatinstate discovery and characterization". Nature Methods 9 (3): 215– 216. doi:10.1038/nmeth.1906. PMC 3577932. PMID 22373907.
10. Pal A.R., Saha D.J Word sense disambiguation: A survey, 2015.
11. Rizayev, S. 2006. O'zbek tilshunosligida lingvostatistika asoslari. Toshkent: Fan
12. Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova Z., Abuzalova M. Hozirgi o'zbek adabiy tili. – T.: Fan va texnologiya, 2010.
13. S.Sorokin. B.N Sintez rechi. -M.: Nauka, 1992, 390-392
14. Stanford Log-linear Part-Of-Speech Tagger. – 2016. – Access mode: <https://nlp.stanford.edu/software/tagger.html>
15. Stanford Named Entity Recognizer (NER). – 2016. – Access mode: <https://nlp.stanford.edu/software/CRF-NER.shtml>.
16. Stanford Named Entity Recognizer (NER). –2016. – Access mode: <https://nlp.stanford.edu/software/CRF-NER.shtml>.
17. Starner, Thad, Pentland, Alex. 1995. Real-Time American Sign Language Visual Recognition From Video Using Hidden Markov Models. Master's Thesis, MIT, Program in Media Arts. Li, N;



18. Stephens, M. 2003. “Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data”. *Genetics* 165 (4): 2213-2233. doi:10.1093/ genetics/165.4.2213.
19. К Кравченко Ю.А. Мансур, А.М. Мохаммад, Ж.Х. Модифицированный метод устранения неоднозначности смысла слова, основанный на методах распределенного представления DOI 10.18522/2311-3103-3-92-101 УДК 04489
20. K.Toutanova Chr. D. Manning. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. – Access mode: <https://nlp.stanford.edu/~manning/papers/emnlp2000.pdf>.
21. Natural Language Toolkit. – Access mode: <http://www.nltk.org/>.
22. <https://asu-analitika.ru/analiz-nastroenij-tipy-instrumenty-i-scenarii-ispolzovaniya>
23. https://en.wikipedia.org/wiki/Hidden_Markov_model.
24. <https://www.freecodecamp.org/news/an-introduction-to-part-ofspeech-tagging-and-the-hidden-markov-model-953d45338f24>.
25. <https://coderlessons.com/tutorials/akademicheskii/obrabotkaestestvennogo-iazyka/pometka-chasti-rechi-pos>.
26. <https://habr.com/ru/post/125988>.
27. https://ru.wikipedia.org/wiki/Частеречная_разметка.
28. https://en.wikipedia.org/wiki/Part-of-speech_tagging
29. https://ru.wikipedia.org/wiki/Часть_речи