

MATNLARGA DASTLABKI ISHLOV BERISH VA SEMANTIK TAHLIL QILISHGA OID AYRIM MULOHAZALAR

SOBIRJONOV B.Q., MAHAMMADOV A.A.

FarDU o'qituvchisi, bekzodbekqahromonovich@gmail.com

FarDU talabasi, a04704941@gmail.com

Annotatsiya. Ushbu maqolada hozirgi kunda ilmiy dolzarbligi bilan tabiiy tillarni qayta ishlash sohasi mutaxassislari e'tiborida bo'lgan masalalardan biri - tabiiy tillarni qayta ishlashning birinchi komponenti hisoblanuvchi morfologik va semantik tahlil, ularning ahamiyati haqida ma'lumotlar keltirilgan.

Kalit so'zlar: tabiiy tillar, kompyuter lingvistikasi, semantik tahlil, morfologik tahlil, sintaktik tahlil, pragmatik tahlil.

Kirish: Hozirgi globallashuv va raqamli texnologiyalar sharoitida axborot hajmining keskin ortishi natijasida matnli ma'lumotlarni avtomatik qayta ishlashga bo'lgan ehtiyoj tobora kuchayib bormoqda. Ayniqsa, internet tarmoqlari, ijtimoiy platformalar va elektron hujjat aylanishi tizimlarida katta hajmdagi matnlar bilan ishlash zarurati tabiiy tillarni kompyuter yordamida tahlil qilishni dolzarb ilmiy yo'nalishga aylantirdi. Shu nuqtayi nazardan Natural Language Processing (TTQI) zamonaviy axborot texnologiyalari va hisoblash lingvistikasining eng muhim yo'nalishlaridan biri hisoblanadi.

Tabiiy tillarni qayta ishlash jarayonida matnni chuqur tahlil qilish bir nechta bosqichlardan iborat bo'lib, ulardan eng muhimi morfologik, sintaktik, semantik va pragmatik tahlil bosqichlaridir. Ushbu bosqichlar matnning mazmunini to'liq tushunish, so'zlar o'rtasidagi grammatik va semantik bog'liqlikni aniqlash hamda kontekstga mos ma'noni ajratib olish imkonini beradi. Xususan, morfologik va semantik tahlil TTQI tizimlarining asosiy poydevori bo'lib xizmat qiladi.

Bugungi kunda dunyo miqyosida sun'iy intellekt texnologiyalarining rivojlanishi, mashinali o'qitish va chuqur neyron tarmoqlarning keng qo'llanilishi matnlarni avtomatik tahlil qilish samaradorligini sezilarli darajada oshirdi. Shu bilan birga, o'zbek tilidagi matnlarni qayta ishlash, uning morfologik xususiyatlarini modellashtirish hamda semantik tahlil algoritmlarini ishlab chiqish masalalari ham dolzarb ilmiy muammolardan biri bo'lib qolmoqda. Shu bilan birga, tabiiy tillarni qayta ishlash tizimlarida matnni to'liq tushunish uchun lingvistik va matematik yondashuvlarning uyg'unligi muhim ahamiyat kasb etadi. Matnning to'g'ri tahlil qilinishi nafaqat uning grammatik tuzilmasini aniqlash, balki uning semantik mazmunini to'liq ochib berish imkonini ham yaratadi. Bu esa axborot qidiruv tizimlari, avtomatik tarjima, matnlarni tasniflash va intellektual savol-javob tizimlarining aniqlik darajasini sezilarli darajada oshiradi.

Bundan tashqari, hozirgi kunda katta hajmdagi matnlar bilan ishlashda ma'lumotlarni oldindan tayyorlash, ya'ni dastlabki ishlov berish bosqichi (preprocessing) muhim o'rin tutadi. Ushbu bosqichda matnni tozalash, tokenizatsiya qilish, stop-so'zlarni olib tashlash, lemmatizatsiya va normalizatsiya kabi amallar bajarilib, keyingi tahlil jarayonlari uchun sifatli ma'lumotlar bazasi shakllantiriladi. Bu jarayonlar semantik tahlilning aniqligini oshirishga bevosita ta'sir ko'rsatadi.

Shuningdek, zamonaviy yondashuvlarda mashinali o'qitish va chuqur neyron tarmoqlar asosida matnni avtomatik tushunish tizimlari ishlab chiqilmoqda. Ushbu texnologiyalar katta hajmdagi ma'lumotlardan qonuniyatlarni aniqlash va ularni model asosida qayta ishlash imkonini beradi. Natijada, tizimlar nafaqat mavjud ma'lumotlarni tahlil qiladi, balki yangi ma'nolarni ham kontekst asosida aniqlash qobiliyatiga ega bo'ladi.

Yurtimizda bugungi kunda tabiiy tillarni qayta ishlash jarayonini nazariy va amaliy jihatdan o'rganish, o'zbek kompyuter lingvistikasini rivojlantirish va til muammolarini hal qilish ustida ko'plab ilmiy tadqiqotlar olib borilmoqda. Davlat tilining zamonaviy axborot texnologiyalari va

kommunikatsiyalariga faol integratsiyalashuvini ta'minlash maqsadida respublikamiz rahbariyati tomonidan "o'zbek tilini Internet jahon axborot tarmog'ida ommalashtirish, unda munosib o'rin egallashini ta'minlash, dasturiy mahsulotlarning o'zbekcha ilovalarini yaratish, o'zbek tilini o'rgatuvchi kompyuter dasturlarini keng miqyosda amaliyotga tatbiq qilish; o'zbek tilidagi matnlar tahririga mo'ljallangan kompyuter dasturlarini yaratish"1 vazifalari belgilangan.

Bu vazifalarni muvafaqqiyatli amalga oshirilishi o'zbek tili so'z shakllarining morfologik tahlil qilish modellari va algoritmlarini, lotin va kirill alifbosini e'tiborga olgan holda o'zbek tilidagi matnlarni samarali qayta ishlashga mo'ljallangan dasturiy vositalarni ishlab chiqish texnologiyalari rivojlantirish bugungi kunning dolzarb masalalaridan biri hisoblanadi.

Jamiyatni axborotlashtirishning rivojlanishi va kengayishi turli ko'rinishdagi video, audio, matnli va boshqa ko'rinishdagi ma'lumotlardan foydalanish qonuniyatlarini puxta ishlab chiqish ko'nikma va malakalarini talab etmoqda. Bugungi kunda hech kimga sir emaski, jamiyatning aksariyat katta qismi matnli ko'rinishdagi axborotlardan foydalaniladi. Bu esa matnni avtomatik qayta ishlash (tabiiy tilni qayta ishlash) va hisoblash lingvistikasi sifatida tanilgan ilmiy tadqiqot sohaning rivojlanishini sezilarli darajada tezlashtirdi.

Tabiiy tillarni qayta ishlashning rivojlanish bosqichi 1950-yilda boshlandi. Bunga Alan Tyuringning "Hisoblash mashinalari va ong" nomli maqolasini chop etishi turtki bo'ldi. Olim o'z maqolasida mashinaning inson kabi fikrlash jarayoni intepretatsiyasini tasvirlaydi. Unga ko'ra inson kompyuter yoki boshqa inson bilan muloqotga kirishda kim bilan suhbat qilayotganligini aniqlash kerak edi. Kompyuter dasturi esa iloji boricha insonni chalkashlikka olib borishi kerak edi. Olimning ushbu gipotezasi jamiyatni ko'proq axborotlarni o'zlashtirishga bo'lgan ehtiyojini orttirib boraverdi. Buning natijasida turli tillardagi ma'lumotlarni o'zlashtirish istagi insoniyatni mashinaviy tarjima bo'yicha tadqiqotlar olib borishga majbur etdi. 1954-yil AQSHda rus tilidan ingliz tiliga 250 ta so'zni tarjima qiladigan avtomatlashtirilgan tarjima tizimining birinchi ommaviy taqdimoti bo'lib o'tgan. Bunda to'g'ridan-to'g'ri tarjima qilish usuli ishlatilgan. Ushbu birinchi avlod arxitekturasi lug'atga asoslangan va manba tilini maqsadli til bilan so'zma-so'z moslashtirishga, ya'ni to'g'ridan-to'g'ri tarjima qilish asosida ishlagan. Tarjima sifati past bo'lishiga qaramay, loyiha yaxshi qabul qilingan va AQSH hamda Sobiq Sovet Ittifoqi tomonidan ushbu sohadagi tadqiqotlar moliyalashtirgan. Bu tadqiqotlarning rivojlanishi tillarning formal grammatikasining yaratilishiga olib keldi. 1960-yilda N.A. Chomsky va boshqa olimlar tomonidan formal til nazariyasi va til sintaksisini yaratish bo'yicha olib borilgan ishlar bunga misol bo'lishi mumkin. Bunda Chomsky tomonidan formal tillar uchun til grammatikasi taklif qilingan. Formal grammatika - bu formal tilni tavsiflash usuli, ya'ni ma'lum bir chekli alifboning barcha so'zlari to'plamidan ma'lum bir to'plamostini tanlab, formal tilni ifodalash usuli hisoblanadi. Yaratiladigan grammatikalar tildagi har qanday so'zni yasash uchun mumkin bo'lgan qoidalarni belgilaydi, taniydigan grammatika esa so'zning ma'lum bir tilga tegishli yoki yo'qligini aniqlash imkonini beradi. Chomsky til grammatikasini quyidagi turlarga bo'lgan.

Xomsk til grammatikasining tasnifi

Grammatika turi	Qabul qilinadigan grammatika	Qabul qilinadigan til	Avtomat
0-tur	Cheklanmagan grammatika	Rekursiv qayta ishlanadigan til	Tyuring mashinasi

1-tur	Kontekstga bog'liq grammatika	Kontekstga bog'liq til	Chiziqli chegaralangan avtomat
2-tur	Kontekstli grammatika	Mustaqil kontekstli til	Pushdown avtomat yoki xotirali avtomat
3-tur	Oddiy grammatika	Oddiy til	Chekli avtomat

1980 - yillar oxirida tabiiy tilni qayta ishlash uchun mashinali o'qitish algoritmlari paydo bo'lishi bilan inqilob yuz berdi. Bu hisoblash kuchining muttasil o'sib borishi va Xomsk lingvistik nazariyasining ustunligini asta-sekin pasayib borishi mashinali o'qitish algoritmlarining rivojlanishiga olib keldi. 1980-yillar oxiri va 1990- yillar o'rtalarida "statistik inqilob" deb nomlangan davrdan beri tabiiy tilni qayta ishlash bo'yicha ko'plab tadqiqotlarda asosan mashinali o'qitishga asoslangan algoritmlar qo'llanila boshlandi. 2000-yilga kelib katta miqdordagi nutq va matnli ma'lumotlarga ishlov berish uchun chuqur neyron tarmoqlar yoki mashinali o'qitish algoritmlari qo'llanila boshlandi.

Tabiiy tillarni qayta ishlash(TTQI) beshta komponentdan iborat. Bularga morfologik tahlil(MT), sintaktik tahlil, semantik tahlil, pragmatik tahlil, nutq integratsiyasi kiradi. Bu komponentlar mashinaviy tarjima, savol va javob tizimlari, nutqni matnga o'girish, orfografik va grammatik xatolarni tuzatish, matnlarni klassifikatsiyalash masalalarida qo'llaniladi.

MTda matn abzasga, so'z va gaplarga bo'linadi, alohida olingan so'zlarning tarkibiy qismlari tahlil qilinadi va matndan tinish belgilari ajratiladi. Sintaktik tahlil. Ushbu komponentning 2 ta maqsadi bor: gap to'g'ri shakllantirilganligini tekshirish va gapda ishtirok etgan so'zlar orasidagi sintaktik munosabatlarni ko'rsatish.

Semantik tahlil- bu qiymatlarni belgilaydigan sintaktik analizator tomonidan yaratilgan struktura. Ushbu komponent so'zlarning chiziqli ketmaketligini struktura ko'rinishiga keltiradi va so'zlar bir-biri bilan qanday bog'liqligini ko'rsatadi. Semantikada so'z, ibora va jumalarning faqat to'g'ridan to'g'ri ma'nosiga e'tibor qaratiladi. So'zning kontekstdagi lug'aviy ma'nosi ajratiladi.

Pragmatik tahlil ma'lum bir kontekstda mavjud bo'lgan real obyektlarga, hodisalarga murojaat qiladi, semantik tahlil natijasida olingan jumlaning ikkita semantik talqinidan biri tanlab olinadi. Matnlarning to'laqonli tahlilini amalga oshirish uchun yuqorida nomlari keltirilgan tabiiy tillarni qayta ishlash komponentlarining hammasi to'liq bajarilishi kerak. Ushbu jarayonda MTning ahamiyati shundaki, birinchi komponentning to'g'ri ishlab chiqilishi keyingi bosqichlarning samarali natijasini ta'minlaydi.

Xulosa: Ushbu maqolada matnlarga dastlabki ishlov berish va semantik tahlil qilish jarayonlarining zamonaviy Natural Language Processingdagi o'rni va ahamiyati yoritildi. Tahlillar shuni ko'rsatadiki, matnni avtomatik tushunish va qayta ishlash tizimlarining samaradorligi avvalo uning morfologik va semantik darajadagi to'g'ri tahlil qilinishiga bog'liqdir. Morfologik tahlil matnni eng kichik til birliklariga ajratish va ularning grammatik xususiyatlarini aniqlash orqali keyingi bosqichlar uchun asos yaratadi. Semantik tahlil esa so'zlar va jumalar o'rtasidagi ma'no munosabatlarini ochib berib, matnning haqiqiy mazmunini aniqlash imkonini beradi. Shu sababli ushbu ikki bosqich bir-biri bilan uzviy bog'liq bo'lib, TTQI tizimlarining asosiy poydevorini tashkil etadi.

Bugungi kunda sun'iy intellekt va mashinali o'qitish texnologiyalarining rivojlanishi natijasida matnlarni qayta ishlash tizimlari tobora takomillashib bormoqda. Biroq o'zbek tilining

morfoloqik murakkabligi va semantik xususiyatlarini to'liq hisobga oluvchi algoritmlarni yaratish hali ham dolzarb ilmiy muammo bo'lib qolmoqda. Shu bilan birga, matnlarni to'g'ri tahlil qilish faqat lingvistika uchun emas, balki zamonaviy axborot texnologiyalari, sun'iy intellekt tizimlari, avtomatik tarjima, matnlarni tasniflash va savol-javob tizimlari kabi amaliy yo'nalishlar uchun ham katta ahamiyatga ega. Aynan shu sababli TTQI sohasida olib borilayotgan tadqiqotlar ilm-fan va texnologiya integratsiyasining muhim yo'nalishlaridan biri sifatida e'tirof etiladi.

Kelajakda ushbu yo'nalishda yaratiladigan algoritmlar va intellektual tizimlar matnni yanada chuqur tushunish, ma'no aniqligini oshirish va inson tiliga yaqin ishlov berish imkoniyatini beradi. Bu esa o'z navbatida raqamli texnologiyalar rivojini yangi bosqichga olib chiqadi. Umuman olganda, matnlarga dastlabki ishlov berish va semantik tahlil qilish zamonaviy ilm-fan va amaliyot uchun strategik ahamiyatga ega yo'nalish bo'lib, uning rivoji kelajakda intellektual tizimlarning yanada mukammallashishiga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR:

1. Абжалова М.А. Ўзбек тилидаги матнларни тахрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (расмий ва илмий услубдаги матнлар таҳрири дастури учун). Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси автореферати. – Тошкент, 2019. – 49 б.
2. Mohinur U., Sitorabonu E. Graduonimiya hodisasi va uning ayrim o'zbek xalq maqollaridagi ifodasi // *Innovations in Technology and Science Education*. – 2023. – T. 2. – № 9. – S. 1757–1763.
3. Mohinur U. Muloqotning pragmatik vaziyat bilan bog'liq ayrim xususiyatlari // *XXI asrda ilm-fan taraqqiyotining rivojlanish istiqbollari va ularda innovatsiyalarning tutgan o'rni mavzusidagi Respublika ilmiy onlayn konferensiya materiallari*. – S. 76.
4. Бахроновна Р.Г. Историко-мифологические основы образов, связанных с деревьями в фольклоре // *World Science: Problems and Innovations: сборник статей LII Международной научно-практической конференции*. – 2021. – T. 52. – S. 157–160.
5. Шербекова Г.Я. Афсона ва ривоятларда “Хаёт дарахти” образи ифодаланишига оид айрим мулоҳазалар // *Scientific Progress*. – 2022. – T. 3. – № 2. – S. 244–250.
6. Рустамова Г.Б. Образнинг семантик таққинлари // *Тошкент давлат шарқшунослик университети*. – 2021. – S. 319.