

MATNNI AVTOMATIK TAHLIL QILISH JARAYONI

Jabbarova Feruza Xojimuratovna,
jabborovaf2@gmail.com

Termiz davlat universiteti, Filologiya fakulteti, Magistratura bo'limi
Lingvistika yo'nalishi 321- guruh magistranti

Annotatsiya: Ushbu maqolada matnni tahlil qilish asosida olib borish mumkin bo'lgan intellektli tizimlar haqida ma'lumot berilgan. Tahlil qilinayotgan matnlar ustida olib boriladigan ishlarni tahlil qilish jarayonining murakkablik omillari keltirilgan.

Kalit so'zlar: *xotira tezligi, avtomatik tahlil, mashinali o'rganish, matn.*

THE PROCESS OF AUTOMATIC TEXT ANALYSIS

Jabbarova Feruza Khojimuratovna
jabborovaf2@gmail.com

Termez State University, Faculty of Philology, Master's Department
Master of Linguistics, Group 321

Annotation: This article provides information on intelligent systems that can be based on text analysis. Factors of complexity of the process of analysis of the work carried out on the analyzed texts are given.

Keywords: *memory speed, automatic analysis, machine learning, text.*

Rivojlanib borayotgan fan va texnika davrida inson juda katta hajmdagi axborot ustida tezkorlik bilan ishlashiga to'g'ri keladi. Bunday jarayonlarni boshqarishda insonning imkoniyatlari (xotira tezligi) chegaralangan bo'ladi. Axborotni qabul qilish va uning ustida ishlash samaradorligini oshirishning yagona yo'li sun'iy intellekt – kompyuterdan foydalanib yechim topishdir. Bu o'z navbatida hamma sohaning kompyuter uslubini yaratishni talab qiladi.

Matnni avtomatik tahlil qilish. Insoniyatning rivojlanish tarixida qo‘lga kiritgan barcha yutuqlari zahirida o‘zining hayotiy faoliyatini yaxshilash, qisqa fursatlarda ko‘p ishlarni amalga oshirish, aql-zakovatini tuganmas imkoniyatlarini namoyon qilish maqsadi yotadi. Kompyuter texnologiyalari deb nom olgan asrimizda fan-texnika yutuqlari deyarli barcha sohaning ajralmas qismiga aylanib ulgurgan. Jumladan, tilshunoslik sohasi ham ayni shu jarayondan mustasno emas[Rahimov,2011:56]. Matematika, informatika, lingvistika fanlarining kesishgan nuqtasida paydo bo‘lgan yangi soha – kompyuter lingvistikasini axborot texnologiyalarisiz, kompyuter dasturlarisiz tasavvur qilib bo‘lmaydi. Endilikda bu hodisa lingvistika oldiga yangi talablarni qo‘yib, fan doirasida turli yo‘nalishlarning paydo bo‘lishiga zamin tayyorladi. Biz ma’lumotlarni tez fursatlarda samaradorligi yuqori bo‘lgan holatda olishni xohlaymiz. Qaysi manbaga murojaat qilmaylik, unda ifodalangan asosiy fikrni tezlikda anglash, matn mazmunini o‘zlashirish diqqatimiz markazida bo‘lgan. Ayni shu narsani kompyuter dasturi yordamida savol-javob tarzida amalga oshirish mumkin. Ya’ni dasturning o‘zi savol qo‘yadi, javob beradi va matnni tahlil qiladi[Areeb,2008:25-40].

Avtomatik matnlarni tahlil qilish quyidagi asosiy ikki turdagi yondashuvlar yordamida amalga oshiriladi: lug‘at va qoidalarga asoslangan yondashuvlar (lingvistik muhandislik yondashuvi) va mashinani o‘rganishga asoslangan yondashuvlar. Lug‘atlar va qoidalarga asoslangan usullar baholovchi so‘z va iboralarning maxsus yaratilgan lug‘atlaridan foydalanish va so‘zlarni qo‘llash kontekstini hisobga oladigan lingvistik qoidalarni qo‘llashdan iborat. Avtomatik matnlarni tahlil qilish tizimlari uchun ishlatiladigan mashinani o‘rganish usullari ikki turga bo‘linadi: nazorat qilinadigan mashinani o‘rganish usuli va nazoratsiz yoki qisman boshqariladigan mashinani o‘rganish usullari deb atash mumkin.

Masalan, qaysidur dastur muayyan matndagi har bir xat boshini tahlil qiladi: asosiy savollarni qo‘yadi, javob beradi. BORIS – darak gapli matnlarni chuqur tushunishga mo‘ljallangan amaliy tizim. Matnlarni chuqurroq tushunib, uning

savollariga javob beruvchi tajribaviy dastur. U bilimlar manbayiga asoslanib ishlaydi. Faqat mazmunni emas, matndagi barcha darak gaplarni tahlil qiladi.

BORIS tizimini takomillashtirish kompyuter lingvistikasining istiqbollaridan biri bo'lib, matn mazmunini o'zlashtirishdagi yuqori samaradorligi bilan ahamiyatlidir. Nafaqat kompyuter lingvistikasida, balki istalgan badiiy matn tahlilida BORIS tizimidagi savol-javob usulini qo'llash mazmunni o'zlashtirishga, aytilmagan o'rinlarga e'tibor qaratishga xizmat qiladi [Большакова, 2017:269].

Tahlil qilinayotgan matnning uzunligi ham matnlarni sentimental tahlil qilish xususiyatlariga katta ta'sir ko'rsatadi. Qisqa matnlar, masalan, Twitter postlari, qisqacha sharhlar juda aniq tahlilni talab qiladi. Kattaroq matnlarda bildirilgan fikr turli versiyalarda bir necha marta takrorlanishi mumkin, bu tahlilni osonlashtiradi. Biroq, uzun matnlarda baholanadigan ob'ektlarning xilma-xilligi ortadi. Uzun matnlar boshqa odamlarning fikrlarini o'z ichiga olishi mumkin. Agar vazifa ko'rsatilgan ob'ektlarga nisbatan bo'linma topish bo'lsa, u holda hisob-kitoblar hajmini aniqlash muammosi paydo bo'ladi. Misol uchun, ko'pincha baholash bir xil jumlada eslatib o'tilgan shaxs bilan bog'lanadi. Lekin muallif predmetga olmoshlar kabi murojaat vositasida murojaat qilishi mumkin. Bundan tashqari, agar butun matn bitta ob'ektni muhokama qilishga bag'ishlangan bo'lsa, unda uni baholash joyidan yetarlicha uzoqda aniq aytib o'tish mumkin.

Foydalanilgan adabiyotlar:

1. Rahimov A. Kompyuter lingvistikasi asoslari. -Toshkent, 2011. 58-bet.
2. Агеев М. С., Добров Б. В., Лукашевич Н. В. Автоматическая рубрикация текстов: методы и проблемы // Учёные записки Казанского государственного университета. Серия Физико-математические науки. —2008. — Т.150, №4. — С. 25–40.
3. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. Автоматическая обработка текстов на естественном языке и анализ данных. —М.: Изд-во НИУ ВШЭ, 2017. —269 с.