



## NOMUHIM SO‘ZLAR TUSHUNCHASI VA UNING AHAMIYATI

**Mahmudjonova Gulshahnoz Ulug‘bek qizi**

*Alisher Navoiy nomidagi o‘zbek tili va adabiyoti universiteti, kompyuter lingvistikasi yo‘nalishi 1-kurs magistranti*

**Annotatsiya:** *Ushbu maqolada matnda nomuhim so‘zlar tushunchasi va uning ahamiyati, nomuhim so‘zlar matnga ta’siri: ijobiy va salbiy tomonlari haqida so‘z yuritiladi.*

**Kalit sozlar:** *nomuhim so‘zlar, tabiiy tilni qayta ishlash, semantik ma’no, chastota, grammatik ma’no, hujjat.*

Nomuhim so‘zlarga tadqiqotchilar tomonidan turlicha ta’riflar berilgan. Kam ma’noli ma’lumotlarga ega bo‘lgan, yoki mustaqil ma’noga ega bo‘lmagan, yoki barcha matnlarga xos keng tarqalgan so‘zlar nomuhim so‘zlari deb ataladi[1].

Nomuhim so‘zlar – bu oz semantik ma’lumotga ega bo‘lgan so‘zlar, ya’ni berilgan hujjatdagi matnga semantik ma’no kam yuklaydi. Matndan bunday so‘zlar olib tashlanganda matn ma’nosiga putur yetmaydi. Dastlab hujjatlarni qayta ishlash, va qidiruv tizimi algoritmlarini tezlashtirish uchun foydalanilgan nomuhim so‘zlar endi modellarni o‘qitish kabi mashinali o‘rganishni o‘z ichiga olgan tabiiy tilni qayta ishlash (NLP) vazifalarida qo‘llanilib kelinmoqda. Nomuhim so‘zlarni olib tashlash ma’lumotlar to‘plami hajmini kamaytirishi mumkin, bu esa ma’lumotlarni tez va oson qayta ishlashda yordam beradi

Nomuhim so‘zlar – bu hech qanday ma’lumot qiymatiga ega bo‘lmagan va hujjatning katta qismini tashkil etuvchi matn tarkibidagi so‘zlar. Darhaqiqat, bu so‘zlarning umumiy xususiyati shundaki, ular hujjatda muhim ma’lumotga ega emas; ular faqat grammatika tufayli ishlatiladi. Hujjatdagi matndan nomuhim so‘zlar tushirib qoldirilsa, matn mazmuni o‘zgarmaydi.

1. Matnning asl varianti: Farzandingizga ertaklardan birini o‘qing. Matn haqida bir nechta savol bering. Agar bola o‘qiy olsa, uni o‘zi qisqa hikoya o‘qishga taklif qiling va keyin uni qayta aytib bering.

2. Nomuhim so‘zlar matndan olib tashlangandan keyingi holati: Farzandingizga ertaklardan o‘qing. Matn haqida savol bering. Bola o‘qiy olsa, qisqa hikoya o‘qishga taklif qiling, qayta aytib bering.

Nomuhim so‘zlar aniqlash usullarini ikkita asosiy toifaga bo‘lish mumkin:

1. Grammatika qoidalariga asoslanib,
2. Statistik usullar[2].

Nomuhim so‘zlar ro‘yhati asosan mustaqil so‘z turkumlaridan sifat darajalari, miqdor-daraja, sabab, maqsad ravishi, olmoshlar, yordamchi so‘zlar, modal, taqlid, undov so‘zlardan tuzilishi mumkin. Sababi, bunday so‘zlarning matn mazmuniga ta’siri kam bo‘ladi, ya’ni gapda grammatik ma’no ifodalash uchun foydalaniladi.



Nomuhim soʻzlarini olib tashlashning ijobiy va salbiy tomonlari bor. Ijobiy tomoni nomuhim soʻzlar matndan chuqur oʻrganish (deep learning) va mashinali oʻrganish (machine learning) modellarini qoʻllashdan oldin olib tashlanadi. Chunki nomuhim soʻzlar hujjatda tez-tez uchraydi va bu soʻzlarni tasniflash yoki klasterlash uchun ishlatilishi mumkin boʻlgan muhim maʼlumotlarni saqlamaydi va bu soʻzlar olib tashlanganda maʼlumotlar toʻplamining hajmi kamayadi va modelni qoʻllashga ketgan vaqt qisqaradi.

Salbiy tomoni notoʻgʻri tanlangan va olib tashlangan nomuhim soʻzlar matn maʼnosini oʻzgartirishi mumkin. Shuning uchun nomuhim soʻzlarini tanlashda ehtiyot boʻlish kerak. Quyida misol bilan tushintiriladi. Siz bu ishni bugun qilishingiz shart emas.

Agar bu gapda nomuhim soʻz sifatida “emas” soʻzi olib tashlansa, gapga teskari maʼno beradi. Yaʼni ish bugun qilinishi shart boʻlib qoladi. Gap mazmuni butunlay oʻzgarib ketadi.

Nomuhim soʻzlarning taʼsiri aniq boʻlsa-da, nomuhim soʻzlar roʻyxatini ishlab chiqish uchun aniq koʻrsatma yoʻq. Shuning uchun turli tizimlar turli tillar uchun turli oʻlchamdagi nomuhim soʻz roʻyxatidan foydalanadi. Ingliz tili uchun SMARTsystem 571 soʻzdan iborat nomuhim soʻz roʻyxatini taklif qildi, Fox 421 soʻz hajmini taklif qildi va DIALOG axborot xizmati kabi tijoriy tizim faqat 9 soʻzdan iborat kichikroq nomuhim soʻzlar roʻyxatidan foydalandi.[3]

Bakhyt Bakiyev tadqiqot ishida[4] qozoq tili uchun 430 ta qozoqcha nomuhim soʻzlar roʻyxatini tuzgan.

Tabiiy til maʼlumotlarini qayta ishlash, toʻliq matn qidirish yoki maʼlumotlarni indekslashdan oldin yoki keyin filtrlanishi kerak boʻlgan toʻliq va yangilangan 337 turkcha nomuhim soʻzlari roʻyxati githab platformasida mavjud.

Nomuhim soʻzlari matnni ajratib olish (text mining), maʼlumotlarini qidirish(IR), matnni sarhisob qilish, indekslash, ijtimoiy media platformalari, klasterlash, qidiruv va qidirish tizimlari, tabiiy tilni qayta ishlash va boshqalarda keng qoʻllaniladi

### FOYDALANILGAN ADABIYOTLAR:

1. Oʻzbek tili matnlaridagi nomuhim soʻzlar, Madatov Xabibulla Axmedovich, Sharipov Maksud Siddikovich, Bekchanov Shukurla Kurbanbayevich, <https://www.researchgate.net/publication/351951567>
2. K. Madatov, S. Bekchanov, J. Vičič. “Automatic Detection of Stop Words for Texts in the Uzbek Language”, Preprints, MDPI, 2022
3. Ljiljana Dolamic and Jacques Savoy. When stopword lists make the difference. Journal of the American Society for Information Science and Technology, 61(1):200–203, 2010.



4. Bakhyt Bakiyev, Method for Determining the Similarity of Text Documents for the Kazakh language, Taking Into Account Synonyms: Extension to TF-IDF, 2022 International Conference on Smart Information Systems and Technologies (SIST), 28-30 April, 2022, Nur-Sultan, Kazakhstan

5. <https://github.com/tkorkunckaya/Turkish-Stopwords>