

NOMUHIM SO‘ZLAR TUSHUNCHASI VA UNING AHAMIYATI

Mahmudjonova Gulshahnoz Ulug‘bek qizi

ToshDO‘TAU, kompyuter lingvistikasi

yo‘nalishi 1-kurs magistranti,

gulshaxnozmahmudjonova@gmail.com

Annotatsiya: Ushbu maqolada nomuhim so‘zlar tushunchasi va uning ahamiyati haqida so‘z yuritiladi. Grammatik qoidalarga asoslangan nomuhim so‘zlar ro‘yxati mustaqil so‘zlardan, yordamchi so‘zlardan ishlab chiqilib, amaliyotga taklif etiladi.

Kalit so‘zlar: *nomuhim so‘zlar, semantik ma‘no, tabiiy tilni qayta ishlash, mashinali o‘qitish, grammatik usul, chastota, grammatik ma‘no, hujjat, yordamchi so‘zlar.*

Annotation: This article discusses the concept of stop words and its importance. A list of stop words based on grammatical rules is developed from independent words and auxiliary words and offered for practice.

Key words: *stop words, semantic meaning, natural language processing, machine learning, grammatical method, frequency, grammatical meaning, document, auxiliary words.*

Kirish. Nomuhim so‘zlar barcha tillarda mavjud bo‘lib, semantik ma‘noga ega bo‘lmagan so‘zlarga aytiladi. 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 [Madatov X.A, Sharipov M.S, Bekchanov Sh.K, O‘zbek tili matnlaridagi nomuhim so‘zlar]. 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: ma‘lumotlarini qidirish (IR), matnni sarhisob qilish, indekslash, ijtimoiy media platformalari, klasterlash va qidirish tizimlarida qo‘llanilib kelinmoqda. Nomuhim so‘zlar matnga semantik ma‘no deyarli yuklamasligi sabab, hujjatdan ularni olib tashlash ma‘lumotlar to‘plami hajmini kamaytirishi mumkin, bu esa ma‘lumotlarni tez va oson qayta ishlashda yordam beradi. Nomuhim so‘zlarni olib tashlash orqali hisoblash tez va ma‘lumotlar xotiradan ko‘p joy olmasligiga erishish mumkin. Ular **hech qanday mazmunli ma‘lumot bermaydi**, ayniqsa, matn tasniflash modellarida foydalanish yaxshi natija beradi. TF-IDF asosida o‘zbek tili matnlarini tasniflashda matnga oldindan ishlov berish jarayonida, ya‘ni lemmatizatsiya qilishdan oldin nomuhim so‘zlar matndan olib tashlangan [Mahmudjonova G.U. TF-IDF asosida o‘zbek tili matnlarini tasniflash, O‘zbek tilining milliy korpusi: muammolar va vazifalar mavzusidagi xalqaro ilmiy-amaliy anjumani, 2023.25.03]

Asosiy qism. Nomuhim soʻzlar hujjatning katta qismini tashkil etuvchi matn tarkibidagi soʻzlar boʻlib, ularning umumiy xususiyati shundaki, hujjatda muhim maʼlumotga ega emas; faqat grammatika tufayli ishlatiladi. Hujjatdagi matndan nomuhim soʻzlar tushirib qoldirilsa, matn mazmuni oʻzgarmaydi.

Matnning asl varianti: Farzandingizga ertaklardan birini oʻqing. Matn haqida bir nechta savol bering. Agar bola oʻqiy olsa, uni oʻzi qisqa hikoya oʻqishga taklif qiling va keyin uni qayta aytib bering.

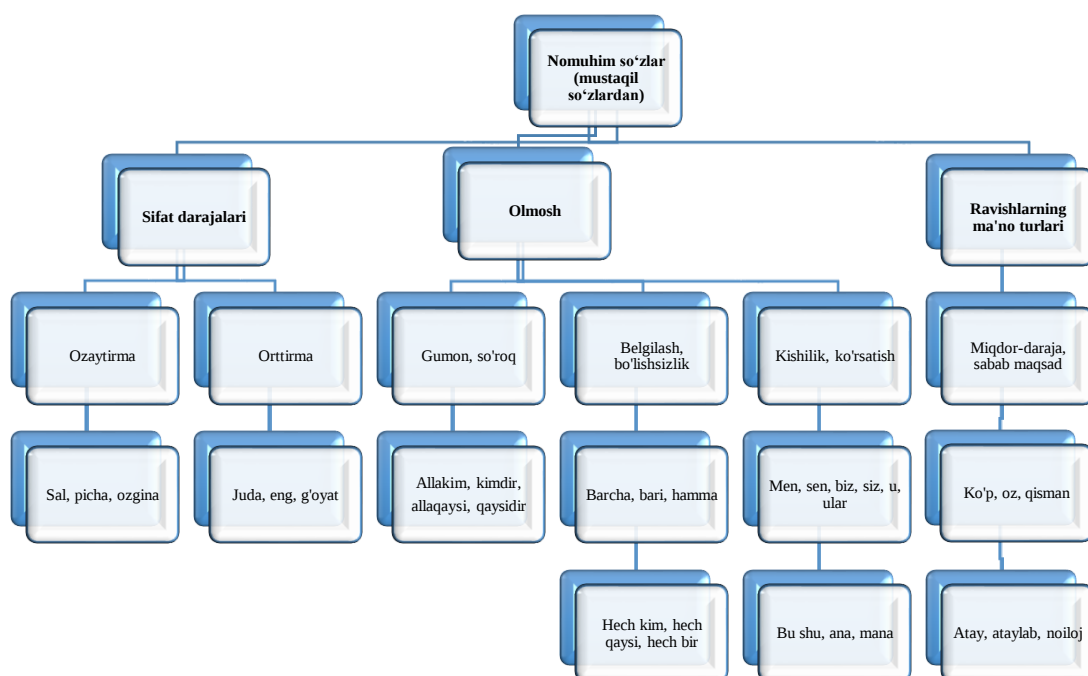
Matndan nomuhim soʻzlar olib tashlangandan keyingi holati: Farzandingizga ertaklardan oʻqing. Matn haqida savol bering. Bola oʻqiy olsa, qisqa hikoya oʻqishga taklif qiling, qayta aytib bering.

Nomuhim soʻzlarini aniqlash usullarini ikkita asosiy toifaga boʻlinadi:

1. Grammatika qoidalariga asoslanib
2. Statistik usullar.

Ushbu maqolada grammatik qoidalarga asoslanib aniqlash mumkin boʻlgan nomuhim soʻzlar taklif etiladi

Nomuhim soʻzlar roʻyxati, asosan, mustaqil soʻz turkumlaridan sifat darajalari, miqdor-daraja, sabab, maqsad ravishi, olmoshlar, yordamchi soʻzlar, modal, taqlid, undov soʻzlardan tuzilishi mumkin. Sababi, bunday soʻzlarning matn mazmuniga taʼsiri kam boʻladi, yaʼni gapda grammatik maʼno ifodalash uchun foydalaniladi. Quyida barchasi misolar orqali tushuntiriladi.



1-jadval. Nomuhim soʻzlar roʻyxati

Predmet, narsa-buyumning belgisini bildirib, qanday, qanaqa soʻroqlariga javob boʻluvchi soʻz turkumi sifat deyiladi. Sifatlar darajalanadi. Belgining meʼyoridan, ortiq yoki kam darajada ekanligini koʻrsatish sifat darajalari deyiladi.

Sifat darajalari 3 xil: oddiy, qiyosiy, orttirma daraja. Leksik usul bilan hosil qilingan sifatning qiyosiy va orttirma darajalari nomuhim soʻz boʻla oladi.

1	<i>Qiyosiy daraja sifat ishtirok etgan gap</i>	Bemorning sogʻligʻi kechagidan sal yaxshi
2	<i>Qiyosiy daraja sifat ishtirok etmagan gap</i>	Bemorning sogʻligʻi kechagidan yaxshi

Yuqorida berilgan gapni orttirma daraja shakli bilan kelganda:

1	<i>Orttirma daraja sifat ishtirok etgan gap</i>	Bemorning sogʻligʻi kechagidan gʻoyatda yaxshi
2	<i>Orttirma daraja sifat ishtirok etmagan gap</i>	Bemorning sogʻligʻi kechagidan yaxshi

Ot, sifat, son, ravish oʻrnida qoʻllanib, ularni koʻrsatishga xizmat qiluvchi va ularning vazifasini bajaruvchi soʻz turkumi olmosh deyiladi. Olmoshlar narsani, uning belgisi va miqdorini anglatmaydi, balki faqat ularni koʻrsatish, ularga ishora qilish uchun xizmat qiladi. Olmosh gapda ega, toʻldiruvchi, aniqlovchi boʻlib keladi. Gapda olmoshlar aniqlovchi vazifasida kelganda, gapda ularni tushurib qoldirish mumkin.

1	<i>Koʻrsatish olmoshi ishtirok etgan gap</i>	Bu qalam juda sifatli.
2	<i>Koʻrsatish olmoshi ishtirok etmagan gap</i>	Qalam juda sifatli

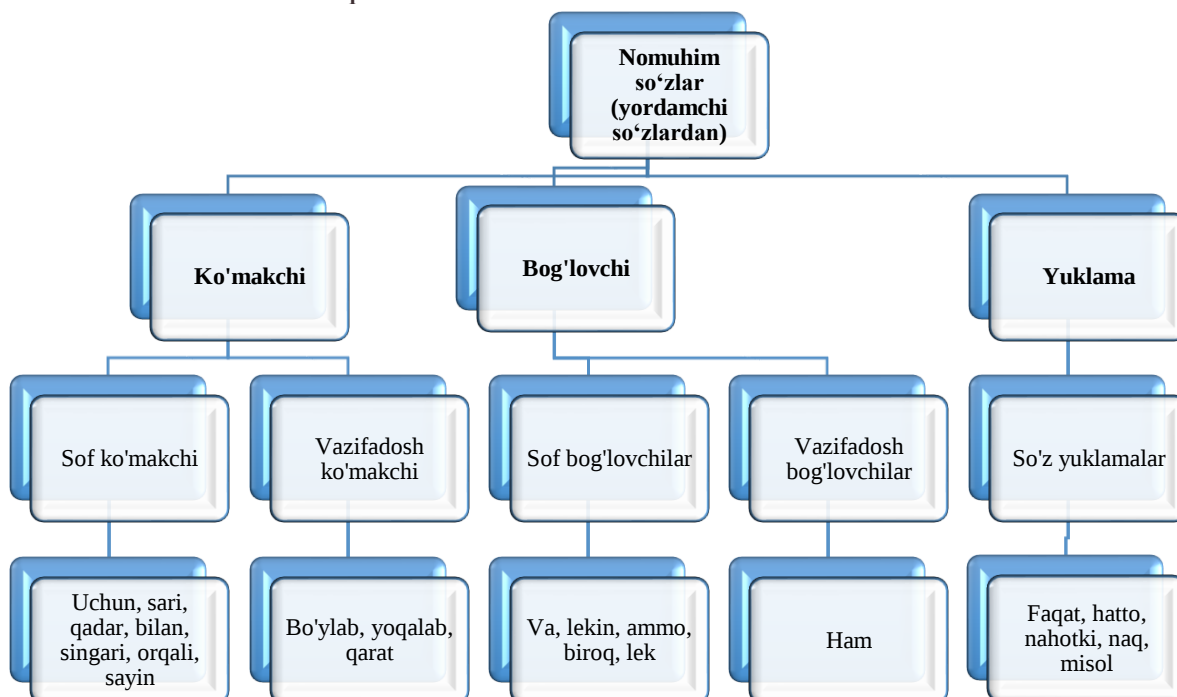
Gumon, belgilash, boʻlishsizlik, kishilik olmoshlari ham gapda aniqlovchi boʻlib kelganda nomuhim soʻzlar roʻyxatiga kiritish mumkin [Sayfullayeva.R, Mengliyev.B va boshqalar, Hozigi oʻzbek tili, Toshkent 2009, 345-bet].

1	<i>Belgilash olmoshi ishtirok etgan gap</i>	Gaplaringiz barchasi yolgʻon
2	<i>Belgilash olmoshi ishtirok etmagan gap</i>	Gaplaringiz yolgʻon

Ish-harakatning, holatning belgisini, belgining belgisini bildiruvchi soʻz turkumi ravish deyiladi. Ravish morfologik jihatdan oʻzgarmas boʻlib, asosan feʼllarga bogʻlanib keladi, hol vazifasini bajaradi. Ravishlarning maʼno turlaridarn miqdor-daraja, sabab, maqsad ravishlarini ham nomuhim soʻzlar roʻyxatiga kiritish mumkin.

1	<i>Miqdor-daraja ravish ishtirok etgan gap</i>	Otabek koʻp ishlardi
2	<i>Miqdor-daraja ravish ishtirok etmagan gap</i>	Otabek ishlardi.

Yordamchi so'zlar turkumiga ko'makchi, bog'lovchi, yuklama kiradi. Yordamchi so'zlar mustaqil (leksik) ma'no bildirmaydi, faqat grammatik ma'no iodalash uchun xizmat qiladi.



2-jadval. Nomuhim so'zlar (yordamchi so'zlar)

Gapda vosita, maqsad, sabab, vaqt, makon kabi sintaktik munosabatlarni ifodalash uchun qo'llanadigan so'zlar ko'makchilar deb ataladi. Ko'makchi o'zi bog'lanib kelgan so'zdan keyin kelib, so'zlar orasidagi birgalik, payt, sabab, maqsad kabi munosabatni ifodalashga xizmat qiladi. Ko'makchi ikki xil bo'ladi: 1. Sof ko'makchi 2. Ko'makchi vazifasidagi so'z (funktional ko'makchi) [6].

1	<i>Sof ko'makchi ishtirok etgan gap</i>	Chapga buriling, keyin to'g'riga yuring
2	<i>Sof ko'makchi ishtirok etmagan gap</i>	Chapga buriling, to'g'riga yuring

Qo'shma gap tarkibidagi sodda gaplar va gap bo'laklarini bog'lash uchun xizmat qiladigan yordamchi so'zlar bog'lovchi deyiladi. Bog'lovchilar qo'llanishiga qarab yakka bog'lovchilar va takroriy bog'lovchilarga ajratiladi. Bog'lovchilar vazifasiga ko'ra esa ikki xil bo'ladi: teng bog'lovchilar; ergashtiruvchi bog'lovchilar [6].

1	<i>Yakka bog'lovchi ishtirok etgan gap</i>	Onam va singlimga gul oldim.
2	<i>Yakka bog'lovchi ishtirok etmagan gap</i>	Onam, singlimga gul oldim.

Xuddi shu gap takroriy bog'lovchi bilan kelganda:

1	<i>Takroriy bog'lovchi ishtirok etgan gap</i>	Ham onamga, ham singlimga gul oldim
2	<i>Takroriy bog'lovchi ishtirok etmagan gap</i>	Onam, singlimga gul oldim.

Matndan teng va ergashtiruvchi bog'lovchilar olib tashlansa, gapning semantik ma'nosi deyarli o'zgarmaydi.

Ayrim so'z yoki gapga qo'shimcha ma'no ifodalash uchun qo'llanadigan yordamchi so'z yuklama deyiladi. Yuklama tuzilishi jihatidan ikki xil: a) affiks holidagi yuklamalar:

-chi, -a (-ya), -da, -u (-yu), -oq (-yoq), -gina (-kina,-qina): b) so'z holidagi yuklamalar: axir, faqat, xuddi, nahotki, hatto, hattoki[6].

1	<i>So'z holidagi yuklama ishtirok etgan gap</i>	Axir , kelaman degandingiz
2	<i>So'z holidagi yuklama ishtirok etmagan gap</i>	Kelaman degandingiz.

Yuqoridagi misollardan ko'rinib turibdiki, yordamchi so'zlar nomuhim so'zlar bo'la oladi.

Fikrning voqelikka munosabatini bildirgan so'zlar modal so'zlar deyiladi. Modal so'zlar ma'no jihatdan quyidagi turlarga bo'linadi:

1. Fikrning aniqligini, realligini tasdiqlovchi modal so'zlar: albatta, shubhasiz, so'zsiz

2. Fikrning taxminligini, gumonligini, noaniqligini bildiruvchi modal so'zlar: ehtimol, chamasi, shekilli, aftidan, haytovur, chog'i.

3. Fikrning o'zaro munosabatini va tartibini bildiradi(yoki fikrning oldingi fikrga bog'liqligini bildiradi): demak, xususan, aksincha, masalan, jumladan, chunonchi,

4. Modallikni ifodalaydi: lozim, mumkin, shart, zarur.

5. Tasdiqni bildiradi: bor, mayli, xo'p, ha.

6. Inkorni bildiradi: yo'q, aksincha.

7. Afsuslanish, ajablanishni bildiradi: afsuski, attang, ajabo.

1	<i>Modal so'z ishtirok etgan gap</i>	Demak , ertaga natijalar aniq bo'ladi.
2	<i>Modal so'z ishtirok etmagan gap</i>	E, Ertaga natijalar aniq bo'ladi.

Demak, nomuhim so'zlar ro'yxatiga modal so'zlar ham kiradi. Shu jumladan, undov so'zlarni ham nomuhim so'zlar qatoriga kiritish mumkin.

His-hayajonni, xitobni bildirishga xizmat qiladigan soʻzlar undov soʻzlar deyiladi: undovlar quyidagi xususiyatlarga ega:

- 1) leksik (lugʻaviy) maʼno anglatmaydi;
 - 2) gap boʻlagi vazifasida kelmaydi va gap boʻlaklari bilan grammatik jihatdan bogʻlanmaydi;
 - 3) undovda ohang muhim oʻrin tutadi;
 - 4) kelishik, egalik, shaxs, zamon kabi grammatik maʼnolarni bildirmaydi.
- Masalan, undov soʻz gapda ishtirok etganda:

1	<i>Undov soʻz ishtirok etgan gap</i>	Oh , buncha shirin olma.
2	<i>Undov soʻz ishtirok etmagan gap</i>	Buncha shirin olma

Taqlid soʻzlar ikki xil:

1. Holatga (obrazga) taqlidni bildiruvchi soʻzlar: yalt, yarq, milt-milt, bilq, dir-dir.
 2. Tovushga taqlidni bildiruvchi soʻzlar: tars, turs, chars, gumbur-gumbur.
- Taqlid soʻzlarni ham nomuhim soʻzlar roʻyxatiga kiritish mumkin.

1	<i>Taqlid soʻz ishtirok etgan gap</i>	U dir-dir qaltirardi
2	<i>Taqlid soʻz ishtirok etmagan gap</i>	U qaltirardi

Yuqoridagi misollardan koʻrinib turibdiki, nomuhim soʻzlar gap tarkibini grammatik shakllantirishda ishlatilinish, semantik maʼno ahamiyati kam.

Nomuhim soʻzlarni hujjatdan 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. Har doim ham ular matnda olib tashlanmaydi. Nomuhim soʻzlarni olib tashlash matn ustida bajarayotgan vazifaga va erishmoqchi boʻlgan maqsadga bogʻliq. Misol uchun, agar his-tuygʻularni tahlil qilish (sentiment analiz) vazifasini bajara oladigan modellarda bunday soʻzlarini olib tashlash notoʻgʻri xulosaga olib keladi. Shuning uchun nomuhim soʻzlarini tanlashda ehtiyot boʻlish kerak. Quyida misol bilan tushintiriladi.

1	Siz bu ishni bugun qilishingiz shart emas .
---	--

2

Siz bu ishni bugun qilishingiz shart.

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 [Tijani O.D. Akinwale A.T. Onashoga S.A. Adeleke E.O, An auto generated approach of stop words using Aggregated analysis, Conference: 2017 Nigeria Computer Society International Conference]. Lekin ingliz tilida doimiy ravishda “a, the, and” [<https://gist.github.com/sebleier>] kabi soʻzlar foydalaniladi. Bundan tashqari rus tilida “с, а, аля” [<https://github.com/stopwords-iso/stopwords-ru>], turk tilida “belki, chunki, dahi” [<https://github.com/tkorkunkaya/Turkish-Stopwords>] kabi soʻzlar nomuhim soʻzlar sifatida githab platformasiga yuklangan. Oʻzbek tili uchun nomuhim soʻzlar roʻyxati Elov B., Hamroyeva Sh., Abdullayeva O., Xusainova Z. tomonidan ishlab chiqilgan boʻlib githab platformasiga [<https://github.com/Gulshahnoz/UzbekStopwords.git>] yuklandi. Nomuhim soʻzlar hujjatlarning konteksti yoki maʼlumotlariga hissa qoʻshmaydi va ular indekslash paytida, shuningdek, maʼlumot olish tizimi tomonidan soʻrov qilishdan oldin olib tashlanishi kerak. Biroq, turli hujjatlar toʻplamida yagona sobit nomuhim soʻzlar roʻyxatidan foydalanish qidiruv samaradorligiga zarar yetkazishi mumkin [Dolamic L. and Savoy J., When stopword lists make the difference. Journal of the American Society for Information Science and Technology, 61(1):200–203, 2010]. Chunki bir hujjatda nomuhim hisoblangan soʻz boshqa bir hujjatda ahamiyatli boʻlishi mumkin [Sankar S.S., Pal S., Effect of stopwords in Indian language IR, Indian Academy of Science, 2021. 25 August]

Nomuhim soʻzlarning xususiyatlari

1. Matnda juda kam qiymatiga ega boʻladi.
2. Matnda paydo boʻlish chastotasi yuqori boʻladi
3. Kamdan-kam hollarda soʻrov soʻzlari/qidiruv soʻzlari sifatida ishlatiladi, yaʼni maʼlumotlarni qidirishda bunday soʻzlardan foydalanilmaydi.
4. Kirish soʻzlar, olmosh, modal, taqlid, undov soʻzlar yordamchi soʻzlar: bogʻlovchi, koʻmakchi yuklama kabi soʻzlar boʻlishi mumkin.
5. Matn uchun umumiy soʻzlar boʻlib, muayyan sohada maxsus ishlatililmaydi.
6. Tilning qurilishi uchun zarurdir.

Xulosa. Nomuhim soʻzlar – bu hech qanday maʼlumot qiymatiga ega boʻlmagan va hujjatning katta qismini tashkil etuvchi matn tarkibidagi soʻzlar. Bunday soʻzlardan mashinali oʻqitish modellarida maʼlumotlar hajmini qisqartirish va modellar tez ishlashi uchun foydalanish mumkin. Matnlarni tasniflovchi

modellarda nomuhim soʻzlarni olib tashlash yaxshi natija beradi. Biroq his-tuygʻularga asoslangan modellarda bunday soʻzlarni olib tashlash notogʻri xulosaga olib kelishi mumkin.

Foydalanilgan adabiyotlar:

1. Dolamic L. and Savoy J., When stopword lists make the difference. Journal of the American Society for Information Science and Technology, 61(1):200–203, 2010
2. Madatov X.A., Sharipov M.S., Bekchanov Sh.K., Oʻzbek tili matnlaridagi nomuhim soʻzlar, <https://www.researchgate.net/publication/351951567>.
3. Elov B., Xusainova Z., Xudayberganov N. Oʻzbek tili korpusi matnlari uchun TF-IDF statistik koʻrsatkichni hisoblash. SCIENCE AND INNOVATION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8 UIF-2022: 8.2 | ISSN: 2181-3337. 1774-1785.
1. Mahmudjonova G.U. TF-IDF asosida oʻzbek tili matnlarini tasniflash, Oʻzbek tilining milliy korpusi: muammolar va vazifalar mavzusidagi xalqaro ilmiy-amaliy anjumani, 2023.25.03
2. Sankar S.S., Pal S. Effect of stopwords in Indian language IR, Indian Academy of Science, 2021. 25 August
3. Sayfullayeva R., Mengliyev B. va boshqalar. Hozigi oʻzbek tili, Toshkent 2009, 345-bet
4. Tijani O.D. Akinwale A.T. Onashoga S.A. Adeleke E.O, An auto generated approach of stop words using Aggregated analysis, Conference: 2017 Nigeria Computer Society International Conference
5. <https://github.com/tkorkuncakaya/Turkish-Stopwords>
6. <https://gist.github.com/sebleier>
7. <https://github.com/stopwords-iso/stopwords-ru>
8. <https://github.com/Gulshahnoz/UzbekStopwords.git>