

SO‘Z MA’NOSINI ANIQLASHDA MASHINALI O‘QITISH YONDASHUVLARI

Axmedova Xolisxon Ilxomovna

Alisher Navoiy nomidagi Toshkent davlat
o‘zbek tili va adabiyoti universiteti tayanch doktoranti
xolisa0629@gmail.com

Mirzaxmedova Parizoda

Alisher Navoiy nomidagi Toshkent davlat
o‘zbek tili va adabiyoti universiteti bakalavr talabasi
mpari0621@gmail.com

Annotatsiya. Tabiiy tilni qayta ishlash jarayonidagi ko‘plab masalalar bugungi kunda o‘z yechimini topmoqda. So‘z ma’nosini aniqlashning muhim vazifalaridan biri omonimiyani aniqlash masalasidir. O‘zbek tilidagi omonim so‘zlarning gapdagi aynan ma’nosini aniqlash usullari bo‘yicha ma’lumot berish maqolamizning asosiy vazifasi. Ushbu maqolada bir so‘z turkumi doirasida omonimlik hosil qiluvchi so‘zlarning gapdagi ma’nosini aniqlashda mashinali o‘qitish usulidan foydalanish va uning turlari, vazifalari haqida ma’lumot berilgan.

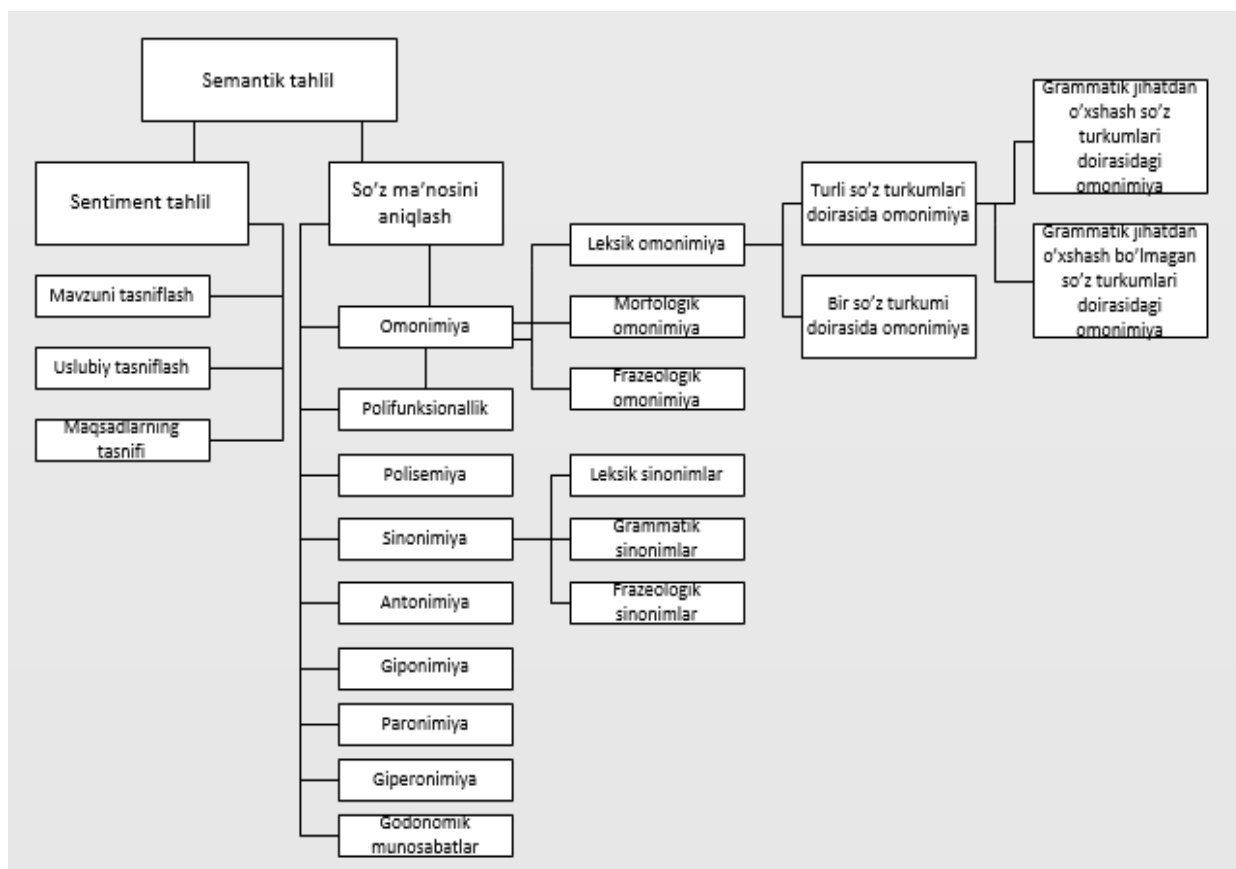
Kalit so‘zlar. *NLP, semantik tahlil, omonimiya, WSD, mashinali o‘qitish usuli, nazorat ostidagi va nazoratsiz usullar.*

Abstract. Many problems in natural language processing are being solved today. One of the important tasks of determining the meaning of a word is the issue of determining homonymy. The main task of our article is to provide information on the methods of determining the exact meaning of homonyms in the Uzbek language. This article provides information on the use of the machine learning method and its types and tasks in determining the meaning of homonymous words in a sentence.

Key words. *NLP, semantic analysis, homonymy, WSD, machine learning method, supervised and unsupervised methods*

Аннотация: В настоящее время решается множество задач обработки естественного языка. Одной из важных задач определения значения слова является вопрос определения омонимии. Основная задача нашей статьи – предоставить информацию о способах определения точного значения омонимов в узбекском языке. В данной статье представлена информация об использовании метода машинного обучения и его видах и задачах при определении значения омонимичных слов в предложении.

Ключевые слова. *НЛП, семантический анализ, омонимия, WSD, метод машинного обучения, контролируемые и неконтролируемые методы*



Tabiiy tilni qayta ishlash (Natural Language Processing (NLP)) – bu sun’iy intellekt va matematik tilshunoslikning umumiy yo‘nalishi bo‘lib, tabiiy tillardagi gaplarni kompyuterda analiz va sintez qilish muammolarini o‘rganadi. Ushbu muammoni hal qilish inson va kompyuter o‘rtasida o‘zaro aloqaning yanada qulay shaklini yaratishni anglatadi. NLP sohasining muhim vazifalaridan biri bu gaplarni semantik tahlil qilish bo‘lib, semantik tahlil ham o‘z navbatida 2 ta guruhga bo‘linadi.

So‘z ma’nosini aniqlash. So'z ma'nosini aniqlash (WSD) usullari kirishning semantik talqinini talab qiladigan ko'plab NLP vazifalari uchun foydalidir. Bundan tashqari, bunday usullar leksikografik tadqiqotlar va tilni o'rganish resurslari uchun muhim bo'lgan turli xil korpuslardagi so'zlarni turli ma'nolarini aniqlash chastotalarini baholashga yordam beradi. Ma'lumotlarga ko'ra, xorijiy tabiiy tillardagi omonimiyani aniqlashda *yarim nazorat ostida* (semi-supervised) so'z ma'nosini aniqlashni amalga oshiradigan har tomonlama baholangan usul yo'q.

Sentiment tahlil qilish- qayta ishlash texnologiyalari(NLP) va Informationsion Texnologiyalar(IT)sohasida yaxshi natijalarga erishilmoqda. Ushbu ilmiy sohaning eng qiziqarli va samarali usullari orasida sentiment tahliliajralib turibdi. Sentiment tahlil qilish kompyuter lingvistikasi uchun samarali tadqiqot yo‘nalishlaridan hisoblanadi.Dastlab, sentiment nima o‘zi degan savolga javob bersak. Bu-matndagi

emotsional rangdagi soʻz boyligini avtomatlashtirilgan tarzda aniqlash uchun ishlatiladi. Sentiment tahlil yana quyidagicha nomlanadi:

- Fikrni izlash (Opinion extraction)
- Fikrni oʻrganish (Opinion mining)
- Hissiy tahlil qilish (Sentiment analysis)
- Subyektiv tahlili (Subjectivity analysis)

Xissiyotlarni tahlil qilish va oʻrganish – bu yozma tildan odamlarning fikrlari, baholari, munosabati va hissiyotlarni tahlil qiladigan soha boʻlib dunyo boʻylab sentiment tahlil deb nomlanadi.

Ushbu maqolamizda soʻz manosi aniqlash boʻyicha jahon tajribasida qoʻllanilgan usullar haqida soʻz boradi. Jahon tajribasi shuni koʻrsatadiki, soʻz maʼnosini aniqlashda usullarini quyidagi guruhlariga ajratish mumkin.

- Qoidalarga asoslangan usul
- Statistik maʼlumotlarga asoslangan usul
- Mashinali oʻqitishga asoslangan usul
- Neyron tarmoq

Usullardan foydalanish samarali ekanligi tajribalardan maʼlum [1],[2], [3],[4]. Navbatdagi dozarb masala gapda uchragan bir soʻz turkumi doirasidagi omonim soʻzlarni semantik farqlash.

1-jadval. Bir va turli soʻz turkumlari doirasidagi omonim soʻzlar va ularning maʼnolari

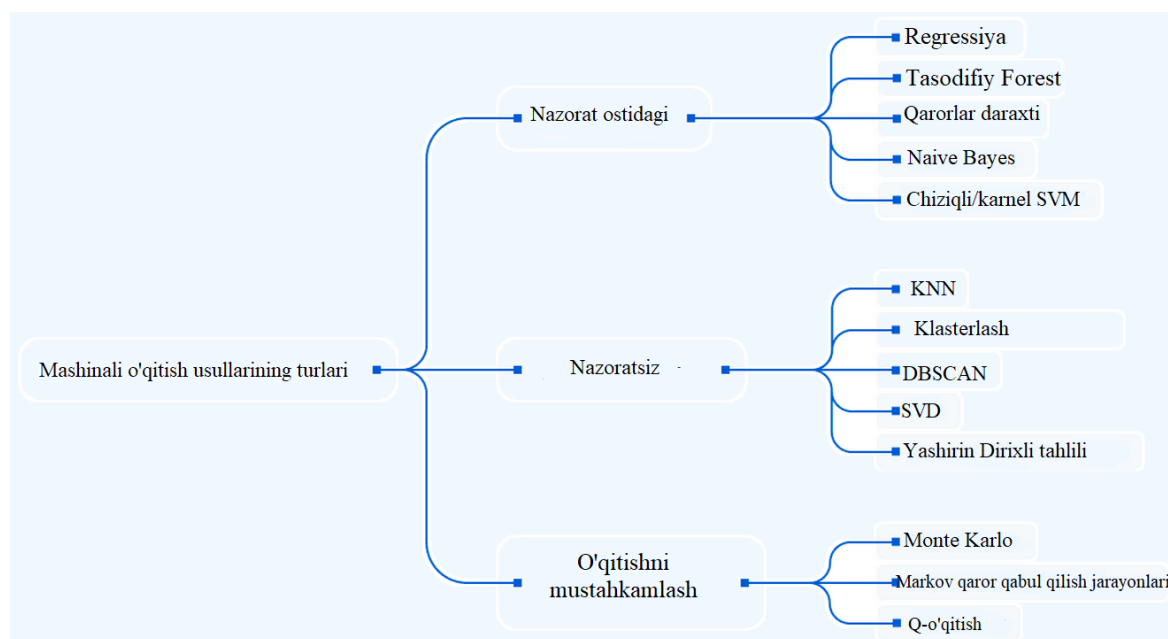
Soʻz	Soʻz turkumi	Maʼnosi
Oʻt	Feʼl	Oʻtmoq feʼli
	Ot	Maysa, oʻt-oʻlan
	Ot	Inson oʻrgani
	Ot	Olov
Oz	Ravish	Kam, miqdori nisbatan koʻp boʻlmagan.
	Feʼl	Oriqlamoq, etidan yoʻqotmoq.
	Feʼl	Noxush boʻlmoq, kuchsizlanmoq, holsizlanmoq.
	Feʼl	Adashmoq, toʻgʻri yoʻldan chetga chiqmoq
Boʻy	Ot	Uzunlik oʻlchovi.
	Ot	Hid, is
...	...	...

1-jadvaldagi omonim soʻzlarning soʻz turkumlarini aniqlash masalasini *qoidalarga asoslangan* hamda *statistik maʼlumotlarga asoslangan usullar* yordamida aniqlash tajribalaridan maʼlum. Navbatdagi masala bir soʻz turkumidagi omonim soʻzning maʼnosini aniqlash.

Masalan:

1. *Bemorning oʻtida tosh borligi aniqlandi*
2. *Qirdagi oʻtlarning yam-yashilligidan insonni quvontiradi.*
3. *Bolaxonadagi somonga ketgan oʻt oʻchirildi.*

Ushbu keltirilgan gaplardagi oʻt soʻzi har bir gapda turli maʼnolarni anglatmoqda. Bu omonim soʻzning gapdagi aynan maʼnosini aniqlash masalasi kunning dolzarb masalalaridan biridir. Bir soʻz turkumidagi omonim soʻzlarni maʼnolarini aniqlash masalasini yechishda mashinali oʻqitish usuli va uning algoritmlaridan foydalanish haqida soʻz yuritsak. Mashinali oʻqitish usullarini 1-rasmda keltirilgani kabi turlarga ajratish mumkin.



1-rasm: Mashinali oʻqitish usulining turlari

1-rasmda keltirilgan usullarga taʼrif beramiz.

Nazorat ostida oʻrganish - bu maʼlumotlarni tasniflash va natijalarni bashorat qilish algoritmlarini oʻrgatish uchun belgilangan maʼlumotlar toʻplamidan foydalanish orqali aniqlanadigan mashinali oʻqitish usuli. Oldindan belgialb qoʻyilgan maʼlumotlar toʻplamidan nimani qidirish kerakligini tushunish uchun mashina kirish maʼlumotlariga mos keladigan natija chiqadi. Nazorat ostidagi oʻqitishda foydali boʻlgan ikkita asosiy yoʻnalish mavjud: *tasniflash muammolari* va *regressiya muammolari*.

Tasniflash–kirish qiymatini olish va uni diskret qiymatga aylantirishni anglatadi. Tasniflash parametrlarini aniqlash orqali omonimiyani aniqlashda Naïve Bayes usulidan foydalanish masalasini [7] da ko‘rib chiqqan edik.

Regressiya–uzluksiz ma'lumotlar (qiymat funktsiyalari) bilan bog'liq. Regressiyada bashorat qilingan so‘zning ma’nosining miqdor ko‘rinishi haqiqiy sonlardan iborat bo‘ladi.

Nazoratsiz o‘qitish modellari–nazorat ostidagi o‘qitish modellariga qaraganda murakkabroq vazifalarni bajarishi mumkin. Ushbu yondashuvdan foydalanadigan asosiy vazifalar: klasterlash, uyushgan (association), hajmlarni qisqartirishdan iborat.

Klasterlash –bu nazoratsiz o‘qitish turi bo‘lib, ma'lumotlarda ularning o'xshashliklari yoki farqlari asosida yashirin farqlarni aniqlaydi. Ushbu farqlar omonim so‘zning ma’nolariga qarab turlicha bo‘lishi mumkin va ma'lumotlar elementlarini guruhlash yoki klasterlar yaratish uchun ishlatiladi.

Uyushgan yoki assotsiatsiya –bu ham nazoratsiz o‘qitish turi bo‘lib, unda so‘zning bir ma'nosi bilan uning boshqa ma’nolari orasidagi aloqasini, qisqa qilib aytganda semantik o‘xshashlinigi topish mumkin.

O‘qitishni mustahkamlash– nazorat ostidagi va nazoratsiz o‘qitish usullari yordamida olingan ma'lumotlarning aniqlik darajasini qayta tekshirishdan iborat.

Tajribalar 23 ta noaniq so'zlar, har biri oltita kontekstda (138 juft so'z) nisbatan nozik so'zlardan foydalangan holda to'g'ri ma'noni aniqladi

Xulosa o‘rnida shuni aytish mumkinki, omonim so‘zlarni aniqlashda qoidalarga asoslangan usullar, statistik ma'lumotlarga asoslangan usullardan foydanish natijasida olinga natijalarning aniqlik darajasi mos ravishda 75% , 82.3% ni tashkil etgan bo‘lsa, bu aniqlikni yanada yuqori darajaga yetkazish mashinali o‘qitish usullaridan foydalanish samarali natija berishiga ishonamiz.

Foydalanilgan adabiyotlar:

1. Elov B.B, Axmedova X.I. Uchta so‘z turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish//O‘zbekiston respublikasi innovatsion rivojlanish vazirligining, ilm-fan va innovation rivojlanish ilmiy jurnal 2022 / 1, 150-161-b.
2. Ш.К.Гулямова, Х.И. Ахмедова. Омоним сўзларнинг лингвистик асоси, математик модели ва алгоритмлари (от ва феъл ҳамда сифат ва феъл туркуми доирасида)// Qo‘qonDPI. Ilmiy xabarlar,2021/1
3. Axmedova X. Ravish/fe‘l va ravish/yuklama so‘z turkumlari orasidagi omonimlikni farqlovchi matematik modellar // “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 18.10.2021. – 452 b.
4. Axmedova X., Xusainova Z. Fe‘l/ko‘makchi va sifat/son so‘z turkumlari orasidagi omonimlikni farqlovchi matematik modellar //“Zamonaviy leksikografiya,



til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 18.10.2021. – 452 b.

5. Aarts, “Corpus linguistics: An appraisal.” In Hammesse, Jacqueline and Zampolli, Antonio (Eds.), Computers in Literary and Linguistic Research, Champion Slatkine, Paris-Genève: 1990/January, 13-28

6. Adriaens, Geert (1987). “WEP (word expert parsing) revised and applied to Dutch.” Proceedings of the 7th European Conference on Artificial Intelligence, ECAI’86, July 1986.

7. Elov B.B, Axmedova X.I. Polisemantik so‘zlarni semantik tahlil qilishning matematik modellar. O‘zbekiston: til va madaniyat (Amaliy filologiya), 2021, 1(4). 77-85-b.

8. Elov B.B, Axmedova X.I. Business Process Modeling That Distinguishes Homonymy Within Three Parts of Speeches in Uzbek Language // 7th International Conference on Computer Science and Engineering, 14-16 September 2022, Turkey. 289-294 -b.